Social Choice and Welfare
https://doi.org/10.1007/s00355-018-1165-x

ORIGINAL PAPER

Upward and downward bias when measuring inequality of opportunity

Paolo Brunori[1,2] · **Vito Peragine**[2] · **Laura Serlenga**[3]

Received: 5 March 2018 / Accepted: 17 November 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract
Estimates of the level of inequality of opportunity have traditionally been proposed as lower bounds due to the downward bias resulting from the partial observability of circumstances that affect individual outcome. We show that such estimates may also suffer from upward bias as a consequence of sampling variance. The magnitude of the latter distortion depends on both the empirical strategy used and the observed sample. We suggest that, although neglected in empirical contributions, the upward bias may be significant and challenge the interpretation of inequality of opportunity estimates as lower bounds. We propose a simple criterion to select the best specification that balances the two sources of bias. Our method is based on cross-validation and can easily be implemented with survey data. To show how this method can improve the reliability of inequality of opportunity measurement, we provide an empirical illustration based on income data from 31 European countries. Our evidence shows that estimates of inequality of opportunity are sensitive to model selection. Alternative specifications lead to significant differences in the absolute level of inequality of opportunity and to the re-ranking of a number of countries, which confirms the need for an objective criterion to select the best econometric model when measuring inequality of opportunity.

✉ Paolo Brunori
paolo.brunori@unifi.it

Vito Peragine
vitorocco.peragine@uniba.it

Laura Serlenga
laura.serlenga@uniba.it

[1] Dipartimento di Scienze per l'Economia e l'Impresa, University of Florence, Via delle Pandette, 32, 50127 Florence, Italy

[2] University of Bari, Bari, Italy

[3] University of Bari and IZA, Bari, Italy

Published online: 23 November 2018

1 Introduction

The measurement of inequality of opportunity (IOp hereafter) is a growing topic in economics, and in the past two decades, the number of empirical contributions to this literature has increased substantially: see Ferreira and Peragine (2016), Roemer and Trannoy (2015), and Van de gaer and Ramos (2016) for a review. The vast majority of these contributions are based on the approach proposed by Roemer (1998) and follow a two-step procedure. First, a counterfactual distribution is derived from an outcome distribution (typically income or consumption). This counterfactual distribution reproduces only unfair inequalities, i.e. inequalities due to circumstances beyond individual control, and does not reflect inequality arising from individual choice and effort. Second, a suitable inequality measure is used to quantify inequality in the counterfactual distribution.

The empirical literature has extensively used two methods to compute counterfactual distributions of survey data: parametric and non-parametric methods. One of the main drawbacks of both approaches is that, unless all the circumstances beyond an individual's responsibility are observable, they produce biased estimates of IOp. While the magnitude of this bias may be impossible to determine (Bourguignon et al. 2013), under some assumptions discussed in the literature, it can be shown that the sign of the bias is negative (Roemer 1998; Ferreira and Gignoux 2011; Luongo 2011). This explains why IOp estimates are generally interpreted as lower-bound estimates of the " true" IOp, whereas the true IOp is interpreted as the estimate one would obtain if all circumstances were observable. The usefulness of those lower-bound measures has been challenged in recent literature, see Kanbur and Wagstaff (2016), Balcazar (2015), and Wendelspiess (2015). In particular, Balcazar (2015) and Ibarra et al. (2015) suggest that the downward bias may lead to a substantial underestimation of the true IOp in empirical applications.

Typically, authors address this problem by using richer data sources and by adopting a variety (or a combination) of empirical strategies: (i) by increasing the number of circumstances, as in Björklund et al. (2012); (ii) by introducing interaction terms among different circumstances, as in Hufe and Peichl (2015); (iii) by splitting the population into finer partition of types.

These empirical strategies reduce the downward bias, increasing the explained variability attributable to IOp. In this paper, we emphasize that these procedures are not riskless and might lead to an upward distortion of IOp estimates. Indeed, the reliability of the estimates depends not only on the number of circumstances and the population partition, but also on the sample distribution across types.

In both parametric and non-parametric approaches, we recognize a trade-off between the downward bias resulting from the observability of circumstances and the upward bias related to the sampling variance of the estimated counterfactual distribution. Although this topic is not new to econometricians and practitioners, our concern over upwardly biased IOp estimates has been neglected in the empirical literature of IOp measurement. This is surprising because, as shown in the empirical section, such a distortion is likely to be far from negligible. We show that the magnitude of the upward distortion depends upon the strategy used to obtain the counterfactual distribution. This problem is particularly straightforward when applying a parametric approach but can easily be generalized to the non-parametric method. The number of explanatory variables involved and the division into types may lead to distortions in both directions. Overfitted models result in upward bias, whereas underfitted models reinforce the well-known downward bias caused by partial observability. We suggest that, when choosing among alternative specifications, scholars should opt for the best balance between the two sources of bias, and we propose a method to select the best econometric specification that minimizes the sum of the two biases. Our method is based on cross-validation (CV hereafter), which is a methodology commonly adopted by statisticians to evaluate the performance of predictive models and is increasingly used by economists (Varian 2014). CV directly provides a nearly unbiased measure of the out-of-sample prediction error. The major interest of CV lies in the minimal assumptions required to obtain unbiased measures of model performance (Arlot and Celisse 2010).

The out-of-sample prediction error is estimated by dividing the original sample into training and test sets. The association between circumstances and outcome is first estimated on the training sample under a large number of meaningful model specifications. Next, the derived coefficients are used to predict the outcome on the test sample. The specification selected is the model that, on average, minimizes the prediction error in the test sample.

Because minimizing the prediction error is equivalent to maximizing the ability of the model to explain the variability of the dependent variable out-of-sample, the proposed criterion does minimize the downward bias due to partial observability out-of-sample. In other words, using CV, we select the model specification that minimizes the downward bias without overfitting the data.

To demonstrate the usefulness of our approach, we apply our method to income data from 31 European countries using the European Union Survey on Income and Living Conditions (EU-SILC) 2011 database. Our evidence shows that IOp estimates are extremely sensitive to model selection. Alternative specifications lead to significant differences in the absolute level of IOp, and in many cases, to the re-ranking of countries. Since our preferred specification is different from what is typically used in the literature, our estimates differ from those provided by other authors who use the same data to estimate IOp.

The rest of this paper is organized as follows: Sect. 2 introduces the canonical model used to measure IOp, presents the estimation methods used to implement it, and clarifies the two possible sources of distortion. Section 3 proposes a criterion to balance the trade-off between the two types of bias when selecting the specification to estimate IOp. Section 4 presents an empirical implementation, and Sect. 5 concludes.

2 Downward and upward biased IOp

The canonical equality of opportunity model can be summarized as follows (see Ferreira and Peragine 2016). Each individual in a society realizes an outcome of interest, y, by means of two sets of characteristics: circumstances beyond individual control, C, belonging to a finite set $\Omega = \{C_1, \ldots, C_J\}$, and a responsibility variable, e, typically treated as scalar. A function $g : \Omega \times \Re_+ \rightarrow \Re_+$ defines the individual outcome:

$$y = g(C, e). \tag{1}$$

For all $j \in \{1, \ldots, J\}$, let us denote by K_j the possible values taken by circumstance C_j and by $\left|K_j\right|$ the cardinality K_j. For instance, if C_j denotes gender, then $K_j = \{male, female\}$. We can now define a partition of the population into T types, where a type is a selection of values, one for each circumstance, that is, $T = \Pi_{i=j}^{J} \left|K_j\right|$. Let us denote by Y the overall outcome distribution.

The IOp is then defined as the inequality in the counterfactual distribution, $\tilde{Y}$, which reproduces all inequalities due to circumstances and does not reflect any inequality due to effort. A number of methods have been proposed to obtain $\tilde{Y}$, and in general, the selected method affects the resulting IOp measure (Ferreira and Peragine 2016; Roemer and Trannoy 2015; Van de gaer and Ramos 2016). In what follows, we focus on the *ex ante* approach introduced by Bourguignon et al. (2007) and Checchi and Peragine (2010), which is by far the most commonly adopted method in the empirical literature (Brunori et al. 2013).[1] This approach interprets the type-specific outcome distribution as the opportunity set of individuals belonging to each type. Then, a given value v_t of the opportunity set of each type is selected. Finally, $\tilde{Y}$ is obtained by replacing the outcome of each individual belonging to type t with the value of her type v_t, for all $t = 1, \ldots, T$.

2.1 Counterfactual estimation

Ex ante IOp can be estimated by either a non-parametric or a parametric approach. Checchi and Peragine (2010) propose non-parametric estimation of $\tilde{Y}$ following the typical two-stage method: (i) after partitioning the sample into types on the basis of all observable circumstances, they choose the arithmetic mean of the outcome of type t, denoted by μ_t, as the value v_t of type t; (ii) for each individual i belonging to type t, they define $\tilde{y}_i = \hat{\mu}_t$—where $\hat{\mu}_t$ is the sample estimate for μ_t—and measure the inequality in $\tilde{Y}$.

Alternatively, Bourguignon et al. (2007) propose parametric measurement of *ex ante* IOp by estimating $\tilde{Y}$ as the prediction of the following reduced form regression:

$$y_i = \sum_{j=1}^{J} \sum_{k=1}^{K_j} \chi_{j_k} c_{i j_k} + u_i, \tag{2}$$

---

[1] Other well-established approaches can be used to measure IOp. Approaches differ in how they define the principle of equal opportunity and in the way the counterfactual distribution is constructed (Roemer 1998; Lefranc et al. 2009; Fleurbaey and Schokkaert 2009; Checchi and Peragine 2010). However, because the construction of these alternative counterfactual distributions generally requires the observation or identification of effort (an extremely difficult variable to measure), they are less frequently adopted in the empirical literature.

where c_{ij_k} identifies each category of the observable characteristics by means of a dichotomous variable, and χ_{j_k} is the corresponding coefficient.[2] In the original specification, the parametric approach consists of ordinary least squares regression where the total outcome variability is explained by a linear combination of regressors with no interaction terms.[3] Hence, the parametric approach does not estimate the counterfactual distribution, $\tilde{Y}$, by directly identifying types. It linearly approximates the types' average outcome by the predictions of a regression of circumstances on outcome. This approach has the main advantage of being more parsimonious than the non-parametric approach. In practice, parametric estimations have been proposed as a reasonable alternative to non-parametric estimation when few observations are available, see Ferreira and Gignoux (2011) and Ibarra et al. (2015). However, parsimony comes at the cost of imposing the effect of the circumstances on outcome to be fixed and additive. For example, being a women is assumed to have an effect on earning that is independent of all the other circumstances, such as socioeconomic background or race. This assumption constrains the ability of regressors to capture outcome variability.

Recently, Hufe and Peichl (2015) discuss the importance of considering interaction terms in estimating IOp. They estimate *ex ante* IOp using the Child & Young Adults Supplement of the National Longitudinal Survey of Youth and alternative model specifications. They implement both a linear model, as in (2), and a non-linear model, where circumstances fully interact, and they acknowledge a critical divergence of the IOp estimates among the different specifications.

Indeed, it is important to note that the parametric and the non-parametric methods coincide when all explanatory variables are categorical, and the parametric counterfactual distribution is obtained by the prediction of a regression model where y is regressed on all possible combinations of circumstance values, i.e. all values of all regressors interact with each other to obtain a model with $T = \Pi_{i=j}^{J} \left|K_j\right|$ dummies. In this particular case, each regressor captures the effect of belonging to one of all the possible circumstance combinations, which is the effect of belonging to a given type. The estimated model becomes:

$$y_i = \sum_{t=1}^{T} \beta_t \pi_{it} + u_i, \tag{3}$$

where π_{it} are T binary variables obtained by interacting all values of the circumstances. Clearly, the typical (linear) parametric approach, (2), explains less inequality than the non-parametric approach, (3), simply because model (3)—by construction—allows variability to be explained by the full set of interactions.

Here, a trade-off emerges: while the linear specification might be too restrictive, the inclusion of the full set of combinations of the circumstances' values might lead to very large sampling variance of the estimated counterfactual distribution, especially when a limited number of observations is available for certain types.

---

[2] In principle, if cardinal circumstances are observed, regressors might be non-categorical. However, to the best of our knowledge in the empirical literature, this is never the case. Even if cardinal measures are available, i.e. parental income, authors tend to use categorical regressors for the quantiles of the continuous distribution (see, for example, Björklund et al. 2012).

[3] Analogously to the Mincer equation, a log-linear specification is preferred by the majority of the authors. (Ferreira and Gignoux 2011)

Following the same reasoning, the sampling variance of the estimated counterfactual distribution is also influenced by alternative population partitions: a broadest partition might, again, lead to larger variance in the case of a limited number of observations per type.

Indeed, the reliability of both parametric and non-parametric IOp estimates requires a sufficient number of observations characterizing each circumstance. Specifically, the limitation might be more severe in the case of the non-parametric approach, where a sufficient number of observations for each combination of circumstances is required. This might represent a serious constraint in empirical applications; in survey data, individuals are unlikely to be uniformly distributed across types and across population partitions. For example, a typical argument arises when considering Western countries in which researchers observe both parental education and parental occupation as circumstances. Those variables are usually strongly correlated with each other, i.e. there are very few individuals whose parents are highly educated and employed in elementary occupations or who have no education but work as managers. To overcome this drawback, scholars tend to consider a limited number of circumstances in the definition of types (using either parental education or parental occupation) or aggregate the different values that a circumstance might take (using blue and white collars rather than more specific occupation value). These are clearly ad hoc solutions, which might greatly affect the shape of the counterfactual distribution and lead to misleading IOp estimates. In what follows, we propose a statistical criterion to properly select among different model specifications or alternative population partitions.

2.2 Bias-variance trade-off in estimating IOp

A number of methodological contributions have shown that if the ‘true’ set of circumstances is not fully observable, the estimated *ex ante* IOp will be lower than the ‘real’ IOp (Roemer 1998; Ferreira and Gignoux 2011; Luongo 2011). This result follows from the assumption of orthogonality between circumstances and effort (see Roemer 1998) and explains why IOp measures are generally interpreted as lower-bound estimates of IOp.

Authors often attempt to solve this problem by using rich datasets that contain the largest possible number of circumstances, including outcome obtained during childhood (Björklund et al. 2012; Hufe et al. 2017). Recently, Niehues and Peichl (2014) endorse an extreme perspective. By exploiting longitudinal datasets, they measure IOp, including individual fixed effects among circumstances beyond individual control, implying that any unobservable individual characteristic that persists over time is considered a source of IOp. Understandably, this method has been, proposed as an ‘upper-bound’ estimate of the true IOp.

However, interpreting IOp estimates as lower bounds is correct only if the entire population of interest is observed. When using survey data, attempting to reduce the downward bias by increasing the number of circumstances or the number of values within each circumstance results in a counterfactual distribution based on a finer partition into types. By construction, this process results in a smaller number of observations in each type,[4] which might increase the sampling variance when estimating the counterfactual distribution.

Surprisingly, the empirical literature on IOp estimation has neglected this second implication thus far. Only recently, the issue has gained importance in the debate on the IOp measurement. Brunori et al. (2016) note that the use of very detailed circumstances, such as hundreds of 'villages of birth' in Madagascar or hundreds of 'ethnic groups' in Congo, tends to dramatically increase the IOp estimates.[5]

Crucially, when measuring inequality, higher sampling variance of the estimated distribution implies an upward-biased IOp measure. This result is easily shown by applying what Chakravarty and Eichhorn (1994) proved for the case of inequality estimation when the variable of interest is measured with error. It turns out that in the IOp framework, higher sampling variance of the estimated type mean might be due to a finer population partition rather than to the classic measurement error. A formal proof based on Chakravarty and Eichhorn (1994) is available in Appendix A.[6]

This result has two interesting consequences in the measurement of IOp. First, it states that if all circumstances are observable and IOp is measured on an appropriate subsample of the original population, such as a typical representative survey, IOp is upward biased. Second, whenever circumstances are not fully observable, two opposite distortions might bias our estimates; hence we can no longer claim that the estimated IOp is a lower-bound of the true IOp.

When the sample size is large relative to the number of circumstances included in the model, the downward bias is likely to be considerable. However, when the sample size is small relative to the number of types/regressors, upward bias might prevail. Simulations in Appendix B illustrate the possible relevance of the upward bias in small samples. Indeed, the absolute and relative sizes of the two biases depend upon a number of factors: the sample size, the joint distribution of outcome and circumstances, and the model specification used to estimate the counterfactual distribution. In other words, it is ultimately an empirical issue.

This discussion should clarify that when estimating IOp, we should aim at minimizing two different sources of distortion that bias our estimates in opposite directions: partial observability and sampling variance of the counterfactual distribution. The solution to minimize the downward distortion cannot consist of ad hoc strategies such as simply including a larger number of circumstances or considering a broad partition of the population. The choice of the researcher should be based on a statistical criterion. In the following section, we propose a simple method for selecting the best model to measure IOp, a method that exploits the information contained in survey data and minimizes the distortion due to the two biases.

---

[4] Or, if adopting a parametric approach, regression with a larger number of controls and fewer degrees of freedom.

[5] Note also that the approach proposed by Li Donni et al. (2015), although not explicitly discussed by the authors, represents a possible strategy to address this issue. They define Roemerian types using latent class analysis. That is, they assume that observable circumstances are manifestations of an unobservable membership to a number of latent groups. Their method reduces the number of types and hence avoids large sampling variance in the counterfactual distribution.

[6] In a framework where the outcome is measured with error and the sampling variance of the counterfactual distribution is ignored, Wendelspiess (2015) predicts the opposite direction of bias.

3 Model selection for measuring IOp

Since Bourguignon et al. (2007), IOp has been measured using a reduced form model, with no assumptions about the functional form of Eq. (1). We follow this strand of the literature and do not impose any a priori restriction on the effect that circumstances might have on outcome. Hence, in this section, we propose a method to select the most suitable model among all possible alternative specifications.

The aim is to select the model specification producing the most reliable measure of inequality in $\tilde{Y}$. To avoid downward bias, and not knowing which circumstance determine the outcome and how, we would like to specify the most flexible possible models including all observable circumstances and their interactions. That is, we would like to maximize the variability in y that can be explained by C and their interactions. In the observed sample, the outcome variability that can be explained by a set of controls is monotonically increasing with the number of estimated parameters. With a minimal number of observable circumstances, we can easily obtain a model that perfectly fits the data. In this way, IOp is equal to total inequality, and the downward bias is eliminated. However, researchers are generally interested in estimating the level of IOp in a population and not in a particular observed sample. Using a model with zero degrees of freedom will produce unreliable estimates of $\tilde{Y}$ and, as proved above, upward biased IOp estimates. We claim here that the most appropriate model is a specification that minimizes the out-of-sample downward bias due to partial observability. That is, the model that maximizes the variability of y that can be explained by C if a different random sample from the same population was observed.

Viewed from this perspective, the estimation of IOp can be understood as a prediction problem. When willing to predict an outcome based on a set of controls, we try to fully exploit the informative content of the observed data. However, with a finite set of observations, increasing the complexity of the model implies loosing confidence on the parameters estimated. Therefore, increasing the model complexity, we improve our predictive performance only if the gain in terms of model flexibility is larger than the loss in terms of parameters reliability. Too complex models will precisely explain the outcome variability in sample, but will poorly predict out-of-sample. Similarly, too simple models will neglect important information contained in the controls and will imprecisely predict both in sample and out-of-sample. The aim of a statistical learning model is to maximize out-of-sample predictive performance trading-off the need of minimizing both sources of error.

Machine learning practitioners select the model that minimizes the out-of-sample (squared) error. This model selection approach is gaining attention in the social sciences literature and increasingly adopted by economists (Varian 2014; Mullainathan and Spiess 2017; Athey 2018). Similarly, we propose to select the most appropriate model to estimate IOp minimizing the same loss function. In fact, maximizing the predictive accuracy of a model is equivalent to maximizing its ability to explain the variability of the dependent variable out-of-sample.[7]

The problem can be formally illustrated exploiting the decomposability of the mean squared error (MSE). MSE is defined as:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left(y_i - \hat{g}(C_i)\right)^2,$$

where y is the dependent variable, C is the vector of controls, and $i = 1, \ldots, n$ are the observations. For given out-of-sample observations y_0 and C_0, the MSE can be understood as the sum of three components:

1. the *variance*, $Var(\hat{g}(C_0))$, which depends on how much the estimated relationship between dependent variable and controls would change if a different sample from the same population was observed;
2. the *bias* squared, $\left[Bias\left(\hat{g}(C_0)\right)\right]^2$, which depends on how the chosen model, adopting simplifying assumptions about the data generating process, constraints our ability to correctly capture the relationship between C and y;
3. the variance of the *irreducible error term*, $Var(u)$, which captures the variability in y independent from C and represents the lowest possible error we can make in predicting y.

$$E\left(y_0 - \hat{g}(C_0)\right)^2 = Var(\hat{g}(C_0)) + \left[Bias\left(\hat{g}(C_0)\right)\right]^2 + Var(u), \tag{4}$$

Note that Eq. (4) formally connects the two sources of bias when measuring IOp with the expected prediction error out-of-sample of a model explaining y as a function of C. The downward bias due to unobservable circumstances (and neglected interactions), largely discussed by the literature, is captured by the *bias*. The upward distortion, discussed in Sect. 2.2 and proved in Appendix A, is taken into account by the *variance*. The most reliable IOp estimate should be based on the model that minimizes the sum of the first two sources of error. Since the variance of the prediction error, $Var(u)$, cannot be reduced, this is equivalent to minimizing the out-of-sample MSE of the model.

Minimizing MSE is the fundamental problem of statistical learning. The magnitude of *bias* and *variance* depends on the complexity of the model specified. On the one hand, to minimize the *variance* we should maximize the available degrees of freedom. For a fixed set of observations, this implies specifying the simplest possible model. The model with lowest possible variance is a model that contains a single parameter. In such a case, the only parameter estimated is the sample mean, circumstances have no effect on outcome, and IOp is zero (we are minimizing the upward bias at the cost of the largest possible downward bias). On the other hand, in order to reduce the *bias*, we would like to specify the most complex possible model. This would result in a very general model, estimated with low level of confidence (the upward bias will be large). This trade-off, known as the bias-variance dilemma, can be solved weighting equally the first two components of Eq. (4) and minimizing the sum of the two, that is, minimizing the out-of-sample MSE. This criterion selects the model that maximizes the ability of circumstances to explain the outcome variability out-of-sample, and produces the largest possible IOp estimate out-of-sample.

---

[7] Based on our conclusion Brunori et al. (2018) have recently compared popular econometric approaches to estimate IOp. Their analysis shows that conditional inference random forests, a machine learning algorithm introduced by Hothorn et al. (2006), outperforms other methods in predicting IOp out-of-sample.

The implementation of such a criterion for selecting the best model is comparative and involves two steps: first, we estimate a number of alternatives, such as model (2), model (3), and all the specifications obtained by both interacting only a subset of circumstances and using different population partitions; second, we choose the best specification by means of cross validation (CV).[8]

The CV originates from the validation approach, which is a data-driven model selection criterion. Historically, the validation approach has been proposed as a method to assess models' performance by avoiding to incur in overoptimistic conclusions (Larson 1931). Once a model has been estimated, the assessment of its performance should ideally rely on new data, such as out-of-sample observations. However, since those observations are rarely available, other solutions have been proposed. In the validation approach, for instance, observed data are randomly divided into two subsamples, one is used to estimate the models (training set) and the other used to evaluate models' accuracy (validation or test set). Indeed, the test set can play the role of unseen observations as long as it has been randomly drawn from the original sample. Hence, in the validation approach, the model is fitted on the training set, saving the estimated prediction function, and then used to predict the dependent variable into the hold-out test set. The MSE estimated in the hold-out set directly provides a nearly unbiased measure of the out-of-sample prediction error (Arlot and Celisse 2010). However, as widely acknowledged, this validation approach has, at least, two main drawbacks. First, the MSE estimates tend to heavily depend on the observations included in the training set and those that are held-out. Second, it does not fully exploit all available information. In fact, only observations included in the training set are used to fit the model (Gareth et al. 2013).

In order to overcome these two downsides other resampling assessment methods, based on the same validation idea, have been proposed in the literature. For instance, in k-fold CV, the sample is randomly divided into k equal-sized parts. Leaving out part k (test sample), the model is fitted to the other $k-1$ parts (training sample), and out-of-sample predictions are obtained for the left-out kth part. For each specification, the average of the k MSEs is stored and the best specification is selected by minimizing the average MSE. CV is commonly adopted by statisticians to evaluate the performance of predictive models as it lies in the minimal assumptions required to obtain unbiased measures of model performance (Hastie et al. 2009).

Also, notice that minimizing MSE estimated by CV is asymptotically equivalent to minimizing the Akaike's Criterion and (Stone 1977). Similarly, the Schwarzs Bayesian

---

[8] We are aware that the number of alternative models exponentially increases when circumstances are interacted. Moreover, researchers might have the choice to consider some circumstances with different levels of aggregation, e.g. country/region/district of birth. In these cases, our method should be complemented with an algorithm that can restrict the number of models considered, for example, best subset selection or stepwise selection, see Gareth et al. (2013).

Information Criterion is asymptotically equivalent to k-fold CV for a particular value of k (Shao 1997).

A not negligible aspect of using CV is the choice of k. The validation approach is close to CV when $k = 2$. At the opposite extreme leave-one-out CV uses a number of folds equal to the number of observation in the original sample. The choice of k leads again to a trade-off: a low k tends to select models characterized by low variance and high bias. A large k tends to select flexible model with relatively high variance. As a rule of thumb machine learners practitioners tend to choose a value of k between five and ten. The exact choice depending on the sample size, the number of models to test, and the type of problem under scrutiny (Kohavi 1995; Hastie et al. 2009; Rodríguez et al. 2010).

Using *k-fold* CV to select the most appropriate specification to measure IOp might imply the use of an alternative model for data sourced by the same country but in different time periods and, in general, each time the country's sample differs. As a consequence, when comparing different countries in terms of IOp, we might compare measures obtained with different specifications. This is in contrast with what is generally proposed in the literature. In practice, so far, when the same source of data is available for different countries, comparable measures of IOp have usually been computed using the same model specification for all countries, see Marrero and Rodrguez (2012), Brzenziński (2015), Checchi et al. (2016), and Suárez and Menéndez (2017). Here, we suggest a different approach: comparable IOp measures should be calculated using the best performing model given the observable circumstances. As a simple example, let us consider the comparison between France and Belgium in terms of IOp. Including 'mother tongue' among circumstances in France would probably make little sense: it would not explain much of the outcome inequality in the country, but regressors although not statistically significant, will not be exactly equal to zero. The model will therefore capture too much variability in the sample, this variability is due to sampling variance, but will be interpreted as IOp. However, the same circumstance is likely to be an important source of opportunity inequality in Belgium. Hence, we might infer that 'mother tongue' should be excluded for France and included for Belgium.

We consider our method to be preferable when the intent is to compare the level of IOp in two populations. The derived IOp measures would be the two most reliable estimates of the effect of circumstances on outcome, given the information available and the statistical relevance of the characteristics that influence IOp. We believe that the specification used may differ for at least two reasons: first, because the set of available information may not be the same for the two populations; second, and most importantly, because the nature of opportunity inequality, i.e. how circumstances affect individual outcomes, may differ in the analysed populations.

4 An empirical illustration

In this section we provide an empirical illustration based on the EU-SILC 2011 dataset. We show that our method is easily implementable and can substantially improve our understanding of IOp. The EU-SILC is a reference source for comparative statistics on income distribution in the European Union. Because of special module on the intergenerational transmission of poverty included in a number of EU-SILC waves, the same data have been exploited for other estimates of IOp in the past, see Suárez and Menéndez (2017), Marrero and Rodrguez (2012), Brzenziński (2015), and Checchi et al. (2016). The 2011 is the most recent wave, which contains information on family of origin and socioeconomic background. The data refer to 31 European countries.[9] In this analysis, we restrict the EU-SILC sample to households whose head is between 26 and 60 years old. The outcome variable is the equivalized disposable income, obtained by dividing total household disposable income by the square root of the household size. The circumstances are categorical and identify area of birth and family background (summarized by retrospective questions about parental education and occupation when the respondent was 14 years old). In selecting the best specified model, we consider all possible models ranging from the most parsimonious to a non-parametric partition based on a large number of types.

In the most parsimonious model, we regress the outcome on four regressors with no interactions: country of origin (a binary variable that takes the value of one if the respondent was born in the country of residence), father's and mother's occupation (white or blue collar),[10] and parental education (low/high).[11]

These variables are initially coded into a larger number of values: mother's and father's occupation in 10 values,[12], mother's and father's education in five values[13] and area of birth in three values (native, born in Europe, and born outside Europe). Interacting all variables coded under the maximum level of detail would result in 7500 types, a number far greater than the average sample size in EU-SILC. Hence, under the broadest sample partition, we opt for a more compact definition, where country of origin is divided in two values, father's occupation into 10 values, mother's occupation into two values, father's education into four values, and mother's education into four values. This model results in 640 possible types. Table 1 shows descriptive statistics. Intermediate models include subsets of interactions.

Figure 1 shows the level of IOp in the 31 countries. Each bar indicates the mean logarithmic deviation (MLD) of the counterfactual distribution. For each country, the three bars refer to the following cases: (i) the model described in Eq. (1), $(linear)$; (ii)

---

[9] Austria (AT), Belgium (BE), Bulgaria (BG), Switzerland (CH), Cyprus (CY), Czech Republic (CZ), Germany (DE), Denmark (DK), Estonia (EE), Greece (EL), Finland (FI), France (FR), Croatia (HR), Hungary (HU), Ireland (IE), Italy (IT), Iceland (IS), Latvia (LV), Lithuania (LT), Luxembourg (LU), Malta (MT), the Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Spain (ES), Slovakia (SK), Slovenia (SI), Sweden (SE), and the United Kingdom (UK).

[10] Those are based on the International Standard Classification of Occupations, published by the International Labour Office ISCO-08. Blue collar includes parents that who do not work or were occupied as: clerical support workers; service and sales workers; skilled agricultural, forestry and fish; craft and related trades workers; plant and machine operators; elementary occupations.

[11] Education categories are based on the International Standard Classification of Education 1997 (ISCED-97). When coded into two, low includes ISCED below level 3.

[12] ISCO-08 1-digit: armed forces occupations; managers; professionals; technicians and associate professionals; clerical support workers; service and sales workers; skilled agricultural, forestry and fish; craft and related trades workers; plant and machine operators; elementary occupations; did not work/unknown father/mother

[13] Unknown father/mother, could neither read nor write; low level (ISCED 0-2); medium level (ISCED 3-4); high level (ISCED 5-6).

Table 1 Descriptive statistics . Source: EU-SILC (2011)

Country	Sample size	Eq. income	Tot. Ineq. (MLD)	Age	Native	Mother			Father		
						White collar	Elementary edu.	Secondary edu.	White collar	Elementary edu.	Secondary edu.
AT	6630	17,024.37	0.167	43.96	0.788	0.283	0.563	0.350	0.363	0.379	0.437
BE	5122	14,581.69	0.150	43.14	0.825	0.260	0.524	0.222	0.454	0.463	0.211
BG	6651	2114.25	0.246	44.25	0.996	0.421	0.452	0.382	0.236	0.448	0.361
CH	7212	28,180.78	0.186	44.58	0.683	0.359	0.398	0.413	0.486	0.226	0.490
CY	5135	11,758.22	0.172	43.48	0.787	0.198	0.654	0.191	0.310	0.638	0.201
CZ	7068	5576.24	0.133	44.26	0.963	0.523	0.650	0.277	0.322	0.597	0.219
DE	11,473	15,797.95	0.173	45.27	0.881	0.371	0.274	0.511	0.439	0.120	0.529
DK	2233	22,189.83	0.229	45.76	0.928	0.570	0.512	0.290	0.461	0.341	0.429
EE	5464	3783.51	0.234	43.95	0.869	0.554	0.303	0.409	0.268	0.279	0.366
EL	6430	7416.41	0.320	43.65	0.892	0.162	0.585	0.155	0.310	0.579	0.156
ES	16,188	10,231.74	0.303	43.65	0.835	0.137	0.801	0.057	0.342	0.761	0.071
FI	3439	18,087.89	0.137	43.90	0.949	0.525	0.458	0.266	0.313	0.413	0.210
FR	11,286	15,448.28	0.174	44.15	0.890	0.336	0.721	0.089	0.397	0.700	0.083
HR	6720	3557.79	0.234	45.36	0.884	0.234	0.648	0.225	0.283	0.474	0.360
HU	13,533	3240.24	0.167	44.43	0.989	0.379	0.636	0.262	0.241	0.595	0.263
IE	3360	15,220.27	0.188	43.08	0.780	0.212	0.524	0.338	0.360	0.562	0.268
IS	1619	12,815.69	0.104	43.20	0.907	0.494	0.609	0.281	0.449	0.318	0.499
IT	22,696	12,036.84	0.299	43.88	0.875	0.165	0.762	0.125	0.315	0.693	0.147
LT	4935	3033.07	0.290	46.36	0.945	0.380	0.493	0.350	0.220	0.550	0.269

Table 1 continued

Country	Sample size	Eq. income	Tot. Ineq. (MLD)	Age	Native	Mother			Father		
						White collar	Elementary edu.	Secondary edu.	White collar	Elementary edu.	Secondary edu.
LU	7244	22,727.56	0.170	43.18	0.484	0.233	0.582	0.256	0.374	0.475	0.327
LV	6748	3081.47	0.288	44.19	0.880	0.486	0.382	0.429	0.201	0.359	0.324
MT	4467	7327.47	0.158	44.23	0.950	0.064	0.662	0.154	0.436	0.571	0.189
NL	5884	17,126.15	0.125	44.32	0.880	0.259	0.503	0.301	0.517	0.355	0.294
NO	2568	27,770.11	0.107	43.81	0.915	0.562	0.343	0.440	0.507	0.309	0.391
PL	14,595	3363.28	0.221	44.14	0.999	0.313	0.480	0.448	0.221	0.426	0.483
PT	6355	6235.56	0.222	44.42	0.907	0.177	0.646	0.034	0.268	0.712	0.037
RO	6549	1504.92	0.264	43.82	0.998	0.180	0.760	0.141	0.147	0.760	0.103
SE	543	16,165.57	0.127	42.79	0.863	0.593	0.473	0.235	0.436	0.491	0.183
SI	5234	7883.03	0.102	43.25	0.872	0.347	0.729	0.166	0.287	0.676	0.175
SK	7241	4139.78	0.151	43.36	0.989	0.490	0.418	0.516	0.287	0.334	0.526
UK	6329	14,177.45	0.251	44.05	0.851	0.443	0.667	0.111	0.462	0.507	0.238

Descriptive statistics refer only to respondents with no missing information. White collar occupations include: Managers; Professionals; Technicians and associate professionals (ISCO-08)

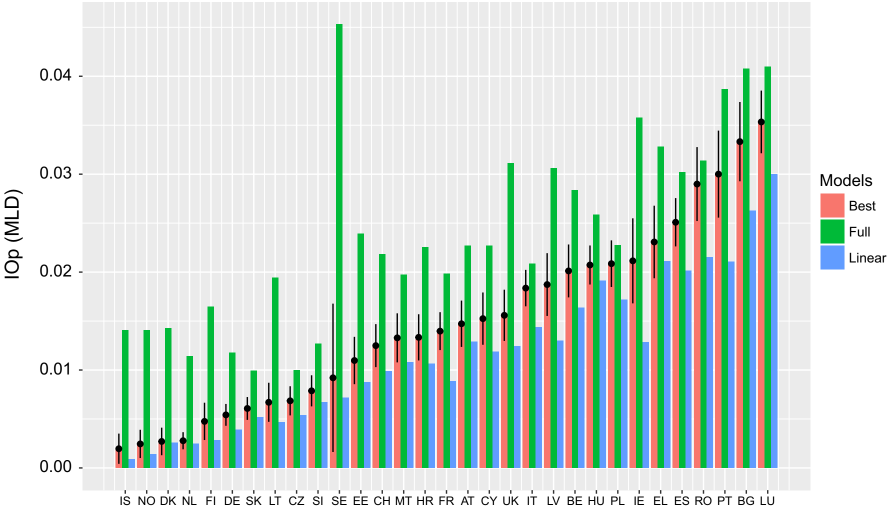

Fig. 1 IOp in 31 European countries under different model specifications. The Figure shows each country's IOp measure obtained with the three alternative methods: (i) the linear, most parsimonious case ($linear$), (ii) the fully interacted model ($full$); (iii) the best model selected ($best$). Countries are ordered according to the IOp level based on the $best$ model specification with 95% confidence intervals. Table 2 in the Appendix contains IOp estimates and relative bootstrapped standard errors based on 500 replications for the three alternative model specifications . *Source: EU-SILC, 2011*

the model described in Eq. (2), ($full$); (iii) an intermediate measure computed from the best model selected by CV (best). k-fold CV is performed by the routine written by Daniels (2012). The number of folds is five for all countries. However, increasing the number of folds up to 10, the model specifications selected appear to be rarely affected.

The three alternative measures clearly differ among each other, and in some cases (mostly on the left), the best model is very close to the linear model (Denmark and Netherlands, for example). These are mainly Nordic countries characterized by a low level of IOp. Note also that for the same countries, the difference in IOp measured with the $linear$ specification and IOp measured with the $full$ model is substantial. This large gap between the two extremes, together with the low level of covariance of circumstances and outcome, is due to the small sample sizes for these countries. When the sample size is limited, such as for Sweden and Iceland, overfitting occurs, even for relatively simple model specifications, and the upward bias discussed above tends to be more pronounced. Interestingly, for Italy, Poland, and Hungary, the three countries with the largest sample sizes, the difference between the two models tends to be small. It might be the case that with a sample larger than 12,000, the problem of upward bias becomes less relevant. The role of sample size in determining the magnitude of the bias is yet analysed in Appendix B by means of simulations.

In other cases (concentrated on the right-hand side of the graph), the best model is far from the linear specification and rather close to the most flexible specification. In particular, in Italy, Poland, Romania, Portugal, Bulgaria, and Luxembourg, our preferred estimate is closer to the full model than to the linear.

Table 2 IOP (MLD) Estimates of 31 countries . Source: EU-SILC 2011

Country	Best	Best low	Best high	Linear	Linear low	Linear high	Full	Full low	Full high
AT	0.0147	0.0124	0.0171	0.0129	0.0108	0.0149	0.0227	0.0197	0.0256
BE	0.0201	0.0174	0.0228	0.0164	0.0137	0.0191	0.0284	0.0247	0.0320
BG	0.0333	0.0293	0.0374	0.0263	0.0221	0.0304	0.0408	0.0361	0.0454
CH	0.0125	0.0103	0.0147	0.0099	0.0080	0.0117	0.0218	0.0189	0.0247
CY	0.0152	0.0126	0.0179	0.0119	0.0092	0.0145	0.0227	0.0194	0.0260
CZ	0.0069	0.0054	0.0083	0.0054	0.0040	0.0068	0.0100	0.0082	0.0118
DE	0.0054	0.0043	0.0065	0.0039	0.0030	0.0048	0.0118	0.0103	0.0133
DK	0.0027	0.0013	0.0041	0.0026	0.0011	0.0041	0.0143	0.0110	0.0176
EE	0.0110	0.0086	0.0134	0.0087	0.0064	0.0111	0.0239	0.0200	0.0279
EL	0.0231	0.0194	0.0268	0.0211	0.0176	0.0247	0.0328	0.0281	0.0375
ES	0.0251	0.0226	0.0276	0.0201	0.0179	0.0224	0.0302	0.0276	0.0329
FI	0.0048	0.0029	0.0067	0.0029	0.0012	0.0045	0.0164	0.0133	0.0196
FR	0.0140	0.0120	0.0159	0.0088	0.0072	0.0104	0.0198	0.0175	0.0221

Table 2 continued

Country	Best	Best low	Best high	Linear	Linear low	Linear high	Full	Full low	Full high
HR	0.0133	0.0110	0.0157	0.0106	0.0085	0.0128	0.0226	0.0193	0.0258
HU	0.0207	0.0187	0.0227	0.0191	0.0170	0.0212	0.0258	0.0234	0.0283
IE	0.0212	0.0168	0.0255	0.0128	0.0094	0.0163	0.0358	0.0312	0.0403
IS	0.0020	0.0004	0.0035	0.0009	0.0001	0.0019	0.0141	0.0104	0.0178
IT	0.0184	0.0165	0.0202	0.0144	0.0128	0.0159	0.0208	0.0190	0.0227
LT	0.0067	0.0047	0.0087	0.0047	0.0028	0.0065	0.0194	0.0156	0.0232
LU	0.0353	0.0321	0.0385	0.0300	0.0270	0.0331	0.0410	0.0376	0.0445
LV	0.0187	0.0155	0.0219	0.0130	0.0103	0.0157	0.0306	0.0259	0.0354
MT	0.0133	0.0108	0.0158	0.0108	0.0083	0.0133	0.0197	0.0162	0.0232
NL	0.0028	0.0019	0.0037	0.0025	0.0015	0.0034	0.0114	0.0095	0.0134
NO	0.0025	0.0010	0.0039	0.0014	0.0002	0.0026	0.0141	0.0112	0.0170
PL	0.0209	0.0185	0.0232	0.0172	0.0151	0.0193	0.0228	0.0204	0.0251
PT	0.0300	0.0256	0.0344	0.0211	0.0170	0.0251	0.0387	0.0338	0.0436
RO	0.0290	0.0252	0.0328	0.0215	0.0180	0.0250	0.0314	0.0275	0.0352
SE	0.0092	0.0016	0.0168	0.0072	0.0022	0.0121	0.0453	0.0310	0.0596
SI	0.0079	0.0063	0.0095	0.0067	0.0052	0.0083	0.0127	0.0105	0.0149
SK	0.0061	0.0049	0.0072	0.0052	0.0041	0.0063	0.0099	0.0082	0.0116
UK	0.0156	0.0130	0.0182	0.0124	0.0096	0.0152	0.0311	0.0270	0.0352

IOp (MLD) estimates derived from (i) the linear most parsimonious case (linear); (ii) the fully interacted model (full); (iii) the best model selected (best). 95% normalized bootstrap confidence intervals are based on 500 replications

An immediate implication is that the countries' rankings clearly depend on the model specification chosen by the researcher. Consider again Fig. 1, where countries are ordered according to the IOp level based on the $best$ model specification. The non-monotonicity of the other two series of bins, $linear$ and $full$, indicates that the countries' rankings vary with the model specified. For instance, France ranks 19th according to the best model specification but would do much better, ranking 12th, if we consider the most parsimonious specification.

To further investigate the problem of IOp sensitivity to alternative econometric specifications, we consider the measures of IOp proposed in two recent papers, see Brzenziński (2015) and Suárez and Menéndez (2017). Both analyses use the same 2011 EU-SILC data, follow the *ex ante* approach and use equivalized disposable income as outcome variable. The measures obtained by these authors differ because they use different model specifications. Suárez and Menéndez (2017) estimate IOp parametrically considering the following circumstances: gender, nationality, urban density, parental education, and parental occupation. Brzenziński (2015) adopts a parametric approach that includes parental education, parental occupation, and nationality. Both models are estimated using log-linear OLS regression with no interactions.

Figure 2 shows the rank correlation of our $best$ measure and the two alternative estimates for the 24 countries considered in both studies. We note that the final assessment differs substantially in both cases. Although the rank-correlation is clearly positive and significant, a number of countries lie outside the 45 degree line. Indeed, the re-ranking is substantial in a few cases. For example, in Suárez and Menéndez (2017), Ireland ranks 17th and Belgium ranks 7th, whereas with our $best$ measure, they rank first and 17th, respectively. Additionally, in Brzenziński (2015), Portugal ranks 13th, whereas if our $best$ specification is adopted, it ranks 23rd.[14]

We believe that this exercise provides convincing evidence that the variance-bias trade-off in IOp measurement is far from negligible in empirical applications. Hence, it is crucial to introduce a statistical criterion to select the best model among a very large number of possible specifications.

5 Conclusions

The past two decades have seen growing interest from scholars and policy makers in the measurement of inequality of opportunity. A number of methodological contributions have shown that estimates of inequality of opportunity are mostly downward biased. This is a consequence of the partial observability of circumstances beyond individual control that affect individual outcome. This issue has typically been addressed by resorting to rich datasets and adopting broad econometric specifications. However, since IOp is measured as inequality in a counterfactual sample distribution, a second possible source of bias might be related to the sampling variance of the estimated counterfactual distribution. In this paper, we discuss this additional source of bias, which has surprisingly been neglected by the empirical literature on IOp measurement.

---

[14] Figure 4 in Appendix C shows a closer but far from perfect ranking correlation between the estimates of Brzenziński (2015) and Suárez and Menéndez (2017).

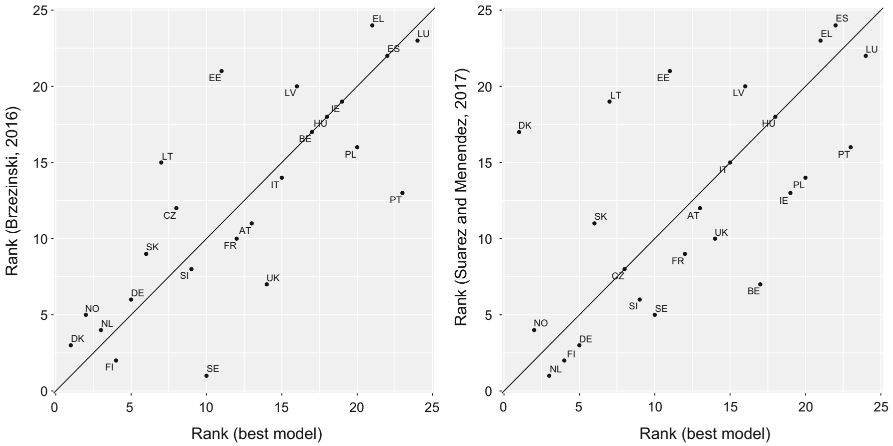

Fig. 2 IOp estimates in 24 European countries from different studies. The Figure shows the rank correlation of countries in terms of IOp. Our best model specification is compared with Suárez and Menéndez (2017) and Brzenziński (2015) . *Source: EU-SILC, 2011*

We show that it implies an upward bias of IOp, which challenges the interpretation of IOp estimates as lower-bound estimates of the real IOp.

We stress that because the empirical specification used to estimate IOp largely influences its magnitude, we require a reasonable statistical criterion to select among alternative models. We suggest that this criterion minimize the two sources of bias.

We interpret this problem as a typical variance-bias trade-off and propose a simple CV method to find the best-fitting model. Cross-validation methods assess the predictive performance of alternative models to estimate a dependent variable out-of-sample. Overfitted models tend to be extremely accurate in explaining variability in sample but perform poorly in predicting on a test sample not used to estimate the model. By providing an unbiased assessment of the relative predictive performance of each possible model specification, CV can be used by researchers as a guide to choose the best model to estimate IOp.

The models selected by the algorithm typically differ across countries in terms of the variables considered and the interactions included, suggesting that, when attempting to produce comparable IOp estimates, scholars may abandon the idea of specifying the same model for all countries in all time periods. By contrast, comparable estimates may be obtained using the model specification that best captures the correlation of individual outcome and circumstances beyond individual control separately for each country and time period.

Finally, we show the empirical relevance of our intuition and implement the proposed method to measure IOp in 31 European countries. Our empirical evidence illustrates that the choice of model specification strongly affects the estimated IOp and demonstrates the importance of having a widely accepted criterion to identify the best possible specification.

A Upward bias when estimating IOp with survey data

Chakravarty and Eichhorn (1994) distinguish between the true distribution of income, y, and the observed distribution, $\tilde{y}$, where $\tilde{y} = y + e$ and e is commonly defined as the measurement error such that $e \sim iid(0, \sigma^2)$. By considering a strictly concave von Neumann–Morgenstern utility function, U, they prove by analogy that if we measure inequality $I(\tilde{y})$ with an inequality index I that satisfies symmetry and the Pigou-Dalton transfer principle, then the inequality of the true y distribution is smaller than inequality in the observed distribution.

Without loss of generality, we apply their result to the case of non-parametric IOp measurement (Eq. 2).

Proposition *Let $\tilde{Y}$ be the counterfactual distribution estimated with Eq. 2. Assume that $\tilde{Y}$ is estimated by observing the full set of circumstances and the entire population. Let $\hat{\tilde{Y}}$ be the same counterfactual distribution estimated by observing the full set of circumstances but considering only a proper subsample of the entire population. Let IOp and $I\hat{O}p$ be any measure of inequality that satisfies symmetry and the Pigou-Dalton transfer principle applied to $\tilde{Y}$ and $\hat{\tilde{Y}}$ respectively. Then, $E(I\hat{O}p) > IOp$.*

Proof Let $\mathbf{M} = \mu_1, \ldots, \mu_T$ be the vector of types' mean outcomes in the population. Let $\hat{\mathbf{M}} = \hat{\mu_1}, \ldots, \hat{\mu_T}$ be the estimates of types' means based on a proper subsample of the population. Then, for each $t = 1, \ldots, n$, $\hat{\mu}_t = \mu_t + \eta$, where $\eta = \frac{\sigma}{\sqrt{N_t}} \sim (0, \chi^2)$ is the standard error of $\hat{\mu}_t$.

Following Chakravarty and Eichhorn (1994), we assume that U is a strictly concave function. By Jensen's inequality, we have

$$E\left(U\left(\hat{\mathbf{M}}|\mathbf{M}\right)\right) < U\left(E\left(\hat{\mathbf{M}}|\mathbf{M}\right)\right). \tag{5}$$

Note that $E\left(\hat{\mathbf{M}}|\mathbf{M}\right) = \mathbf{M}$, so:

$$E\left(U\left(\hat{\mathbf{M}}|\mathbf{M}\right)\right) < U\left(\mathbf{M}\right). \tag{6}$$

By taking expectations with respect to $\mathbf{M}$ on both sides, (4) becomes:

$$E\left(U\left(\hat{\mathbf{M}}\right)\right) < U\left(E(\mathbf{M})\right). \tag{7}$$

Because $E(\eta) = 0$, the two distributions have the same mean. If U is a strictly concave function, then (5) is equivalent to saying that the distribution of $\mathbf{M}$ Lorenz dominates the distribution of $\hat{\mathbf{M}}$, which implies that $E(I\hat{O}p) > IOp$. □

Corollary *When one or more of the relevant circumstances is not used to partition the population into types (partial observability) and $\tilde{Y}$ is estimated on a proper subsample of the population, $I\hat{O}p$ cannot be interpreted as a lower bound of IOp.*

B A simulation to assess the magnitude of the upward bias

The reader may wonder whether the upward bias discussed in this paper actually represents a non-negligible issue in empirical implementations. To provide an idea of the possible magnitude of the bias, we perform a simulation. When estimating inequality of opportunity, the data generating process is typically unknown. We therefore prefer to base the simulation on the entire EU-SILC dataset instead of creating an ad hoc dataset.

Assume that the entire EU-SILC dataset is our population of interest. A population composed of 202,843 individuals aged between 26 and 60 years (more than the same age population in Iceland and approximately the same population in Luxembourg). Additionally, assume that a few observable circumstances are the only circumstances that determine inequality of opportunity. Individual outcome is assumed to be the result of the interactions of three circumstances: parental education, parental occupation, and origin. Individuals in the same type share the same highest parental education (five values), same immigration history (a dummy that takes the value of one if the respondent is a first- or second-generation immigrant), and the same highest parental occupation (ISCO 1 digit).

Under our assumptions, we can observe the real partition of the population into types. The observed between-type inequality is then the real IOp in the population. The residual inequality is assumed to be due to effort. Measured by MLD, IOp in the entire sample is 0.0314, approximately 7% of the total variability.

Our aim is then to understand the circumstances under which an estimate of inequality of opportunity based on a random subsample of this population results in upward bias. To this end, we estimate IOp using samples of increasing size. We start with 500, which is approximately the sample size of the smallest country in EU-SILC (Sweden). We then add 500 observations in each step until we have a sample of 20,000 observations (not far from Italy's sample size, the largest country in EU-SILC). Each sample is randomly drawn 500 times to obtain normalized bootstrap confidence intervals around the point estimate.

Figure 3 shows the IOp estimates for samples of increasing size. In grey, we provide a histogram showing the frequency of countries' sample size (reported on the right y-axis) in EU-SILC 2011.[15]

The estimates show a marked upward bias for the smallest samples. The average IOp based on the samples is more than 1.2 times higher than the IOp in the population for samples smaller than 4000. These are not unrealistically small samples: six of the 31 countries have smaller sample sizes. Interestingly, the confidence intervals of the estimates do not contain the population's estimate for all samples smaller than 3000 (Sweden, Iceland, Denmark, and Norway have smaller sample sizes). Moreover, the upward bias is less than 10% only for sample sizes larger than 9000. Only France, Germany, Hungary, Poland, Spain, and Italy have larger sample sizes.

Estimates based on the samples approach the IOp in the population rather slowly; at the extreme right of the graph, the bias is approximately 4%. This may be considered

---

[15] Note that these are the sample sizes used in the regression; they include only individuals with non-missing information.

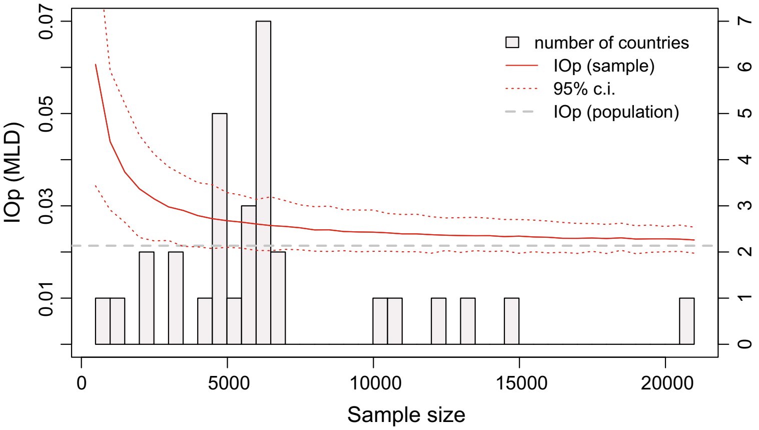

Fig. 3 IOp estimated on samples of increasing size . Source: EU-SILC, 2011

a negligible distortion. Interestingly, the reader may recall that in Fig. 1 of Sect. 4, we found a relatively small difference between the IOp estimated with the two extreme specifications for countries with sample sizes larger than 10,000. However, in our simulation, a sample size of 20,000 observations is extremely large as it represents slightly less than 10% of the population.

C Additional tables and figures

See Table 3 and Fig. 4.

Table 3 Model specifications . Source: EU-SILC 2011

Country	Regressors included in the 'best' model specification					
AT	Father occ. (10)	Mother occ. (10)	Mother edu. (5)	(Birth area) × (father white)	(Father white) × (highest par. edu.)	(Mother white) × (highest par. edu.)
BE	Father occ. (10)	Mother occ. (10)	Mother edu. (5)	(Birth area) × (highest par. edu)	(Father white) × (mother white)	(Mother white) × (highest par. edu.)
BG	Birth area (3)	Father occ. (10)	Father edu. (5)	Mother edu. (5)	(Father white) × (mother white)	(Father white) × (highest par. edu.)
CH	Father white	Father edu. (5)	Mother edu. (5)	(Birth area) × (mother white)	(Birth area) × (highest par. edu)	(Mother white) × (highest par. edu.)
CY	Father occ. (10)	Father edu. (5)	Mother edu. (5)	(Birth area) × (mother white)		
CZ	Father occ. (10)	Mother occ. (10)	Father edu. (5)	(Birth area) × (father white)	(Birth area) × (mother white)	
DE	Father edu. (5)	Mother edu. (5)	(Birth area) × (mother white)	(Father white) × (highest par. edu.)		
DK	Birth area (3)	Father white	Mother white	Mother edu. (5)		
EE	Mother white	Father edu. (5)	Mother edu. (5)	(Birth area) × (father white)		
EL	Mother occ. (10)	(Birth area) × (mother white)	(Birth area) × (highest par. edu)	(Father white) × (highest par. edu.)		
ES	Father occ. (10)	Mother occ. (10)	Father edu. (5)	Mother edu. (5)	(Birth area) × (mother white)	(Father white) × (highest par. edu.)
FI	Mother occ. (10)	Father white	Mother edu. (5)	(Birth area) × (mother white)	(Birth area) × (highest par. edu)	
FR	Father occ. (10)	Mother white	Father edu. (5)	Mother edu. (5)	(Birth area) × (father white)	

Table 3 continued

Country	Regressors included in the 'best' model specification							
HR	Birth area (3)	Father occ. (10)	Mother white	Father edu. (5)	Mother edu. (5)			
HU	Birth area (3)	Mother white	Father edu. (5)	Mother edu. (5)	(Father white) × (highest par. edu.)			
IE	Birth area (3)	Father occ. (10)	Mother occ. (10)	Father edu. (5)	Mother edu. (5)	(Father white) × (mother white)	(Father white) × (highest par. edu.)	(Mother white) × (highest par. edu.)
IS	Mother occ. (10)	(Birth area) × (father white)	(Birth area) × (mother white)	(Father white) × (mother white)				
IT	Father occ. (10)	Father edu. (5)	Mother edu. (5)	(Birth area) × (mother white)	(Birth area) × (highest par. edu)	(Father white) × (mother white)	(Mother white) × (highest par. edu.)	
LT	Father white	Father edu. (5)	Mother edu. (5)	(Birth area) × (mother white)				
LU	Father white	Father edu. (5)	Mother edu. (5)	(Birth area) × (highest par. edu)	(Mother white) × (highest par. edu.)			
LV	Father occ. (10)	Mother occ. (10)	Father edu. (5)	Mother edu. (5)	(Birth area) × (highest par. edu)			
MT	Father occ. (10)	Father edu. (5)	Mother edu. (5)	(Birth area) × (father white)	(Father white) × (mother white)			
NL	Mother edu. (5)	(Birth area) × (father white)	(Birth area) × (mother white)	(Father white) × (mother white)				
NO	Father occ. (10)	Mother edu. (5)	(Birth area) × (father white)	(Birth area) × (highest par. edu)	(Father white) × (highest par. edu.)			
PL	Mother occ. (10)	Father edu. (5)	(Birth area) × (mother white)	(Birth area) × (highest par. edu)	(Father white) × (mother white)			

Table 3 continued

Country	Regressors included in the 'best' model specification						
PT	Mother occ. (10)	Father edu. (5)	Mother edu. (5)	(Birth area) × (father white)	(Birth area) × (highest par. edu)	(Father white) × (highest par. edu.)	(Mother white) × (highest par. edu.)
RO	Mother occ. (10)	Mother edu. (5)	(Birth area) × (mother white)	(Father white) × (mother white)	(Father white) × (highest par. edu.)	(Mother white) × (highest par. edu.)	
SE	Birth area (3)	Mother edu. (5)	(Father white) × (mother white)	(Father white) × (highest par. edu.)			
SI	Birth area (3)	Mother occ. (10)	Father edu. (5)	Mother edu. (5)	(Father white) × (mother white)		
SK	Father occ. (10)	Mother occ. (10)	Highest par. edu. (5)	(Birth area) × (father white)	(Birth area) × (mother white)	(Father white) × (mother white)	
UK	Mother edu. (5)	(Birth area) × (father white)	(Father white) × (mother white)	(Father white) × (highest par. edu.)	(Mother white) × (highest par. edu.)		

Numbers in parentheses refer to the number of categories. Complete regression tables are available upon request

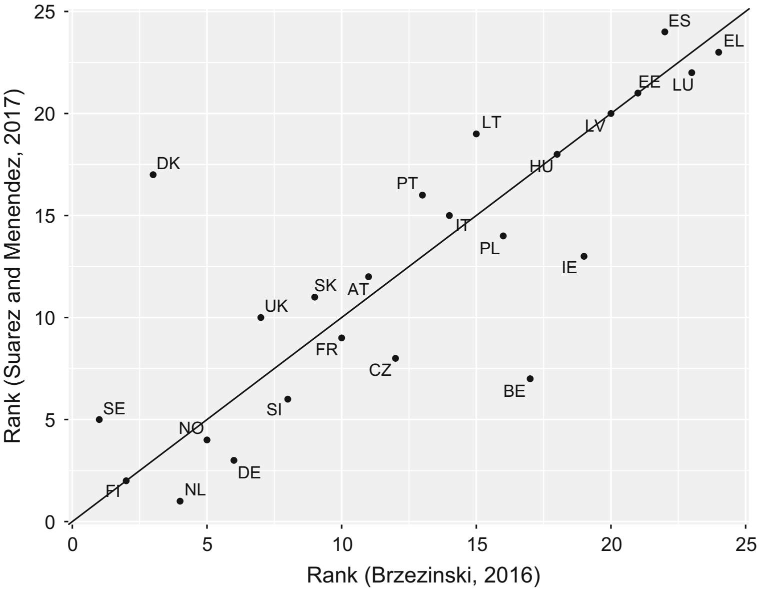

Fig. 4 IOp estimates for 24 European countries from different studies. The figure shows the rank correlation of countries in terms of IOp. The ranking proposed by Suárez and Menéndez (2017) is compared with the ranking proposed by Brzenziński (2015) . *Source: EU-SILC, 2011*

References

Arlot S, Celisse A (2010) A survey of cross-validation procedures for model selection. Stat Surv 4:40–79

Athey S (2018) The impact of machine learning on economics. In: Agrawal AK, Gans J, Goldfarb A (eds) Chapter 21 in the economics of artificial intelligence: an agenda. University of Chicago Press, Chicago

Balcazar C (2015) Lower bounds on inequality of opportunity and measurement error. Econ Lett 137:102–105

Björklund A, Jäntti A, Roemer J (2012) Equality of opportunity and the distribution of long-run income in Sweden. Soc Choice Welf 39:675–696

Bourguignon F, Ferreira F, Ménendez M (2007) Inequality of opportunity in Brazil. Rev Income Wealth 53:585–618

Bourguignon F, Ferreira F, Ménendez M (2013) Inequality of opportunity in Brazil: a corrigendum. Rev Income Wealth 59:551–555

Brunori P, Ferreira F, Peragine V (2013) Inequality ofopportunity, income inequality and mobility: some internationalcomparisons. In: Paus E (ed) Getting development right: structural transformation, inclusion and sustainability in the post-crisis era. Palgrave Macmillan

Brunori P, Hufe P, Mahler GD (2018) The roots of inequality: estimating inequality of opportunity from regression trees. In: World bank policy research working papers 8349

Brunori P, Palmisano F, Peragine V (2016) Inequality of opportunity in Sub Saharan Africa. In: World bank policy research working papers 7782

Brzenziński M (2015) Inequality of opportunity in Europe before and after the Great Recession. In: Working Paper n. 2/2015 (150). Faculty of Economic Sciences, University of Warsaw

Chakravarty SR, Eichhorn W (1994) Measurement of income inequality: observed versus true data. In: Eichhorn W (ed) Models and measurement of welfare and inequality. Springer, Berlin

Checchi D, Peragine V (2010) Inequality of opportunity in Italy. J Econ Inequal 8:429–450

Checchi D, Peragine V, Serlenga L (2016) Inequality of opportunity in Europe: is there a role for institutions? In: Cappellari L, Polachek S, Tatsiramos K (eds) Inequality: causes and consequences, research in labor economics, vol 43. Emerald, Bingley

Daniels B (2012) “CROSSFOLD: stata module to perform k-fold cross-validation,” Statistical Software Components S457426. Boston College Department of Economics

Ferreira F, Gignoux J (2011) The measurement of inequality of opportunity: theory and an application to Latin America. Rev Income Wealth 57:622–657

Ferreira F, Peragine V (2016) Equality of opportunity: theory and evidence. In: Adler M, Fleurbaey M (eds) Oxford handbook of well-being and public policy. Oxford University Press, Oxford

Fleurbaey M, Schokkaert E (2009) Unfair inequalities in health and health care. J Health Econ 28:73–90

Gareth J, Witten D, Hastie T, Tibshirani R (2013) An introduction to statistical learning with applications in R. Springer, New York

Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning data mining, inference, and prediction, 2nd edn. Springer

Hothorn T, Hornik K, Zeileis A (2006) Unbiased recursive partitioning: a conditional inference framework. J Comput Graph Stat 15(3):651–674

Hufe P, Peichl A, Roemer J, Ungerer M (2017) Inequality of income acquisition: the role of childhood circumstances. Soc Choice Welf 49:499–544

Hufe P, Peichl A (2015) Lower bounds and the linearity assumption in parametric estimations of inequality of opportunity. In: IZA working papers, DP No. 9605

Ibarra L, Martinez C, Adan L (2015) Exploring the sources of downward bias in measuring inequality of opportunity. In: World bank policy research working paper no. WPS 7458. Washington

Kanbur R, Wagstaff A (2016) How useful is inequality of opportunity as a policy construct? In: Basu K, Stiglitz JE (eds) Inequality and growth: patterns and policy. International economic association series. Palgrave Macmillan, London

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation and model selection. In: Proceedings of the 14th international joint conference on artificial intelligence, vol 2, pp 1137–1143

Larson SC (1931) The shrinkage of the coefficient of multiple correlation. J Educ Psychol 22(1):45–55

Lefranc A, Pistolesi N, Trannoy A (2009) Equality of opportunity and luck: definitions and testable conditions, with an application to income in France. J Public Econ 93(11–12):1189–1207

Li Donni P, Rodriguez JG, Rosa Dias P (2015) Empirical definition of social types in the analysis of inequality of opportunity: a latent classes approach. Soc Choice Welf 44:673–701

Luongo P (2011) The implication of partial observability of circumstances on the measurement of inequality of opportunity. In: Rodriguez J (ed) Research on economic inequality, vol 19, pp 23-49

Marrero G, Rodrguez J (2012) Inequality of opportunity in Europe. Rev Income Wealth 58:597–621

Mullainathan S, Spiess J (2017) Machine learning: an applied econometric approach. J Econ Perspect 31(2):87–106

Niehues J, Peichl A (2014) Upper bounds of inequality of opportunity: theory and evidence for Germany and the US. Soc Choice Welf 43:63–79

Rodríguez JD, Pérez A, Lozano JA (2010) Sensitivity analysis of kappa-fold cross validation in prediction error estimation. IEEE Trans Pattern Anal Mach Intell 32(3):569–575

Roemer J (1998) Equality of opportunity. Harvard University Press, Cambridge

Roemer J, Trannoy A (2015) Equality of Opportunity. In: Atkinson AB, Bourguignon F (eds) Handbook of income distribution, vol 2. Elsevier, New York

Shao J (1997) An asymptotic theory for linear model selection. Stat Sin 7(1997):221–264

Stone M (1977) An asymptotic equivalence of choice of model by cross-validation and akaike’s criterion. J R Stat Soc Ser B 39(1):44–47

Suárez AA, Menéndez AJL (2017) Income inequality and inequality of opportunity in Europe. Are they on the rise? ECINEQ WP 2017-436

Van de Gaer D, Ramos X (2016) Empirical approaches to inequality of opportunity: principles, measures, and evidence. J Econ Surv 30(5):855–883

Varian HR (2014) Big data: new tricks for econometrics. J Econ Perspect 28(2):3–27

Wendelspiess FCJ (2015) Measuring inequality of opportunity with latent variables. J Hum Dev Capab 16(1):106–121