The estimation of inequality of opportunity

Inequality: Measurement, analysis and policies
CEPAL, Santiago de Chile - 26 & 27 August 2019

Paolo Brunori
University of Florence & University of Bari

based on material co-authored with: Paul Hufe, Daniel G. Mahler, Guido Neidhöfer and Jaquín Serrano

Unfair inequalities

PSJM, American Colours, 2010

Unfair inequalities

PSJM, American Colours, 2010

Roemer's Model

$$y_i = g(C_i, e_i)$$

- y_i: individual's i outcome (income, health, education, access to basic services, ...);
- C_i: circumstances beyond individual control;
- e_i: effort.

Types and effort tranches

- Romerian type: set of individuals sharing exactly the same circumstances;
- effort tranche: set of individuals exerting the same effort;
- there is equality of opportunity if:

$$e_i = e_j \iff y_i = y_j \ , \ \forall i, j \in 1, ..., n$$

$\Rightarrow$ inequality of opportunity (IOP) = within-tranche inequality.

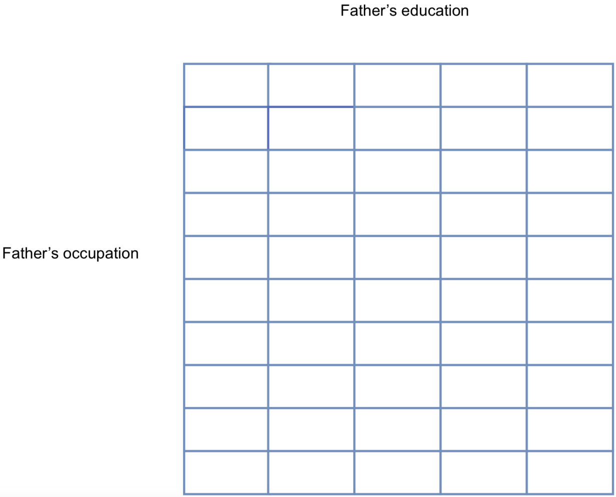
Measurement strategy

- a substantial part of the empirical contributions ignore effort;
- Roemer's identification assumptions:
  1. observability: we correctly assign individuals to types;
  2. orthogonality: $e \perp\!\!\!\perp C$;
  3. monotonicity: $\frac{\partial g}{\partial e} \geq 0$.
- degree of effort = quantile of the type-specific outcome distribution;

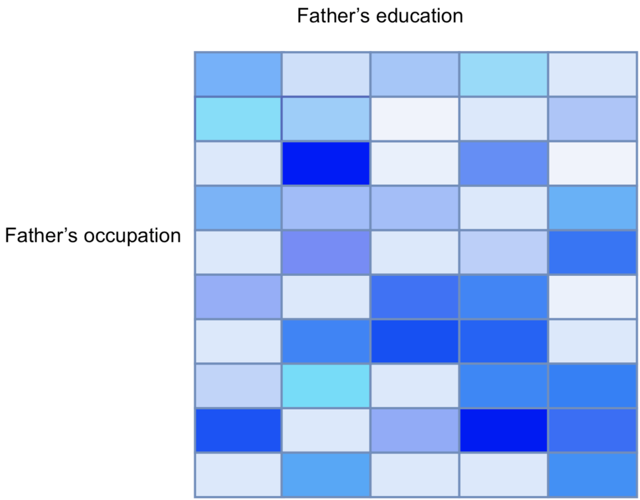
3-step estimation

- identification of Romerian types;
- estimation of the type-specific outcome distribution;
- $\text{IOP} = I\left(\frac{y_i}{\mathbf{E}[y|_{e_i}]}\right)$.

Simplified model



Non-parametric (Checchi and Peragine, 2010)



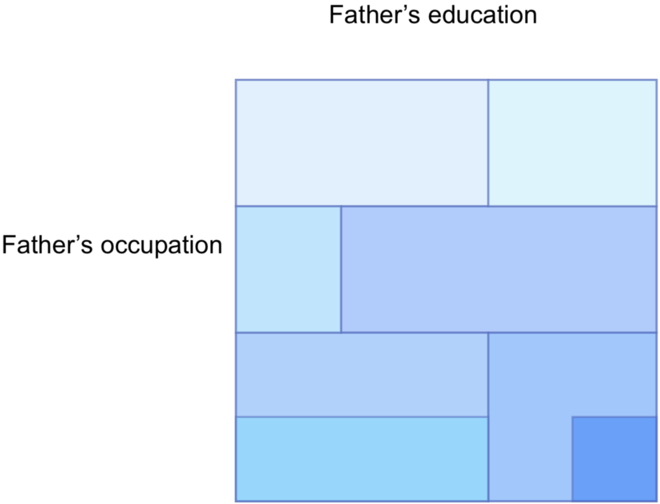
Types

- sparsely populated types $\rightarrow$ uncertain and upward-biased estimates;
- arbitrary reduction of types $\rightarrow$ downward biased estimates;
- assumptions on the DGP (Ferreira and Gignoux, 2011) $\rightarrow$ downward-biased estimates;
- recent contributions use machine learning (ML).

Can supervised ML help?

- yes: we have no idea of the data generating process;
- yes: we increasingly have larger set of controls and larger sample size;
- yes: ML provides a non-arbitrary criterion to select model complexity (minimize MSE out-of-sample).

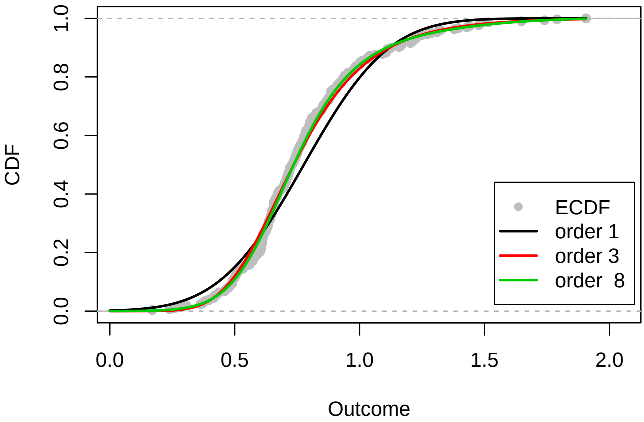
Supervised learning: conditional inference trees



Conditional inference trees (Hothorn et al., 2006)

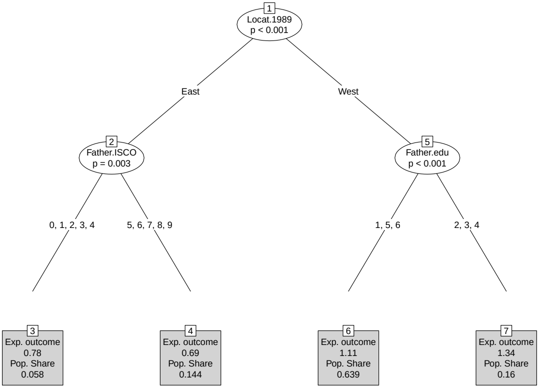
- choose α;
- $\forall p$ test the null hypothesis of independence: $H^{C_p} = D(Y|Cp) = D(Y), \forall C_p \in \mathbf{C}$;
- if no (adjusted) p-value $< \alpha \rightarrow$ exit the algorithm;
- select the variable, $C^{\star}$, with the lowest p-value;
- test the discrepancy between the subsamples for each possible binary partition based on $C^{\star}$;
- split the sample by selecting the splitting point that yields the lowest p-value;
- repeat the algorithm for each of the resulting subsample.

Effort identification by ECDF approximation



Bernstein polynomials of optimal degree are used to estimate $\mathbf{E}[y|_e]$ in each type.

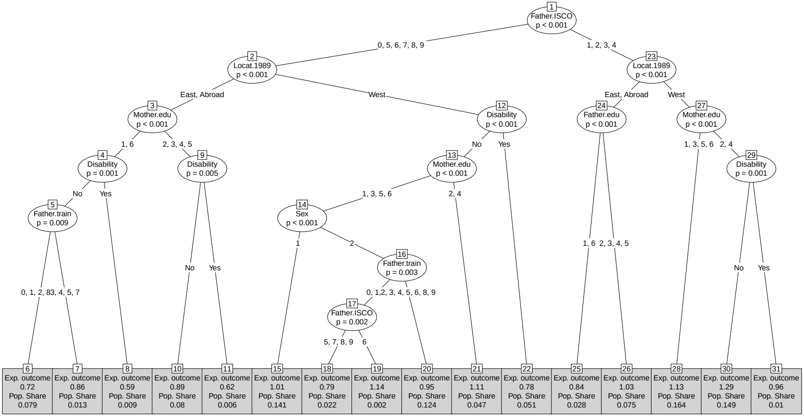
Opportunity tree in Germany: 1992



source: SOEP

Edu: 1=Sec., 2=Interm., 3=Tech., 4=Upper sec., 5=Other degr., 6=No degr., 7=Not atteded

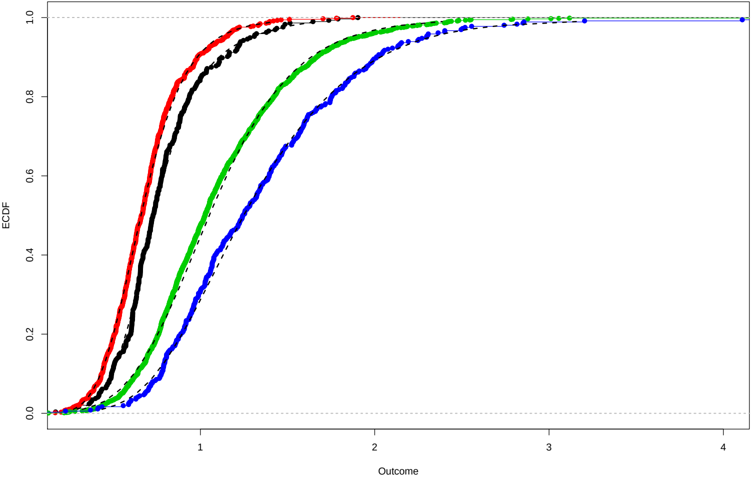
Opportunity tree in Germany: 2016



source: SOEP

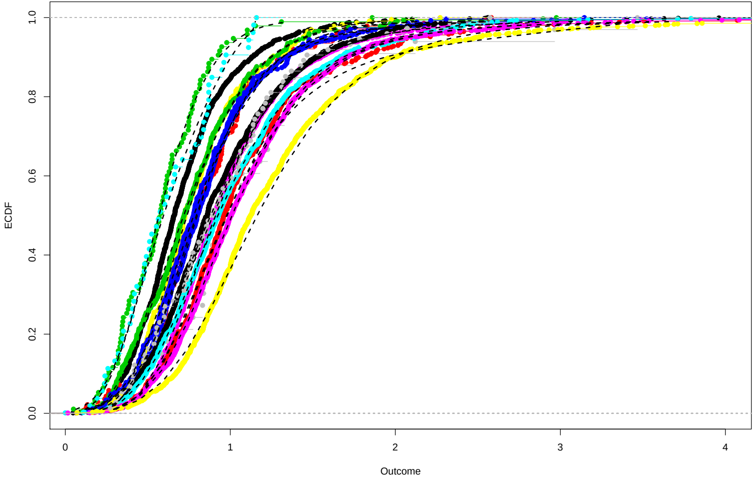
Edu: 1=Sec., 2=Interm., 3=Tech., 4=Upper sec., 5=Other degr., 6=No degr., 7=Not atteded

IOP in 1992



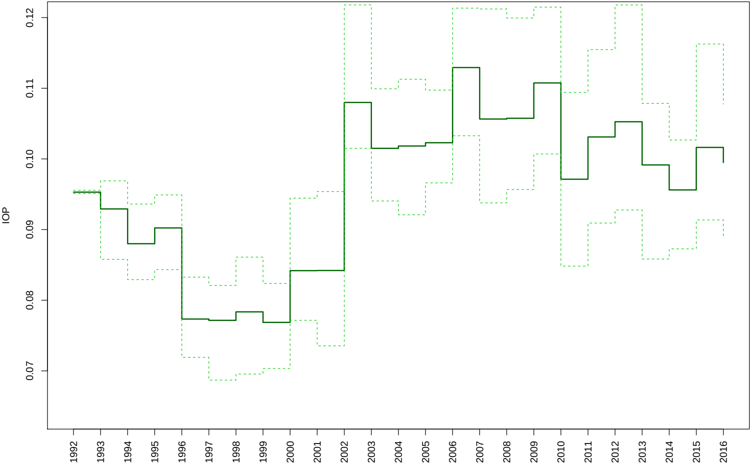
source: SOEP

IOP in 2016



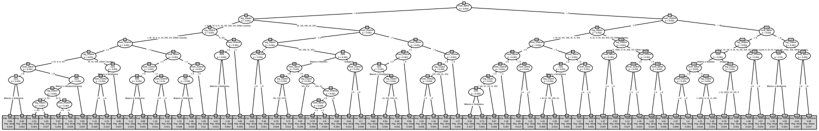
source: SOEP

IOP overtime in Germany



source: SOEP

Opportunity tree: Chile 2009

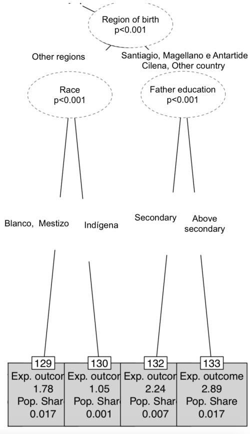


source: CASEN

Circumstances considered:

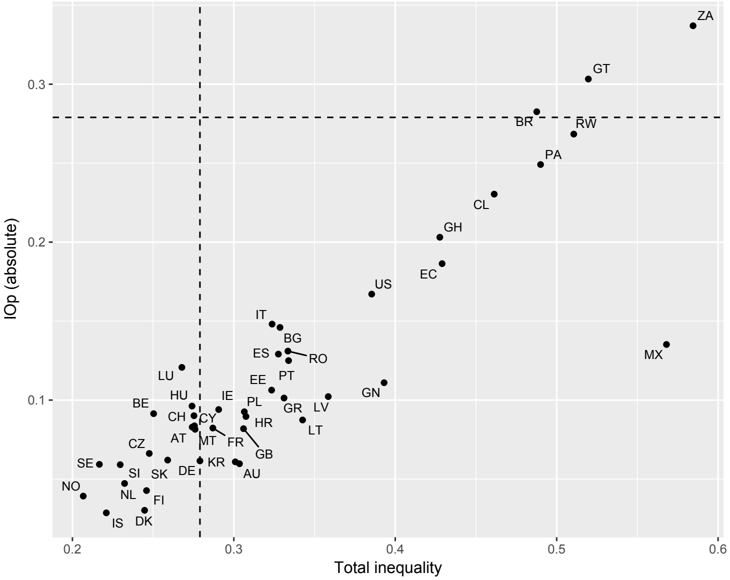
race (3), region of bird (16), mother education (7), father education (7).

Opportunity tree: Chile 2009



source: CASEN

Both parents have at least secondary education



source: EqualChances.org