# Measuring Income Inequality of Opportunity

## Accounting for Dynamic Complementarity

Aman Desai
CUNY Graduate Center

GLO Job Market Session VI

December 7, 2024

# Introduction

# Motivation

**"The rise in inequality in the United States over the last three decades has reached the point that inequality in incomes is causing an unhealthy division in opportunities, and is a threat to our economic growth" (Alan Krueger, Center for American Progress, 12 January 2012)**

Rigorous treatment to measurement of inequality of opportunity (IOp hereafter) is vital from policy perspective.

# Main Results

- About 40-45% of inequality in individual's adult income is unfair.
- About 31–34% of total inequality in an individual's adult income could be attributed to unequal circumstances faced in their childhood up to age 5.

# Contribution

- Categorization of circumstance and effort factors using the age of majority at 18 years.
- Accounting for the role of dynamic complementarity by constructing age-based circumstance sets in measuring the inequality of opportunity.
- Using supervised machine learning to construct counterfactual distribution of adult incomes based on circumstances.

# Related Literature

# Inequality of Opportunity

- Seminal work by Roemer (1993). Success in adult life is considered to be influenced by
    - **Circumstance** : Beyond individual's control, hence for those the individual should not be held responsible and should be compensated for inequalities generated due to those.
    - **Effort** : Individual is in control of their effort and hence should be rewarded in the market economy.
- Several empirical approaches in last twenty years. (Bourguignon, Ferreira, and Menéndez 2007; Pistolesi 2009; Ferreira and Gignoux 2011; Niehues and Peichl 2014; Hufe et al. 2017). The estimated shares of IOp in outcome inequality varies largely from 10% to as high as 70%.
- Usage of machine learning algorithms to model IOp (Brunori, Hufe, and Mahler 2023).
- Fixed set of circumstances where measurement of IOp is dependent on researcher's value judgements.
- Lower bound measures of IOp.

# Inequality of Opportunity

Consider a population $\mathscr{N} = \{1, 2, \ldots, N\}$. Each individual in the population is characterized by a triple $(y, C, e)$ where $C \in \Omega^c$, $e \in \Omega^e$, and $y = g(C, e)$, with $g : \Omega^c \times \Omega^e \Longrightarrow \mathbb{R}$.

- An individual in the population is identified by a *type* and a *tranch*.
- A *type* consists of individuals with the same circumstances beyond their control.
- A *tranch* consists of individuals with the same effort.
- According to Roemer, equality of opportunity is achieved when inequality generated due to differential circumstances is eliminated (between *types*),that is $F(y|C) = F(y)$.
- Inequality of opportunity is measured by the extent to which this principle is violated, that is $F(y|C) \neq F(y)$.

# Technology of Skill Formation

Cunha and Heckman (2007) model technology for skill formation, conceptualized as a law of motion.

$$\omega_{i,t+1} = f(\omega_{i,t}, x_{i,t}, \omega_i^p, \epsilon_{i,t}) \tag{1}$$

- $f(.)$ is assumed to be twice continuously differentiable, increasing in all arguments, and concave in $x_{i,t}$.
- $x_{i,t}$ is the parental investment for the child $i$ at age $t$.
- $\omega_i^p$ is parental human capital at time $t$.
- $\epsilon_{i,t}$ is an iid unobserved individual component.

**Insight**

Investment in period $t + k$ and investment in any prior years $t$ are always complements as long as $\omega_{i,t+k}$ and $x_{i,t+k}$ are complements.

## Idea

If a child can not consent before the age of 18, all the measurable data on the child including her achievements, before she turns 18, can be thought of beyond her control and hence should be considered circumstances.

**Critical Stages in Childhood**

To incorporate the idea of dynamic complementarity, age cutoffs are chosen based on critical stages in childhood.

- 2 years : A child starts to speak.
- 5 years : A child enters K-12 system.
- 14 years : A child enters high school.
- 18 years : A child becomes an adult and can consent.

Four datasets are constructed according to four age cutoffs.
i.e. $C^2 \subseteq C^5 \subseteq C^{14} \subseteq C^{18} \subseteq \Omega^c$

# Data

# Analytical Sample

Ideally, one would have an entire biography of the individual's childhood experiences.

- Database : Panel Study of Income Dynamics (Main Interview, FRM[1], FIMS[2]).
- Cohorts : 1978-1983.
- Number of Individuals : 639 (SRC sample[3]), 1022 (Full sample[4]).
- Outcome Variables : Individual labor income at age 35 years, Average age adjusted labor income over four years[5].

The data in consideration is in wide format. Every observation reflects information on measurable factors for an individual over the first 18 years of their life.

[1] Family Relationship Matrix.

[2] Family Identification Mapping System.

[3] Survey Research Center sample is representative of the US population.

[4] Includes both SRC and SEO samples. The Survey of Economic Opportunity (SEO) sample includes a disproportionately higher number of poor households.

[5] Individual labor income excludes farm and unincorporated business income. All monetary variables including adult incomes are adjusted to 2018 dollars and individual cross sectional weights from 2013-2019 are used in the analyses.

# Circumstances

: Selected Circumstances

| Family/Demographic | Market/Monetary | Government/Community |
|---|---|---|
| Race, sex of the individual | Family income | Usage of foodstamps |
| Race of the family head, spouse | Childcare cost | Medicaid/Medicare usage in the family |
| Sex of the head | Homeownership | Help from family members, others, insiders |
| Education of the family head, spouse | Marginal tax rate on family income | Any outside dependents for head? |
| Occupation of the family head, spouse | Value of family home | Union membership of the family head, spouse |
| Number of children to father, mother | | Availability of a car |
| Marital status of mother when individual was born | | |
| Number of rooms in family home | | |
| State of residence of family | | |
| Birthweight | | |
| Birthcohort | | |

- Choice of circumstances is informed by theory.
- All these circumstances are measured across the first 18 years of a child's life. As I allow these circumstance sets to expand with critical stages in childhood, some circumstances may appear in multiple sets.

# Measurement

**Parametric Specification** (Bourguignon, Ferreira, and Menéndez 2007; Ferreira and Gignoux 2011; Niehues and Peichl 2014)

$$ln(y_i) = \alpha_0 + \sum_{l=1}^{L} (\alpha_l C_{i,l}^s) + u_i \tag{2}$$

where $y$ is the adult income, $C$ is the collection of factors that are categorized as circumstance belonging to a finite set $\Omega^c$, $s \in \{2, 5, 14, 18\}$ reflecting four different sets of circumstances based on chosen age cutoffs.

$$\hat{y}_i = exp\left[\alpha_0 + \sum_{l=1}^{L} (\hat{\alpha}_l C_{i,l}^s)\right] \tag{3}$$

The measurement of inequality of opportunity can be thought of as a two-step procedure: first, the actual distribution of $y_i$ is transformed into a counterfactual distribution (obtain $\hat{y}_i$) that reflects only and fully the unfair inequality in $y_i$, while all the fair inequality is removed. In the second step, a measure of inequality[6] is applied to $\hat{y}_i$. I use mean logarithmic deviation as an inequality measure[7].

$$Absolute\ IOp = I(\hat{y}_{EA}) \tag{4}$$

where $I(\hat{y}_{EA})$ is the ex-ante measure of inequality of opportunity.

$$Relative\ IOp = \frac{I(\hat{y}_{EA})}{I(y)} \tag{5}$$

The value of relative IOp ranges from 0 to 1. If all income differences are solely due to circumstances, relative IOp will be 1.

[6] any standard measure of inequality that satisfies anonymity, the principle of transfers, population replication, and scale invariance.

[7] $MLD(x) = ln(\bar{x}) - \overline{ln(x)}$.

Obtaining $\hat{y}_i$ is a prediction problem.

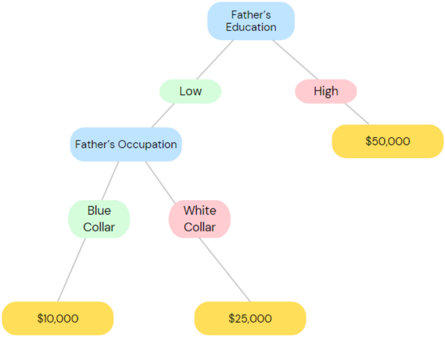


Figure 1: An Example of a Regression Tree

I use supervised machine learning method : Random Forest, an ensemble of decision trees.

- Better at dealing with high dimensional data, unlike OLS.
- Ensemble of regression trees reduces overfitting.

# Algorithm

I fit the models on training data, tune the hyper parameters on validation data, and then use the best model(with the lowest rmse) on the full data set. The algorithm runs as follows:

- Execute the random forest algorithm and use 5-fold cross validation for hyperparameter tuning. Select the models with hyperparameters that yield the lowest *rmse*. In each fold, the data is divided into $N_{train} = \frac{4}{5}N$ and $N_{validation} = \frac{1}{5}N$.
- Store the prediction functions $\hat{f}_{train}(\hat{\Omega}^c)$.
- Obtain final predictions using the full data $\hat{y} = \hat{f}_{train}(\hat{\Omega}^c_{fulldata})$.

# Results

| Characteristic | Full Sample N = 1,022 | SRC Sample N = 639 |
|---|---|---|
| **Individual labor income at age 35 (in natural logarithms)** | 10.5 (0.9) | 10.7 (0.9) |
| **Family income during the child's first year (in natural logarithms)** | 10.8 (0.8) | 11.0 (0.8) |
| **Sex** | | |
| Male | 474 (46%) | 311 (49%) |
| Female | 548 (54%) | 328 (51%) |
| **Race** | | |
| White | 559 (55%) | 554 (87%) |
| Black | 446 (44%) | 72 (11%) |
| Other | 17 (1.7%) | 13 (2.0%) |
| **Occupation of the head during the child's first year** | | |
| Other | 178 (17%) | 60 (9.4%) |
| Professional, Technical, and Kindred Workers | 168 (16%) | 157 (25%) |
| Managers and Administrators, except Farm | 72 (7.0%) | 62 (9.7%) |
| Sales Workers | 22 (2.2%) | 20 (3.1%) |
| Clerical and Kindred Workers | 51 (5.0%) | 28 (4.4%) |
| Craftsman and Kindred Workers | 219 (21%) | 151 (24%) |
| Operatives, except Transport | 128 (13%) | 72 (11%) |
| Transport Equipment Operatives | 45 (4.4%) | 23 (3.6%) |
| Laborers, except Farm | 41 (4.0%) | 24 (3.8%) |
| Farmers and Farm Managers | 13 (1.3%) | 12 (1.9%) |
| Farm Laborers and Farm Foremen | 5 (0.5%) | 2 (0.3%) |
| Service Workers, except Private Household | 79 (7.7%) | 28 (4.4%) |
| Private Household Workers | 1 (<0.1%) | |
| **Years of education of the head during the child's first year** | 12.4 (2.6) | 13.1 (2.5) |
| **Years of education of the spouse during the child's first year** | 10.4 (5.1) | 12.0 (4.1) |

[1] Mean (SD); n (%)

Figure 2: Descriptive Statistics for Selected Variables

- Baseline circumstances include individual's sex, race as well as the occupation of the family head, total family income, education of the head and the spouse (all measured during child's first year).
- Using OLS, relative IOp is estimated to be about 19-23%.

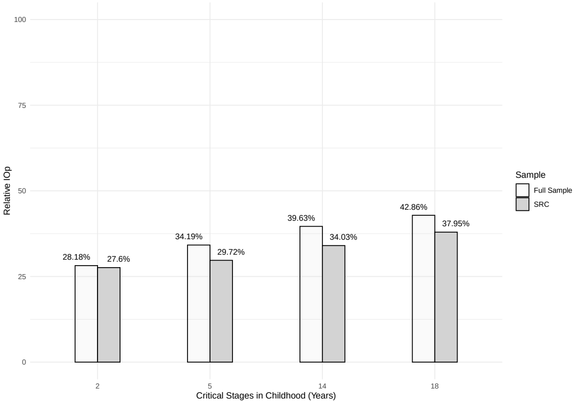


Figure 3: Relative IOp Estimates Across Age Cutoffs (Using Individual Labor Income at Age 35)

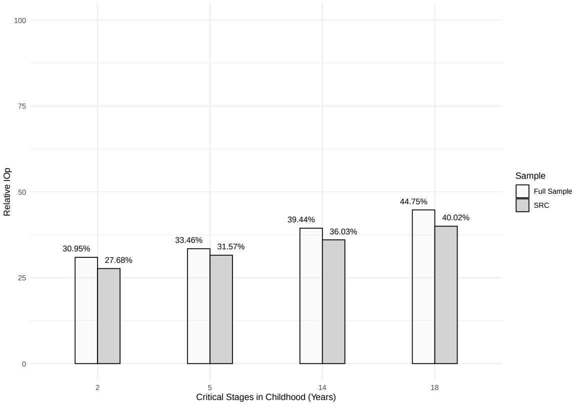


Figure 4: Relative IOp Estimates Across Age Cutoffs (Using Averaged Age-adjusted Incomes Across 2013-2019 Waves)

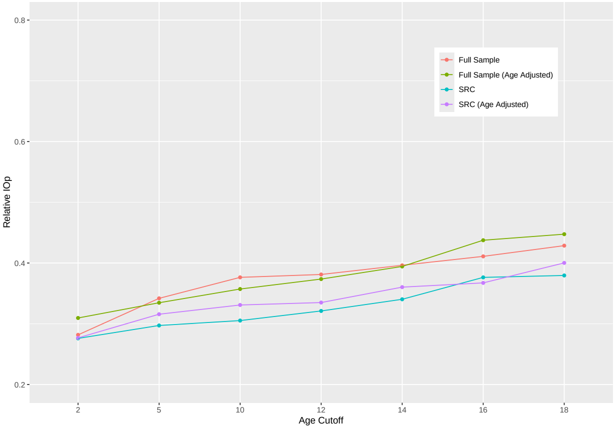


Figure 5: Relative IOp Profiles Across All Age Cutoffs

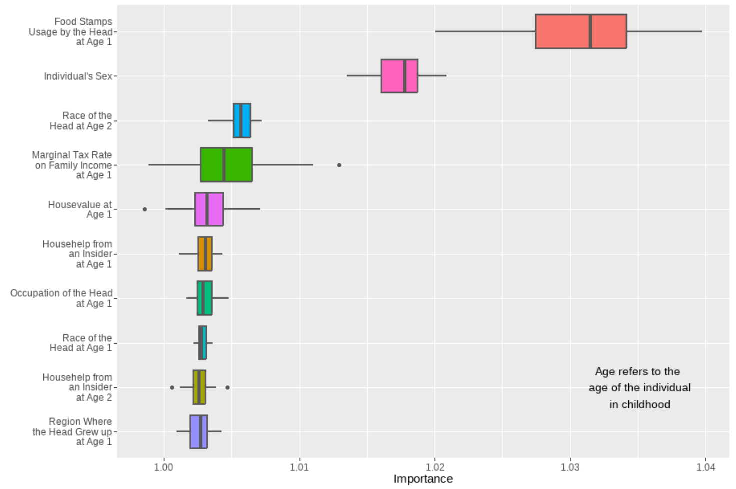


Figure 6: Variable Importance Scores for Circumstances up to Age 2 (Full Sample)

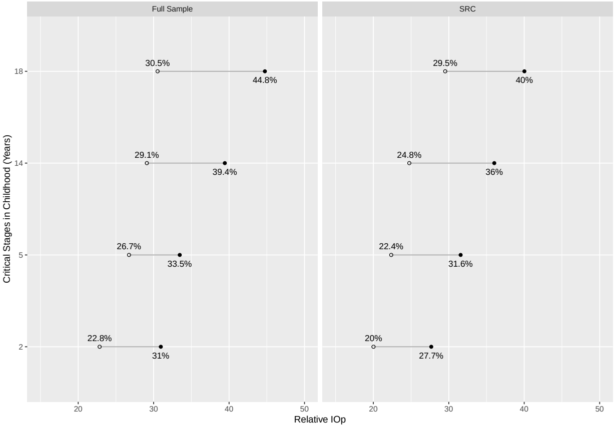


Figure 7: Lower and Upper Bounds of Relative IOp Estimates (Age-adjusted Average Incomes)

# Conclusion

- I evaluate inequality of opportunity through the lens of childhood circumstances.
- 31–34% of total income inequality can be attributed to unequal circumstances up to age 5, which is about 22-27% while using only the selected circumstances based on variable importance scores.
- I argue that these are upper-bound estimates, given the small number of circumstances that contribute most to unfair inequality.
- From a policy perspective, whether considering ex-post compensation or ex-ante investments (or both), I demonstrate the importance of accounting for dynamic complementarity in measurement rather than relying on a fixed set of circumstances.

# Thank You

# References

Bourguignon, François, Francisco H. G. Ferreira, and Marta Menéndez. 2007. "Inequality of Opportunity in Brazil." *Review of Income and Wealth* 53 (4): 585–618. https://doi.org/10.1111/j.1475-4991.2007.00247.x.

Brunori, Paolo, Paul Hufe, and Daniel Mahler. 2023. "The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Forests*." *The Scandinavian Journal of Economics* 125 (4): 900–932. https://doi.org/10.1111/sjoe.12530.

Cunha, Flavio, and James Heckman. 2007. "The Technology of Skill Formation." *American Economic Review* 97 (2): 31–47. https://doi.org/10.1257/aer.97.2.31.

Ferreira, Francisco H. G., and Jérémie Gignoux. 2011. "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America." *Review of Income and Wealth* 57 (4): 622–57. https://doi.org/10.1111/j.1475-4991.2011.00467.x.

Hufe, Paul, Andreas Peichl, John Roemer, and Martin Ungerer. 2017. "Inequality of Income Acquisition: The Role of Childhood Circumstances." *Soc Choice Welf* 49 (3-4): 499–544. https://doi.org/10.1007/s00355-017-1044-x.

Niehues, Judith, and Andreas Peichl. 2014. "Upper Bounds of Inequality of Opportunity: Theory and Evidence for Germany and the US." *Social Choice and Welfare* 43 (1): 73–99. https://www.jstor.org/stable/43662521.

Pistolesi, Nicolas. 2009. "Inequality of Opportunity in the Land of Opportunities, 1968–2001." *J Econ Inequal* 7 (4): 411–33. https://doi.org/10.1007/s10888-008-9099-7.

Roemer, John E. 1993. "A Pragmatic Theory of Responsibility for the Egalitarian