Inequality of Opportunity in Education, Occupation, Income, and Wealth

Sanni Kuikka (sanni.kuikka@sociology.su.se)*

Max Thaning (max.thaning@sofi.su.se)**

Background

One of the most foundational questions asked in social science is to what extent children's life chances are constrained by their background and upbringing. While the rich history of social mobility in sociology, and income mobility in economics attempt to answer parts of this question, the discrepancy in results between approaches remain an intriguing research puzzle. However, a more all-encompassing framework is the (in)equality of opportunity approach, developed by John (Roemer, 1998). This approach conceptualises inequality as stemming form both circumstances as well as efforts of individuals. Within this framework, equality is often defined as providing different groups equal opportunities to mitigate the effects of the circumstances outside one's control, while recognizing that individual efforts to achieve equal outcomes can differ, and thus should not be accounted for when levelling the playing field through policy-making. While distinguishing between these two areas – circumstances and efforts – has it's challenges, as well as political, philosophical, and conceptual debates (see e.g. Byskov et al., 2023; Grätz, 2023; Roemer & Trannoy, 2015), some important aspects of the life one is born into, that can be labelled as circumstances, are as follows: the home situation and family constellation, parental socioeconomic status, the neighbourhood one lives in, the school one goes to, sex, race, migration background, and (dis)ability.

Regardless of the way the distinction between circumstances and efforts is made, measuring the (in)equality of opportunity (IOp) relies heavily on collecting as much information on potential indicators of circumstances as possible, and analysing the joint effects of these. Recent methodological literature on IOp measurement has advanced the estimation discussion by relying on computational methods designed to understand complex data, namely, machine learning. Machine learning (ML) methods are particularly well-suited to understand large amounts of data, and the complex associations between several variables.

* Department of Sociology, Demography Unit, Stockholm University, Sweden; Institute for Analytical Sociology, Linköping University, Sweden

** Swedish Institute for Social Research, Stockholm University, Sweden

ML approaches for measuring IOp, more specifically regression trees and forests, have been proposed by Brunori et al. (2023), to account for estimate biases. Recent applications of integrating ML methods into estimating IOp include the case of income transmission in Germany (Brunori & Neidhöfer, 2021), income transmission in Latin America (Brunori, Ferreira, et al., 2023), household consumption in Sub-Saharan Africa (Brunori et al., 2019), wealth in regions of Mexico (Plassot et al., 2022), and income in a cross-national comparison in Europe (Carranza, 2022).

The current examples of the literature share some commonalities: they are often using cross-sectional survey data, and primarily analyzing income. However, it is crucial to account for the multiple domains of SES (i.e. multidimensional SES), both as the input, as well as the outcome, to understand the complex dynamics of socioeconomic transmissions (see e.g. Thaning, 2023). Additionally, relying solely on survey data poses certain challenges in terms of, for example, coverage, missingness, and variable availability. Hence, we propose to extend this literature with leveraging a multidimensional SES approach, and using longitudinal population register data of high definition and quality.

Using egalitarian Sweden as a test case, the aim of this paper is to obtain a range of estimates for IOp, that complement one another. This way we can estimate the lower and upper bounds of IOp in four SES outcomes, namely, education, occupation, income, and wealth. In order to obtain a range of estimates describing the lower and upper bounds of IOp, we integrate supervised machine leaning methods and dynamic individual and sibling fixed effects models. To expand the measurement of inherited circumstances beyond intergenerational transmission of SES, we include neighbourhood level, as well as school level predictors in the model. This approach in combination with high-quality population register data allows us to capture the wider conditions children are brought up in, and that are likely to influence their life courses and socioeconomic outcomes. Specific research question is as follows:

> What are the lower and upper bounds of inequality of opportunity, for each four SES outcomes, in the context of egalitarian Sweden?

Data

We use Swedish register data, containing longitudinal individual level information on full populations, that is linked intergenerationally. Our sample comprises of all individuals in the population registers born in the years 1967-1969, amounting to N ~300 000. Outcome variables are computed from longitudinal data, and defined as follows. Income is operationalized as lifetime disposable income from 20-years-old until the year 2018 (49-51-years-old), derived from the Income and Taxation register. Occupation is operationalized through occupational status scores, ISEI (see Ganzeboom & Treiman, 1996), derived from employer-reported occupational registers. Education information is derived from the educational registers, operationalized as highest education level obtained, expressed as years of education. Finally, wealth – an understudied SES dimension – is derived from the Wealth Register (only available between 1999-2007), and calculated as average wealth across years. As predictors, we use the above-mentioned SES indicators from both parents, as well as municipality ID and school ID, to represent the physical, social, and educational circumstances the children grow up in. We also derive sex and country of birth from the population registers.[1]

Analytical Approach

We study four SES outcomes: education, occupation, income, and wealth. We use a machine learning approach, namely the Random Forest algorithm, as well as a conventional OLS, sibling correlations, and individual fixed effects models. To evaluate the different models in a similar framework, we follow previous literature (e.g. Carranza, 2022), and make use of different inequality indices (e.g., Gini and MLD). Hence, our target quantity is calculated by dividing, for example, the Gini of the predicted values (for each respective model) by the Gini of the observed values. In this sense, models that predict the child outcome in superior ways, will result in target quantities that are closer to 1.

We present five ways of modeling IOp:

- m1: Linear OLS of observable circumstances
- m2: Non-parametric regression of observable circumstances
- m3: Random Forest of observable circumstances
- m4: Sibling fixed effects model based on individuals clustered in families
- m5: Individual fixed effects model based on time points clustered in individuals

Our contribution is thus three-fold. First, we estimate improved bounds and range of the inequality of opportunity, using a variety of methods. Compared to previous literature, we augment the conventional individual fixed effects model with longer time series, which reduce the problems related to measurement error (e.g. life-cycle bias and year-to-year volatility). Moreover, ML algorithms using more extensive and reliable data than in previous research result in more reliable predictions. Second, we extend the outcomes of study to occupation, education, and wealth, to provide a fuller picture of IOp compared to previous research only estimating one outcome. Finally, we contribute to the increasing demand of integrative modelling practices in the social sciences, combining conventional and novel methods to better measure and understand the social world (see e.g. Hofman et al., 2021).

[1] We have access to these data through an established project at Stockholm University Sociology Department, and have an ethical approval from the Swedish Ethical Review Authority (approval number 2023-00820-02).

Bibliography

Brunori, P., Ferreira, F. H. G., Neidhöfer, G., & UNU-WIDER. (2023). *Inequality of opportunity and intergenerational persistence in Latin America* (WIDER Working Paper 2023; WIDER Working Paper, Vol. 2023). UNU-WIDER. https://doi.org/10.35188/UNU-WIDER/2023/347-5

Brunori, P., Hufe, P., & Mahler, D. (2023). The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees and Forests. *The Scandinavian Journal of Economics*. https://doi.org/10.1111/sjoe.12530

Brunori, P., & Neidhöfer, G. (2021). The Evolution of Inequality of Opportunity in Germany: A Machine Learning Approach. *Review of Income and Wealth*, *67*(4), 900–927. https://doi.org/10.1111/roiw.12502

Brunori, P., Palmisano, F., & Peragine, V. (2019). Inequality of opportunity in sub-Saharan Africa. *Applied Economics*, *51*(60), 6428–6458. https://doi.org/10.1080/00036846.2019.1619018

Byskov, M. F., Kramm, M., & Östlund, S. (2023). Capabilities as Substantive Opportunities and the Robustness of Conversion Factors. In M. Sardoč (Ed.), *Handbook of Equality of Opportunity* (pp. 1–25). Springer International Publishing. https://doi.org/10.1007/978-3-319-52269-2_14-1

Carranza, R. (2022). Upper and Lower Bound Estimates of Inequality of Opportunity: A Cross-National Comparison for Europe. *Review of Income and Wealth*. https://doi.org/10.1111/roiw.12622

Ganzeboom, H. B. G., & Treiman, D. J. (1996). Internationally Comparable Measures of Occupational Status for the 1988 International Standard Classification of Occupations. *Social Science Research*, *25*(3), 201–239. https://doi.org/10.1006/ssre.1996.0010

Grätz, M. (2023). Empirical Approaches to Measuring Equality of Opportunity. In M. Sardoč (Ed.), *Handbook of Equality of Opportunity* (pp. 1–9). Springer International Publishing. https://doi.org/10.1007/978-3-319-52269-2_83-1

Hofman, J. M., Watts, D. J., Athey, S., Garip, F., Griffiths, T. L., Kleinberg, J., Margetts, H., Mullainathan, S., Salganik, M. J., Vazire, S., Vespignani, A., & Yarkoni, T. (2021). Integrating explanation and prediction in computational social science. *Nature*, *595*(7866), 181–188. https://doi.org/10.1038/s41586-021-03659-0

Plassot, T., Soloaga, I., & Torres, P. (2022). Inequality of Opportunity in Mexico and its Regions: A Data-Driven Approach. *The Journal of Development Studies*, *58*(9), 1857–1873. https://doi.org/10.1080/00220388.2022.2055465

Roemer, J. E. (1998). *Equality of Opportunity*. Harvard University Press.

Roemer, J. E., & Trannoy, A. (2015). Equality of Opportunity. In A. B. Atkinson & F. Bourguignon (Eds.), *Handbook of Income Distribution* (pp. 217–300). Elsevier B.V. https://doi.org/10.1016/B978-0-444-59428-0.00005-9

Thaning, M. (2023). *Social Mobilities—Multidimensionality, Operationalization, and Subgroup Heterogeneity*. Department of Sociology, Stockholm University.