ORIGINAL RESEARCH

Causal bias in measures of inequality of opportunity

Lennart B. Ackermans[1] ⬛

Abstract

In recent decades, economists have developed methods for measuring the country-wide level of inequality of opportunity. The most popular method, called the *ex-ante* method, uses data on the distribution of outcomes stratified by groups of individuals with the same circumstances, in order to estimate the part of outcome inequality that is due to these circumstances. I argue that these methods are potentially biased, both upwards and downwards, and that the unknown size of this bias could be large. To argue that the methods are biased, I show that they ought to measure causal or counterfactual quantities, while the methods are only capable of identifying correlational information. To argue that the bias is potentially large, I illustrate how the causal complexity of the real world leads to numerous non-causal correlations between circumstances and outcomes and respond to objections claiming that such correlations are nonetheless indicators of unfair disadvantage, that is, inequality of opportunity.

Keywords Inequality of opportunity · Measurement · Causal modelling · Confounding · Facts and values

1 Introduction

Equality of opportunity, as it is typically described, occurs when differently situated individuals have similar outcome prospects. Outcomes may depend on choices or personal characteristics for which people are held responsible, but they should not depend on their circumstances. The measurement of inequality of opportunity is an active research area within economics.[1] Various methods have been proposed to measure the level of inequality of opportunity within a country, which have been used to compare this level to other countries (Lefranc et al., 2008; Ferreira & Gignoux,

---

[1] For an overview of the literature, see Ramos and Van de gaer (2016), Roemer and Trannoy (2015), and Ferreira and Peragine (2016).

✉ Lennart B. Ackermans
 ackermans@esphil.eur.nl

1   Erasmus School of Philosophy, Erasmus University Rotterdam, Rotterdam, The Netherlands

2011; Checchi et al., 2016; Hufe et al., 2017; Brunori et al., 2019a) or to ascertain the percentage of inequality within a country that is the result of unequal opportunities (Bourguignon et al., 2007; Checchi & Peragine, 2010; Pistolesi, 2008; Almås et al., 2011; Björklund et al., 2012; Davillas & Jones, 2020). Studies using these methods typically consider opportunities for income attainment, but other outcomes such as health are also considered.

The measurement of inequality of opportunity is complicated and a large number of related methods have been developed. I discuss the dominant methodological approach for measuring inequality of opportunity, which has been labeled the *ex-ante* approach. Ex-ante methods are considered favorable, for two main reasons. First, they can be applied even with limited data. Second, the estimates produced by these methods are believed to only suffer from downward bias—and can thus be interpreted as lower bounds on the true level of inequality of opportunity (Ferreira & Gignoux, 2011; Lara Ibarra & Martinez-Cruz, 2015; Juárez, 2015; Balcázar, 2015).[2]

I argue that the problem of bias is more severe than has previously been recognized: ex-ante measures of inequality of opportunity can be biased both upwards and downwards, and the size of the bias is potentially large. This puts pressure on the interpretation of these measures as lower bounds. To make this case, I first argue that normatively appropriate concepts of inequality of opportunity involve causal (or counterfactual) notions, and that the proper measurement of inequality of opportunity requires the proper measurement of causal effect sizes. Mere correlations between circumstances and outcomes are insufficient to establish (normative) claims about the level of inequality of opportunity. Since ex-ante methods measure correlations rather than causes, they are biased when these correlations do not indicate underlying causal effects. I make use of the causal modeling literature (particularly Pearl, 2009) to illustrate that this bias exists and that it can be both upwards and downwards.

Second, I consider whether the size of this bias can nevertheless be assumed to be small. Since the causal mechanisms behind economic outcomes are complex and our knowledge about these mechanisms is very limited, it is hard to make a judgment on the size of the bias. I consider two arguments in favor of a small bias based on very particular views of determinism and responsibility. I argue that these views fail on normative and metaphysical grounds. This leads me to conclude that the size of the bias is unknown and potentially large. Hence, the methods are of limited use in practice.

While the paper is mainly concerned with the ex-ante approach, there exists a different approach called *ex-post* (see Fleurbaey & Peragine, 2013). *Ex-post* methods that use what is called Roemer's Identification Assumption face similar problems as those described in this paper, on which I briefly reflect at the end of the paper.

The philosophy of science literature has not yet extensively addressed the inequality of opportunity measurement project discussed in this paper, but my analysis is related to recent debates in causal modeling of the measurement of discrimination (Bright et al., 2016; Kincaid, 2018; Hu & Kohler-Hausmann, 2020; Weinberger, 2022). Likewise, these authors have shown that the measurement of discrimination requires complicated judgments about causal structure that are partly dictated by normative considerations.

---

[2] An exception is Brunori et al. (2019b), who identify a possible source of upwards bias.

My findings are further demonstration of fact-value entanglement in economics (Reiss, 2017). Inequality of opportunity is a so-called thick ethical concept, a term that mixes factual and normative components. The measurement of such terms poses problems for social scientists (Alexandrova & Fabian, 2022). In particular, as I show, reporting on the degree of inequality of opportunity implies taking a normative stance on a variety of issues. It is still a matter of debate how scientists should decide which normative position to adopt when measuring thick ethical concepts.

Section 2 introduces the ex-ante approach to measuring inequality of opportunity and argues that it should measure causal counterfactual quantities. Section 3 shows that the parametric methods within the ex-ante approach are causally biased, using insights from causal modeling. Section 4 argues that the problem is severe, and responds to the objection that the problem may be less severe under appropriate classifications of circumstance and responsibility factors. Section 5 shows that non-parametric ex-ante methods face similar problems as parametric methods. Section 6 discusses the ex-post approach. Section 7 concludes and gives some thoughts about future research. Appendix A criticizes the Monte-Carlo method used by Bourguignon et al. (2007) to explore the size of omitted variables bias for their inequality of opportunity measure.

2 Measurement approaches and normative foundations

In this section, I summarize the primary normative foundations underlying the popular ex-ante approach for measuring inequality of opportunity. I then argue that this measurement approach succeeds in measuring the true level of inequality of opportunity only if it is based on measuring causal or counterfactual quantities. This sets the stage for the later sections, which are concerned with the size and sign of the bias when causal parameters are not appropriately measured.

As the principle is typically expressed in the economic literature, equality of opportunity obtains when differential outcomes are the result of factors for which individuals are responsible, labeled *effort*, but not the result of *circumstances*, which are factors for which individuals are not responsible. It should be noted that the word 'effort' is used as a shorthand for the combined matters of responsibility. Effort is typically a scalar variable or vector that is assumed to reflect all factors that individuals are responsible for. Effort in the usual sense of the word (exertion) may but does not need to be one of these factors. Circumstances are typically defined as factors outside of individuals' control, but other definitions are possible. The question of what should by classified as effort and circumstance is not of concern in this paper.

These normative foundations are heavily inspired by works of philosophers in the tradition of luck egalitarianism, such as Cohen (1989), Dworkin (2002), and Arneson (1989). A strength of the economic methods that have been developed, however, is that they are not tied to luck egalitarianism and can be used to measure inequality of opportunity as defined by a variety of normative frameworks, including, for example, Rawlsian fair equality of opportunity. The methods can also be applied to different settings, such as opportunities in education (Ferreira & Gignoux, 2014) or health (Fleurbaey & Schokkaert, 2009), in which different normative foundations and different classifications of variables into *effort* and *circumstance* may be required.

For the purposes of this paper, I will focus on the negative formulation of inequality of opportunity: inequalities that are caused by circumstances are unacceptable. Call this principle *circumstance egalitarianism*.[3]

A second conception of equality of opportunity, which I discuss later in the paper, is that each individual should face the same set of opportunities (introduced in the economic literature by Van de gaer, 1993). Call this principle *opportunity egalitarianism*. Opportunity sets consist of combinations of effort choices and associated outcomes that individuals are able to choose from. It is typically assumed that people who share the same circumstances have the same opportunity set.

Both positions define an optimal state which is called equality of opportunity. The distance between the actual state of affairs and the optimal state is the degree of inequality of opportunity.

The economic literature identifies two broad categories of measurement approaches, the ex-ante and ex-post approach (Fleurbaey & Peragine, 2013). The ex-ante approach focuses on inequality between *types*, which are groups of individuals who share the same circumstances. The ex-post approach focuses on inequality between *tranches*, which are groups of individuals who are the same in matters of responsibility. While these two approaches are frequently formulated in such a way that they could be seen as normative principles of equality of opportunity, they should rather be understood as different methodological approaches to measuring inequality of opportunity. This paper is mainly concerned with the ex-ante approach, but Sect. 6 reflects on what the findings imply for the ex-post approach.

Ex-ante inequality of opportunity can be defined as follows.

> **Ex-ante inequality of opportunity:** Let $v(t)$ be a measure of the advantage of type t. Inequality of opportunity decreases if inequality between types, $v(t_1) - v(t_2)$, decreases.

Studies differ with respect to the measure $v(t)$ they adopt, as well as what $v(t)$ is thought to be a measure *of*. Most commonly, $v(t)$ is taken to be the average outcome of individuals within type t, and most commonly, $v(t)$ is thought of as measuring the value of a type's opportunity set. In that case, the normative justification of the measurement approach is based on opportunity egalitarianism. A version of the ex-ante method based on opportunity egalitarianism is discussed in Sect. 5. Alternatively, $v(t)$ can be thought to measure the causal effect of the type's circumstances on outcomes, in which case the normative justification is based on what I called circumstance egalitarianism. A version of the ex-ante method based on circumstance egalitarianism is discussed in Sect. 3.

Two related methods that are not the subject of this paper should be mentioned. Bourguignon et al. (2007, 2013) develop a Monte-Carlo method that extends the ex-ante method and estimates a lower and upper bound on the true level of inequality of opportunity. I criticize this method in appendix A. Second, Niehues and Peichl (2014) develop a method that is argued to give an upper bound estimate of inequality of

---

[3] Circumstance egalitarianism may differ from the "positive formulation" that inequalities are only acceptable if they are the result of matters of responsibility, if circumstances (in part) cause matters of responsibility. Circumstance egalitarianism would not hold people responsible for the part of their effort that is caused by circumstances. See Hild and Voorhoeve (2004) for a discussion of the normative merits of both positions.

opportunity, to use in combination with the (supposed) lower bound methods discussed in this paper.

2.1 Measuring equality of opportunity requires measuring causal effects

Circumstance egalitarianism involves a causal notion, since it requires that circumstances do not *cause* outcomes. Opportunity egalitarianism involves a (related) counterfactual notion, since it requires that individuals' counterfactual options (their alternative effort choices) should lead to the same outcomes. Hence, the methods should measure causal or counterfactual quantities, and not just correlations.

To see this more clearly, consider two examples. First, suppose two groups, A and B, are applying to a prestigious college. We suppose that one's group membership (A or B) is a circumstance, and that the quality of an applicant's application is an effort variable (i.e., a matter of responsibility). The outcome is fully determined by group membership and quality of the application. In this situation, in order for there to be equality of opportunity supposing circumstance egalitarianism, group membership should not cause inequalities in application decisions. In order for there to be equality of opportunity supposing opportunity egalitarianism, opportunity sets, which consist of all possible application qualities and associated outcomes, should be equal for each group. Suppose that the application committee is blinded from group membership and that group membership has no causal effect on application quality. In that case, group membership does not cause unequal outcomes and opportunity sets are identical. Hence, there is equality of opportunity according to both conceptions.[4]

Nevertheless, it could be the case that there is a correlation between outcomes and group membership, because there is a common cause of group membership and application quality (we assume this common cause is not a circumstance). There could be many such common causes. For example, one's upbringing may be a cause of religious group membership, as well as a cause of valuing academic performance. Suppose $v(t)$ is a correlational statistic reflecting the probability of acceptance conditional on group membership t. Then $v(A)$-$v(B)$ is nonzero, so the ex-ante measure of inequality of opportunity reports there is inequality of opportunity. However, by assumption, there is equality of opportunity. A version of this example (common cause bias or confounding bias) is discussed in greater detail in Sect. 3.1 below.

As a second example, suppose that a society consists of workers with two occupations, A and B, which are circumstances. Workers' wages are entirely determined by hours spent working, which is an effort variable. Hence, there is equality of opportunity in income. Suppose that, initially, there is no correlation between occupation and wage, such that a correlational version of the ex-ante measure correctly reports that there is equality of opportunity. Now suppose that an unforeseen accident leads to the death of all occupation A workers who are present at their workplace late in the

---

[4] This conclusion can be contested if one has a different normative position about what is effort and circumstance, or if one rejects the assumption that all effort levels (i.e., all application qualities) are equally available to both groups. I ask the reader to put aside these concerns: the example is intended to illustrate the role of causality. If the reader rejects the particular assumptions used here, one can probably come up with a different example with the same causal structure that better matches the reader's favorite normative theory.

day. The deceased workers happen to habitually spend a higher amount of hours than those not present during the accident. As a result, wage will now be correlated with occupation, so a correlational ex-ante method (based on wages of living workers) will report that there is inequality of opportunity. Yet, opportunities for workers after the accident are the same as before the accident. Moreover, it is still the case that wages (and therefore, inequalities in wages) are entirely caused by hours spent working. Hence, by assumption, there is equality of opportunity according to both conceptions. A version of this example (sample selection bias) will be discussed in greater detail in Sect. 3.2.

These examples show that when $v(t)$ is a correlational quantity such as the average outcome for individuals of type t, there are cases in which the ex-ante method gives incorrect conclusions. These incorrect conclusions would be avoided if $v(t)$ captures a causal relation between circumstances and outcomes, instead of a mere correlation.

It is not always appreciated in the literature that a normatively valid measure of inequality of opportunity should be based on the measurement of causal and counterfactual quantities. For example, Roemer and Trannoy (2015) assert that we should worry about a lack of causal interpretation only if the aim is to advise policy-makers who want to reduce inequality of opportunity. On the other hand, Roemer and Trannoy do not believe the absence of a causal interpretation is problematic for measuring the degree of inequality of opportunity: "if one merely wants to measure the degree of inequality of opportunity—that is inequality due to circumstances—a correlation (with variables which occurred in the past) is already something that is relevant" (Roemer & Trannoy, 2015, p. 274). It is unclear how the words "due to" should be understood if not as *caused by*, in which case a correlation is not necessarily relevant (as the above examples show). On other hand, if "due to" should be understood as *correlated with*, the offered definition of inequality of opportunity as "inequality due to circumstances" is deficient: in both examples above, there is no inequality of opportunity, while there is a correlation between circumstances and outcomes.

Given that we should measure causal or counterfactual quantities, it is of course possible that in practice correlational measures are sometimes relevant because under the circumstances they *do* indicate causal effects. For example, it is possible that taking $v(t)$ to be a simple statistic such as the average outcome of type t is an empirically good choice because it is often sufficiently close to the causal effect of t on outcomes. I will explore this possibility extensively in Sect. 4, in which I argue that, in most cases, there is no guarantee that $v(t)$ is sufficiently unbiased.

3 Parametric ex-ante: measuring the causal effect of circumstances on outcomes

In this section, my arguments of the previous section are expanded on with respect to one particular method for measuring inequality of opportunity, introduced by Bourguignon et al. (2007), abbreviated here as BFM.[5] This method is based on an estimation

---

[5] Due to a programming error the estimates in BFM are faulty. See the corrections in Bourguignon et al. (2013).

of parameters that reflect the (causal) contribution of observed circumstances to wages, and is thus called the parametric ex-ante approach. In Examples 3.1 and 3.2 below, I show that this method does not capture the true level of inequality of opportunity because its parameter estimates do not reflect causal effects.[6]

The normative foundation of the parametric ex-ante approach as used by BFM is what I called circumstance egalitarianism in Sect. 2, the principle that circumstances should not cause unequal outcomes.[7] In this spirit, BFM set out to measure the total effect of circumstances on outcomes (which is supposed to include a direct effect of circumstances on outcomes, as well as an indirect effect via their effect on effort[8]). Equality of opportunity is said to occur if the total effect is the same for each type, that is, if circumstances do not cause inequality between types.

The estimates are made by an ordinary least squares (OLS) regression on

$$\ln(w_i) = \psi C_i + \varepsilon_i. \tag{1}$$

Here the regressand w_i is the wage of the i'th individual, the regressor C_i is a vector of observed circumstance values (plus a constant), ψ is a vector of coefficients, and ε_i an error term with mean 0. The error term ε_i is assumed to reflect the influence of effort on outcomes. (BFM are aware that C_i and ε_i may be correlated, which biases the result. This will be discussed in greater detail below.)

Based on the estimated coefficients $\hat{\psi}$ and residuals $\hat{\varepsilon}_i$, BFM calculate a "counterfactual" earnings distribution X^C, which is interpreted as the distribution that would arise if everyone had the same circumstances. That is, it is assumed that in the counterfactual state of affairs in which there is equality of opportunity, X^C is the earnings distribution. This distribution $X^C = (\tilde{w}_i)$ is calculated based on the model $\ln \tilde{w}_i = \bar{C}\hat{\psi}_i + \hat{\varepsilon}_i$. Here $\bar{C}$ denotes the population mean of the circumstance variables. The counterfactual earnings distribution is compared to the actual earnings distribution $X = (w_i)$ to obtain a measure of inequality of opportunity:

$$\Theta_I := I(X) - I(X^C). \tag{2}$$

---

[6] The method discussed in this section uses a parametric approach. Most empirical studies of ex-ante inequality of opportunity use the non-parametric approach described in Sect. 5, but Ferreira and Gignoux (2011) show that the parametric BFM estimate of inequality of opportunity is essentially the same quantity (measured differently) as the non-parametric ex-ante estimate that will be discussed below in Sect. 5. Hence, the arguments in this section apply to the non-parametric approach as well in case it is interpreted as measuring the same type of inequality of opportunity (defined by circumstance egalitarianism).

[7] This is implied by BFM's definition of inequality of opportunity as the inequality that remains if circumstances are equal. Moreover, BFM consider the causal effect of circumstances on outcomes to be problematic, even if the effect is indirectly via the effect of circumstances on effort.

[8] There is much discussion in the empirical and social choice literature concerning direct and indirect effects (Roemer, 1998; Bourguignon et al., 2007; Ferreira & Gignoux, 2011; Björklund et al., 2012; Ramos & Van de gaer, 2016). The indirect effect of circumstances is their effect on outcomes via their effect on effort, whereas the direct effect is their effect not mediated by effort. The combination of the two (the total effect) is thought to be captured by ψ in the "reduced form" (1), while direct effects are measured using a regression equation containing effort variables as well as circumstances (which BFM also do separately from their total effect analysis). This paper focuses on the total effect as measured by the reduced form, but different contexts (in which circumstance egalitarianism is not the right principle) might require a different analysis (Hild & Voorhoeve, 2004).

Here *I* denotes an inequality index, which summarizes the amount of inequality in the distribution as a single number, such as the Gini index or Theil index. BFM use the Theil index for their own calculations.

There are various issues with this methodology that would lead estimates of inequality of opportunity to have a bias of unknown size. Sections 3.1 and 3.2 show two ways in which the results can be biased both upwards and downwards. Section 3.3 then argues that there is no clear way to improve on current methods to remove the bias or estimate its size. This provides the framework for Sect. 4, in which I argue that the bias might be large.

3.1 Example: common causes

As I argued in Sect. 2.1, the coefficients ψ need to be given a causal interpretation if they are used to measure inequality of opportunity. However, regression coefficients only represent causal contributions under strict conditions. In general, a causal interpretation requires that the measured variables (C) are not confounded by unobserved variables (see e.g., Pearl, 2009). Confounding occurs if there is a common cause of a circumstance and wages that is not adjusted for by other regressors. The following example illustrates that the BFM method needs to measure causal parameters in order to measure the true level of inequality of opportunity, and that its bias when the parameter estimates diverge from the causal parameters can be both upwards and downwards.

Consider a population of individuals who go to separate schools of different levels of prestigiousness, spend some time on schooling, and then enter the labor market. An empirical researcher examines this population and seeks to estimate the effect of the school's level of prestigousness—considered a circumstance—on wages, in order to measure income inequality of opportunity.

Suppose that individuals in our population differ with respect to their ambition (A), which we suppose is not classified as a circumstance variable, but is a cause of school prestigiousness. (That is, we assume that ambition is neither a circumstance nor an effort variable.[9]) For mathematical simplicity, suppose that school prestigiousness (S) is a continuous variable. Individuals are assigned to a school based on their ambition according to

$$S_i = r A_i, \tag{3}$$

where *r* is some unknown constant.

---

[9] Some might reject that an individual characteristic such as ambition can be neither a circumstance nor an effort variable, insisting that any characteristic must be either. I believe this position is untenable. When you go back in the causal history—for example, considering the causes of ambition and the causes of the causes of ambition—there will be a point at which you arrive at factors which can't reasonably be classified as either circumstance or effort. One may attempt to avoid this conclusion using an expansive conception of circumstances that considers all causes of circumstances to be circumstances themselves (disregarding, for the moment, the question how we should think of causes of effort). This conception also seems undesirable: I discuss two versions of it in Sects. 4.1 and 4.2 below.

Fig. 1 Causal graph of example 3.1
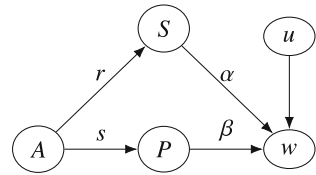
After finishing their education, individuals enter the labor market. We suppose that their productivity P is an effort variable. It is determined linearly from A as

$$P_i = s A_i, \tag{4}$$

for some constant s. Employers set wages according to

$$w_i = \alpha S_i + \beta P_i + u_i. \tag{5}$$

Here α is the causal contribution of school prestigiousness to wages, and β the causal contribution of productivity to wages. We suppose that u is normally distributed about 0 and uncorrelated with S and P. The full structural model is depicted as a causal graph in Fig. 1.

Suppose the researcher is unaware of the way in which S and P are determined from A and is able to measure only S. She proposes to use the BFM method, which requires her to estimate the contribution of S to w by regressing

$$w_i = \psi S_i + \varepsilon_i. \tag{6}$$

As an aid to determine the bias of (6), we can rewrite the structural model by consecutive substitution of (4) and (3) into (5), leading to

$$w_i = (\alpha + \frac{\beta s}{r}) S_i + u_i. \tag{7}$$

BFM reason that the regression estimate of ψ is biased only if there is econometric endogeneity, that is, if there is no econometric exogeneity. Econometric exogeneity of S means that ε and S are uncorrelated and that ε has mean 0, such that $\mathbf{E}[\varepsilon C] = 0$. Econometric exogeneity is commonly considered a sufficient condition for the regression coefficients to be unbiased due to "omitted variables." Whether there is econometric exogeneity, however, depends on the interpretation of the error term ε (see e.g., Pratt & Schlaifer, 1984), and BFM have little discussion of this interpretation. Putting aside for a moment the question how ε_i *should* be interpreted, consider the interpretation $\varepsilon_i = u_i$.[10] In this interpretation of ε, a regression on (6) is a regression

---

[10] This interpretation would follow from a purely statistical understanding of regression analysis in which regressed parameters represent statistical relations rather than structural relations. In this paradigm, ε is defined as $\varepsilon_i = w_i - \mathbf{E}[w_i \mid S_i]$, the difference between the actual wage and its expected value conditional on the regressor S. This statistical interpretation can be contrasted with a causal interpretation, in which ε reflects the influence of omitted causal variables. See also the discussion in appendix A.

on (7); that is, we have $\varepsilon_i = u_i$ and $\psi = \alpha + \beta s/r$. Since $\mathbf{E}[S\varepsilon] = \mathbf{E}[Su] = 0$, that is, there is econometric exogeneity under this interpretation, it follows that the OLS estimator $\hat{\psi}$ is an unbiased estimate of ψ in the narrow, statistical sense of *unbiased*, which in this case means that we have $\mathbf{E}[\hat{\psi}] = \psi = \alpha + \beta s/r$. Hence, while $\hat{\psi}$ is an unbiased estimate of ψ, under this interpretation of ε it is a biased estimate of α, the causal effect of S on w. This bias is given by $\beta s/r$.

It is easy to see that the quantity of interest is indeed α and not $\alpha + \beta s/r$ (see also the discussion in the previous section). A situation in which equality of opportunity occurs—interpreted based on circumstance egalitarianism—is when school assignment does not cause inequalities. This situation would occur if $\alpha = 0$, but when $\alpha = 0$ the regression of (6) would lead to an expected estimate of the contribution of schooling to wages of $\mathbf{E}[\hat{\psi}] = \beta s/r$, which may not be 0.

The following shows that Θ_I is a biased estimator of the true level of inequality of opportunity. If every individual were assigned the same school level $\bar{S}$, that is, if Eq. (3) would be replaced by $S_i = \bar{S}$, then wages would satisfy

$$w_i = \alpha \bar{S} + \beta P_i + u_i. \tag{8}$$

If we follow the spirit of BFM, an estimator of inequality of opportunity should measure the difference between the actual distribution generated by (5) and the counterfactual distribution generated by (8), which is the distribution that would arise if circumstances were equal. The estimator Θ_I, on the other hand, measures the difference between the distribution generated by (5) and the distribution generated by

$$\tilde{w}_i = (\alpha + \frac{\beta s}{r})\bar{S} + u_i. \tag{9}$$

Since $(\alpha + \beta s/r)\bar{S}$ is a constant, a distribution generated by (9) contains only inequalities that are the result of the random disturbances u_i. The appropriate counterfactual distribution generated by (8), on the other hand, also contains inequalities that are the result of differences in productivity P, which by assumption is an effort variable. The inequality due to P is considered acceptable, so it should be included in the counterfactual distribution in which there is equality of opportunity. Hence, in this example, Θ_I overestimates the level of inequality of opportunity.

BFM are not unaware of such problems. They propose to explore the likely magnitude of bias due to omitted variables (relevant variables which are not included as regressors). Omitting relevant variables could bias the estimation results by creating econometric endogeneity. It can be shown that causal bias can indeed be reflected in econometric endogeneity if the error term ε is "causally" interpreted as representing the effect of causal variables other than S, which in our case implies $\varepsilon_i = \beta P_i + u$. If this is how ε is defined, then P is an omitted variable and ε correlates with C (which also depends on P), so there is econometric endogeneity. Such endogeneity is known to create a bias in the regression estimate $\hat{\psi}$ that can be shown to be equal to $\beta s/r$, which is the causal bias identified above.

One would be able to estimate the size of this bias $\beta s/r$ if one knew the correlation coefficient $\rho_{S\varepsilon}$ between S and the error term, where ε is interpreted as above. Since

$\rho_{S\varepsilon}$ is unknown, BFM introduce a Monte-Carlo method (a computational method based on random sampling) that considers a wide range of guesses of values for $\rho_{S\varepsilon}$. These guesses are used to create an interval in which the true degree of inequality of opportunity likely lies. However, I show in appendix A that such a method is unable to reduce the bounds on the bias of $\hat{\psi}$ further than what one already knows given one's knowledge of the underlying causal structure. Hence, BFM's Monte Carlo method is not capable of mitigating the problem discussed in this section.

The BFM Monte Carlo method has not been used in subsequent empirical studies, in part because the method turned out to be less useful than initially thought (Bourguignon et al., 2013) and in part due to a formal result from Ferreira and Gignoux (2011), which is thought to show that inequality of opportunity measures can only be biased downwards. This has led to a consensus in the literature that inequality of opportunity measures should be interpreted as lower bounds (Bourguignon et al., 2013; Balcázar, 2015; Roemer & Trannoy, 2015; Ferreira & Peragine, 2016; Ramos & Van de gaer, 2016; but potential upward bias is discussed by Brunori et al., 2019b). However, the result from Ferreira and Gignoux (2011) merely shows that the exclusion of circumstance variables leads to an estimate of inequality of opportunity that is smaller than the estimate would have been if those variables were included. As the above example demonstrates—where all circumstances were measured—measures could still be biased upwards due to confounding bias, which persists even if one measures all circumstances. Hence, the existing literature does not address the problems of causal bias raised in this section.

3.2 Example: 'sample selection' bias

The following example gives a different way in which inequality of opportunity might be overestimated due to the regression coefficients not matching the causal effects.

Existing studies of inequality of opportunity use large data sets that might be argued to be representative of the population, as do Bourguignon et al. (2007). However, a similar sort of problem is that the population itself has been altered—by death and migration—in a way that creates a correlation between circumstances and effort. The structure of this problem is the same as ordinary sample selection bias. See Bareinboim et al. (2014) and Bareinboim and Pearl (2016) for discussions of sample selection bias in the context of measuring causal effects.

Suppose we have one circumstance variable C, and one effort variable E, which are initially independently normally distributed. A lower value of C means a worse circumstance, and a lower value of E means lower effort. Suppose now that a number of high effort individuals with bad circumstances migrate out of the measured population. The result will be that E and C are positively correlated in the sample.

Suppose wages are set according to

$$w_i = \alpha C_i + \beta E_i. \tag{10}$$

Here $\alpha > 0$ is the causal contribution of circumstances to wages, and $\beta > 0$ is the causal contribution of effort to wages. The structural model is graphically depicted in

Fig. 2 Causal graph of example
3.2. The node M indicates
migration. Following
Bareinboim et al. (2014), the
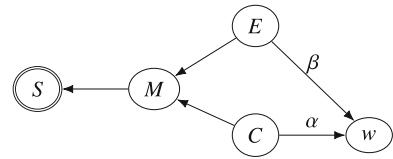sample selection node S is given
two circles

Fig. 2. In that graph, E and C are causes of both wages and migration M. M in turn
is a cause of sample selection S.

Similarly to the previous example, suppose E is not measurable. The empirical
researcher estimates ψ by OLS on

$$w_i = \psi C_i + \varepsilon_i. \tag{11}$$

As shown in the previous example, the quantity of interest is the causal contribution
α. However, we have[11]

$$\mathbf{E}[\hat{\psi}] = \frac{\mathrm{cov}(w, C)}{\mathrm{var}(C)} = \alpha + \beta \frac{\mathrm{cov}(E, C)}{\mathrm{var}(C)}. \tag{12}$$

Hence, since E and C are positively correlated, the OLS estimator $\hat{\psi}$ overestimates
the causal effect of circumstances on wages by $\beta \, \mathrm{cov}(E, C)/\mathrm{var}(C)$. (Similarly, if
effort and circumstances are inversely correlated, $\hat{\psi}$ underestimates the causal effect
of circumstances on wages.)

3.3 Can we adjust for confounding?

One might wonder whether it is possible to improve upon the methodology used
by BFM in order to adjust for or reduce confounding bias. This section focuses on
back-door adjustment, the simplest method to do this. I argue that back-door adjust-
ment achieves little with the limited observational data that most studies of (income)
inequality of opportunity have been using so far.

A criterion for sufficiently adjusting for confounding called the back-door criterion
is given by Pearl (2009). Application of this criterion requires knowledge of the causal
mechanism that is described by a causal model using a directed acyclic graph (DAG).
To measure the causal contribution of a variable X to Y one needs a set of adjustment
variables Z that satisfy two conditions (the back-door criterion below).

First, we need some definitions. In a DAG, a path of three variables is called a chain
if they are connected by directed edges that go in one direction, such as the chain
$A \rightarrow S \rightarrow w$ in Fig. 1. It is called a fork if the arrows are directed outwards from the
middle variable, such as the fork $S \leftarrow A \rightarrow P$ in Fig. 1. It is called a collider if the
arrows point to the middle variable, such as the collider $P \rightarrow w \leftarrow S$ in Fig. 1. When
adjusting using the back-door criterion, confounding causal paths need to be *blocked*
by a set of adjustment variables. Blocking is defined as follows.

---

[11] This is a standard result in econometrics. See the omitted variable formula in (Greene, 2018, p. 59).

> ***d*-separation or blocking** (Pearl, 2009)**: a path p is blocked by a set of variables Z if and only if
>
> (i) p contains a chain $i \rightarrow m \rightarrow j$ or a fork $i \leftarrow m \rightarrow j$ such that the middle node m is in Z; or
> (ii) p contains a collider $i \rightarrow m \leftarrow j$ such that the middle node m is not in Z and such that no descendant of m is in Z.

One adjusts for confounding by conditioning on the variables in a set Z that satisfies the back-door criterion (below). In a linear regression model, this conditioning is carried out by using the variables in Z as regressors alongside X. If Z satisfies the back-door criterion, then the partial regression coefficient of X conditional on Z is a reliable estimate of the total causal effect of X on the dependent variable, Y (Pearl, 2009, p. 152).

> **Back-door criterion** (Pearl, 2009)**: Z satisfies the back-door criterion relative to a cause X, effect Y and a DAG G if
>
> (i) no node of Z is a descendant of X; and
> (ii) Z blocks every path between X and Y that contains an arrow into X.

Condition (i) ensures that we don't condition on the wrong variables, which lie on the causal path we want to estimate. Conditioning on such a variable would exclude the causal effect via that variable from our estimation of the causal effect. Condition (ii) requires that variables on the paths between X, Y and a common cause of X and Y are included in Z to adjust for.

As an example, suppose we want to measure the causal effect of P on w in Fig. 1. Both $Z = \{A\}$ and $Z = \{A, S\}$ satisfy the back-door criterion. The empty set $Z = \emptyset$ does not, since the path $P \leftarrow A \rightarrow S \rightarrow w$ would not be blocked.

Turning to the BFM method, it is theoretically possible to estimate the causal contribution of the circumstance vector C to income using back-door adjustment. That is to say, conditioning on the variables that the back-door criterion tells us to condition on yields effect sizes that are in theory causally unbiased . For the simple example of confounding bias in Sect. 3.1—for which it is conceivable to have knowledge of the underlying causal structure—this would be an easy task. However, to execute this procedure correctly in the real world, one needs extensive knowledge of the underlying causal mechanism and comprehensive data to make the required adjustments. Specifically, one needs to have information about each path between C and w via a common cause. Such paths are likely plentiful (see Sect. 4), and it seems impossible that we could ever unveil the full structure, let alone gather data for each of the paths that we need to adjust for to satisfy the back-door criterion.

Even if we had data for all relevant variables, it would be easy to mistakenly condition on the wrong variables if the full causal model is not known. This could happen when one conditions on a collider, which could unblock paths that would otherwise be blocked, failing condition (ii) of the back-door criterion (Elwert & Winship, 2014). This is similar to what happens in example 3.2. In that example, since the data is gathered after migration has taken place, we are effectively conditioning on migration, a collider. In other words, sample selection S is part of the adjustment set Z; but Z

Fig. 3 Example in which
conditioning on X_1 blocks and
unblocks a path at the same time
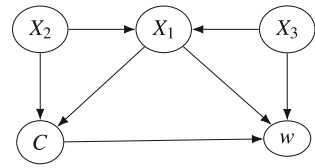
does not satisfy the back-door criterion: S is a descendant of M, and M is a collider. Therefore, conditioning on S unblocks the path $C \rightarrow M \leftarrow E \rightarrow w$. And this, Pearl's backdoor criterion teaches us, is liable to bias one's causal measurements.

Sometimes, however, conditioning on a collider is necessary. Consider the causal structure depicted in Fig. 3, in which a common cause X_1 of C and w is also a collider in a different path between C and w. In this model, conditioning on X_1 is necessary to block the path $C \leftarrow X_1 \rightarrow w$, and it would also block the other previously unblocked paths; but at the same time, adjustment for X_1 would unblock the path $C \leftarrow X_2 \rightarrow X_1 \leftarrow X_3 \rightarrow w$, in which it is a collider. Hence, to satisfy the back-door criterion one should additionally condition on X_2, X_3, or both. Given the complexity of the real causal mechanism behind wages, it is likely that the selection of suitable adjustment variables is a non-trivial matter for which one needs complex knowledge of the causal mechanism, as in this example.

When we have incomplete knowledge of the causal mechanism and limited data, back-door adjustment is not possible. Causal effects may still be identifiable under special conditions, such as when the available data allows for an instrumental variable approach (Angrist et al., 1996). Existing studies of inequality of opportunity all use observational data, and none uses an instrumental variables approach. Possibly, existing data do not allow the causal effect of circumstances on wages to be identified at all.

4 How bad is the causal bias?

In the previous section, I showed that the inequality of opportunity measures—known as parametric ex-ante measures—proposed by Bourguignon et al. (2007) and Ferreira and Gignoux (2011) may be biased upwards and downwards—but it could still be maintained that the problem is small. I argue that this is not the case: the bias could be very large. As the extent of the problem cannot be measured, we are left in the dark about how accurate measures of inequality of opportunity really are.

As examples 3.1 and 3.2 show, there are two kinds of problems that bias the parametric ex-ante measure of inequality of opportunity. First, individuals' responsibility characteristics could have common causes with their circumstances. Second, circumstances and effort variables could be correlated as a result of individuals leaving the measured population.

First, consider the second problem. Deaths and migration occur all the time and should typically lead to correlations between circumstances and effort variables. This would result in significant bias unless (a) the changes are independent of circumstances and effort variables or (b) these changes in the population are very small. Option (a) is

implausible, as it requires a lack of interaction between individual characteristics and circumstances. Such interactions are likely to exist. For example, adventurous people living in circumstances with low prospects may be more likely to migrate than adventurous people with better circumstances, so if adventurousness is an effort variable, (a) is not satisfied. Similarly, hard-working people living in a miner's community may be more likely to die. Or, ambitious people of wealthy families may be more or less likely to migrate to another country than poor ambitious people. On the other hand, (b) is more likely to be be satisfied in some cases, such as when there is no migration and mortality in the measured population is sufficiently low. Hence, the second problem can potentially be overcome if researchers have good data. However, researchers should put effort into demonstrating that (b) is the case, which as far as I'm aware has never been done in the measurement literature.

Now consider the first problem, of common causes. Variables that (in the measurement literature) are typically considered circumstances are parental socioeconomic status, race, gender, and locality. Effort variables, on the other hand, could be choice of education, occupation, and hours worked per week. If you would trace the causal history of these circumstances and effort variables, you would find many common causes. Individual characteristics such as interests and ambition are heavily influenced by parental upbringing, which in turn is associated with parental socioeconomic status. Parents' genetic characteristics could be a common cause of circumstances such as parental socioeconomic status and individual genetic traits that are not circumstances (if one's normative view allows for genetic non-circumstances). And so forth. This lead me to believe that the type of common cause bias discussed in Sect. 3.1 is potentially severe.

Some people would object along the lines that these common causes (such as genetic and biological factors) are themselves circumstances. This would lead to a more expansive conception of a *circumstance* that would, perhaps, make inequality of opportunity measures more accurate, as explained below. Two versions of this objection will be discussed in Sects. 4.1 and 4.2.

4.1 The free will objection

A special kind of objection is available to libertarians in the free will debate. According to the position I will call *cause-exclusive libertarianism*, individuals make free choices for which they are morally responsible only if these choices are entirely uncaused, and moreover, individuals are in fact capable of making free choices of this type. Now consider a version of circumstance egalitarianism which holds individuals responsible only for such free choices and takes circumstances to be all things that are neither a free choice nor caused by a free choice. It follows that circumstances and effort variables have no common causes, eliminating any common cause bias. (Note that there might still be "selection bias" of the type introduced in Example 3.2, and inequality of opportunity might also be underestimated due to unmeasured circumstances.)

However, cause-exclusive libertarianism is not an attractive position. First, the view that free actions must be entirely uncaused is controversial even among libertarians (Capes, 2017). Most would agree that actions that have some causes but are not causally

determined can be free choices of the kind that people are fully responsible for. As Capes argues, the choice of a poll worker to rig an election after being bribed is caused in part by the bribe offer, but the poll worker may still be fully responsible.

Second, even if choices are only free if they are entirely uncaused, such choices may be rare. If the world is causally deterministic, then uncaused choices would not exist at all. Even if there is some indeterminism, and allowing for the possibility that *partially* uncaused choices exist, *entirely* uncaused choices should expected to be much less common. The environment in which people grow up and live has a profound influence on their preferences, habits, and views. If these environmental factors have only a small effect on the choices that people make, these choices would no longer be entirely uncaused. Consequently, free choices of the kind required in the argument would be rare or would not exist at all, which means that a view of equality of opportunity based on it collapses to outcome egalitarianism—which is something that most advocates of equality of opportunity as a normative ideal want to resist.

4.2 The 'common causes are circumstances' objection

A couple of authors claim that if circumstances and effort are correlated, it is not ethically acceptable to hold people responsible for their effort (Fleurbaey, 1998, p. 221; Checchi & Peragine, 2010, p. 433). It is then often proposed—seemingly as a solution to this problem but without much justification—to assume that effort variables (which can be measured variables or theoretical variables) are in fact uncorrelated. (A similar point, not discussed here, is made by Roemer 2002, who argues that the outcome distribution of a type should be seen as a characteristic of the type, and therefore, a circumstance.) If this assumption is warranted it might be argued that common cause bias (which implies a correlation between effort and circumstances) need not be a problem. I will attempt to formulate a charitable interpretation of this objection.[12] This version of the objection does not depend on libertarianism about free will; it would be valid even if all outcomes, circumstances, and effort variables are causally determined by earlier factors.

A causal (and more reasonable) version of the above objection proceeds from a particularly expansive conception of circumstances, according to which causes of circumstances should be considered circumstances as well. In the context of circumstance egalitarianism—according to which people are not responsible for variables to the extent that they are influenced by circumstances—this view implies that one is not fully responsible for a proposed effort variable that has a common cause with a circumstance. From this perspective, a correlation between *these* effort variables and circumstances is unproblematic for estimating inequality of opportunity (as will be shown below).

In what follows, I will refer as an *anti-circumstance* to any factor that is not a circumstance, not caused by circumstances and not in part caused by circumstances.

---

[12] A literal interpretation of the objection does not work. Consider the example, discussed in Sect. 2, in which an unforeseen accident leads to the death of all high-effort individuals within one type. As a result, outcomes will be correlated with circumstances, but it does not follow that after the accident we can no longer hold people responsible for their effort.

Anti-circumstances are similar to effort variables, since under circumstance egalitarianism, inequalities that are caused by anti-circumstances are unproblematic.[13]

To see why this view allows for unbiased estimation of the causal effect of circumstances, consider a procedure that generates a *sufficient set of circumstances* and a *sufficient set of anti-circumstances*, which are sets such that: (a) circumstances and anti-circumstances are causally unrelated to each other (they do not cause each other and have no common causes) and (b) the combined variables from the two sets explain all variation in outcomes. If one were to include all variables in a sufficient set of circumstances in one's regression of wages, one would expect to end up with an unconfounded estimate of the total effect of all circumstances on outcomes. This procedure assumes full knowledge of the entire causal mechanism behind outcomes, so it can't be used in practice. But the procedure demonstrates that it is theoretically possible to find a sufficient set of circumstances, and therefore, that it is theoretically possible to regress on a set of circumstance variables resulting in a causally unbiased estimate of inequality of opportunity.

Step 1. List a set of factors (causal variables) that one initially regards as circumstances and add these to a provisional circumstance set. All other direct causes of the outcome of interest (typically all effort variables) are added to a provisional set of anti-circumstances.

Step 2. Add all the common causes of the provisional circumstances and provisional anti-circumstances to the circumstance set. (One does not need to add all causes of circumstances to obtain a sufficient set, since variables that only cause already included circumstances but no anti-circumstances do not affect outcomes apart from the effect that is already accounted for by the included circumstances.)

Step 3. Remove each factor from the set of provisional anti-circumstances that is now caused (in part) by a circumstance. For each removed anti-circumstance, add all causes of this removed anti-circumstance that are not a provisional circumstance or caused by a provisional circumstance to the provisional set of anti-circumstances. Repeat from step 2 until there are no common causes left between the two sets.

The procedure is illustrated in Fig. 4a. Suppose that c_1 and c_2 are initially thought of as circumstances, and nc_1 is initially thought of as an effort variable. First, c_1 and c_2 are selected as provisional circumstances and nc_1 as a provisional anti-circumstance. Then, f_5 is added to the circumstance set, since it is a common cause of c_1 and nc_1. Then, nc_1 is removed from the anti-circumstance set, and f_4 is added to the anti-circumstance set. Now that the procedure has finished, $\{c_1, c_2, f_5\}$ is identified as a sufficient set of circumstances, and $\{f_4\}$ as a sufficient set of anti-circumstances.

When this process has finished, one ends with a sufficient set of circumstances and a sufficient set of anti-circumstances that are causally unrelated to each other. The sufficient anti-circumstances have no common causes with circumstances, so they are variables that people may be held fully responsible for. The variables in the causal mechanism that end up without a classification (the pink and white variables in Fig. 4)

---

[13] It is useful to distinguish effort and anti-circumstance, since effort variables are usually explicitly defined by a normative theory about what individuals can be held responsible for, whereas anti-circumstances are only implicitly variables that individuals can be held responsible for (given circumstance egalitarianism). Moreover, under the view discussed here individuals may in fact *not* be held responsible for variables initially deemed effort variables, while they are (implicitly) held responsible for anti-circumstances.
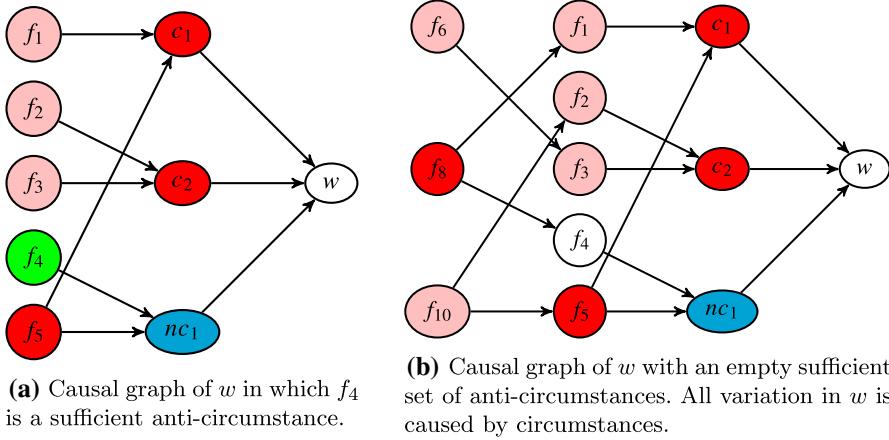
(a) Causal graph of w in which f_4 is a sufficient anti-circumstance.

(b) Causal graph of w with an empty sufficient set of anti-circumstances. All variation in w is caused by circumstances.

Fig. 4 An illustration of the circumstance selection procedure. Circumstances in the sufficient set are colored red, and other circumstances are colored pink. The direct causes of w that are provisional anti-circumstances are colored blue. Anti-circumstances in the sufficient set are colored green

do not cause additional variation in outcomes to the classified variables, so they can be disregarded when calculating the amount of inequality of opportunity.

The procedure cannot be executed in practice, since we typically do not have full knowledge of the causal mechanism behind an outcome. However, with this conception of circumstances, common cause bias will be eliminated even if one only measures a subset of circumstances and no effort variables or anti-circumstances. Existing arguments that inequality of opportunity measures can be seen as lower bounds if not all circumstances are measured are more plausible in the absence of common cause bias. Ferreira and Gignoux (2011), for example, show that, when not all circumstances are measured, the BFM method for measuring inequality of opportunity yields an estimate of inequality of opportunity that is lower than it would have been if all circumstances had been included in the estimation. Under the expansive conception of circumstances, a hypothetical estimation which includes all circumstances from a sufficient set would not suffer from common cause bias with anti-circumstances, and would therefore more likely be close to the true level of inequality of opportunity. Since the actual estimation with omitted circumstance variables yields a lower value than the hypothetical estimation, it can be interpreted as a lower bound on inequality of opportunity.

However, there is a big danger associated with the expansive conception of circumstances. It is likely that circumstance factors become so numerous that no anti-circumstances remain, or that their effect on outcomes is negligible. In that case, circumstances will be responsible for nearly 100% of inequality, such that one's position of inequality of opportunity collapses to outcome inequality. The problem is illustrated in Fig. 4b, which adds an additional layer of causes to Fig. 4a. In the more complex version, nc_1 now also has a common cause with c_1: namely, f_8. As a result, no anti-circumstances remain.

This example illustrates that when adding additional causal history to a graph, it becomes less likely that anti-circumstances remain. Anti-circumstances arise only

if there are paths ending in the outcome variable of interest that never cross a circumstance variable. This situation is especially unlikely if the world is causally deterministic, since outcome variables (such as wages) that are typically measured have a very complicated causal history. The more of this causal history is included in a graph, the less likely that any anti-circumstances remain.[14]

Perhaps an objector will find it more plausible than I do that anti-circumstances remain at the end of our procedure, even if one has access to the full causal history. But a second problem with the objection is that it is unattractive in a normative sense. The position implies that individuals are fully held responsible only for choices that are entirely uncaused by circumstances *and* all causes of circumstances. As I pointed out in the discussion above of cause-exclusive libertarianism, a similar position that individuals can be held responsible only for choices that are entirely uncaused is implausible. Consider that the procedure may start with a list of well-established circumstances, whose effect on outcome inequalities is generally agreed to be undesirable. It does not follow that *causes of* well-established circumstances have equally undesirable effects on outcomes. This should at least shift the burden of proof to the objector to show that the view is normatively acceptable. The categorization of circumstances is usually determined by reference to principles about choice and desert. The objector needs to show that the set of circumstances which the procedure generates is plausible given such principles.

Moreover, economists that measure inequality of opportunity have generally not used very expansive conceptions of circumstances, and often, like Roemer (1998), want to allow societies to decide for themselves what they consider circumstances and matters of responsibility. A society will not typically choose a set of circumstances that includes all the circumstances that ought to be added if one were to follow the above procedure. Hence, if the objector persists, she must concede that society does not have the last word on what circumstances are.

5 Non-parametric ex-ante: measuring opportunity sets

This section discusses an approach known as the non-parametric ex-ante approach, used by Checchi and Peragine (2010) and Ferreira and Gignoux (2011) and inspired by Van de gaer (1993). As this approach is typically described, its purpose is to measure the value of each type's opportunity set and estimate the inequality between these opportunity sets. (Recall that a type is a collection of individuals who have the same circumstances.) Hence, the normative principle underlying this approach is what I

---

[14] The following combinatorial argument makes this point more formally. Consider DAGs with a number of *periods*. In each period, there are q factors with arrows to two factors in the next period, with the requirement that each factor after the first period has at least one cause (it is allowed that both arrows point towards the same factor in the next period). In the last period (the most recent period before the outcome), there is one provisional anti-circumstance and there are $q-1$ circumstances. It can be shown that the fraction of DAGs of this type for which there are non-empty sufficient sets of anti-circumstances, out of all DAGs of this type, shrinks quickly towards zero when periods are added. (The fraction of DAGs of this type, with $n+1$ periods and q causes in each period, that contains sufficient anti-circumstances is $(q-1)^{n(q-1)}/q^{n(q-1)}$.) Hence, if reality is somewhat like this example, with a sufficient number of periods, it is unlikely that anti-circumstances exist.

called opportunity egalitarianism in Sect. 2. Most commonly, a non-parametric ex-ante method uses the arithmetic mean of a type's outcome distribution as a proxy for the value of its opportunity set, and it is then called the *utilitarian ex-ante* method.

The non-parametric approach is similar to the parametric approach discussed in Sect. 3, but since the approach—when interpreted as measuring inequality of opportunity sets—does not involve the estimation of causal parameters, one might think that the issues discussed in the preceding sections do not apply to the non-parametric approach. I show that this is not the case. While opportunity egalitarianism does not involve causal conditions, it does involve counterfactual conditions. An opportunity set contains the outcomes that an individual *would have* achieved if she had chosen differently. These counterfactual conditions are as hard to measure as causal effects. As a result, I argue, the measurement of inequality of opportunity sets as done in the non-parametric ex-ante approach suffers from to the same sort of bias as the parametric approach was shown to have.

In what follows I first give the formal definition of this approach. One imagines that each individual in society faces an opportunity set, which contains possible effort levels (which reflect choices or other matters of responsibility) and associated expected outcomes. It is assumed that two individuals of the same type face the same opportunity set. In the measurement literature, it is also often assumed, as I do below, that all effort choices are available for all types.

We need a valuation function in order to compare opportunity sets. An obvious choice is to sum all outcome values in the set, leading to the valuation function $U(t)$ for the opportunity set of type t. Let $f(e, t)$ be the outcome for an individual of type t with effort level e. This opportunity set valuation is defined as

$$U(t) = \sum_{e \in E} f(e, t).$$

(This valuation appears in Bossert, 1997 and Ooghe, Schokkaert, & Van de gaer, 2007. A different approach to valuing opportunity sets not discussed here is used by Lefranc et al., 2008.)

Since effort is usually not measured, it is difficult to calculate $U(t)$ directly. Therefore, $U(t)$ is typically replaced by a proxy, a different function that can be estimated from data about circumstances only. The hope is that this proxy function is a good estimate of $U(t)$. A popular proxy of the value of a type's opportunity set, due to Van de gaer (1993), is the mean outcome $\mu(t)$ of that type, defined as

$$\mu(t) = \frac{1}{N_t} \sum_{i \in t} x_i,$$

where N_t is the amount of people that are of type t and x_i is the outcome value of individual i. Note that we have assumed that the value of outcomes is determined by effort and type, so we assume $x_i = f(e_i, t_i)$, where e_i and t_i are the agent's effort level and type, respectively. Using $\mu(t)$ as valuation of opportunity sets leads to what is called the utilitarian ex-ante approach for measuring inequality of opportunity.

Assuming that all circumstances and outcomes are measured, $\mu(t)$ can be computed from the data (which consists of the outcomes x_i stratified by type). Inequality of opportunity can then be calculated as follows. We create a "counterfactual" distribution by replacing each individual's outcome with their type's opportunity set valuation $\mu(t)$, yielding

$$X^M = (\mu(1)1_{N_1}, \ldots, \mu(n)1_{N_n}).$$

Here N_t is the amount of people that are of type t and 1_{N_t} is a vector of 1's of length N_t. This distribution can be seen as the distribution of opportunity set values. Inequality of opportunity can be estimated as the inequality within this distribution. That is, if I is some inequality index, inequality of opportunity is measured by

$$\tilde{\Theta}_I = I(X^M).$$

5.1 Problems with using mean income to value opportunity sets

The use of a type's mean income to value opportunity sets comes not without a cost.[15] Understood as a normative claim that the value of opportunity sets is given by $\mu(t)$ regardless of the underlying outcome distribution, it is deficient. This is because it is theoretically possible that two types face identical opportunity sets, but that individuals of one type happen to choose different effort levels than individuals of the other type. This would lead to different estimations of the opportunity sets' value by $\mu(t)$, though by assumption, the opportunity sets are the same and must therefore have the same value. This shows that $\mu(t)$ is not universally acceptable as a measure of opportunity sets, while it may still be empirically acceptable under suitable conditions, which I discuss below.

First, consider a simple example of a society in which there are two types with identical opportunity sets, depicted in Table 1. Each individual has the option to choose either effort level e^1 with outcome 5 or effort level e^2 with outcome 10. However, 50 individuals of type t^1 choose e^1, while 100 individuals of type t^2 choose e^1. Moreover, 100 individuals from t^1 choose e^2, while 50 individuals of t^2 choose e^2. By hypothesis, there is equality of opportunity (as the opportunity sets are identical), and this would also be our conclusion if we valued the opportunity sets using $U(C_t)$, which yields a value of 15 for both types. However, the use of $\mu(t)$ leads to a valuation of t^1's opportunity set of 8.34 and a valuation of t^2's opportunity set of 6.67.

The valuations U and μ do agree on there being equality of opportunity if effort is distributed identically across types. In what follows I make the case that the use of $\mu(t)$ for valuation is acceptable in general only if effort and circumstances are distributed independently.

To make the case that $\mu(t)$ is or is not an empirically acceptable measure, one first needs to find a normatively appropriate valuation of opportunity sets $U'(C_t)$ which $\mu(t)$ is supposed to track. A starting point for finding such a valuation is to note what the use of $\mu(t)$ gets right (or may get right). There is something to be said for giving

---

[15] See also Hild and Voorhoeve (2004), who make a related observation about Van de gaer's policy rule.

Table 1 Outcomes in a society with two effort levels and two types, with identical opportunity sets

	Type t^1	Type t^2
Effort e^1	5 (50)	5 (100)
Effort e^2	10 (100)	10 (50)

The values in parentheses are the amount of individuals having the effort level and type of that cell

Table 2 A society with two types and two effort levels and identical effort distributions

	Type A	Type B
Effort e^1	5 (100)	7 (100)
Effort e^2	10 (50)	8 (50)

less weight in one's opportunity set valuation to effort levels which individuals are less likely to choose. For example, suppose opportunity set A offers a popular effort choice with very low rewards and an unpopular effort choice with very high rewards. The opportunity set B, on the other hand, offers rewards that are only slightly below average for the popular effort choice, and rewards that are only slightly above average for the unpopular effort choice. It would make sense to value B above A, since more individuals would be benefited if they were given opportunities from B rather than A. This example is depicted in Table 2. Using mean incomes as valuations concurs with this intuition: the mean income of type A is 6.67, while the mean income of type B is 7.33. On the other hand, we have both $U(A) = 15$ and $U(B) = 15$.

This consideration would favor the following valuation of a type's opportunity set. Let $P(e)$ be the probability that a given individual (regardless of type) has effort level e in the measured population. The valuation is given by

$$U'_P(C_t) = \sum_{e \in E} f(e, t) P(e). \tag{13}$$

This valuation scales an effort-outcome combination by the same factor $P(e)$ for each type. Since two identical opportunity sets always have the same value, it is not susceptible to the kind of normative problem discussed in the first example above with regard to the use of $\mu(t)$. If we apply this valuation to Table 2, we get $U'_P(A) = 6.67$ and $U'_P(B) = 7.33$, the same values that the mean income assigns in this situation.

The assumption that $U'_P(C_t)$ is the right valuation can of course be questioned, but someone who does so should explain which other valuation, if not $U'_P(C_t)$, ought to be tracked by $\mu(t)$, and I am aware of no such alternative. It seems $U'_P(C_t)$ comes closest to a valuation that (a) is normatively acceptable and (b) could reasonably motivate the use of $\mu(t)$ as an empirical surrogate.

With this assumption, a condition under which $\mu(t)$ *is* empirically acceptable can be formulated. The expected value of $\mu(t)$, given the probability distribution of effort, should be equal to $U'_P(C_t)$. This can be written as follows, where the left hand side

equals $\mathbf{E}[\mu(t) \mid t]$ and the right hand side equals $U_P'(C_t)$:

$$\sum_{e \in E} f(e, t) P(e \mid t) = \sum_{e \in E} f(e, t) P(e). \tag{14}$$

For Eq. (14) to hold in general for all t, we need that effort conditional on type is distributed identically for each type, that is, we need $P(e \mid t) = P(e)$. (If not, the only other way in which (14) is satisfied is when different terms on the left side of the equation perfectly offset each other to equal their counterparts on the right, which will be rare.) In other words, effort and type should be statistically independent. This we might call the assumption of randomness.[16]

> **The assumption of randomness.** For each type t_1 and t_2, the conditional distributions of effort are identical, that is, we have $P(e \mid t_1) = P(e \mid t_2)$.

This is assumption 2 in Checchi and Peragine (2010, p. 433), but it is typically left implicit in studies using an ex-ante approach. (It commonly appears in ex-post studies; see Sect. 6.)

By Reichenbach's common cause principle, the assumption of randomness holds if there is no causal connection between effort and circumstances. If, to the contrary, there is a causal connection between effort and circumstances, either because one causes the other or if they have common causes, then the assumption of randomness is satisfied only under special conditions (see chapter 6 in Pearl, 2009). Similarly, the assumption of randomness will fail to be satisfied if the population has changed (by death or migration) in a way that creates a correlation between effort and circumstances. Hence, the same sort of problems arise as with the parametric ex-ante approach. The arguments given in Sect. 4 imply that the scope of the problem may be large for the non-parametric approach as well.

6 Ex-post approaches and Roemer's Identification Assumption

I will now briefly consider the ex-post approach, an alternative approach to measuring inequality of opportunity. The ex-post approach is used less often in empirical studies, but it has been given considerable theoretical attention (see e.g., Fleurbaey, Peragine, & Ramos, 2017). Unlike the ex-ante approach, the ex-post approach requires measuring effort variables as well as circumstances. Roughly speaking, there is ex-post equality of opportunity if individuals who are the same in matters of responsibility have the same outcomes. This section briefly considers what the previous findings may imply for the ex-post approach.

More formally, ex-post inequality of opportunity can be defined as follows.

---

[16] The name *assumption of randomness* is inspired by what Sowell (1990) calls the "randomness assumption", the assumption that groups would be evenly represented by various outcomes measures, i.e., have identical outcome distributions, in the absence of unequal treatment of those groups.

Ex-post inequality of opportunity: Let $f(e, t)$ be the outcome of an individual with effort level e and type t. Inequality of opportunity decreases if outcome inequality between individuals with the same effort level, $f(e, t_1) - f(e, t_2)$, decreases.

Ex-post approaches face the problem that effort is typically hard to measure. A "solution" to this problem, adopted by many studies, is to measure effort indirectly via Roemer's Identification Assumption (RIA), which states that an individual's effort level is her position in her type's outcome distribution. That is, if individual i and j are both in quantile q of their respective type's outcome distribution, they are assumed to have the same effort level. With this assumption, the effort distribution of each type is the same. According to Roemer (2002), RIA ensures that we take into account the effect of circumstances on choices, since we hold people responsible only for their relative effort level within their type.

When using RIA, the ex-post approach becomes similar to the ex-ante approach, and in some cases identical (see also Ramos & Van de gaer, 2016). For example, one could use RIA and choose to measure inequality only in the median effort tranche. This is identical to an ex-ante approach in which the type valuation function $v(t)$ (see Sect. 2) is chosen to be the median income of type t (instead of the more common arithmetic mean).

It has been recognized in the literature that RIA depends on the assumption of randomness (Fleurbaey, 1998; Ramos & Van de gaer, 2016). The defense of RIA starts from the assumption that individuals' true effort level for which we actually hold them responsible—called *propensity to exert effort* by Roemer (1998)—is distributed identically for each type. If you combine this assumption with the assumption that outcomes are a strictly increasing function in propensity to exert effort, you can derive RIA. Criticism in the literature of RIA has generally focused on the latter assumption, whereas the assumption of randomness is often assumed to be unproblematic. Given my arguments against it above, this assumption seems unwarranted.

Since ex-post methods do not necessarily depend on the measurement of causal quantities, my arguments against the ex-ante approach do not extend to all ex-post methods. However, most ex-post studies use RIA. Given its dependence on the assumption of randomness, we should expect ex-post methods that depend on RIA (and by extension on the assumption of randomness) to be biased for similar reasons.

7 Conclusion

I argued that ex-ante approaches to measuring inequality of opportunity are biased due to causal confounding. This is the case both with parametric ex-ante methods, interpreted in accordance with circumstance egalitarianism, as well as non-parametric ex-ante methods, interpreted in accordance with opportunity egalitarianism.

The size of the bias may be large or small, and with current data, it is impossible to tell. Based on my arguments in Sect. 4, I suggest that the bias can be assumed to be small only with very expansive conceptions of circumstances that make a principle of equal opportunity close to a principle of outcome equality. In cases of less expansive

conceptions of circumstances, the bias may be very large, although it may also be small. This uncertainty makes country-wide measures of unequal opportunity less useful in practice as a way to compare inequality of opportunity in different countries.

These problems come on top of the known problem that many measures of inequality of opportunity are biased downwards due to unobserved circumstance variables. This has lead some to argue that the value of these measures for policy is limited (Kanbur & Wagstaff, 2016). My finding that these measures may be biased upwards as well further diminishes their policy relevance.

It is unlikely that the problem of causal bias can be mitigated with realistically obtainable data. Hence, I suggest the following alternative routes for future research.

First, the field can alter its goals to the measurement of phenomena for which more reliable methods are available, such as experimental methods utilizing randomization, or semi-experimental methods using instrumental variables. For example, a randomized controlled trial in which children were randomly assigned to preschool (Schweinhart et al., 1993) could be used to assess the effect of preschool on inequalities later in life. As another example, to test the effect of school quality on inequality in academic outcomes, one can use a regression discontinuity design that exploits a cut-off point for scores in entrance exams (see e.g., Hoekstra, Mouganie, & Wang, 2018). Such methods do not allow one to create a single country-wide measure of inequality of opportunity, but they do allow one to gain insight about the effect on inequality of particular differences in opportunities. The combination of many different studies of different types can potentially create a bigger picture as well.

Second, the existing methodology can be used to identify groups in society that may have particularly bad opportunities, instead of creating a single country-wide measure that can be used for country rankings. This might then prompt further research using different methods to (a) confirm that they have particularly bad opportunities and (b) find out what causes these bad opportunities and what can be done about them. When taking this way forward, the observational methods discussed in this paper have a merely exploratory function that I believe is more appropriate.

Declarations

Appendix A The BFM Monte Carlo method

Bourguignon et al. (2007) propose a method to explore the scope of bias due to omitted variables (which could be both circumstances and effort variables) in the regression equation. This method is of potential interest, since it could theoretically be used to estimate the scope of confounding bias. This would be remarkable, since many authors in the causal modeling literature have argued that confounding bias cannot be estimated with statistical tools without having knowledge of the causal structure. However, as I show below, the BFM method does rely on tacit causal assumptions, and it does not achieve results that improve one's knowledge of the bias on top of what is implicitly assumed.

I will summarize the method here; for a complete description, see the original paper (2007). For readers unfamiliar with vector notation for regression equations: the formalism below is not essential for understanding the issues I raise. After introducing the complex version of the method as used by BFM, I switch to a simpler version which applies to the example of Sect. 3.1.

The method is based on the insight that the bias B due to econometric endogeneity can be written as a function of the vector of correlation coefficients $\rho_{C\varepsilon}$ between the circumstances C and the error term ε:

$$B = \mathbf{E}[\hat{\psi}] - \psi = \Sigma_C(\underline{\rho_{C\varepsilon}\sigma_C})\sigma_\varepsilon. \tag{A1}$$

Here Σ_C denotes the underlying variance-covariance matrix of the circumstance vector C, and σ_x denotes the vector of standard deviations of a vector of random variables x. The underlined term is a vector with elements $\rho_{C_k\varepsilon}\sigma_{C_k}$. The values for Σ_C and σ_C can be estimated from the data, and σ_ε can be estimated on the basis of the data and an estimate of the bias $\tilde{B}$. The value of $\rho_{C\varepsilon}$ is unknown and will be "guessed" as $\tilde{\rho}_{C\varepsilon}$. This leads to a system of two equations which BFM solve for $\tilde{B}$, which becomes:

$$\tilde{B} = \frac{N(\underline{\tilde{\rho}_{C\varepsilon}\hat{\sigma}_C})\hat{\sigma}_\varepsilon}{(C'C)\sqrt{1 - (\underline{\tilde{\rho}_{C\varepsilon}\hat{\sigma}_C})'(C'C)(\underline{\tilde{\rho}_{C\varepsilon}\hat{\sigma}_C})}}, \tag{A2}$$

where N is the sample size and $\hat{\sigma}_x$ denotes the sample standard deviation of x. Hence, the quantity $\tilde{B}$ can be calculated from sample statistics and the guesses $\tilde{\rho}_{C\varepsilon}$.

BFM then draw random values for the coefficients $\tilde{\rho}_{C\varepsilon}$ from a uniform distribution on $(-1, 1)$.[17] Drawings that lead to a covariance matrix that is not positive semi-definite, or fail a number of conditions on the coefficients, are discarded. The first 100 drawings of $\tilde{\rho}_{C\varepsilon}$ that survive are used to calculate values of $\tilde{B}$. This creates a distribution of 100 values for each circumstance coefficient. For each coefficient, the five highest and five lowest values are then discarded. The next highest and lowest values are denoted the upper and lower bounds on the coefficient. According to BFM, the resulting interval can be seen as a 90% confidence interval for the coefficient values.

---

[17] The elements of $\rho_{C\varepsilon}$ (Pearson correlation coefficients) take values between -1 and 1.

The method as described has as of yet not been successfully used. In the original paper (Bourguignon et al., 2007), a programming error led to bounds on the coefficients that were much smaller than they should have been. Bourguignon et al. (2013) repeat the procedure and find that the bounds on circumstance coefficients become too wide to be useful. The bounds on the level of counterfactual inequality (the estimated level of inequality if circumstances were equal) are reasonably small, but only when the correlation coefficients are drawn from $(-0.2, 0.2)$ instead of $(-1, 1)$. The authors acknowledge that they have no rationale for this restriction.

In what follows, I apply the method to Example 3.1 and examine whether it is capable of correcting for the common cause bias of that example. I identify two problems with the method. First, there is no rationale to discard the 5% lowest and highest coefficient values. Without the deletion of these extremes, the method would produce extremely wide bounds on B. Second, the method relies on assumptions about underlying model parameters including the bias B, and the bounds that the method produces cannot improve on the assumed bounds on B. B is a causal parameter, so the method does not actually succeed in obtaining causal knowledge beyond what is implicitly assumed.

Fortunately, the method becomes much simpler when applied to our simple model. For guesses of the correlation coefficient $\tilde{\rho}_{S\varepsilon} \in (-1.1)$, the bias can be estimated as

$$\tilde{B} = \frac{\hat{\sigma}_\varepsilon \tilde{\rho} S\varepsilon}{\hat{\sigma}_S \sqrt{1 - \tilde{\rho}_{S\varepsilon}^2}}. \tag{A3}$$

Here $\hat{\sigma}_\varepsilon$ is the sample standard deviation of the residual and $\hat{\sigma}_S$ is the sample standard deviation of S.[18]

As can be seen from Eq. (A3), there is no need to perform Monte Carlo simulations in this simple case, since $\tilde{B}$ as a function of $\tilde{\rho}_{S\varepsilon}$ can be analytically computed. $\tilde{B}$ goes towards infinity (respectively minus infinity) when $\tilde{\rho}_{S\varepsilon}$ goes towards 1 (respectively -1). Hence, the bounds that one puts on $\tilde{\rho}_{S\varepsilon}$ determine the bounds on $\tilde{B}$. In this light, the choice to draw 100 values from $\tilde{\rho}_{S\varepsilon}$ from a uniform distribution on $(-1, 1)$, and the subsequent deletion of the 5 highest and lowest values of $\tilde{B}$, is suspect. The values of $\tilde{\rho}_{S\varepsilon}$ that lead to the most extreme estimates of $\tilde{B}$ are those closest to 1 and -1. Hence, the deletion of extreme values comes down to a tacit assumption that $\rho_{S\varepsilon}$ is bounded from below by -0.95 and from above by 0.95. No justification for such an assumption is given by BFM. $\tilde{B}$ grows especially large and small outside of these bounds, so the method's result are driven by an undefended assumption.

Might one nevertheless have reason to suppose $\rho_{S\varepsilon} \in (-0.95, 0.95)$? In order to reason about the correlation coefficient $\rho_{S\varepsilon}$, one needs to give ε an interpretation. As is known—but often ignored—any assumptions about the correlation between the error term and regressors are meaningless unless the error term is given an interpretation (Pratt & Schlaifer, 1984). Since we use the method to identify the common cause

---

[18] Note that under the causal interpretation of ε as $\varepsilon = u + \beta P$, $\hat{\sigma}_\varepsilon$ is a biased estimate of σ_ε, although it happens to be an unbiased estimate of σ_u. After all, $\hat{\sigma}_\varepsilon$ is the standard deviation of the residual, *not* the standard deviation of ε, which is unobservable under the causal interpretation. The square root factor in the denominator is supposed to correct for this bias.

bias from Sect. 3.1, the required interpretation is the causal interpretation based on the causal structure given in Fig. 1. According to this interpretation, we have $\varepsilon_i = w_i - \mathbf{E}[w \mid \mathrm{do}(S = S_i)] = \beta P_i + u_i$, from which it follows that $\psi = \alpha$, the causal effect of circumstances on wages. (Here $\mathrm{do}(S = S_i)$ is the notation from Pearl (2009), which means that S is intervened upon to set it to the value of S_i, undoing the relationship between S and P, while the other variables A and P are kept at their naturally occurring values.) Under this interpretation, ε and S are correlated, so a regression on (6) yields a biased estimate of ψ, which equals $B = \beta s / r$. On the other hand, if ε were given a purely statistical interpretation as $\varepsilon_i = w_i - \mathbf{E}[w \mid S_i] = u_i$, we would have $\rho_{S\varepsilon} = 0$ and $B = 0$. From this it can be seen that one cannot have reasoned assumptions about $\rho_{S\varepsilon}$ without interpreting the error term, which could be both zero and nonzero depending on the interpretation. Since the BFM method requires us to identify causal effects, the causal interpretation is the required interpretation. Hence, assumptions about $\rho_{S\varepsilon}$ depend on assumptions about the underlying causal structure (which in the case of this example is given by Fig. 1).

Since BFM do not discuss causal structure when introducing their assumptions about $\rho_{C\varepsilon}$, these assumptions have no justification. Yet, it is still possible in theory to use justified assumptions about the underlying causal structure in order to put bounds on the correlation coefficient. The following shows that such assumptions need to be quite strong, *and* that such assumptions directly imply bounds on the bias B which cannot be improved upon by the BFM method.

Taking the causal interpretation of ε, if one were to know the true correlation coefficient $\rho_{S\varepsilon}$ one could use the BFM method to estimate the extent of the bias. Indeed, given $\tilde{\rho}_{S\varepsilon} = \rho_{S\varepsilon}$, one can show that $\mathbf{E}[\tilde{B}] = B$, that is, the estimate of the bias is itself unbiased. To find reasoned bounds on $\rho_{S\varepsilon}$, consider that $\rho_{S\varepsilon}$ can be written as

$$\rho_{S\varepsilon} = \frac{B\sigma_S}{\sqrt{B^2\sigma_S^2 + \sigma_u^2}}. \tag{A4}$$

Hence, an assumption about $\rho_{S\varepsilon}$ is equivalent to an assumption about the underlying model parameters B and the ratio σ_S / σ_u. This shows how one can justify an assumption about $\rho_{S\varepsilon}$ on the basis of assumptions about parameters of the causal model. One might have reason to assume B is bounded from above at some value B^* and that σ_S / σ_u is bounded from above at σ^*. This then leads to an upper bound for the correlation coefficient of $\rho_{S\varepsilon}^* = B^*\sigma^* / \sqrt{B^{*2}\sigma^{*2} + 1}$. Taking the expected value of (A3) with respect to the chosen upper bound $\tilde{\rho}_{S\varepsilon} = \rho_{S\varepsilon}^*$ yields

$$\mathbf{E}[\tilde{B}] = \frac{\sigma_u}{\sigma_S}\sigma^* B^*.$$

Hence, with a large amount of data the BFM method's upper bound would approximately be the value above. If $\sigma^* = \sigma_S / \sigma_u$, that is, if the chosen upper bound matches the underlying model's parameters and $\tilde{\rho}_{S\varepsilon} = \rho_{S\varepsilon}^*$, we have $\mathbf{E}[\tilde{B}] = B^*$; so the bias estimation method would not improve upon the assumed bounds on the bias. Only if $\sigma_S / \sigma_u > \sigma^*$, that is, if we were wrong to assume σ^* is an upper bound, do we get an upper bound on $\tilde{B}$ that is smaller than the assumed upper bound B^*. (Note that in

our case, σ_S as well as σ_u can be estimated from the data, so there is no reason to be wrong about σ^* if one has knowledge of the causal structure given by Fig. 1.)

In conclusion, in order to execute the BFM method to correct for common cause bias in our example one needs to make (implicit or explicit) assumptions on the bounds on B and the ratio σ_S/σ_u. The BFM method can produce better bounds on B than the assumed bounds only if the assumed bounds on σ_S/σ_u are mistaken; if one is not mistaken about the bounds on σ_S/σ_u, the method will produce bounds on B^* that are at least as wide as those assumed. Hence, the method is useless.

References

Alexandrova, A., & Fabian, M. (2022). Democratising measurement: Or why thick concepts call for coproduction. *European Journal for Philosophy of Science, 12*(1), 7. https://doi.org/10.1007/s13194-021-00437-7

Almås, I., Cappelen, A. W., Lind, J. T., Sørensen, E. Ø., & Tungodden, B. (2011). Measuring unfair (in)equality Measuring unfair (in)equality. *Journal of Public Economics, 95*(7), 488–499. https://doi.org/10.1016/j.jpubeco.2010.11.002

Angrist, J. D., Imbens, G. W., & Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association, 91*(434), 444–455. https://doi.org/10.1080/01621459.1996.10476902

Arneson, R. J. (1989). Equality and equal opportunity for welfare. *Philosophical Studies, 56*(1), 77–93. https://doi.org/10.1007/BF00646210

Balcázar, C. F. (2015). Lower bounds on inequality of opportunity and measurement error. *Economics Letters, 137*, 102–105. https://doi.org/10.1016/j.econlet.2015.10.026

Bareinboim, E., & Pearl, J. (2016). Causal inference and the data-fusion problem. *Proceedings of the National Academy of Sciences, 113*(27), 7345–7352. https://doi.org/10.1073/pnas.1510507113

Bareinboim, E., Tian, J., & Pearl, J. (2014). Recovering from selection bias in causal and statistical inference. *Proceedings of the AAAI Conference on Artificial Intelligence, 28*(1), 2410–2416. https://doi.org/10.1609/aaai.v28i1.9074

Björklund, A., Jäntti, M., & Roemer, J. E. (2012). Equality of opportunity and the distribution of long-run income in Sweden. *Social Choice and Welfare, 39*(2), 675–696. https://doi.org/10.1007/s00355-011-0609-3

Bossert, W. (1997). Opportunity sets and individual well-being. *Social Choice and Welfare, 14*(1), 97–112. https://doi.org/10.1007/s003550050054

Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (2007). Inequality of opportunity in Brazil. *Review of Income and Wealth, 53*(4), 585–618. https://doi.org/10.1111/j.1475-4991.2007.00247.x

Bourguignon, F., Ferreira, F. H. G., & Menéndez, M. (2013). Inequality of opportunity in Brazil: A corrigendum. *Review of Income and Wealth, 59*(3), 551–555. https://doi.org/10.1111/roiw.12045

Bright, L. K., Malinsky, D., & Thompson, M. (2016). Causally interpreting intersectionality theory. *Philosophy of Science, 83*(1), 60–81. https://doi.org/10.1086/684173

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

Brunori, P., Peragine, V., & Serlenga, L. (2019). Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare, 52*(4), 635–661. https://doi.org/10.1007/s00355-018-1165-x

Capes, J. A. (2017). Freedom with causation. *Erkenntnis, 82*, 327–338. https://doi.org/10.1007/s10670-016-9819-5

Checchi, D., & Peragine, V. (2010). Inequality of opportunity in Italy. *The Journal of Economic Inequality, 8*(4), 429–450. https://doi.org/10.1007/s10888-009-9118-3

Checchi, D., Peragine, V., & Serlenga, L. (2016). Inequality of opportunity in europe: Is there a role for institutions? In L. Cappellari, S. W. Polachek, & K. Tatsiramos (Eds.), Inequality: Causes and consequences (pp. 1–44). Emerald Group Publishing Limited, Bingley. https://doi.org/10.1108/S0147-912120160000043008

Cohen, G. A. (1989). On the currency of egalitarian justice. *Ethics, 99*(4), 906–944. https://doi.org/10.1086/293126

Davillas, A., & Jones, A. M. (2020). Ex ante inequality of opportunity in health, decomposition and distributional analysis of biomarkers. *Journal of Health Economics, 69*, 102251. https://doi.org/10.1016/j.jhealeco.2019.102251

Dworkin, R. (2002). *Sovereign virtue: The theory and practice of equality*. Harvard University Press.

Elwert, F., & Winship, C. (2014). Endogenous selection bias: The problem of conditioning on a collider variable. *Annual Review of Sociology, 40*(1), 31–53. https://doi.org/10.1146/annurev-soc-071913-043455

Ferreira, F. H. G., & Gignoux, J. (2011). The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth, 57*(4), 622–657. https://doi.org/10.1111/j.1475-4991.2011.00467.x

Ferreira, F. H. G., & Gignoux, J. (2014). The measurement of educational inequality: Achievement and opportunity. *World Bank Economic Review, 28*(2), 210–246. https://doi.org/10.1093/wber/lht004

Ferreira, F. H. G., & Peragine, V. (2016). Individual responsibility and equality of opportunity. In M. D. Adler & M. Fleurbaey (Eds.), *The Oxford handbook of well-being and public policy*. Oxford University Press.

Fleurbaey, M. (1998). Equality among responsible individuals. In J.-F. Laslier, M. Fleurbaey, N. Gravel, & A. Trannoy (Eds.), *Freedom in economics: New perspectives in normative economics* (pp. 206–234). Routledge.

Fleurbaey, M., & Peragine, V. (2013). Ex ante versus ex post equality of opportunity. *Economica, 80*(317), 118–130. https://doi.org/10.1111/j.1468-0335.2012.00941.x

Fleurbaey, M., Peragine, V., & Ramos, X. (2017). Ex post inequality of opportunity comparisons. *Social Choice and Welfare, 49*(3–4), 577–603. https://doi.org/10.1007/s00355-017-1090-4

Fleurbaey, M., & Schokkaert, E. (2009). Unfair inequalities in health and health care. *Journal of Health Economics, 28*(1), 73–90. https://doi.org/10.1016/j.jhealeco.2008.07.016

Greene, W. H. (2018). *Econometric analysis* (8th ed.). Pearson.

Hild, M., & Voorhoeve, A. (2004). Equality of opportunity and opportunity dominance. *Economics and Philosophy, 20*(1), 117–145. https://doi.org/10.1017/S0266267104001282

Hoekstra, M., Mouganie, P., & Wang, Y. (2018). Peer quality and the academic benefits to attending better schools. *Journal of Labor Economics, 36*(4), 841–884. https://doi.org/10.1086/697465

Hu, L., & Kohler-Hausmann, I. (2020). What's sex got to do with machine learning? In: Proceedings of the 2020 conference on fairness, accountability, and transparency (p. 513). Association for Computing Machinery. https://doi.org/10.1145/3351095.3375674

Hufe, P., Peichl, A., Roemer, J., & Ungerer, M. (2017). Inequality of income acquisition: The role of childhood circumstances. *Social Choice and Welfare, 49*(3), 499–544. https://doi.org/10.1007/s00355-017-1044-x

Juárez, F. W. C. (2015). Measuring inequality of opportunity with latent variables. *Journal of Human Development and Capabilities, 16*(1), 106–121. https://doi.org/10.1080/19452829.2014.907247

Kanbur, R., & Wagstaff, A. (2016). How useful is inequality of opportunity as a policy construct? In: K. Basu & J. E. Stiglitz (Eds.), Inequality and growth: Patterns and policy: Volume i: Concepts and analysis (pp. 131–150). Palgrave Macmillan. https://doi.org/10.1057/9781137554543_4

Kincaid, H. (2018). Debating the reality of race, caste, and ethnicity. *Philosophy of the Social Sciences, 48*(2), 139–167. https://doi.org/10.1177/0048393117742878

Lara Ibarra, G., & Martinez-Cruz, A. L. (2015). Exploring the sources of downward bias in measuring inequality of opportunity (World Bank Policy Research Working Paper No. No. 7458). The World Bank, Washington, DC. https://ssrn.com/abstract=2680470

Lefranc, A., Pistolesi, N., & Trannoy, A. (2008). Inequality of opportunities vs. inequality of outcomes: Are western societies all alike? *Review of Income and Wealth, 54*(4), 513–546. https://doi.org/10.1111/j.1475-4991.2008.00289.x

Niehues, J., & Peichl, A. (2014). Upper bounds of inequality of opportunity: theory and evidence for Germany and the US. *Social Choice and Welfare, 43*(1), 73–99. https://doi.org/10.1007/s00355-013-0770-y

Ooghe, E., Schokkaert, E., & Van de gaer, D. (2007). Equality of opportunity versus equality of opportunity sets. *Social Choice and Welfare, 28*(2), 209–230. https://doi.org/10.1007/s00355-006-0165-4.

Pearl, J. (2009). *Causality* (2nd ed.). Cambridge University Press.

Pistolesi, N. (2008). Inequality of opportunity in the land of opportunities, 1968–2001. *The Journal of Economic Inequality, 7*(4), 411. https://doi.org/10.1007/s10888-008-9099-7

Pratt, J. W., & Schlaifer, R. (1984). On the nature and discovery of structure. *Journal of the American Statistical Association, 79*(385), 9–21. https://doi.org/10.1080/01621459.1984.10477054

Ramos, X., & Van de gaer, D. (2016). Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys, 30*(5), 855–883. https://doi.org/10.1111/joes.12121

Reiss, J. (2017). Fact-value entanglement in positive economics. *Journal of Economic Methodology, 24*(2), 134–149. https://doi.org/10.1080/1350178X.2017.1309749

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

Roemer, J. E. (2002). Equality of opportunity: A progress report. *Social Choice and Welfare, 19*(2), 455–471. https://doi.org/10.1007/s003550100123

Roemer, J. E., & Trannoy, A. (2015). Equality of opportunity. In: A. B. Atkinson & F. Bourguignon (Eds.), Handbook of income distribution (Vol. 2, pp. 217–300). North-Holland. https://doi.org/10.1016/B978-0-444-59428-0.00005-9

Schweinhart, L. J., Barnes, H. V., & Weikart, D. P. (1993). *Significant benefits: The high/scope perry preschool study through age 27 (monographs of the high/scope educational research foundation)*. High/Scope Foundation.

Sowell, T. (1990). *Preferential policies: An international perspective*. William Morrow and Company.

Van de gaer, D. (1993). Equality of opportunity and investment in human capital. Doctoral dissertation, KULeuven, Leuven

Weinberger, N. (2022). Signal manipulation and the causal analysis of racial discrimination. Ergo, forthcoming. http://philsci-archive.pitt.edu/21031/