New estimates of inequality of opportunity across European cohorts (and some insights on long-term impacts of educational policy)

Philippe Van Kerm
University of Luxembourg and LISER
joint with Francesco Andreoli, Alessio Fusco and Iryna Kyzyma
Norface IMCHILD Project — http://dynamicsofinequality.org

LSE International Inequalities Institute, e-London, November 16, 2021

Objectives

What do we (aim to) do?

1. Provide new estimates of inequality of opportunity across EU countries (see, e.g., Brzezinski, 2020, Ramos and Van de gaer, 2021, among others)
   - 'fresh' EU-SILC 2019 data
   - estimates by *birth cohort* and for a range of alternative IOp measures
   - a semi-parametric distribution regression model
   - a simple correction for the 'upward bias' in IOp measures

2. Examine if educational policies affecting parental education are related to IOp in the offspring generation (using variations over time (across cohorts) and across countries in selected educational policy—notably compulsory schooling regulation)

Overview

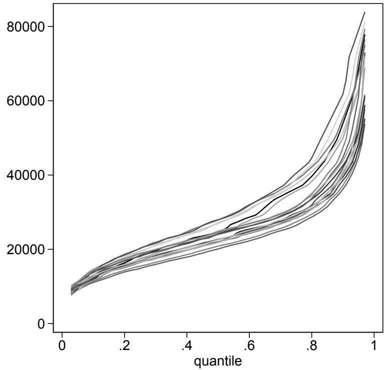
1. The measurement of Inequality of Opportunity: principles and methods
2. EU SILC Data
3. Results I: Aggregate IOp results
4. Results II: Inequality of opportunity across countries and cohorts
5. Results III: A long-term impact of educational policy?

The measurement of Inequality of Opportunity: principles and methods

Empirical IOp measurement in a nutshell

What is the extent of 'unfair' inequality?

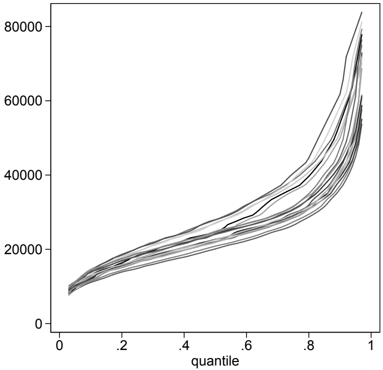
- Partition population by 'circumstances': personal characteristics independent of a person's responsibility
- *Parent's education*, gender, ethnicity, parent's occupation, ...
- NB: not all circumstances ever observed (lower bound)
- Outcome: earnings or income in adulthood



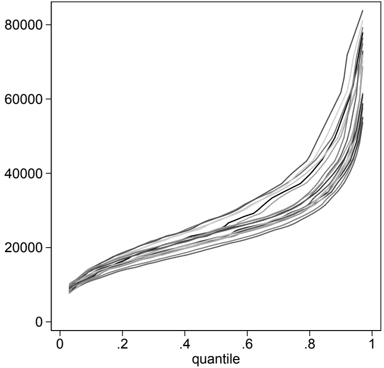
Empirical IOp measurement in a nutshell (ctd.)

What is the extent of 'unfair' inequality?

- Ascribe effort level to position in 'within type' distribution (Roemer et al., 2003)
- IOp: how much of overall differences in income are driven by differences across quantile lines
- Alternative ways to evaluate IOp in a single index
  - *ex ante* vs. *ex post*; direct or indirect; absolute or relative (Ramos and Van de gaer, 2016)
  - based on different counterfactual distributions

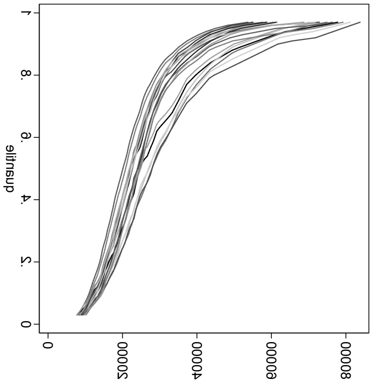


Estimation of conditional quantile functions by ‘distribution regression’



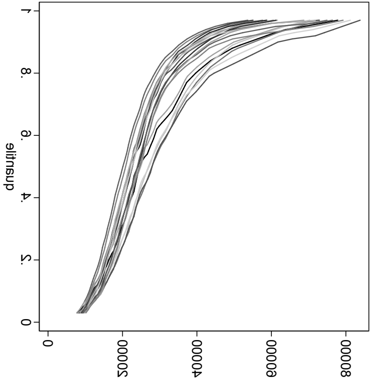
- $F_C(y) = \Pr\{y_i \leqslant y | C\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$).
- Estimate $F_C(y)$ on a grid of values for y spanning the domain of definition of Y by repeated standard binary choice models (e.g., a logit) (Foresi and Peracchi, 1995)
- Invert estimated F_C to get the quantile functions

Estimation of conditional quantile functions by ‘distribution regression’



- $F_C(y) = \Pr\{y_i \leqslant y|C\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$).
- Estimate $F_C(y)$ on a grid of values for y spanning the domain of definition of Y by repeated standard binary choice models (e.g., a logit) (Foresi and Peracchi, 1995)
- Invert estimated F_C to get the quantile functions

Estimation of conditional quantile functions by 'distribution regression'

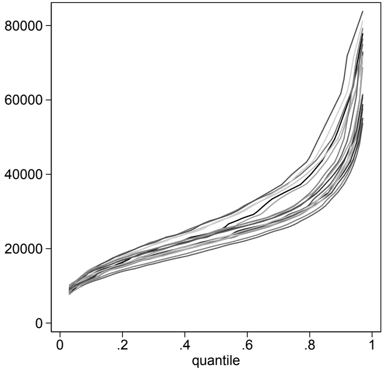


$F_C(y) = \Pr\{y_i \leqslant y|C\}$ is a binary choice model once y is fixed (dependent variable is $1(y_i < y)$).

Estimate $F_C(y)$ on a grid of values for y spanning the domain of definition of Y by repeated standard binary choice models (e.g., a logit) (Foresi and Peracchi, 1995)

Invert estimated F_C to get the quantile functions

Estimation of conditional quantile by ‘distribution regression’



- estimation is straightforward and fast
- often provides better fit than quantile regression (Rothe and Wied, 2013, Van Kerm et al., 2017)
- little assumptions about the overall shape of distribution
- (exchangeable) bootstrap inference (Chernozhukov et al., 2013)

A simple solution to the upward bias in IOp

- ‘upward bias’: Adding circumstances mechanically increases IOp ... even if a circumstance is irrelevant (unrelated to income)
- A solution: penalized regression approaches or cross-validation based selection of circumstances (Brunori et al., 2019)
  - Comparisons of estimates based on different sets?
  - Extension beyond Ex Ante regression models?
- A (simpler) solution: correct IOp estimate by measure of the upward bias

$$\mathrm{AdjIOp} = \mathrm{IOp} - \mathit{Bias}$$

where *Bias* is estimated as IOp in permutations of the data that dissociate income and circumstances while preserving their marginal distributions

A simple solution to the upward bias in IOp

- 'upward bias': Adding circumstances mechanically increases IOp ... even if a circumstance is irrelevant (unrelated to income)
- A solution: penalized regression approaches or cross-validation based selection of circumstances (Brunori et al., 2019)
  - Comparisons of estimates based on different sets?
  - Extension beyond Ex Ante regression models?
- A (simpler) solution: correct IOp estimate by measure of the upward bias

$$\mathrm{AdjIOp} = \mathrm{IOp} - \mathit{Bias}$$

where *Bias* is estimated as IOp in permutations of the data that dissociate income and circumstances while preserving their marginal distributions

A simple solution to the upward bias in IOp

- ‘upward bias’: Adding circumstances mechanically increases IOp ... even if a circumstance is irrelevant (unrelated to income)
- A solution: penalized regression approaches or cross-validation based selection of circumstances (Brunori et al., 2019)
  - Comparisons of estimates based on different sets?
  - Extension beyond Ex Ante regression models?
- A (simpler) solution: correct IOp estimate by measure of the upward bias

$$\mathrm{AdjIOp} = \mathrm{IOp} - \mathit{Bias}$$

where *Bias* is estimated as IOp in permutations of the data that dissociate income and circumstances while preserving their marginal distributions

The treatment of age in empirical IOp measurement

- many papers examine a restricted age range in the population of interest (e.g., 26–50, 30–50) and ignore age
- other papers allow for a broader age range and treat age as a circumstance
- others condition age away by regression

Yet

- age is an important determinant of outcome
- it differs widely in a cross-section of the population
- and it is usually correlated with circumstances (parental background)
- it is obviously beyond one's control… but a circumstance that needs compensation?

The treatment of age in empirical IOp measurement

- many papers examine a restricted age range in the population of interest (e.g., 26–50, 30–50) and ignore age
- other papers allow for a broader age range and treat age as a circumstance
- others condition age away by regression

Yet

- age is an important determinant of outcome
- it differs widely in a cross-section of the population
- and it is usually correlated with circumstances (parental background)
- it is obviously beyond one's control... but a circumstance that needs compensation?

Focus on within cohort IOp

Our approach: concentrate on ‘within cohort’ IOp (Bussolo et al., 2019)

- stratify population by year of birth
- estimate IOp indicators among individuals born in (approximately) same year
- practically, to handle small sample sizes, we apply kernel weighting to semi-parametric conditional distribution estimation
  - varying coefficient model of degree 0
  - Epanechnikov kernel weighting (5 years bandwidth)

EU SILC Data

EU Statistics on Income and Living Conditions

- Micro-data source for EU social indicators (legally binding in all EU countries)—used to produce ‘Laeken indicators’ (and Eurostat statistics on income and living conditions)
- Representative samples of the population of all EU Member States
- Detailed household annual income information
- Largely (yet not perfectly) comparable across countries
- Mix of register-based and survey-based information
- (4-year rotating panel structure)
- Available annually since 2003

2005, 2011 and 2019 special ‘ad hoc’ modules

In these years, EU SILC collects information about parental background (for respondents aged 16+)

- 2005 – Intergenerational transmission of poverty
- 2011 – Intergenerational transmission of disadvantages
- 2019 – Intergenerational transmission of disadvantages

$\Longrightarrow$ parental education and occupation (when teenager), year of birth of parents, presence of siblings, family structure, (financial precarity when teenager–not used) (not always fully comparable)

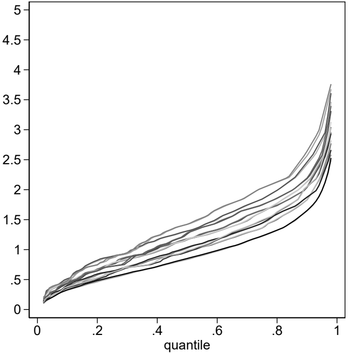
Outcome and circumstances

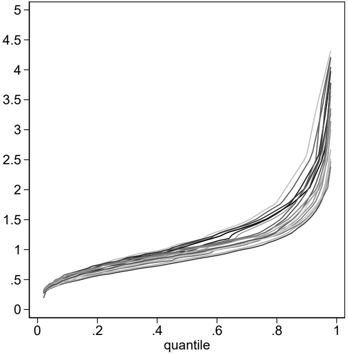
- Outcome: household disposable income (adjusted for family size) — the standard measure of welfare for official poverty or inequality indicators
- Circumstance set 1: sex, father and mother education (low, middle, high, or unknown)
- Circumstance set 2: as above plus country of birth and father and mother occupation presence of siblings and family composition when teenager

The Gini coefficient is used as inequality functional.

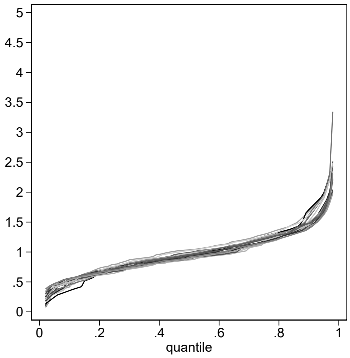
Results I: Aggregate IOp results

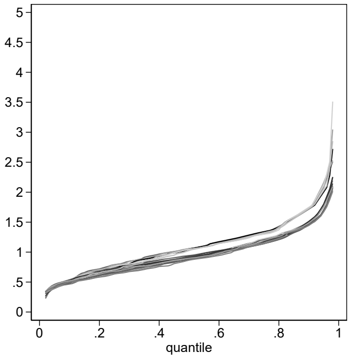
Quantile functions for different circumstances (ex. ES and FR)





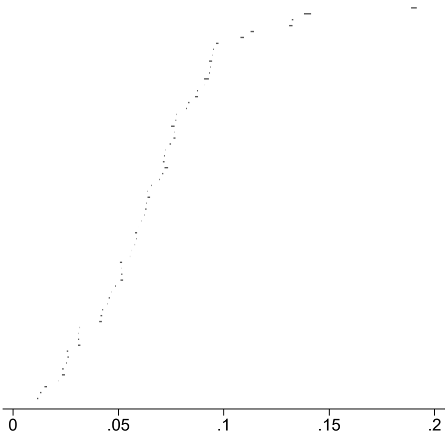
Quantile functions for different circumstances (ex. DK and FI)





How does the distribution regression approach compares with the classic OLS?

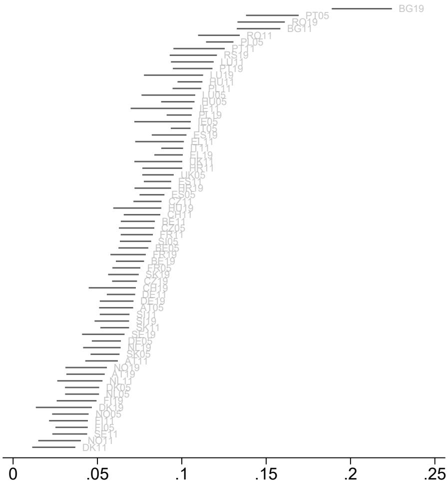
Very little difference on a standard Direct Ex Ante between OLS and our DR approach



What difference does the upward bias correction make?

The upward bias correction is substantial!

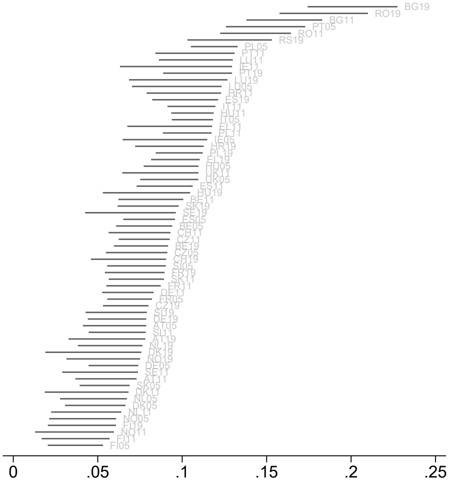
In line with Brunori et al. (2019)



What difference does the upward bias correction make?

The upward bias correction is substantial!

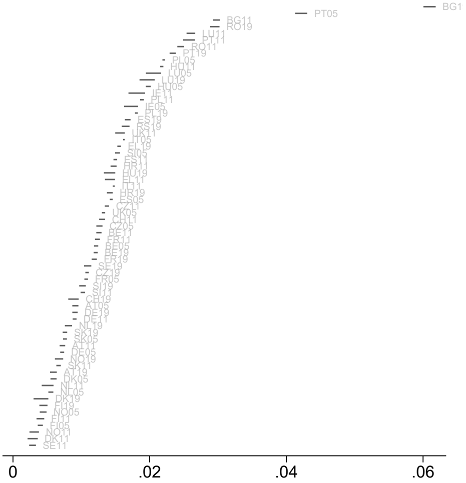
In line with Brunori et al. (2019)



What difference does the upward bias correction make?

The upward bias correction is substantial!

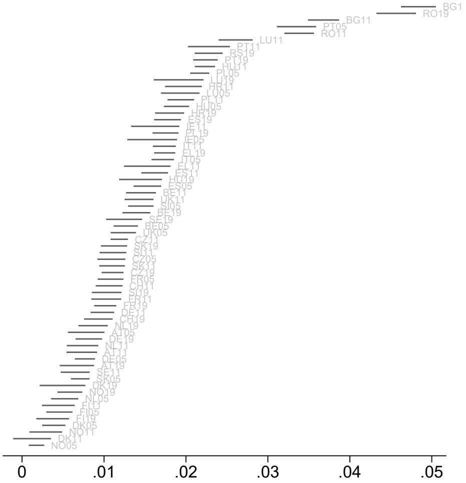
In line with Brunori et al. (2019)



What difference does the upward bias correction make?

The upward bias correction is substantial!

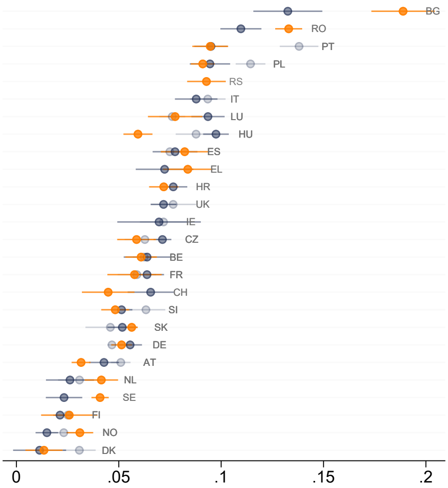
In line with Brunori et al. (2019)



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

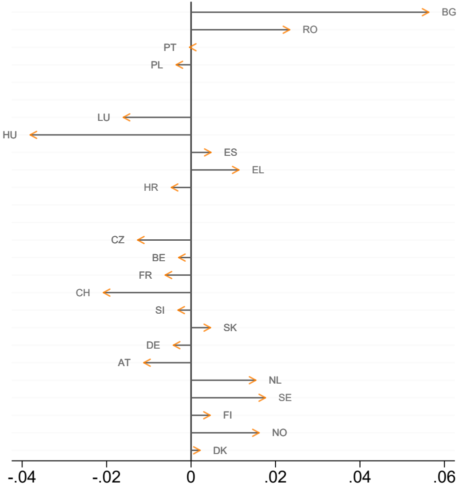
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

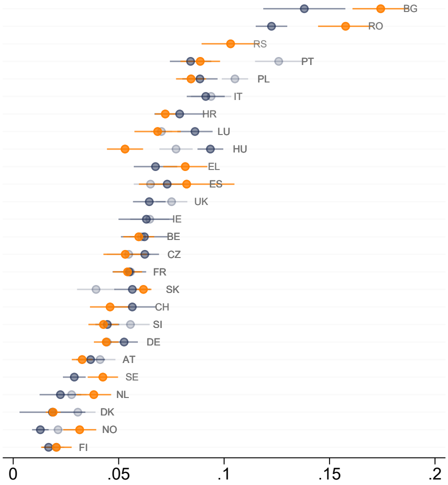
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

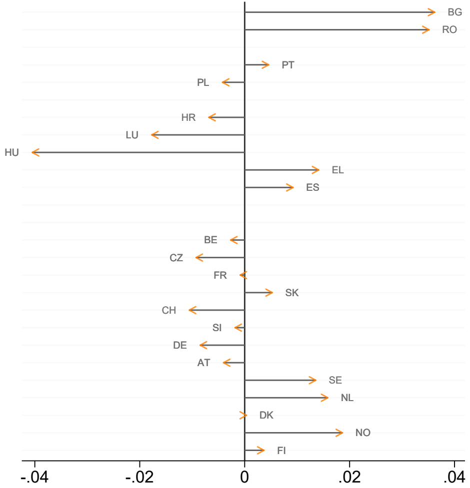
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

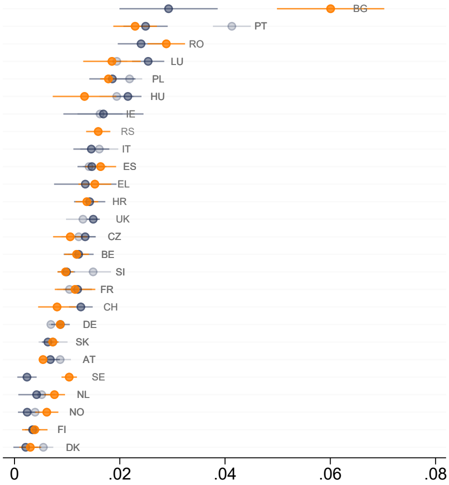
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

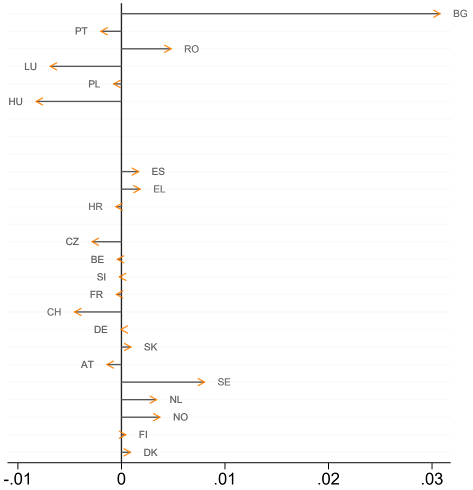
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

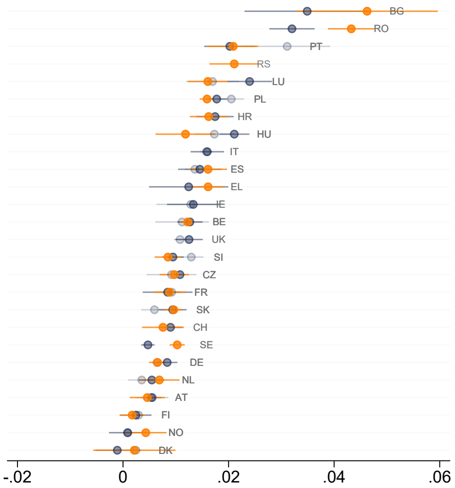
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

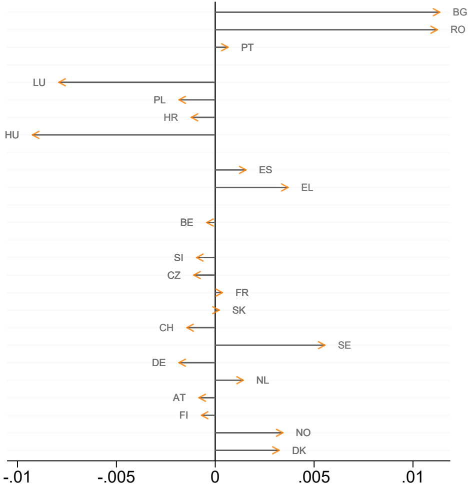
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Evolution (at aggregate level)

New data for 2019 show sharp increases in IOp for a range of countries (BG, RO, SE)

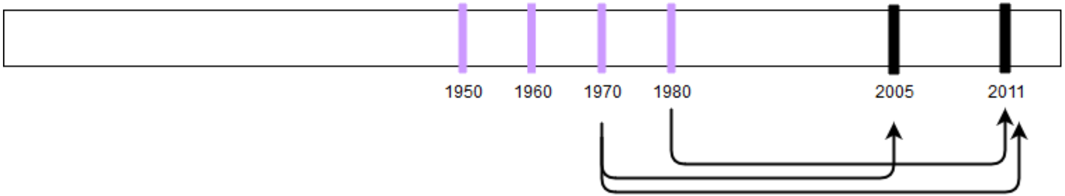
... some regression towards the mean (convergence) across countries but no obvious pattern emerges for the rest



Results II: Inequality of opportunity across countries and cohorts

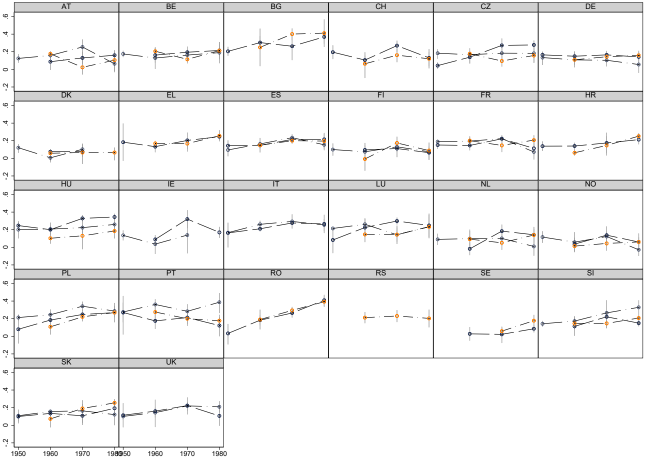
Focus on ‘within cohort’ IOp

- estimate IOp indicators among individuals born in (approximately) same year
  - practically, to handle small sample sizes, we apply kernel weighting to distribution estimation
  - varying coefficient model of degree 0
  - Epanechnikov kernel weighting (5 years bandwidth)



Results: Direct ex ante IOp (relative)

Cohort-level IOp for cohorts born in 1950, 1960, 1970 and 1980



Upward linear trend across cohorts in (almost) all countries after year-effect adjustment

Notable pattern in Eastern EU countries

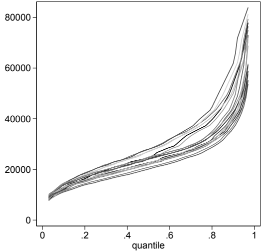
Results III: A long-term impact of educational policy?

Policy transmission and IOp

Do education policies affecting parents educational achievement influence offspring IOp measures?

Many possible mechanisms (Bussolo et al., 2019), e.g.,

- Composition effects (more children of 'better background'—ambiguous)?
- Price effects (increased supply affects return to education—lower IOp)?
- Selection effects (what parents are affected by education policies—ambiguous)?

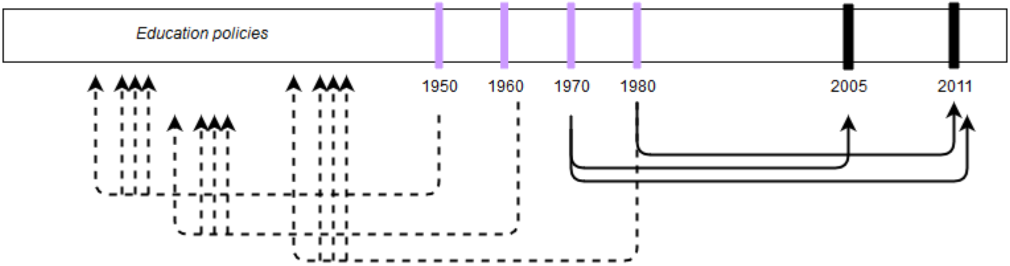


Education policies and IOp estimates

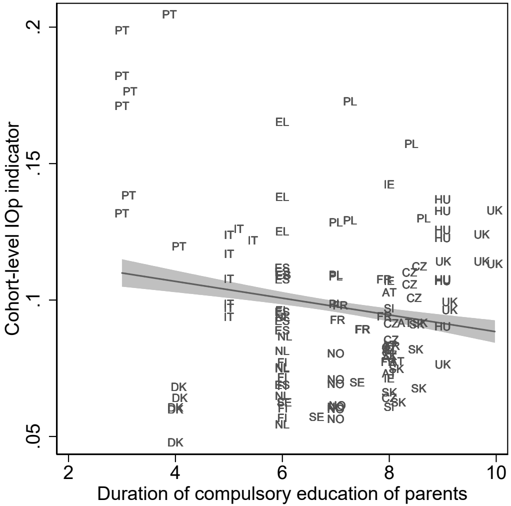
Main education reforms across Europe in first half of 20th century:

- Number of compulsory school years
- Minimum school leaving age

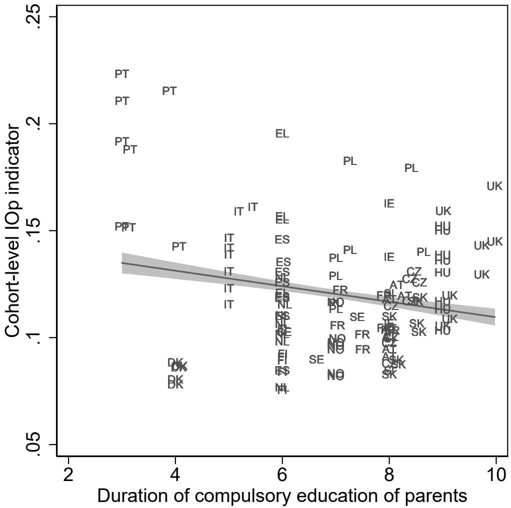
... matched to respondent's parent year of birth (Braga et al., 2013)



Variations in duration of compulsory education



Variations in duration of compulsory education



Cross-country regressions

Direct Ex ante IOp (divided by Gini)

	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) FE
DEA Gini index					
Compulsory schooling	-0.00242		-0.0224	-0.0182	0.00834
Minimum school leaving age		0.00147	0.0273	0.0175	–
Average parental birth year				0.00331*	0.00326*
DEP Gini index					
Compulsory schooling	-0.00412		-0.0218+	-0.0159	-0.00571
Minimum school leaving age		-0.000991	0.0242+	0.0105	–
Average parental birth year				0.00462*	0.00461*
IEA Gini index					
Compulsory schooling	-0.00334		-0.0131	-0.0116	0.00467
Minimum school leaving age		-0.00177	0.0133	0.00997	–
Average parental birth year				0.00114*	0.00109*
IEP Gini index					

In short

- New estimates of IOp across European countries and *cohorts*
- Application of semi-parametric approach to IOp analysis
- ... with correction for upward bias
- IOp higher for more recent cohorts (at least after baby boomers)—contrast with Bussolo et al. (2019)
- Variation in IOp across countries and cohorts related to educational policy variation affecting one's parents
- ... but no very clear evidence of direction so far!

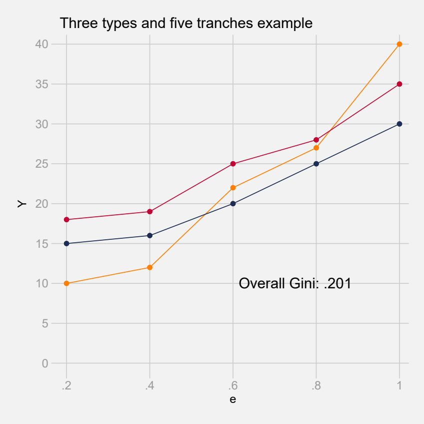
Option 1: Direct, ex ante

Counterfactual outcome is average outcome in the type:

$$y_i^{EA,d}(C_i, e_i) = \mu_{C_i}$$

and

$$IOP^{EA,d} = I(y^{EA,d})$$



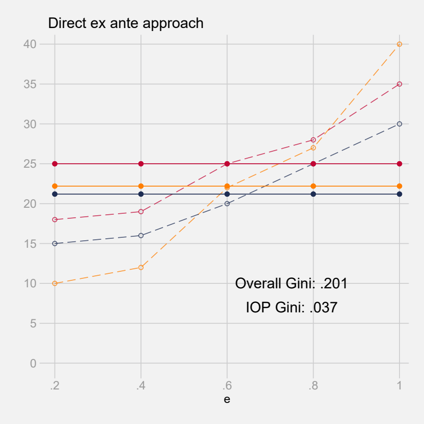
Option 1: Direct, ex ante

Counterfactual outcome is average outcome in the type:

$$y_i^{EA,d}(C_i, e_i) = \mu_{C_i}$$

and

$$IOP^{EA,d} = I(y^{EA,d})$$



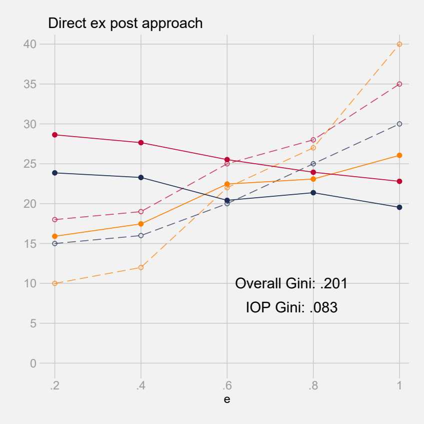
Option 2: Direct, ex post

Counterfactual outcome is outcome scaled (shrunken) by tranche mean relative to grand mean:

$$y_i^{EP,d}(C_i, e_i) = y_i \times \frac{\mu}{\mu_{e_i}}$$

and

$$IOP^{EP,d} = I(y^{EP,d})$$



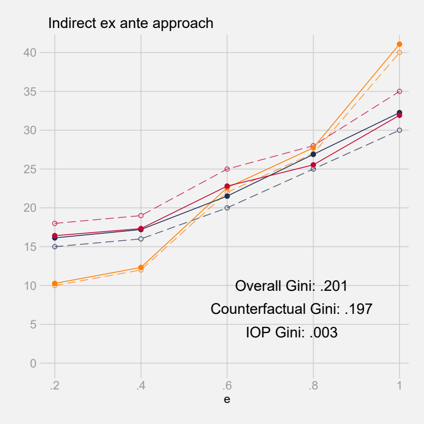
Option 3: Indirect, ex ante

Counterfactual outcome is outcome scaled (shrunken) by type mean relative to grand mean:

$$y_i^{EA,i}(C_i, e_i) = y_i \times \frac{\mu}{\mu_{C_i}}$$

and

$$IOP^{EA,i} = I(y) - I(y^{EA,i})$$



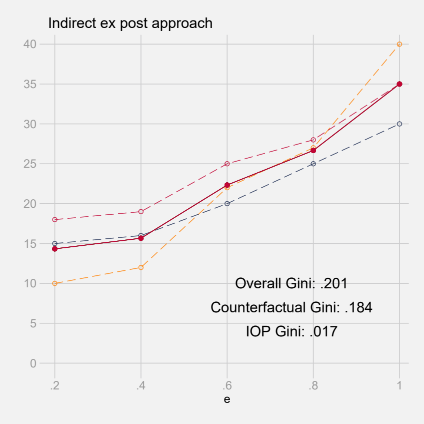
Option 4: Indirect, ex post

Counterfactual outcome is mean outcome within tranche:

$$y_i^{EP,i}(C_i, e_i) = \mu_{e_i}$$

and

$$IOP^{EP,i} = I(y) - I(y^{EP,i})$$



LISER
UNIVERSITÉ DU LUXEMBOURG

References

Braga, M., Checchi, D. and Meschi, E. (2013), 'Educational policies in a long-run perspective', *Economic Policy* **28**(73), 45–100.

Brunori, P., Peragine, V. and Serlenga, L. (2019), 'Upward and downward bias when measuring inequality of opportunity', *Social Choice and Welfare* **52**(4), 635–661.

Brzezinski, M. (2020), 'The evolution of inequality of opportunity in Europe', *Applied Economics Letters* **27**(4), 262–266.

Bussolo, M., Checchi, D. and Peragine, V. (2019), Long-term evolution of inequality of opportunity, Policy Research Working Paper 8700, The World Bank Group, Washington D.C., USA.

References ii

Chernozhukov, V., Fernández-Val, I. and Melly, B. (2013), 'Inference on counterfactual distributions', *Econometrica* **81**(6), 2205–2268.
URL: *http://dx.doi.org/10.3982/ECTA10582*

Foresi, S. and Peracchi, F. (1995), 'The conditional distribution of excess returns: An empirical analysis', *Journal of the American Statistical Association* **90**(430), 451–466.

Ramos, X. and Van de gaer, D. (2016), 'Approaches to inequality of opportunity: Principles, measures and evidence', *Journal of Economic Surveys* **30**(5), 855–883.

Ramos, X. and Van de gaer, D. (2021), 'Is inequality of opportunity robust to the measurement approach?', *Review of Income and Wealth* **67**(1), 18–36.

Roemer, J. E., Aaberge, R., Colombino, U., Fritzell, J., Jenkins, S. P., Lefranc, A., Marx, I., Page, M., Pommer, E., Ruiz-Castillo, J., Segundo, M. J. S., Tranaes, T., Trannoy, A., Wagner, G. G. and Zubiri, I. (2003), 'To what extent do fiscal regimes equalize opportunities for income acquisition among citizens?', *Journal of Public Economics* **87**(3–4), 539–565.
URL: *http://www.sciencedirect.com/science/article/pii/S0047272701001451*

References iii

Rothe, C. and Wied, D. (2013), ‘Misspecification testing in a class of conditional distributional models’, *Journal of the American Statistical Association* **108**(501), 314–324.

Van Kerm, P., Yu, S. and Choe, C. (2017), ‘Decomposing quantile wage gaps: a conditional likelihood approach’, *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **65**(4), 507–527. **URL:** *http://onlinelibrary.wiley.com/doi/10.1111/rssc.12137/pdf*