The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach

Data Science & Social Research
University of Milano - Bicocca

Paolo Brunori
University of Florence & University of Bari

Guido Neidhöfer
ZEW

Literature

- a “third generation” paper on inequality of opportunity:
- first generation (theory): moral philosophers and welfare economists Rawls (1971), Dworkin (1981), Arneson (1989) and Cohen (1989), Roemer (1998);
- second generation (measurement): Lefranc et al. (2009), Checchi and Peragine (2010), Bourguignon et al. (2007), Ferreira and Gignoux (2011);
- third generation (econometric specification): Li Donni et al. (2015), Brunori et al. (2018)

Roemer's Model

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

- y_i: individual's i outcome;
- 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;

- no random component:
  same type and same tranche $\rightarrow$ same outcome;

- there is equality of opportunity if:

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

$\Rightarrow$ IOP = within-tranche inequality.

Effort identification

- effort: observable and not observable choices;
- Roemer's identification strategy, two assumptions:
  1. orthogonality: $e \perp\!\!\!\perp C$
  2. 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;
- measurement of degree of effort exerted;
- IOP = within-tranche Gini.

Roemerian types

- conditional inference trees (Hothorn et al., 2006);
- algorithm to predict a dependent variable partitioning a controls' space into non-overlapping regions;
- Brunori, Hufe, Mahler (2018): outperform standard methods to identify types in terms of out-of-sample MSE.

The algorithm

- 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

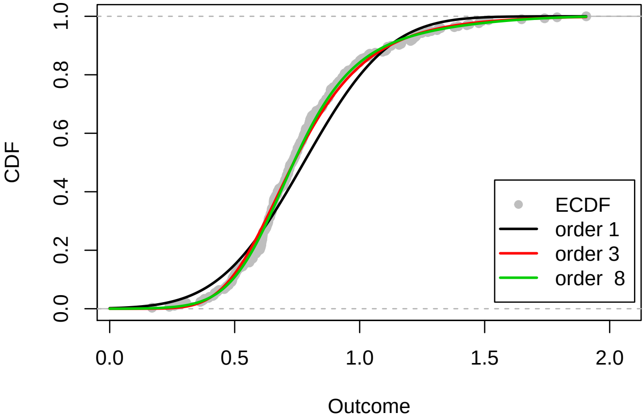
Effort

- standard approach: choose an arbitrary number of quantiles;
- limited comparability across studies;
- our approach: Bernstein polynomial approximation.

Bernstein polynomials

- introduced in 1912 by Sergei Bernstein
- today: mathematical basis for curves' approximation in computer graphics
- outperform competitors (kernel estimators) in approximating distribution functions (Leblanc, 2012)

ECDF approximation by Bernstein polynomials



Choice of the polynomial's degree

- out-of-sample log-likelihood to select the most appropriate order of the polynomial;
- out-of-sample log-likelihood is estimated by 5-fold cross validation;
- the polynomial is estimated with the *mlt* algorithm written by Hothorn (2018).

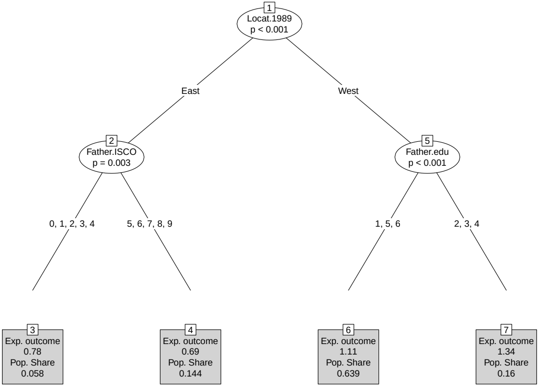
IOP in Germany

- SOEP (v33) including all subsamples apart from the refugee samples;
- adult individuals (30-60);
- $y =$ age-adjusted household equivalent disposable income;
- $IOP = Gini\left(\frac{y_i}{\mu_j}\right), \; \mu_j =$ tranche avg.

Missing information about circumstances

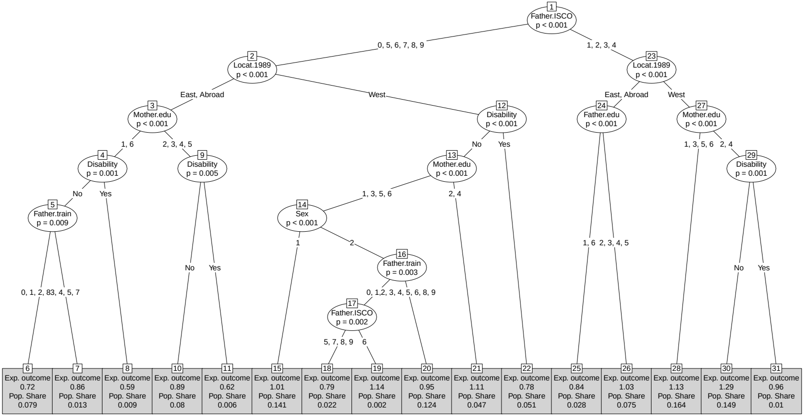
- SOEP provides comprehensive information about circumstances beyond individual control;
- waves considered 1992-2016;
- circumstances considered: migration background, location in 1989, mother's education, father's education, father's occupation, father's training, disability, siblings;

Opportunity tree in 1992



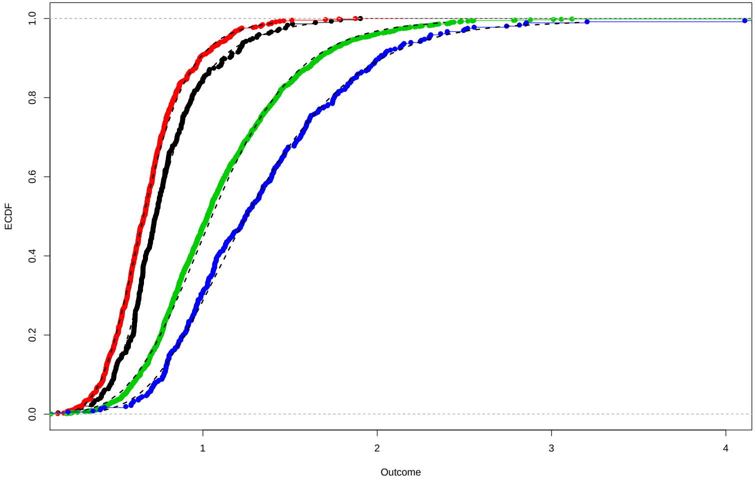
Edu: 1=Sec., 2=Interm., 3=Tech., 4=Upper sec., 5=Other degr., 6=No degr., 7=Not atteded

Opportunity tree in 2016

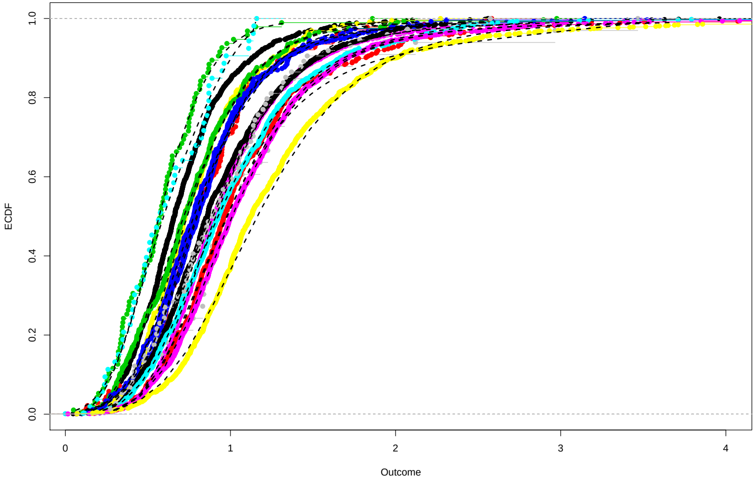


Edu: 1=Sec., 2=Interm., 3=Tech., 4=Upper sec., 5=Other degr., 6=No degr., 7=Not atteded

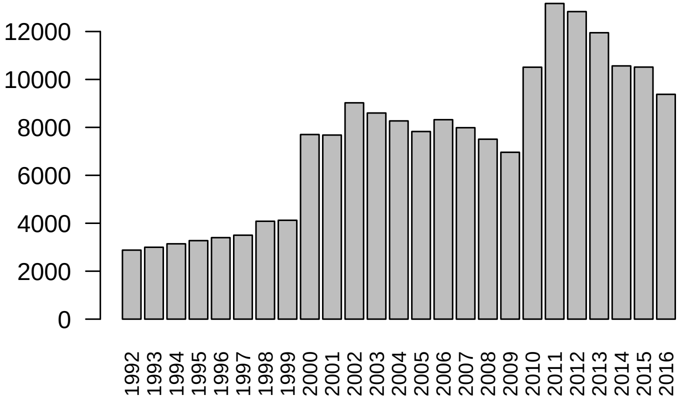
IOP in 1992



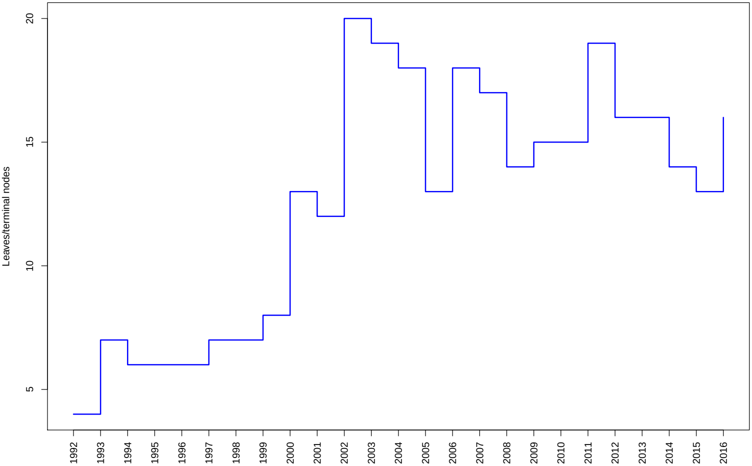
IOP in 2016



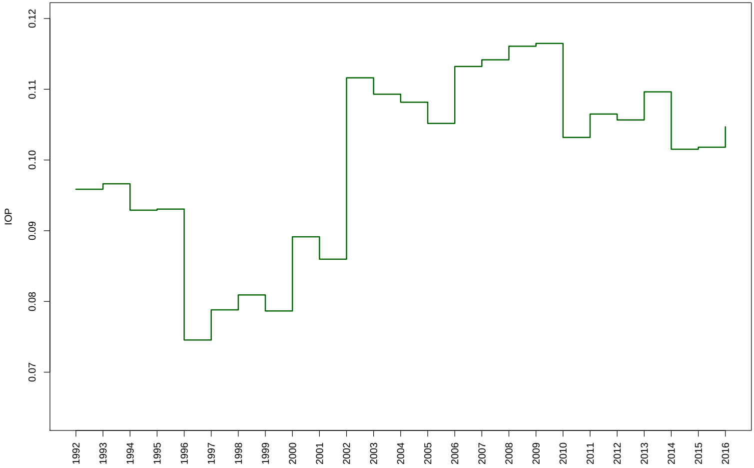
Sample size 1992-2016



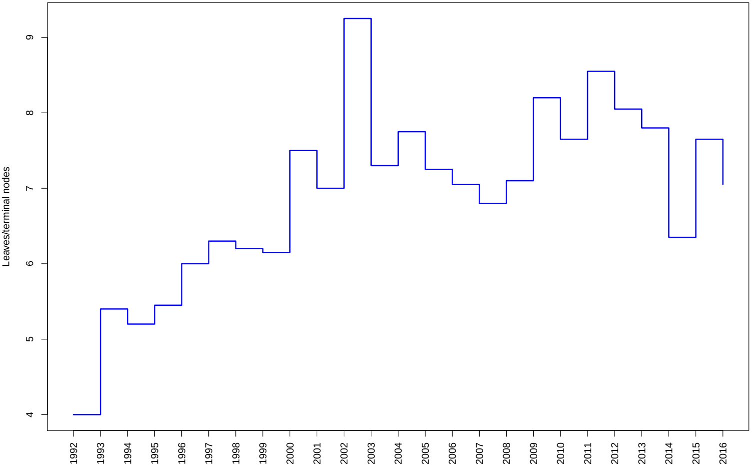
Types 1992-2016



IOP trend 1992-2016



Types 1992-2016 (same sample size)



IOP trend 1992-2016 (same sample size)

Summary

- we propose an approach to estimate IOP fully consistent to Roemer's theory;
- effort identification method maximizes comparability;
- since 1992 in Germany the opportunity structure has become more complex;
- IOP declined after reunification and surged in early '00s;
- $IOP_{1992} \approx IOP_{2016}$