the review of income and wealth

Review of Income and Wealth
Series 67, Number 4, December 2021
DOI: 10.1111/roiw.12502

THE EVOLUTION OF INEQUALITY OF OPPORTUNITY IN GERMANY: A MACHINE LEARNING APPROACH

BY PAOLO BRUNORI*

University of Florence and University of Bari

Leibniz-Zentrum für Europäische Wirtschaftsforschung

AND

GUIDO NEIDHÖFER

ZEW - Leibniz Centre for European Economic Research

We show that measures of inequality of opportunity (IOP) fully consistent with the IOP theory of Roemer (1998) can be straightforwardly estimated by adopting a machine learning approach, and apply our method to analyze the development of IOP in Germany during the past three decades. Hereby, we take advantage of information contained in 25 waves of the Socio-Economic Panel. Our analysis shows that in Germany IOP declined immediately after reunification, increased in the first decade of the century, and slightly declined again after 2010. Over the entire period, at the top of the distribution we always find individuals who resided in West Germany before the fall of the Berlin Wall, whose fathers had a high occupational position, and whose mothers had a high educational degree. East German residents in 1989, with low-educated parents, persistently qualify at the bottom.

JEL Codes: D63, D30, D31

Keywords: Germany, inequality, opportunity, SOEP

1. Introduction

The ideal of equality of opportunity has fascinated mankind for centuries. Its popularity among people from both sides of the political spectrum probably derives from the fact that it encompasses and balances two aspects: equality of outcomes and freedom of choice. In addition, while the normative evaluation of outcome inequality is controversial, nobody would argue against equality of opportunity as an important goal. At the same time, the political rhetoric demanding it might be sufficiently vague that it allows for different interpretations.

Note: We are grateful to Torsten Hothorn, Paul Hufe, Daniel G. Mahler, Julia Heigle, Eduard Brüll, Robin Jessen, Friedhelm Pfeiffer, and John Roemer for useful suggestions. Furthermore, we would like to thank the participants of presentations given in Berlin, Mannheim, Milan, Paris, and Trento for their comments. Guido Neidhöfer gratefully acknowledges funding from the state Baden-Württemberg within the SEEK program. Any errors remain our own.

*Correspondence to: Paolo Brunori, Dipartimento di Scienze per l'Economia e l'Impresa, University of Florence and University of Bari, Via delle Pandette 32 - 50127 Firenze, Italy (paolo.brunori@unifi.it).

[Correction added on 26 March 2021, after first online publication: An affiliation to Leibniz-Zentrum für Europäische Wirtschaftsforschung has been added for Paolo Brunori.]

For a long time, moral philosophers and welfare economists defined and conceptualized the notion of equality of opportunity as well as its implications from a normative point of view. Rawls (1971) proposed a theory of social justice in which redistribution of outcomes and social roles was somehow limited by the need to take into consideration individuals' responsibility. Dworkin (1981) went a step further, focusing on the distinction between preferences and resources. From his perspective, inequality in final conditions is morally objectionable, and calls for redistribution in the case that these differences arise from unequal resources, though not when they arise from preferences and choices. Among economists, the most influential formalization of the principle of equal opportunity is because of Roemer (1998). Roemer's definition of inequality of opportunity (IOP) comprises the interplay between circumstances that individuals are exposed to and the degree of effort they exert; circumstances are exogenous factors outside individual control, whereas effort indicates the result of choices for which the society wishes to hold individuals responsible.

Roemer's theory of equal opportunity has triggered a lively empirical literature. Metrics were proposed to measure IOP based on the distance of a given distribution of individual outcomes from equal opportunity (among others, Lefranc *et al.*, 2009; Checchi and Peragine, 2010; Almås *et al.*, 2011). A popular approach is, for instance, the regression-based method proposed by Ferreira and Gignoux (2011) to quantify the share of total inequality because of opportunity. Two recurrent issues of the empirical literature are as follows: *i*. the need to identify the set of circumstances beyond individual control, and *ii*. the need to assume how these circumstances interact with effort in determining individual outcomes. Both choices have been shown to crucially affect the estimated level of IOP (Brunori *et al.*, 2019).

Recent contributions have proposed approaches to improve the empirical specification of the underlying models, finding consistent econometric methods to identify the relevant circumstances, and eventually estimate IOP. Within this literature, Li Donni *et al.* (2015) and Brunori *et al.* (2018) propose data-driven approaches to identify Roemerian types (i.e., sets of individuals characterized by identical circumstances). Still, besides the identification of circumstances, in the vast majority of empirical contributions so far IOP is estimated without considering the role of effort, which is key in Roemer's theory (Ramos and Van de gaer, 2020).

In this study, we propose a method that builds on a data-driven approach and exploits two machine learning algorithms (namely regression trees and polynomial approximation) to estimate IOP consistent with Roemer's original theory. In a first step, we follow Brunori *et al.* (2018) to identify types and estimate opportunity trees. Then, we develop a new approach to estimate the degree of effort by polynomial approximation of the conditional distribution of household income for each type. This enables us to precisely estimate the relationship between effort and outcome, even for types with a small sample size.

We apply this novel approach to estimate the evolution of IOP in Germany from shortly after the fall of the Berlin Wall to the present. Germany is an interesting case study for our analysis because of the societal changes that the country underwent during the past 30 years. Our application using the Socio-Economic Panel (SOEP) shows that the opportunity structure of the German society is much more complex today than it was back in the 1990s. The number of types identifiable in SOEP has increased substantially over time, a pattern that is not fully explained just by changes in survey characteristics. Despite a substantial change in the structure of opportunities over time, having being located in the Eastern or Western part of Germany in 1989 is, more than two decades after reunification, constantly a significant circumstance defining the subdivision in types over the course of time.

Our analysis uncovers another interesting peculiarity. Although usually societies characterized by a large number of types tend to have much higher levels of IOP, this is not the case for Germany over time. Controlling for the sample size, albeit the number of types increases by roughly 75 percent, the level of IOP in 2016 is just around 7 percent higher than in 1992. Generally, during this period Germany experienced first a slow decrease in IOP after reunification and then a sudden rise. Thenceforth, IOP stayed at this relatively high level. A further increase is observed in coincidence with rising income inequality and the implementation of the *Hartz-reforms*, a set of substantial changes to the German labor market and welfare benefits system that had persistent repercussions for German society. A slight decrease is recorded from 2010 onward.

2. Inequality of Opportunity

Roemer (1998) represents the seminal contribution for the empirical literature on IOP. In his book, Roemer did not explicitly write down a definition of IOP. Rather, his theory proposes a criterion to select the redistributive policy that would equalize opportunity in a society. This theory has been translated into more than one definition of IOP.

Roemer's theory distinguishes between two categories of factors that determine individual outcomes: factors over which individuals have control, which he calls *effort*, and factors for which individuals cannot be held responsible, which he calls *circumstances*. He defines equal opportunity in the distribution of a certain desirable outcome as the scenario in which individuals are compensated for the difference in their circumstances, insofar as those differences affect the advantage they attain. To realize equal opportunity, Roemer proposes a partition of the population into *types*. A type is a set of individuals characterized by exactly the same circumstances (gender, race, socioeconomic background, and so on). When exerting effort, individuals in the same type have the same ability to transform resources into outcomes. Therefore, an equal opportunity policy prescribes to ignore within-type variability in outcomes, which by definition is due to individual effort, and requires the removal of any between-type inequality.

Roemer's definition of equal opportunity can be formalized in a simple model. In a population of 1, …, N individuals, individual i obtains an outcome of interest, y_i, as the result of two sets of traits: a set of circumstances beyond her control, $\mathbf{C}_i$, and a responsibility variable, e_i, called effort:

$$y_i = g(\mathbf{C}_i, e_i), \quad \forall i = 1, \ldots N.$$

$\mathbf{C}_i$ contains $J > 0$ circumstances, and each circumstance, $C^j \in \mathbf{C}$, is characterized by a total of x^j possible realizations. All possible combinations of realizations taken one at a time from $\mathbf{C}$ define a partition of the population into types. This partition is made of a maximum of $K = \Pi_{j=1}^{J} x^j$ nonempty subsets, where every individual is included in one and only one of the subsets. Note that this simple model does not introduce any random component or uncertainty.

Equality of opportunity is realized when individuals exerting same effort obtain the same valuable outcome, independently from the type they belong to. To measure to what extent this principle is violated, one must compare the outcome of individuals belonging to different types but exerting the same effort. Because effort is typically unobservable, Roemer proposes a method for identifying effort. His method is based on three assumptions. First, we fully observe relevant circumstances; that is, we correctly assign individuals to types. Second, the outcome is assumed to be monotonically increasing in effort: in every type, higher effort implies higher outcome:

$$(1) \qquad y^k(e_i) \geq y^k(e_j) \Longleftrightarrow e_i^k \geq e_j^k, \quad \forall k = 1, \ldots, K, \quad \forall e_i, e_j \in \mathbb{R}.$$

Third, the degree of effort exerted is by definition a variable orthogonal to circumstances. In Roemer's view, if individuals belonging to different types face different incentives and constraints in exerting effort, this is to be considered a characteristic of the type and therefore included among circumstances beyond individual control.

For example, a student with well-educated parents may find it much easier to spend hours sitting at her desk, whereas a student growing up in a less favorable environment may find it harder to study. Roemer believes that the distribution of effort is, indeed, a characteristic of the type: *thus, in comparing efforts of individuals in different types, we should somehow adjust for the fact that those efforts are drawn from distributions which are different, a difference for which individuals should not be held responsible*. (Roemer, 2002, p. 458).

Therefore, Roemer distinguishes between the "level of effort" and the "degree of effort" exerted by an individual. The latter is the morally relevant variable of effort and is identified with the quantile of the effort distribution for the type to which the individual belongs. We denote with $G^k(e)$ the distribution of effort within type k and with $\pi \in [0, 1]$ its quantiles.

If effort is not observable, but the outcome is monotonically increasing in e, Roemer suggests to identify the degree of effort exerted by a given individual with the quantile of the type-specific outcome distribution she sits at: $y^k(G^k(e)) = y^k(\pi)$. This definition of effort is insensitive to differences in the absolute level of effort exerted that, in Roemer's view, are due to circumstances beyond individual control, and it permits the comparison of effort exerted by individuals in different types.

Then the requirement of same outcome for individuals exerting same effort can be rewritten in terms of type-specific outcome distributions:

$$(2) \qquad y^k(\pi) = y^l(\pi) \Longleftrightarrow F^k(y) = F^l(y), \quad \forall \pi \in [0,1]; \quad k, l = 1, \ldots, K,$$

where $F^k(y)$ is the type-specific cumulative distribution of outcome in type k.

A measure of IOP quantifies to what extent this equality of opportunity principle is violated. This is done measuring the variability of the outcome distribution within individuals exerting same effort (Lefranc *et al.*, 2009; Checchi and Peragine, 2010; Almås *et al.*, 2011; Ferreira and Gignoux, 2011). By construction, these measures take value zero when equation (2) is satisfied and all individuals exerting the same effort obtain the same outcome, and increase with larger differences in outcomes obtained by individuals exerting the same degree of effort.

For example, the ex-post measure of IOP proposed by Checchi and Peragine (2010) evaluates inequality in the standardized distribution $\tilde{Y}_{EP}$ obtained replacing individual outcome with:

$$(3) \qquad \tilde{y}_i^k(\pi) = y_i^k(\pi)\frac{\mu}{\mu^{\pi}}, \quad \forall i = 1, \ldots, N; \quad k = 1, \ldots, K; \quad \forall \pi \in [0, 1],$$

where $y_i^k(\pi)$ is the outcome of individual i belonging to type k and sitting at quantile π of the type-specific effort distribution, μ^{π} is the average outcome of individuals sitting at quantile π across all types, and μ is the population mean outcome. Note that in the standardized distribution, the average value for individuals sitting at all quantiles is the same; that is, between-quantile inequality has been removed. On the contrary, the within-quantile relative distance of outcome is preserved. IOP, IOP_{EP}, is then inequality in the standardized distribution:

$$(4) \qquad IOP_{EP} = I(\tilde{Y}_{EP}),$$

where I is any inequality measure satisfying the typical properties, including scale invariance.[1]

Ex-post measures of IOP are not frequently implemented in empirical analysis. The majority of applied studies focus on a second, less demanding, definition of equal opportunity. The *ex-ante* equality of opportunity is a "weak equality of opportunity" criterion that allows some inequality within groups of individuals exerting the same effort but requires that mean advantage levels should be the same across types (Ferreira and Gignoux, 2011).

The ex-ante measure of IOP first proposed by Van de gaer (1993) is a measure based on this weaker definition. The approach interprets the type-specific outcome distribution as the opportunity set of individuals belonging to each type. The (utilitarian) value of the opportunity set of each type is the mean outcome of the type. Therefore, IOP in this case is simply between-type inequality, and the counterfactual distribution $\tilde{Y}_{EA}$ is obtained replacing individual outcome with:

$$(5) \qquad \tilde{y}_i^k(\pi) = \mu^k, \quad \forall i = 1, \ldots, N; \quad \forall k = 1, \ldots, K; \quad \forall \pi \in [0, 1],$$

where μ^k is the mean outcome of type k:

$$(6) \qquad IOP_{EA} = I(\tilde{Y}_{EA}).$$

[1]Researchers interested in measuring IOP with a translation invariant inequality measure should replace equation (4) with: $\tilde{y}_i^k(\pi) = y_i^k(\pi) + \mu - \mu^j$.

Adopting the ex-ante approach simplifies the measurement of IOP, which becomes equivalent to a measure of between-group inequality. Furthermore, IOP_{EA} is by far the most popular measure of IOP.[2] However, this approach implies a loss of consistency with the principle of compensation, which, in its original formulation, is the fundamental ethical principle of Roemer's theory of equal opportunity, stating that individuals exerting same effort should obtain same outcome (see Fleurbaey and Peragine, 2013, for a discussion of this incompatibility).[3]

3. Machine Learning Estimation of IOP

The estimation of IOP_{EP} is based on two fundamental tasks: the identification of Roemerian types and the measurement of the degree of effort exerted. We adopt a machine learning approach to accomplish both. The partition into types is obtained estimating regression trees; the degree of effort is measured estimating the type-specific outcome distribution by a polynomial approximation.

3.1. *Identification of Types*

The first step to estimate both equation (4) and equation (6) is the identification of circumstances beyond individual control that define types. Note that other methods, based on a parametric estimation of the function $g()$, have been proposed (Bourguignon *et al.*, 2007; Ferreira and Gignoux, 2011). In what follows, we limit the discussion to methods that do not impose a functional form on the data-generating process and explicitly identify types. The selection of circumstances is a key aspect of any empirical analysis of IOP because estimates have been shown to be sensitive to the number of types considered (Rodríguez, 2008; Ferreira and Gignoux, 2011; Brunori *et al.*, 2019).

In principle, one should include all variables beyond individual control that can affect the outcome. However, this is not a realistic option. First, surveys typically contain only a subset of all the exogenous determinants of individual outcomes. Second, even when a rich data set is available, the sample size constrains the number of circumstances that can be considered if one must reliably estimate the counterfactual distributions.

The so-called non-parametric approach proposed by Checchi and Peragine (2010) attempts to exactly implement Roemer's definition of types: the partition in types is obtained by interacting all circumstances. This typically results in a partition with a large number of types. However, many of these types are sparsely populated making it impossible to estimate with accuracy the type-specific cumulative distribution of the outcome.

Empirical exercises generally address this issue by limiting the number of circumstances used to define types. In addition, the categories that describe circumstances are recoded reducing their variability. For instance, districts of birth

[2]The project Equalchances.org estimated comparable estimates of IOP_{EA} for 51 countries.

[3]Quoting Ramos and Van de gaer (2020): *"Most of the empirical literature continues to treat [ex-ante and ex-post approaches] as interchangeable, by motivating their concern with inequality of opportunity from ex-post intuitions and using ex-ante measures of inequality of opportunity."* p. 2.

are aggregated into macro-region; parental occupations become a binary variable for white- or blue-collar workers; ethnicity becomes a dummy for minorities. The resulting types have large sample sizes but are based on arbitrary choices. These *ad-hoc* methods to identify types severely undermine the interpretability and comparability of estimates of IOP (Brunori *et al.*, 2019).

Two papers have proposed data-driven criteria to identify Roemerian types. Li Donni et al. (2015) specify that types are inherently unobservable and suggest grouping individuals in types using latent class models. Latent class models assign individuals to types based on observed circumstances, interpreted as observable manifestations of the underlying latent types. Type membership is determined in the attempt to maximize local independence. Local independence means that, conditional on class membership, observed items are conditionally independent from each other. Once latent types are identified, the researcher can move to the second step of the analysis, which consists of identification of the effort exerted.

Latent types are an appealing theoretical construct. However, their implementation has two problematic aspects. First, in latent class models the number of classes is exogenously given. Li Donni *et al.* (2015) suggest selecting the number of latent types guided by the Bayesian information criterion (BIC). BIC evaluates the likelihood of the model introducing a penalty term for the number of parameters estimated. The BIC selects the most appropriate model balancing between choosing a model able to closely fit the data in the sample and choosing a model with the lowest possible number of parameters. When estimating models based on an increasing number of latent types, BIC will first rise, indicating that additional classes substantially improve model fit, and then, when the effect of the penalty dominates, BIC will start to decline. According to Li Donni *et al.* (2015), the most appropriate partition is obtained by choosing the number of classes that produces the highest BIC.

A perfect fit is obtained when the distribution of manifest variables is orthogonal to classes, i.e., when local independence is fully satisfied. Therefore, the BIC of a latent class model evaluates the capacity of the model to explain the correlation of manifest variables in the sample. However, when estimating IOP, the aim is not to explain covariance of circumstances, but to identify the partition in types that best explains the outcome variability. A criterion such as the BIC may not be the most appropriate for selecting the number of latent types. This is a specific case of the more general problem of using latent class membership as predictor for a distal dependent variable. As discussed by Lanza *et al.* (2013), such an approach is likely to produce attenuated estimates of the effect of the latent class membership on the outcome. Such downward bias can be attenuated adopting alternative criteria to select the number of latent types Brunori *et al.* (2020).

Another issue concerns the choice of observable circumstances considered in the latent class model. As discussed earlier, the problem of arbitrary selection of circumstances severely undermines the nonparametric estimation of IOP; this problem is attenuated but not completely solved when using latent class models. The number of parameters one needs to estimate when applying a latent class model is, in fact, a function of the number of classes, the number of circumstances considered, and the number of values each circumstance can take. This implies that the choice of circumstances considered, which is arbitrary, will affect the result.

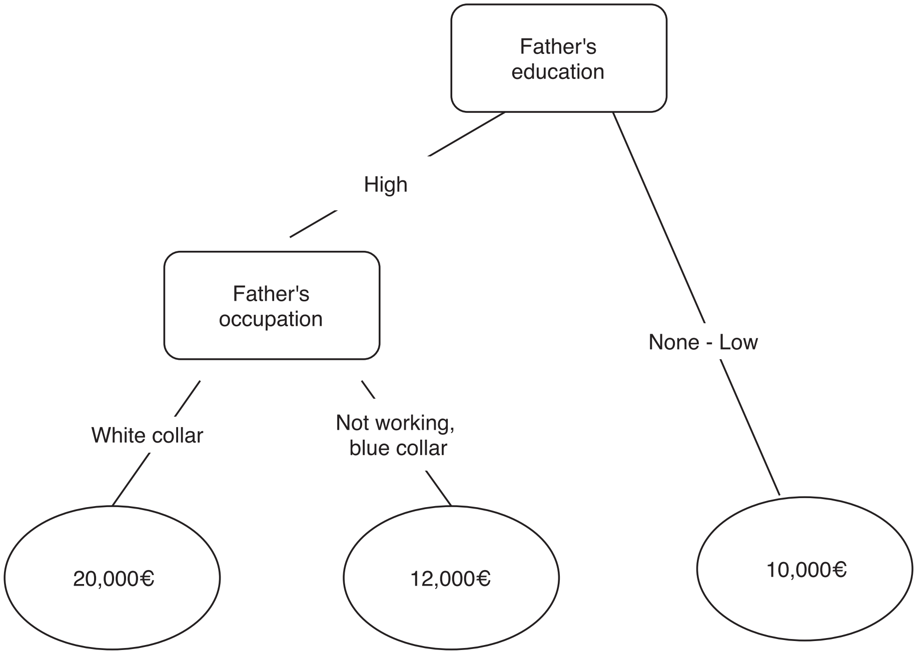

Figure 1. A Regression Tree
Notes: A simplified example of a regression tree explaining individual income variability. The tree is made of two splitting points (father's education and father's occupation) and three terminal nodes (ovals reporting average outcome for each type).

This weakness of the latent types approach highlights that a proper method that aims to estimate IOP needs to comprise both the identification of types based on observed circumstances and a variable selection criterion that would select the most appropriate set of the (possibly) many observable circumstances.

In light of this, Brunori *et al*. (2018) proposed the use of a machine learning algorithm, known as *conditional inference regression trees*, to identify Roemerian types. Regression trees are prediction algorithms introduced by Morgan and Sonquist (1963) and popularized by Breiman *et al*. (1984) almost 20 years later. The algorithm aims to predict an outcome out of sample based on a number of covariates. This is done by partitioning the space of the regressors in non-overlapping regions. The name *tree* comes from the way this algorithm can be graphically represented as an upside-down tree. Figure 1 shows an example of a regression tree for predicting income based on two regressors: parental education and parental occupation. The predicted income is simply the average outcome of individuals assigned to each terminal node (ovals at the bottom of the tree). Regression trees are generally grown in the attempt to maximize the ability of the model to predict out-of-sample. That is, trees aim at maximizing the variability of the dependent variable that can be explained by between-node variability, without overfitting the model. A very deep tree (i.e., an overfitted tree) would result in a very low in-sample error but would poorly perform out-of-sample.

Various methods exist for growing trees while avoiding overfitting. *Weakest link pruning* estimates the mean squared error (MSE) out-of-sample for all trees obtained replacing a sub-tree with a terminal node. This method of pruning is computationally costly in cases with a large set of regressors. Other methods such as *cost complexity pruning* or *conditional inference trees* can be used to prevent trees from growing too deep (James *et al*., 2013). Conditional inference trees introduced by Hothorn *et al*. (2006) prevent overfitting by growing the tree while conditioning the splitting on a sequence of statistical tests. The algorithm follows a stepwise procedure:

A conditional inference regression tree is an algorithm aiming at predicting a dependent variable **Y** based on a set of J regressors **C**. The conditional distribution of **Y** is assumed to depend on a function f of the regressors: $D(\mathbf{Y}|\mathbf{C} = D(\mathbf{Y}|f(C^1, \ldots, C^J)$. The algorithm proceeds with the following steps:

1. select the appropriate confidence level $(1-\alpha)$;
2. for each regressor, $j = 1, \ldots, J$, test the null hypothesis of independence between **Y** and **C**, $H_0^j: D(\mathbf{Y}|C^j) = D(\mathbf{Y})$, and store all the resulting p-values;
3. multiply the p-values by the Bonferroni correction term (J);
4. select the regressor with the lowest p-value (C^*);

$\Longrightarrow$. if for C^*: adjusted p-value $>\alpha \rightarrow$ exit the algorithm.

$\Longrightarrow$. if for C^*: adjusted p-value $<\alpha \rightarrow$, select C^* as splitting regressor;

5. for all possible binary partition splitting point s (observed values of C^*), test the discrepancy between the conditional expectation in the two resulting subsamples and store the p-value associated with each test;

6. 1. $\Longrightarrow$ Split the sample based on C^*, by choosing the splitting point s that yields the lowest p-value.

7. repeat steps 2–5 for all resulting subsamples.

The use of this algorithm presents a number of advantages: first, the choice of circumstances used to construct types is no longer arbitrary. Even when very large sets of observable circumstances are available, the algorithm will use only the characteristics that have the strongest association with the outcome. Second, the model specification is no longer exogenously given: how circumstances interact in determining the outcome is driven by the attempt of the algorithm to explain the variability of the outcome. Third, the algorithm automatically provides a test for the null hypothesis of equality of opportunity. Indeed, it is not impossible that the algorithm stops at step 3; the original sample is not split and the tree is made of a single terminal node. In this particular case, we could not reject the null hypothesis of equal opportunity[4]. Fourth, but no less important, opportunity trees tell us a story about the structure of opportunity that is immediately possible to understand even without formal statistical training.

However, it is important to not overemphasize on the opportunity structure described by a single tree obtained from survey data. One of the main weaknesses of regression trees is that estimates they produce tend to be heavily dependent

[4]From this point of view, the construction of a conditional inference tree can be interpreted as a robust version of a statistical test for the null hypothesis of equal opportunity in the spirit of Lefranc *et al*. (2009).

on the particular sample observed. This implies that estimating the same structure for the same time period using a different survey could result in a different opportunity structure. A second reason of caution when interpreting trees concerns the relatively poor performance of this type of algorithm in handling highly correlated regressors. In fact, when one of the two correlated regressors has been used to determine a split, it is unlikely that the second will play any role in the tree, although it may be nearly as correlated as the first one with the dependent variable.

This second issue can be partly solved using ensemble methods that combine several decision trees to obtain better predictive performance. The same regression tree is estimated a large number of times on a perturbated sample (typically obtained by random sampling with replacement). This allows different opportunity structures to emerge across iterations, making possible for correlated circumstances, that do not appear in the tree estimated in the entire sample, to determine some splitting point. The average of all the predictions obtained across trees has been shown to be a stronger predictor than a single regression tree (Breiman, 1996). In Section 5.3 in which we attempt to maximize the comparability of estimates across years, we turn to this approach using bagging of conditional inference trees.

3.2. *Identification of Effort*

Once types are identified, the second step consists of estimating effort. Adopting Roemer's identification strategy, this is done by estimating the shape of the type-specific outcome distribution in all types. Previous contributions select an arbitrary number of quantiles, generally not larger than 10, and estimate equation (4) setting μ^{π} equal to the average outcome across all individuals belonging to the j-th quantile of their type-specific outcome distribution.

Although not explicitly discussed by these contributions, the need to estimate the distribution of outcomes for each type has a clear impact on the empirical exercise. If in the ex-ante approach the main constraint when identifying types is the need to reliably estimate their mean, following the ex-post approach the need to estimate the type-specific outcome distribution for each type imposes a more severe trade-off. On one hand, a precise description of the data-generating process requires the consideration of a sufficiently large number of types; on the other hand, the estimation of the type-specific outcome distribution requires a large number of observations in each type, much larger than the sample size required for estimating a single parameter for each type. Particularly because Roemer's strategy imposes no restriction on the type-specific outcome distribution, the researcher must estimate a number of parameters equal to the number of quantiles.

As clarified by Luongo (2011), IOP estimates are sensitive to the selected number of quantiles. However, if one can imagine that—theoretically—a precise number of Roemerian types does exist, it is clear that quantiles are used to approximate an intrinsically unknown continuous distribution function. Therefore, there is no true number of quantiles.

This paper proposes a non-arbitrary criterion to approximate the type-specific outcome distribution based on a procedure suggested by Hothorn (2018). Such a criterion makes measures of IOP *à la* Roemer less dependent on discretionary methodological choices and more easily comparable across time and space.

Moreover, our method explicitly addresses the problem of balancing the need to precisely estimate the distribution of outcome in each type and the data constraints, constraints that will typically differ across types of different sample size.

We approximate the shape of the type-specific outcome distribution $F^k(y)$ using the Bernstein polynomial, i.e., a linear combination of Bernstein basis polynomials.[5] The Bernstein basis polynomial of degree m for some positive continuous variable $t \in [a, b]$ is defined as the set of polynomials:

$$\left\{ b_{j,m}(t, a, b) = \frac{1}{(b-a)^m} \binom{m}{j} (t-a)^j (b-t)^{m-j}, \quad \forall j = 1, \ldots, m \right\}. \tag{7}$$

A linear combination of Bernstein basis polynomial has the form:

$$B_m(t,a,b) = \sum_{i=0}^{m} \beta_i b_{j,m}(t,a,b). \tag{8}$$

In each type, we select the degree of the Bernstein polynomial that maximizes the out-of-sample log likelihood in approximating the real cumulative distribution function. Out-of-sample log likelihood is estimated by 10-fold cross-validation. The algorithm proceeds for each type $k = 1, \ldots, K$ with the following steps:

1. partition the population of type k into 10 non-overlapping sets of approximately equal size (folds);
2. for every $b = 1, \ldots, 10$;

a for every fold $f = 1, \ldots, 10$;

(I)obtain the *training* sample by excluding from the sample the f-th fold that will be later used as *test* sample;

(II) estimate the shape of the type-specific outcome distribution with a monotone increasing Bernstein polynomial of order b on the training set;

(III) predict the cumulative distribution of the type based on the Bernstein coefficients $\widehat{F}_b^k(y)$ on the test set;

(IV) estimate the out-of-sample log-likelihood (LL_b^f);

(V) store LL_b^f;

b we calculate and store $LL_b = \sum_{f=1}^{10} LL_b^f$;

3. select the maximum $LL_{b^*} \in [LL_1, \ldots, LL_{10}]$;
4. b^* indicates most appropriate order for the Bernstein polynomial approximating the type-specific distribution function of type k.

[5]Introduced in 1912 by Sergei Bernstein, Bernstein polynomials become known as the mathematical basis of the Bèzier curves, used first to design automobile bodies and, more recently, widely adopted in computer graphics to model smooth curves (Farouki, 2012). We opt for this approximation method as it has been shown to outperform competitors, such as kernel estimators, in approximating distribution functions (Leblanc, 2012).

The algorithm, repeated for all types, produces parametric approximations of the type-specific outcome distribution functions based on the coefficients of Bernstein polynomials of different order. Under Roemer's assumption, the quantiles of such distributions measure the degrees of effort individuals exerted. The estimation of equation (4) can then be conceptualized as inequality between a set of weighted, type-specific outcome distributions, and will depend on population weights and the Bernstein coefficients. Equation (4) can then be precisely approximated using a sufficiently high number of points to approximate each type-specific outcome distribution.

Equation (3) is then estimated by multiplying the outcome of each individual $i = 1, \ldots, N$ belonging to each type $k = 1, \ldots, K$ by the average outcome in the population (μ) divided by the expected outcome of an individual exerting the same degree of effort (μ^{π}), independently from the type she belongs to.

A typical problem, when using a fixed number of quantiles to estimate IOP, is the trade-off between the number of quantiles and the number of types. The ideal situation would be having a sufficiently large number of observations to allow growing a deep tree and having large sample size in each type. Sufficient sample size in each type would make possible to approximate the type-specific outcome distribution in a satisfactory number of points. However, the limited number of observation imposes a trade-off: a finer partition in types leads to a low number of observations per type limiting the number of quantiles one can use to estimate the conditional outcome distribution.

A possible pragmatic approach consists in choosing a reasonable number of quantiles (e.g., 10), and then adjust the partition in types to have sufficient degrees of freedom in each type to estimate 10 parameters. Empirically, in specific cases this may imply the suppression of a certain number of types, and this could lead to lower IOP estimates. This problem is to a large extent solved when using a polynomial approximation of the type-specific outcome distribution. In Section 5.4, we will explore to what extent this issue is empirically salient for the German SOEP survey.

4. Data

4.1. *The SOEP*

Our analysis on the evolution of IOP in Germany, applying the methods explained earlier, is based on the SOEP (SOEP; see Goebel *et al.*, 2019). The SOEP is a representative longitudinal survey of private households in Germany conducted annually since 1984, including the East German population since 1991.

The SOEP is one of the main data sources for distributional analysis in Germany. Furthermore, it includes a remarkable amount of retrospective information on individual characteristics that shall be indicative for the circumstances faced in childhood. For instance, questions on the education and occupation of parents, region of birth, migration background, and country of origin are included in the questionnaire and made comparable across the survey years.

We use the v33 version of SOEP including all subsamples apart from the two newly added refugee samples, which we exclude because of a rather high number of missing values among relevant circumstances. We restrict the sample to individuals in the age range 30–60 with available information on household income and all circumstances that we include in our analysis. Note that the reported income in a survey year refers to the year before the survey was conducted.

Descriptive statistics for the 25 waves used are reported in Appendix A. In the same appendix, we show for each survey wave the share of individuals with missing information for potential circumstances that we identify based on the past literature on IOP and the specific German context. To ensure the maximum possible comparability of estimates across time, we choose the set of circumstances with the lowest number of missing values across all survey waves. The circumstances that we include are as follows: sex, migration background, resident in East or West Germany in 1989, father's and mother's education, father's training and occupation measured by the ISCO-code (one digit), the number of siblings, and an indicator of whether the individual is disabled.

We set 1992 as the first survey wave because information on household income for people in East Germany is available from that year on, and therefore the analysis starts from the year 1991. To warrant the representability of our analysis at the national level in every year, each observation is weighted by the inverse probability of selection. Household income is displayed in Euro at 2011 prices. Our outcome of interest is household disposable equivalent income (applying the square root scale). To avoid life-cycle bias in our estimates, we compute, for every year, the deviation of individual income from its expected value given the respondent's age. Therefore, our outcome of interest is the deviation from what is predicted by a regression in which disposable household equivalent income declared by the respondent is regressed on her age and her age squared. Therefore, individuals with outcome higher than one have a larger equivalized income than the average resident in Germany in their age-specific reference group.[6] Total inequality trends for this residual outcome measure, as well as total equivalized household income, can be found in Appendix A.

4.2. *Sample Size*

A key challenge for our analysis is that our sample size varies considerably across waves of the SOEP, ranging between 2868 and 13,160 (see Appendix A). Conditional inference regression trees have been shown to be sensitive to the sample size because the splitting points are conditioned on a sequence of statistical tests and, when sample size is small, p-values of the tests tend to be higher. Other things held constant, the larger the sample size, the deeper might be the resulting tree. Deeper trees, made of many terminal nodes, usually tend to produce larger IOP estimates.

To overcome this problem and maximize the comparability of estimates over time, we proceed in two steps. First, in Subsections 5.1 and 5.2 we show the partition in types obtained using the original samples. Then, in Subsection 5.3 we

[6]Aware that inequality statistics tend to be heavily influenced by outliers (Cowell and Victoria-Feser, 1996), we adopt a standard winsorization method according to which we scale back all incomes below the 0.1th percentile and exceeding the 99.9th percentile of the year-specific outcome distribution to these thresholds.

proceed as follows: for each year a random subsample of size 2868 is drawn from the original data (the smallest recorded sample size over all survey waves that we use). Then, the opportunity tree and the resulting IOP are estimated and stored. This procedure is repeated 200 times. Doing so, the variables used, the splitting points, and the number of terminal nodes might differ not only across waves, but also across iterations for the same year. Finally, we report the average number of types and IOP values for each year.

Iterating the estimation of regression trees, we get close to what in machine learning is known as *random forest*, a large collection of trees obtained using a subset of the available information. Each tree of a random forest of conditional inference trees has two key characteristics: (1) the confidence level is very low (typically $(1-\alpha) = 0$), and (2) only a subset of regressors is considered for each possible split. Then the prediction is obtained averaging predictions across all trees. The predictive accuracy of random forests tends to outperform single trees.

However, our aim here is not predicting individual outcome as function of a set of regressors, but rather controlling for the effect of sample size in determining the depth of the opportunity tree. Therefore, using a subset of circumstances and growing a deep tree would not be useful. In contrast, fixing the sample size to 2868 observations in all waves will control for the possible effect of different sample sizes making our estimates more comparable across waves.

The use of a subset of observations, sampled without replacement, determines a certain level of heterogeneity of the tree structure across iterations. This reflects the level of uncertainty we have about the real data-generating process, which, in this case, is our uncertainty about the real Roemerian types existing in the German society. Therefore, the level of IOP becomes the average of 200 possible levels of IOP under alternative assumptions about the type partition and different re-samples of the original data.

4.3. *Sample Selection*

As mentioned, we restrict the sample to individuals with available information on circumstances included in the main analysis. Table A5 in Appendix A shows the share of missing values among circumstances for each survey year. This share varies from nought to almost 40 percent for some variables in single years, and decreases substantially over time. On average across all circumstances, the share of missing values is less than 10 percent of all observations.

To get a sense of whether issues of sample selection affect our estimates, we compare the distribution of the outcome variable among our sample with the full sample including observations with at least one missing value among all circumstances. Figure A6 in Appendix A depicts these differences. The distributions are essentially similar with some overlap, whereas household incomes of the sample without missing values are, on average, significantly higher.

We are aware that such systematic difference could bias our estimate in an unknown direction, particularly in years with a high share of missing circumstances. Indeed, observing the entire period from 1992 to 2016, we do find a sizeable negative correlation between the share of observations with missing value in at least one of the relevant circumstances and IOP. We show this relationship in

Figure A7 in Appendix A. A partly reassuring statistic shown in the figure is the low and unsystematic correlation between prevalence of missing information and estimated IOP, measured within the two subperiods 1992–2001 and 2002–2016 separately. However, as evident, we consider the share of missing information as suspect source of downward bias when estimating IOP. Interestingly, while in the past this issue was mostly focused on the problem of unobservable circumstances (Ferreira and Gignoux, 2011), our analysis highlights that the share of non-missing information regarding these circumstances may play a role as well.

5. The Evolution of IOP in Germany, 1992–2016

5.1. *Development of the Opportunity Tree*

Figure 2 shows the opportunity tree in 1992, 2000, and 2016. All partitions are obtained using the full sample and show a different structure of opportunities in Germany society for the three points in time. Appendix B shows the tree for all the other years over this time period.

The just-unified Germany is a polarized society, with one main driver of between-type inequality: place of residence in 1989. For Germans originally from the East, the second significant circumstance is father's occupation. Those having a father with a high occupation, or employed in the armed forces, show a higher level of expected outcome than those with a father in an unskilled occupation. Both East types have an expected outcome below 80 percent of the national average for their age group. For people residing in West Germany, the splitting node is instead defined by fathers' education. The average level of outcome in both western types is way above the national average.

The structure of opportunity is much more complex in 2000, and even more so in 2016. What stands out in the former is the appearance of migration background and disability as relevant circumstances leading to a very low level of income. In the latter almost the entire set of circumstances, excluding the number of siblings and migration background, determines at least one splitting point. Interestingly, the place of residence in 1989 is still a fundamental driver of inequality and the second-most correlated circumstance with the outcome. The first splitting point is determined by father's occupation, with armed forces occupations now together with unskilled occupations at the bottom of the distribution.

In 2016, the German society appears to be composed of a much larger number of types in comparison with the simpler opportunity structure suggested by the data for 1992. The vector of circumstances used in 1992 is location in 1989, father's occupation and education. In addition, for 2016 disability, sex, mother's education, and father's training play a role in splitting the sample. Furthermore, people who were resident abroad in 1989, and migrated to Germany later, form the lower types together with individuals from the East. However, among these subtypes migration background does not contribute to further explaining the opportunity structure beyond other circumstances, like parental background and disability.

It is important to underline here that opportunity trees based on conditional inference regression tree should be interpreted with caution. Notoriously regression trees tend to be sensitive to the particular observed sample, using machine

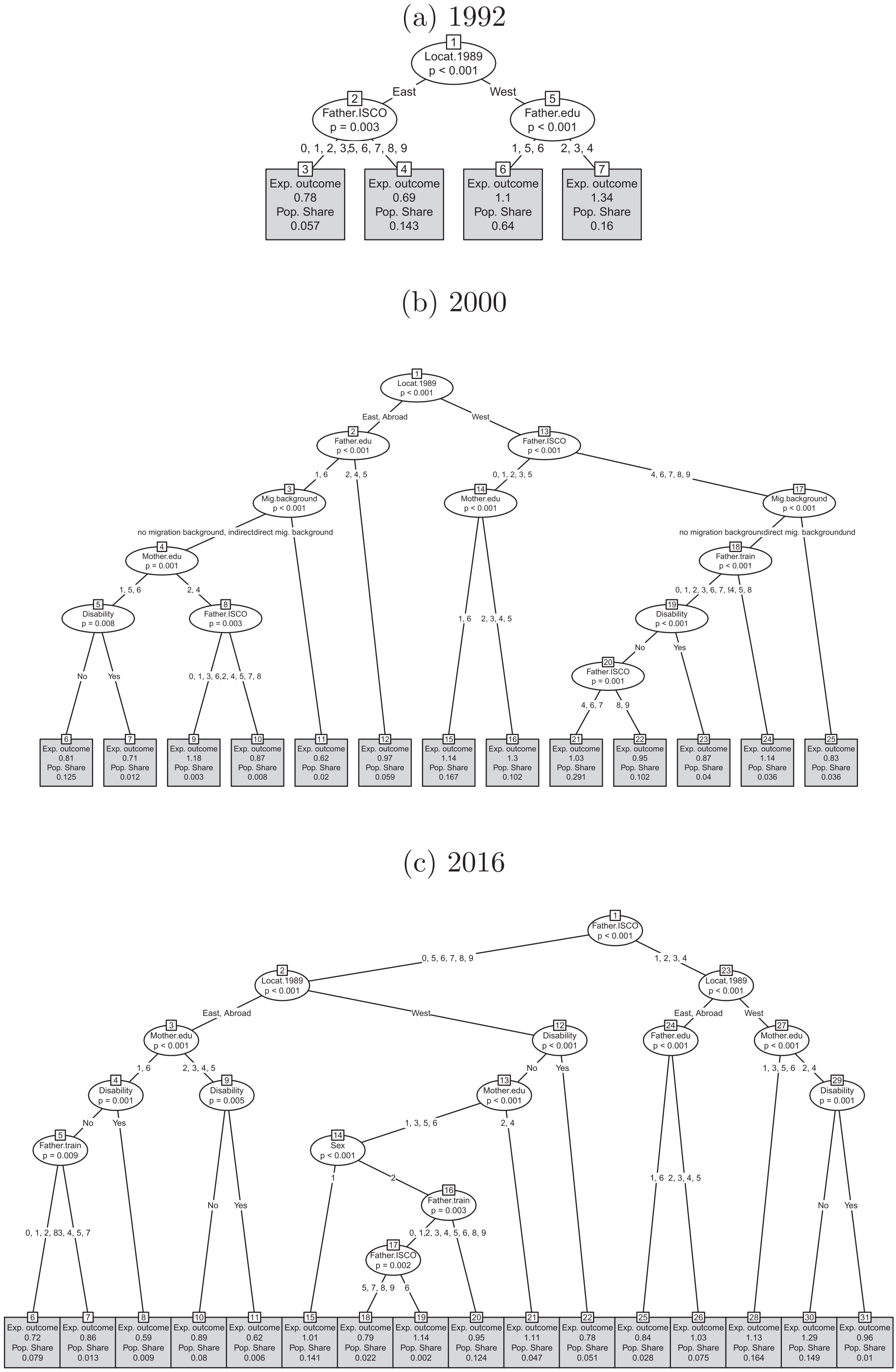
(a) 1992
1
Locat.1989
p < 0.001
East
West
2
Father.ISCO
p = 0.003
5
Father.edu
p < 0.001
0, 1, 2, 3,
5, 6, 7, 8, 9
1, 5, 6
2, 3, 4
3
Exp. outcome
0.78
Pop. Share
0.057
4
Exp. outcome
0.69
Pop. Share
0.143
6
Exp. outcome
1.1
Pop. Share
0.64
7
Exp. outcome
1.34
Pop. Share
0.16
(b) 2000
(c) 2016

Figure 2. Opportunity Tree in 1992, 2000, and 2016

Source: SOEPv33.

Note: Years refer to the survey wave. Incomes reported in the survey wave refer to the year before. **Father/mother education**: 1 = lower secondary, 2 = intermediate secondary, 3 = technical school, 4 = upper secondary, 5 = other school degree, 6 = no school degree, 7 = school not attended. **ISCO**: 0 = armed forces, 1 = managers, 2 = professionals, 3 = technicians, 4 = clerks, 5 = service workers, 6 = skilled agricultural workers, 7 = craftsmen, 8 = plant and machine workers, 9 = elementary occupations. **Training**: 0 = no information, 1 = no vocational degree, 2 = vocational degree, 3 = trade or farming apprentice, 4 = business, 5 = health care or special technical school, 6 = civil service training, 7 = tech engineer school, 8 = college, university, 9 = other training.

learning jargon: they have low bias but high variance. Therefore, it is possible that observing a slightly different sample or observing a different pattern of missing information, one or more types would not show in the tree and/or new types could appear. Nevertheless, without excessively emphasizing the role of each particular type, general characteristics of the tree, such as the number of terminal nodes and the type of variables most frequently used for splitting the sample, are certainly informative about the structure of opportunity in each year in Germany.[7]

Table A7 in Appendix A shows the development of the number of terminal nodes (types) from 1992 to 2016. The number of types gradually increases until the early 2000s with dramatic rises in 2000 and 2002, when the number jumps first to 13 and then to 20. Thereafter, the trend is characterized by ups and downs within this higher range.

A closer look at the development of the opportunity tree from 1992 to 2016 reveals some striking patterns (see Appendix C). The first, most compelling evidence is that, albeit the appearance of some characteristics on the opportunity tree and the rising complexity of German society, the interaction of circumstances that defines the highest and the lowest type in the income distribution is rather constant. At the top of the distribution we always find individuals who resided in West Germany before the fall of the Berlin Wall, whose parents had a high occupational position, and whose mothers had a high educational degree, whereas East Germans with low educated parents persistently qualify at the lowest end.

Until 2001, location in 1989 is the first splitting variable of the tree, and for the rest of the time it is a circumstance that persistently splits German society. For example, in 2014 people from East Germany with high parental occupational background (managers and professionals) have lower average income levels than their West German counterparts with middle parental occupation (technicians and clerks). Having a disability is, particularly from 2002 on, a circumstance that consistently explains inequality in Germany. This does not depend on the relative size of this group in the survey, because the share of respondents reporting disability oscillates without much variation around 9 percent.

Migration background appears relevant to the splitting of the West German population in subtypes from 1999 to 2013. This applies to people with own migration experience (direct migration background) that moved to West Germany before 1989. In contrast, the average income of the children of migrants (indirect

[7]A criterion to assess the accuracy of the partition in types is its ability in predicting the outcome variability out-of-sample. Appendix C presents a discussion of the predictive performance of the estimated trees.

Figure 3. Development of IOP in Germany, 1992–2016

Source: SOEPv33, 1992–2016.

Note: Years refer to the survey wave. Incomes reported in the survey wave refer to the year before. 99 percent confidence intervals are obtained from 500 bootstrap re-samples of the data. [Colour figure can be viewed at wileyonlinelibrary.com]

migration background) is not distinguishable to the average income of the rest of the population; that is, this circumstance mostly does not play a major role after having controlled for parental occupation and education, confirming past findings on the topic (e.g. Krause *et al.*, 2015; Bönke and Neidhöfer, 2018). New migrants, i.e. people who were resident abroad in 1989, mostly belong to the lower types together with people who resided in East Germany. However, only in 1 year, 2001, their average age-adjusted incomes are significantly different from the outcomes of East Germans. In this year, new migrants with low parental education form the type with the lowest age-adjusted income.

5.2. *IOP Estimates*

Figure 3 (and Table A7 in Appendix A) shows the development of IOP, measured by the Gini coefficient, from 1992 to 2016. The IOP trend is characterized by decreasing inequality in the 1990s and a subsequent increase, leading to a rather stable trend with little variation from 2003 onward. Interestingly, despite the 2016 partition has four times the number of types than the 1992 partition, the level of IOP is only slightly higher in 2016 (0.1046) than in 1992 (0.0959). Germany appears a much more complex society, with a complicated interaction of circumstances in producing opportunity. However, the level of the resulting inequality is just slightly higher.

This becomes even more evident looking at how the type-specific empirical cumulative distribution functions (ECDFs) change over time. Figure 4 shows the type-specific ECDFs in 1992, 2000, and 2016; graphs for all the other years can be found in Appendix B. The dots are observed distributions, whereas dashed lines show the interpolation of the distribution obtained applying the Bernstein polynomial approximation.

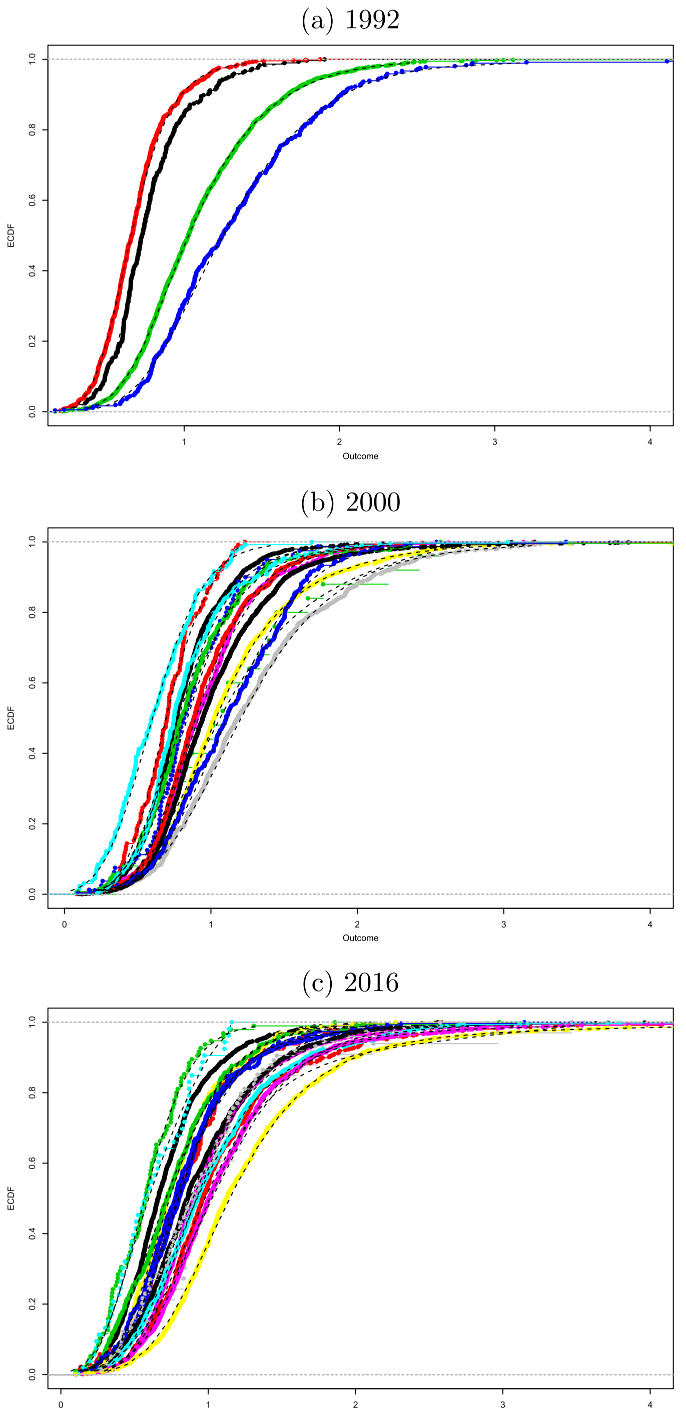

Figure 4. ECDFs in 1992, 2000, and 2016

Source: SOEPv33.

Note: Years refer to the survey wave. Incomes reported in the survey wave refer to the year before. [Colour figure can be viewed at wileyonlinelibrary.com]

The comparison of the ECDFs in 1992 and 2016 is an illustrative example for the spectacular changing of IOP in Germany. In 1992, the two compressed distributions of East residents lie rather close to each other, whereas the more dispersed Western distributions lie far to the right. This polarization is no longer evident in 2000 and 2016. The 13 or 16 type-specific distributions lie close to each other and cross in several points. However, the distance between the highest and the lowest type is remarkable and remains stable over the entire period. This explains why the huge increase in the number of types has not been accompanied by a drastic rise in IOP.

5.3. *Fixed Sample Size*

As already discussed, opportunity trees and corresponding type-specific outcome distributions might not be strictly comparable because of changes in sample size. Therefore, we proceed by fixing the number of observations at 2868 (the minimum sample size) and repeating the entire estimation procedure 200 times. This choice is similar in spirit to estimating a "random forest" of conditional inference trees. There are two key differences: first at each iteration no regressor is excluded, second the confidence level $((1-\alpha))$ is not reduced allowing deeper tree. These differences are justified by the different purpose we have here, which is not maximizing predictive ability but maximizing comparability of estimate across time. Reducing the sample size does affect the predictive performance of regression trees that tend to have a slightly higher MSE when used to predict out-of-sample individual outcome (Figure C8 in Appendix C shows that MSE of trees obtained with subsamples of 2868 observation are on average 4 percent higher than MSE estimated with the entire sample both in the training and in the test set).

Figure 5 shows the average number of terminal nodes and the level of IOP obtained by an iterative procedure with 200 repetitions. The reported bounds show the 0.975 and 0.025 quantiles of the distribution of the estimates.[8] In comparison to the main analysis, in this sensitivity test the number of terminal nodes is reduced. The maximum number of types over the 200 iterations is on average 9.25 against the 20 obtained using the entire sample. Nevertheless, the trends in IOP, as well as in the number of types, are similar and show both an increase in the level of complexity over time and an opportunity structure in 2016 markedly more complex than in 1992. The trend in IOP is also close to what is obtained with the full sample with a decrease during the 1990s and a steep increase in the early 2000s. Again, the level of IOP in 2016 appears slightly above the level of 1992.

Estimating 200 trees with different samples makes it impossible to show opportunity trees for this application. However, such resampling approach can be used to evaluate the relative importance of correlated regressors that is otherwise problematic when using a single regression tree. Depending on the particular sample used, a slightly different opportunity structure may emerge at each iteration,

[8]Note that bounds cannot be interpreted exactly as bootstrap confidence intervals: we do not resample with replacement, and we do not draw samples of the same sample size of the initial distribution. This also explains why there is no variability around the point estimates for 1992.

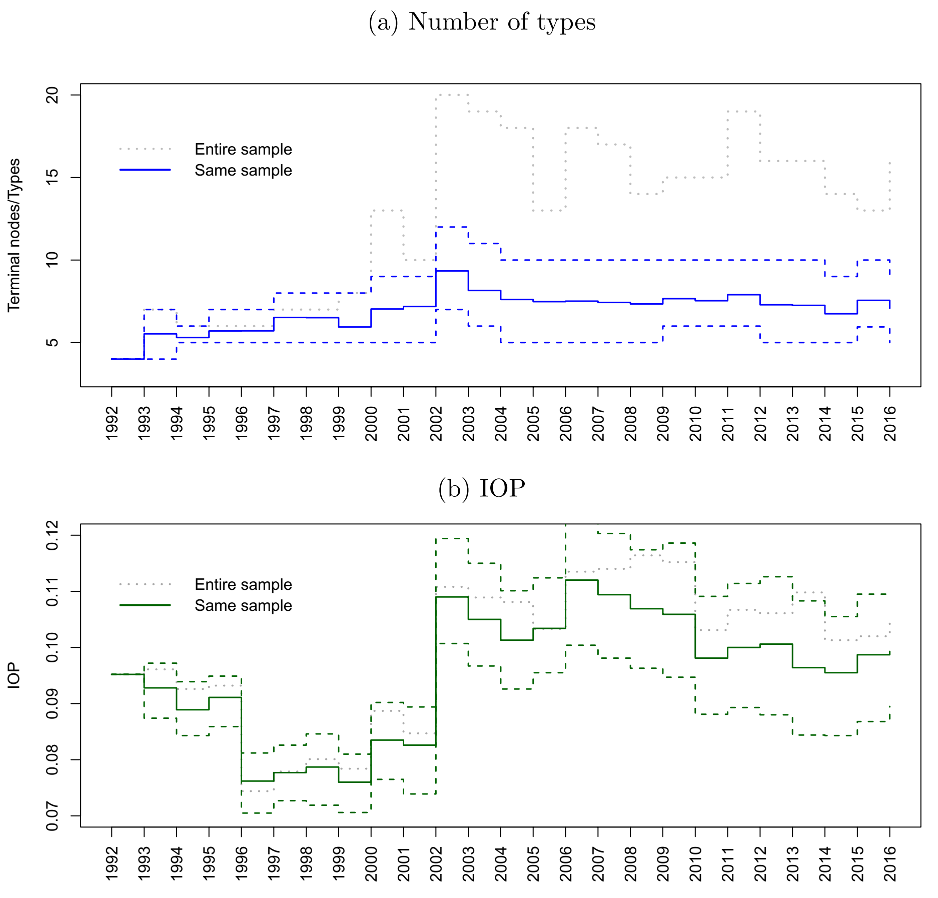

Figure 5. Average Number of Types (Top) and IOP (Bottom) Controlling for Sample Size
Source: SOEPv33, 1992 and 2016.
Note: Averages are calculated over 200 trees based on a sample of 2868 observations drawn without replacement. Bounds show the 0.975 and 0.025 quantiles of the distribution of the estimates [Colour figure can be viewed at wileyonlinelibrary.com]

making possible for correlated circumstances, that do not appear in the tree estimated in the entire sample, to determine some splitting point.

Table A6 in Appendix A shows for each circumstance the share of trees in which the variable determines at least one split. A striking finding is that location in 1989 is used in all 5000 trees estimated for at least one split. Interesting trends can be observed for other circumstances. In Figure 6, we report how often father's education, disability, and sex determine at least one split in the opportunity tree. The first circumstance shows a clear decrease over time, whereas the second and the third, almost absent in the opportunity structure of the early 1990s, appear, respectively, in 80 percent and 20 percent of the trees in most recent waves. Hereby, the increasing role of sex in determining outcome inequality is likely to be explained by the increasing number of single-parent households (from 13 percent in 1992 to 23 percent in 2016). In contrast, the increasing role of disability has no

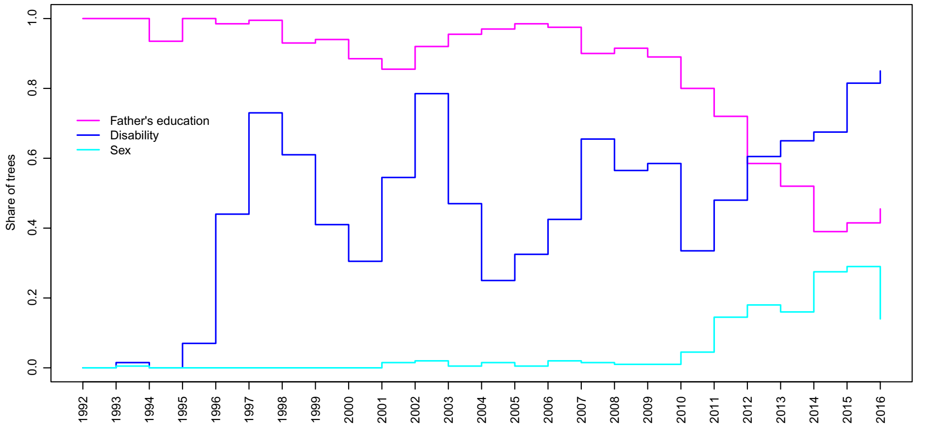

Figure 6. Share of Trees That Use Father's Education, Disability, and Sex to Obtain Roemerian Types
Note: Shares are calculated over 200 trees based on a sample of 2868 observations drawn without replacement. *Source*: SOEPv33, 1992–2016. [Colour figure can be viewed at wileyonlinelibrary.com]

straightforward mechanical explanation, because the share of respondents reporting disability in SOEP is not growing over time (constantly between 7.5 percent and 10 percent from 1992 to 2016). The analysis of this remarkable pattern goes beyond the scope of this work, but should be addressed in future research.[9]

5.4. *Other Robustness Checks*

In this section, we discuss the sensitivity of our results to a number of methodological choices. First, we show how sensitive the IOP trend is to the confidence level used to construct conditional inference trees ($1-\alpha$). Because a lower confidence level implies a deeper tree we expect higher levels of IOP , the lower the imposed confidence level. Figure A3 in appendix confirms this expectation. The level of IOP obtained when the confidence level is set to 0.95 and 0.90 tends to be higher than that obtained imposing $(1-\alpha) = 0.99$. As expected, a higher number of terminal nodes imply a higher level of IOP. However, estimates are rather close and move together following exactly the same trend.

Second, we repeat the entire empirical exercise using the traditional classification of regression trees (CART) grown and pruned as proposed by Breiman *et al*. (1984) and implemented by Therneau and Atkinson (2019). The partition is obtained recursively splitting the sample to maximize impurity reduction. Hereby, the splitting is iterated growing a deep (and certainly overfitted) tree. Then, the resulting tree is cost-complexity pruned using five-fold cross-validation to obtain the tree with the highest predictive accuracy (see Therneau and Atkinson, 2019 for the details). Figure A4 in Appendix A compares the IOP estimates obtained using conditional inference regression tree and CART. Conditional inference regression trees are a little more conservative than CART for the first decade, but produce

[9]The shares of single parents and disabled individuals are based on our own estimates using SOEPv33.

higher IOP estimates in later waves. On the contrary, excluding the five waves between 2010 and 2015, CART estimates tend to be close to our preferred measure of IOP and within its 99 percent confidence bounds. The trend overtime is also very similar.

Third, we abandon Bernstein polynomials and we adopt an alternative approach to approximate the type-specific outcome distribution. We adopt the approach typically used in the literature: we set a fixed number of quantiles for all types and we adjust the partition in types to have sufficient sample size in all groups. We set the number of quantiles to 10. Therefore, when obtaining the partition in types, we need a sufficiently large sample size in each type to subdivide the population in 10 quantiles and estimate the average outcome in each of them.

It is not immediate to calculate how many observations per quantile are necessary to consistently estimate statistical parameters (see e.g. Lenth, 2001). We use a rule of thumb imposing a minimum of 10 observations in each quantile of each type. Therefore, we repeat the entire exercise preventing the algorithm from splitting the data if one of the resulting terminal nodes would result having a size smaller than 100. The resulting IOP trend is reported in Figure A5 in Appendix A.

There are two reasons that can explain differences in the estimated degree of IOP with respect to the use of polynomial approximation: *i*. the tree used may contain a smaller number of types, and *ii*. within type, a different method of approximation of the outcome distribution leads to a different counterfactual distribution.

In terms of the number of types, the minimum sample size requirement turns out to constrain the number of types in more than one case. As shown in Figure A5 in Appendix A, the number of suppressed types is larger than zero in 17 years with a maximum of six types suppressed in 2006. Note also that IOP tends to be slightly lower than when the estimates are obtained by Bernstein polynomial approximation. Nevertheless, they are very close when the number of terminal nodes is the same. The downward distortion appears more evident whenever one or more terminal nodes are suppressed. We conclude that, at least for the German context and data under consideration, differences in IOP with respect to the traditional approach based on quantiles do not depend on the approximation method itself, but on the fact that our method allows to identify small types that would otherwise be ignored.

5.5. *Discussion*

Past studies report a rise in net income inequality in Germany in the 1990s until 2005/2006 and a subsequent stagnation characterized by small ups and downs (e.g. Biewen and Juhasz, 2012; Biewen *et al*., 2019; Jessen, 2019; Peichl *et al*., 2012). These studies identify changes in employment driven by part-time and marginal part-time work, and changes in the tax system as the major driver of this development, as well as the rising dispersion of labor market incomes because of skill-biased technological change (see Dustmann *et al*., 2009). Changes in the household size and structure and reforms of the transfer system have been identified as minor influencing factors.[10]

[10]Besides the economic literature on wage and income inequality in Germany, less attention has been dedicated to IOP. Two exceptions are Peichl and Ungerer (2017) measuring East–West disparities in IOP and Niehues and Peichl (2014) comparing IOP levels in Germany to the US.

Part of the mechanisms described earlier could also explain the first decrease, then the sudden rise, and finally the rather stagnant development of IOP in Germany from 1992 to 2016. The rising wage inequality in the 1990s, evidenced, e.g. by Schündeln *et al*. (2010), was not accompanied by rising IOP, as our analysis shows. Instead IOP drops, particularly from 1995 to 1996, and then slowly rises until 2001. Then, it experiences a sharp rise in 2002 that brings it to a new, higher level. It remains an open question how strongly this sudden increase is associated with the inclusion of a special subsample of high-income households in 2002. This sample was included in SOEP to more adequately capture the upper end of the income distribution. In principle, applying sampling weights, this issue should be corrected for. However, there is no actual way to test it over time, because the provided weights to perform estimations excluding the high-income sample would also exclude all additional samples included after 2002. For 2002, we are able to estimate IOP including the high-income sample and excluding it with consistent weights. IOP measured without the high-income sample is substantially smaller; the difference between the two values is around 34 percent. Therefore, pre-2002 estimates, before the inclusion of the high-income sample, should be evaluated with caution.[11] After 2002, IOP stays rather constant, with ups and downs, until 2016.

The increase from 2002–2005 to 2006–2009 is contemporaneous to major reforms of the tax and transfer system and of the unemployment benefit schemes. Particularly, the changes to the social benefit system also known as Hartz-reforms that were enacted in 2003, 2004, and 2005 as response to steadily rising unemployment had long-lasting and controversially discussed effects on the German society. Simulations of the effect of the reforms estimate a reduction in non-cyclical unemployment by around 1.5 percent (Krebs and Scheffel, 2013). However, past studies have shown an overall small income inequality-reducing effect of the reforms, with a different impact on the middle and bottom of the distribution of income (Biewen and Juhasz, 2012). Particularly the incomes of longer-term unemployed were negatively affected, whereas social assistance receiver slightly gained from the reforms. Generally, in the period 2005–2008, in which the German unemployment rate fell by almost 4 percent, IOP stays rather constant.

We also do not observe a sizeable effect of the 2008–2009 financial crisis on IOP, confirming the conclusion of studies dedicated to income inequality. If any, we observe a downfall of IOP by 2 percentage points from 2009 to 2010. Bargain *et al*. (2017) found that in this period and until 2010 in Germany, policy changes induced a rise in poverty rates, mainly because of the slow adaption of social assistance, decreasing tax allowances, and changes in the taxation of capital income. However, the tax reforms produced also lower marginal tax rates at both ends of the distribution, and particularly for low levels of gross labor incomes the budget constraint rose from 2002 to 2011 (Jessen, 2019). This possible offset of mechanisms could explain why the small rises and falls in IOP in this period are of minor magnitude. Another possible reason for this is that several elements that explained

[11]Surprisingly, although most studies on income inequality using SOEP data that include the high-income sample record this exceptional rise, to the best of our knowledge the issue was not directly addressed so far.

the income inequality increase before 2005 became weaker over time. For instance, the rise in wage inequality became less steep, employment opportunities increased, and the middle and upper parts of the distribution benefited from the employment boom after 2006 (Biewen *et al.*, 2019).

6. Conclusions

Consistent with the theory proposed by Roemer (1998), we have suggested a novel approach to model the role of effort when estimating IOP. In Roemer's view, IOP is inequality because of circumstances beyond individual control. Outcome variability because of variables of choices is, instead, not part of IOP. The implementation of a measure consistent with this theory is complex because it necessitates both to identify relevant circumstances beyond individual control and to measure responsibility variables.

Our analysis borrows from machine learning methods and proposes an improved data-driven approach to the estimation of IOP. The main advantages of our approach are to minimize arbitrary assumptions about the shape of the type-specific outcome distributions.

Roemerian types, i.e., relevant interactions of circumstances, are obtained through conditional inference regression trees. The algorithm selects a partition in types that maximizes the outcome variability that can be consistently explained by between-type inequality. The identification of effort relies on a polynomial approximation of the ECDF of outcomes in each type. The degree of the (Bernstein) polynomial is selected by a 10-fold cross-validation to maximize its out-of-sample log likelihood.

We implement our method to 25 waves of the SOEP to describe the evolution of IOP in Germany between 1992 and 2016. We show that the structure of opportunities has markedly evolved over time. The partition in types detected by our approach in 2016 is much more complex than in 1992. We show that this difference is not only driven by changes in the survey's sample size; the trend persists also applying an iterative procedure that controls for changes in the sample size. Whether other improvements in the SOEP quality that occurred after 2000 can in part explain the trend remains an open question.

Despite the increase in the complexity describing the partition of the German society, the level of IOP we measure in 2016 is only slightly higher than that in 1992. The trend we observe sees a decrease in IOP in the 1990s, a sharp, sustained rise from 1999 to 2003 followed by another, smaller, sudden increase in 2006, and a subsequent stagnation with small ups and downs. The increase in IOP is suggestively contemporaneous with rising income inequality in the period 1999–2005 and the introduction of the Hartz-reforms. In contrast, a clear interconnection with the rising wage inequality in the 1990s and the financial crisis is not visible. Mechanisms that could have offset these developments to cause a stronger role of circumstances to determine outcomes are a topic of great research interest for future studies.

Several further interesting suggestions arise from our analysis of the evolution of IOP in Germany. The most compelling fact is that the East–West divide characterizes, more than two decades after reunification, most of the disparities in the German society. It is remarkable that even in recent times this circumstance divides the society in subtypes at the lower and higher ends of the income distribution. Therefore, over the entire observation period the types with the highest average level of outcome always consist of individuals who resided in West Germany in 1989. Besides, individuals with highly educated mothers, often in combination with fathers in high occupational positions, are constantly part of the type with the highest level of outcome.

To sum up, our novel analysis of the development of IOP in Germany shows that over the past two decades, the type with the highest level of outcome in the entire German income distribution consists of people with highly educated mothers and fathers in high occupational positions who resided in West Germany before reunification. At the bottom of the distribution, we constantly find individuals with low-educated fathers who resided in the former German Democratic Republic until the fall of the Berlin Wall.

References

Almås, I., A. W. Cappelen, J. T. Lind, E. Ø. S⊘rensen, and B. Tungodden, "Measuring Unfair (In) Equality," *Journal of Public Economics*, 95(7–8), 488–99, 2011.

Bargain, O., T. Callan, K. Doorley, and C. Keane, "Changes in Income Distributions and the Role of Tax-Benefit Policy During the Great Recession: An International Perspective," *Fiscal Studies*, 38(4), 559–85, 2017.

Biewen, M., and A. Juhasz, "Understanding rising income inequality in Germany, 1999/2000-2005/2006," *Review of Income and Wealth*, 58(4), 622–47, 2012.

Biewen, M., M. Ungerer, and M. Löffler, "Why Did Income Inequality in Germany Not Increase Further After 2005?" *German Economic Review*, 20(4), 471–504, 2019.

Bönke, T., and G. Neidhöfer, "Parental background matters: Intergenerational mobility and assimilation of Italian immigrants in Germany," *German Economic Review*, 19(1), 1–31, 2018.

Bourguignon, F., F. H. G. Ferreira, and M. Menéndez, "Inequality of Opportunity in Brazil," *Review of Income and Wealth*, 53(4), 585–618, 2007.

Breiman, L., "Bagging Predictors," *Machine Learning*, 2(24), 77–93, 1996.

Breiman, L., J. Friedman, C. Stone, and R. Olshen, *Classification and Regression Trees*, Taylor & Francis, Belmont, https://doi.org/10.1201/9781315139470, 1984. ISBN 9780412048418.

Brunori, P., P. Hufe, and D. G. Mahler, "The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees," World Bank Policy Research Working Paper WPS8349, 2018.

Brunori, P., V. Peragine, and L. Serlenga, "Upward and Downward Bias When Measuring Inequality of Opportunity," *Social Choice and Welfare*, 52(4), 635–61, 2019.

Brunori, P., A. Trannoy, and C. F. Guidi, "Ranking Populations in Terms of Inequality of Health Opportunity: A Flexible Latent Type Approach," *ECINEQ WP 515*, 2020.

Checchi, D., and V. Peragine, "Inequality of Opportunity in Italy," *The Journal of Economic Inequality*, 8(4), 429–50, 2010.

Cowell, F. A., and M.-P. Victoria-Feser, "Robustness Properties of Inequality Measures," *Econometrica*, 64(1), 77–101, 1996.

Dustmann, C., J. Ludsteck, and U. Schönberg, "Revisiting the German Wage Structure," *The Quarterly Journal of Economics*, 124(2), 843–81, 2009.

Dworkin, R., "What is Equality? Part 2. Equality of Resources," *Philosophy and Public Affairs*, 10(4), 283–345, 1981.

Farouki, R. T., "The Bernstein Polynomial Basis: A Centennial Retrospective," *Computer Aided Geometric Design*, 29, 379–419, 2012.

Ferreira, F. H. G., and J. Gignoux, "The Measurement of Inequality of Opportunity: Theory and an Application to Latin America," *Review of Income and Wealth*, 57(4), 622–57, 2011.

Fleurbaey, M., and V. Peragine, "Ex Ante Versus Ex Post Equality of Opportunity," *Economica*, 80(317), 118–30, 2013.

Schündeln, N., D. Krueger, and M. Sommer, "Inequality Trends for Germany in the Last Two Decades: A Tale of Two Countries," *Review of Economic Dynamics*, 13(1), 103–32, 2010.
Goebel, J., M. M. Grabka, S. Liebig, M. Kroh, D. Richter, C. Schröder, and J. Schupp, "The German Socio-Economic Panel (SOEP)," *Jahrbücher für Nationalökonomie und Statistik*, 239(2), 345–60, 2019.
Hothorn, T., basefun: Infrastructure for Computing with Basis Functions, 2018.
Hothorn, T., K. Hornik, and A. Zeileis, "Unbiased Recursive Partitioning: A Conditional Inference Framework," *Journal of Computational and Graphical Statistics*, 15(3), 651–74, 2006.
James, G., D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning with Applications in R*. Springer Texts in Statistics. Springer, 2013.
Jessen, R., "Why has Income Inequality in Germany Increased From 2002 to 2011? A Behavioral Microsimulation Decomposition," *Review of Income and Wealth*, 65(3), 540–560, 2019.
Krause, A., U. Rinne, and S. Schüller, "Kick it Like Özil? Decomposing the Native-Migrant Education Gap," *International Migration Review*, 49(3), 757–89, 2015.
Krebs, T., and M. Scheffel, "Macroeconomic Evaluation of Labor Market Reform in Germany," *IMF Economic Review*, 61(4), 664–701, 2013.
Lanza, S., T. Xianming, and B. Bethany, "Latent Class Analysis with Distal Outcomes: A Flexible Model-Based Approach," *Structural Equation Modeling*, 20(1), 1–26, 2013.
Leblanc, A., "On Estimating Distribution Functions Using Bernstein Polynomials," *Annals of the Institute of Statistical Mathematics*, 64(5), 919–43, 2012.
Lefranc, A., N. Pistolesi, and A. Trannoy, "Equality of Opportunity and Luck: Definitions and Testable Conditions, with an Application to Income in France," *Journal of Public Economics*, 93(11–12), 1189–207, 2009.
Lenth, R. V., "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, 55(3), 187–93, 2001.
Li Donni, P., J. G. Rodríguez, and P. Rosa Dias, "Empirical Definition of Social Types in the Analysis of Inequality of Opportunity: A Latent Classes Approach," *Social Choice and Welfare*, 44(3), 673–701, 2015.
Luongo, P., "The Implication of Partial Observability of Circumstances on the Measurement of Inequality of Opportunity," in *Inequality of Opportunity: Theory and Measurement*, in J. G. Rodriguez (ed.), vol. 19 of Research on Economic Inequality, Emerald, Bingley, 23–49, 2011.
Morgan, J. N., and J. A. Sonquist, "Problems in the Analysis of Survey Data, and a Proposal," *Journal of the American Statistical Association*, 58(302), 415–34, 1963.
Niehues, J., and A. Peichl, "Upper Bounds of Inequality of Opportunity: Theory and Evidence for Germany and the US," *Social Choice and Welfare*, 43(1), 73–99, 2014.
Peichl, A., N. Pestel, and H. Schneider, "Does Size Matter? The Impact of Changes in Household Structure on Income Distribution in Germany," *Review of Income and Wealth*, 58(1), 118–41, 2012.
Peichl, A., and M. Ungerer, "Equality of Opportunity: East vs. West Germany," *Bulletin of Economic Research*, 69(4), 421–27, 2017.
Ramos, X., and D. Van de gaer, "Is Inequality of Opportunity Robust to the Measurement Approach?" *Review of Income and Wealth*, https://doi.org/10.1111/roiw.12448, 2020.
Rawls, J., *A Theory of Justice*. Harvard University Press, https://doi.org/10.2307/j.ctvjf9z6v, 1971.
Rodríguez, J. G., "Partial Equality-of-Opportunity Orderings," *Social Choice and Welfare*, 31, 435–56, 2008.
Roemer, J. E., *Equality of Opportunity*. Harvard University Press, Cambridge, https://doi.org/10.1057/978-1-349-95189-5_2223, 1998.
______, "Equality of Opportunity: A Progress Report," *Social Choice and Welfare*, 19(2), 455–71, 2002.
Therneau, E. J. M. F., and Terry M. Atkinson, *An Introduction to Recursive Partitioning Using the Rpart Routines*, https://stat.ethz.ch/R-manual/R-patched/library/rpart/doc/longintro.pdf, 2019.
Van de gaer, D., "Equality of Opportunity and Investment in Human Capital," *Ph.D. thesis*, University of Leuven, 1993.

Supporting Information

Additional supporting information may be found in the online version of this article at the publisher's web site:

Appendix A Additional tables and figures
Figure A1: Total inequality (Gini coefficient) Notes: Eq. disp. income is the

equivalized household disposable income. Outcome is the deviation of equivalent income from what expected given age (used in the main analysis). *Source: SOEPv33, 1992–2016*.

Figure A2: Number of observations per survey year *Source: SOEPv33, 1992–2016*.

Table A1: Descriptive statistics

Table A2: Descriptive statistics, cnt.

Table A3: Categories for parental education, training and occupation

Table A4: Share of individuals with missing information on circumstances—observations with non-zero weight

Table A5: Share of individuals with missing information on circumstances—full sample

Table A6: Share of trees in which each circumstance determines at least one split

Table A7: Estimates

Figure A3: IOP: alternative confidence levels required to split the sample Note: 99% confidence intervals are obtained from 500 bootstrap re-samples of the data. *Source: SOEPv33, 1992–2016*.

Figure A4: IOP estimated using recursive partitioning routine Source: SOEPv33, 1992–2016. Note: CART estimates are obtained growing the tree by binary splitting maximising impurity reduction and pruned by 5-fold cross-validation. For details of the routine used see Therneau (2019). 99% confidence intervals are obtained from 500 bootstrap re-samples of the data.

Figure A5: IOP estimates using 10 quantiles of the type specific outcome distribution *Source: SOEPv33, 1992–2016. Note: a the bottom of the figure we report the number of types suppressed in order to obtain 100 observations in each type considered the minimum sample size to estimate 10 quantiles. 99% confidence intervals are obtained from 500 bootstrap re-samples of the data*

Figure A6: Comparison of household income among the analytical sample and the full sample including missing values on circumstances) *Source: SOEPv33, 1992–2016*.

Figure A7: IOP estimates using 10 quantiles of the type specific outcome distribution *Source: SOEPv33, 1992–2016. Note: The solid line shows the correlation between the share of observations with missing values and IOP over the entire period 1992-2016. The two dashed lines show the correlations for the sub-periods 1992–2001 and 2002–2016. Sample includes observations with non-zero weight from all SOEP-samples, excluding the two refugee samples added in 2016*.

Appendix B Opportunity Tree and CDFs for single years

Appendix C Predictive accuracy of conditional inference regression trees

Figure C8: In and out-of-sample MSE for conditional inference regression trees *Source: SOEPv33, 1992–2016*.

Figure C9: In and out-of-sample MSE: conditional inference trees Vs. random forest *Source: SOEPv33, 1992–2016*. Note: Prediction based on conditional inference random forests are obtained by running a forest of 500 conditional inference regression trees with $\alpha = 0$. Error bars are 99% confidence interval based on the quantiles of 500 bootstrap replications of the statistics.