“Fair inequality”
Some comments

Maximilian Kasy

4 December, 2024

Normative considerations

Methodological considerations

Different “cuts” of inequality

1. Statistics:
   Predictable vs. residual
2. Economic theory:
   Choice sets vs. preferences
3. Politically philosophy:
   Morally arbitrary vs. fair
4. Normative common sense:
   Opportunities vs. effort

Beware of conceptual slippage between these categories!

Predictions and residuals

A curious symmetry of two literatures:

- *Equality of opportunity*:
  - Predictable = unfair,
  - Residual = fair.
  - cf. Roemer, Chetty, the present project.
- *Discrimination*:
  - Predictiable = fair,
  - Residual = unfair.
  - cf. Oaxaca-Binder.

The limit of predictability

- Add more predictors
  $\Rightarrow$ the explained share of inequality increases.
- Arguably: Add enough predictors
  $\Rightarrow$ the explained share becomes arbitrarily large.
- So maybe all inequality is predictable and thus "unfair?"
- But maybe some "genuine" randomness remains?
  - Why should the luck of a coin toss be more fair
  - than the luck of parental background?

Objective answers to normative questions?

- Any aggregate assessment requires interpersonal tradeoffs.
  - These can be formalized as Pareto weights.
  - Formally: Separating hyperplane – rationalizing any point on the Pareto frontier.
- Economists would love to provide “objective” foundations:
  1. Empirically - risk aversion (e.g. Chetty)
  2. Empirically - policymaker preferences (e.g. Hendren)
  3. Empirically - from surveys (e.g. Stantcheva)
  4. Theoretically - from axioms (e.g. Atkinson)
  5. Theoretically - from thought experiments (e.g. Rawls)
  6. Institutionally - from votes

An alternative

- Do not provide aggregate assessments.
  - Admit that there are no objective foundations for such.
- Make explicit that policy choices entail
  - winners and losers,
  - distributional conflicts.
- Frame normative position as taking sides (e.g. with those who are worse off).
- $\approx$ Marxian position.

Normative considerations

Methodological considerations

Flexible models and overfitting

- Flexible models fitted without regularization
  - attribute too much inequality to predictors,
  - too little to residuals,
  - in finite samples.
- Negligible in linear regressions on 1 variable.
  e.g.: parental income.
- Very salient in regressions on rich fixed effects.
  e.g.: neighborhood FEs (Chetty et al.), family FEs (present project).

Overfitting in a random effects model

- Let Y_{ij} be $\log$ income
  of sibling j in family i,
  where each family has 2 siblings.

- Suppose

$$
\begin{aligned}
Y_{ij} &= \theta_i + \epsilon_{ij} \\
\mathrm{Cov}(\theta_i, \epsilon_{ij}) &= \mathrm{Cov}(\epsilon_{i1}, \epsilon_{i2}) = 0 \\
\mathrm{Var}(\theta_i) &= \sigma_\theta^2 \\
\mathrm{Var}(\epsilon_{ij}) &= \sigma_\epsilon^2
\end{aligned}
$$

Estimation using fixed effects

$$\widehat{\theta}_i = \tfrac{1}{2}(Y_{i1} + Y_{i2})$$

$$\mathrm{Var}(\widehat{\theta}_i) = \sigma_\theta^2 + \tfrac{1}{2}\sigma_\epsilon^2$$

$$\frac{\mathrm{Var}(\theta_i)}{\mathrm{Var}(Y_{ij})} = \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\epsilon^2}$$

$$\frac{\mathrm{Var}(\widehat{\theta}_i)}{\mathrm{Var}(Y_{ij})} = \frac{\sigma_\theta^2 + \frac{1}{2}\sigma_\epsilon^2}{\sigma_\theta^2 + \sigma_\epsilon^2}$$

- Systematically overestimate share of inequality explained by fixed effect. This holds for any flexible predictive model.
- Corrections are possible. In closed form in the present setting.
- Using cross-validation for any predictive model: What share of variation is **predictable out of sample**.

Thank you!