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

Paolo Brunori[*], Guido J. Neidhöfer[†]

February 14, 2018

We adopt a machine learning approach to estimate inequality of opportunity in Germany between 1990 and 2016 using the Socio-Economic Panel.

Inequality of opportunity is measured following Roemer (1998). Roemer's approach distinguishes between two components of inequality: inequality due to effort and inequality of opportunity. Inequality of opportunity is hereby defined as differences in outcomes due to circumstances beyond individual control, like place of birth, race or socioeconomic status of parents. According to Roemer these two components can be separated following a two-step procedure. First, identifying socioeconomic types, i.e. sets of individuals who share the same combination of circumstances. Second, measuring the degree of effort exerted by each individual within the specific types. This procedure allows to estimate inequality of opportunity: i.e. inequality between individuals exerting the same level of effort but characterized by different circumstances beyond their individual control.

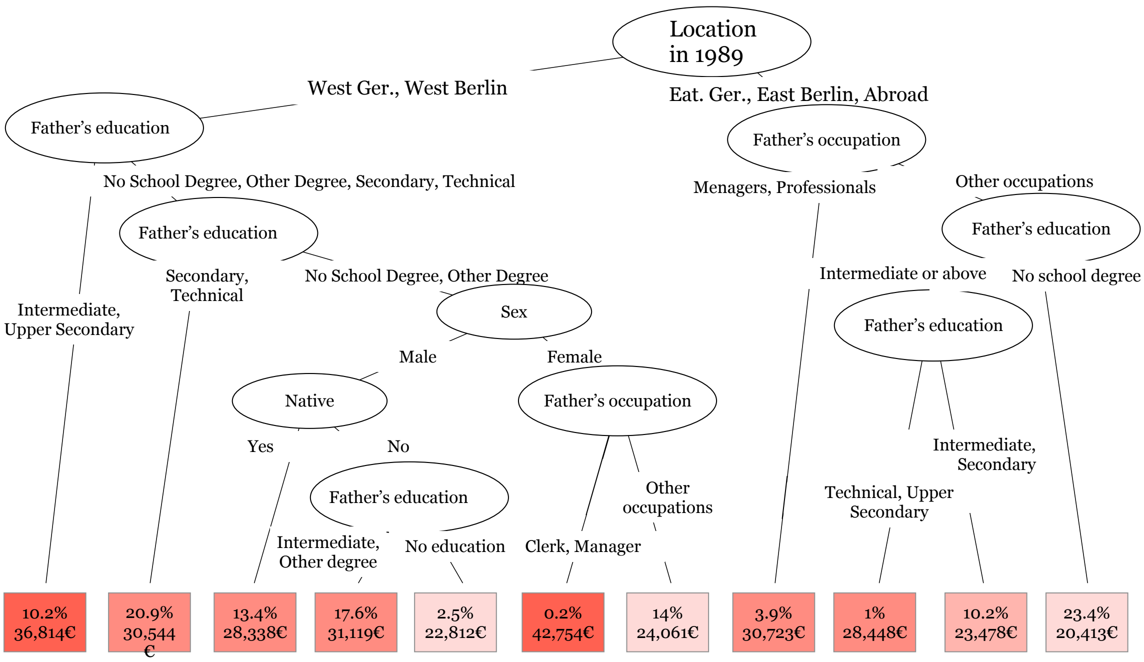
We implement the first step using conditional inference regression trees, a machine learning algorithm based on regression trees (Hothorn et al., 2016). This method has been shown to produce reliable identification of Romerian types in a large sample of European countries (Brunori et al., 2018). A second advantage of using conditional inference trees is that they can be displayed graphically and are easily interpreted (below the German opportunity tree for 2016), allowing an intuitive representation of the evolution of inequality of opportunity in Germany during the three decades after reunification (Figure 1 shows the opportunity tree for 2016).

To identify the degree of effort that individuals exert, we follow Roemer (1998) and extend the procedure proposed by Brunori et al. (2018). We estimate the type-specific outcome distribution of all types, assuming that all individuals in one specific quantile of their within-type income distribution have exerted the same degree of effort. Furthermore, we apply cross-validation techniques to reduce the possible bias deriving from small subsample sizes within types, on the one hand, or low number of types, on the other.

[*]University of Florence, Dipartimento di Scienze per l'Economia e l'Impresa, Via delle Pandette 32 - 50127 Firenze, Italy, paolo.brunori@unifi.it.

[†]DIW Berlin German Institute for Economic Research, gneidhoefer@diw.de.

Figure 1: The German opportunity tree in 2016



Source: SOEP, 2016 own estimates. Note: the figure shows the conditional inference tree for household disposable equivalent income as function of circumstances beyond individual control. Terminal nodes are Romerian types with a level of statistical significance above 99.9%.

References

Brunori, P., Hufe, P., and Mahler, D. G. (2018). The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees. *ECINEQ Working Paper Series*, 2018-455.

Hothorn, T., Hornik, K., and Zeileis, A. (2016). *ctree: Conditional inference trees, Vignette R package partykit version 1.1-1.*

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