The Stata Journal (2014)
14, Number 4, pp. 830–846

iop: Estimating ex-ante inequality of opportunity

Florian Wendelspiess Chávez Juárez
University of Geneva
Geneva, Switzerland
florian.wendelspiess@unige.ch

Isidro Soloaga
Universidad Iberoamericana
Mexico City, Mexico
isidro.soloaga@ibero.mx

Abstract. This article describes the user-written command `iop` to estimate ex-ante inequality of opportunity for different types of variables. Inequality of opportunity is the part of inequality that is due to circumstances beyond the control of the individual. Therefore, it is the ethically offensive part of inequality. Several estimation procedures have been proposed over the past years, and `iop` is a comprehensive and easy-to-use command that implements many of them. It handles continuous, dichotomous, and ordered variables. In addition to the point estimates, `iop` also provides bootstrap standard errors and two decomposition methods.

Keywords: st0361, iop, inequality of opportunity, dissimilarity index, mean log deviation, decomposition

1 Introduction

The concept of inequality of opportunity has received much attention in development economics over the last decade. In his seminal contribution, Roemer (1998) proposed to divide total inequality into inequality due to different effort levels, to luck, and to different opportunities. The idea is that not all types of inequalities are equally bad. Checchi and Peragine (2010) call the part of inequality that is due to different levels of effort the ethically nonoffensive inequality. Different effort should lead to different outcomes; thus inequality due to different levels of effort might be desirable. In contrast, the ethically offensive part of inequality is the part that is due to circumstances beyond the control of individuals. These circumstances are factors that people cannot change through effort and that affect their outcome. Typical examples for circumstances include gender, race, and family background. Hence, in a situation of perfect equality of opportunities, circumstances should not affect the outcome of individuals. Let us use a school exam as an example. If students get different grades because they studied for different amounts of time, we consider the inequality in the grade as something desirable. If the differences in grades were due to only family background and not to different levels of effort, the same inequality would be considered as ethically offensive. Therefore, the goal is to split total inequality into ethically offensive and nonoffensive parts.

We distinguish ex-ante and ex-post inequality of opportunity (Fleurbaey and Peragine 2013). Ex-ante equality of opportunity is achieved when circumstances do not matter for the outcome. The ex-post approach focuses more on effort and states that

 st0361

equality of opportunity is achieved when all people making the same degree of effort achieve the same outcome independently of their circumstances. While the two approaches seem to differ only marginally at first, there are important differences that sometimes make them incompatible.[1]

Conceptually, the two approaches are equally valid, and it is hard to favor one over the other. However, empirically, the ex-ante approach is easier to implement than the ex-post approach. For both approaches, the main challenge is that both effort and luck are not observable; therefore, it is difficult to distinguish them empirically. While the ex-post approach requires at least an estimate of effort, the ex-ante approach does not and can be estimated without it. This is likely to be the main reason why empirical applications focus mostly on ex-ante inequality of opportunity, given that estimating effort requires very strong assumptions. We follow the empirical applications and focus on ex-ante inequality of opportunity.

Researchers have proposed several methods to assess ex-ante inequality of opportunity over the years. The regression approach became very popular and was widely used for studies in different countries and for different outcomes. The main idea of this method is to relate outcome to circumstances by parametric or nonparametric regression methods. The intuition is that in a world of equal opportunities, the circumstances should not matter, so the regression should have a low fit. If we find that circumstances affect the outcome, we consequently have inequality of opportunity. A weakness of this approach is that it provides only lower-bound estimates of inequality of opportunity. This is primarily because the part of inequality due to unobserved circumstances might be wrongly attributed to effort and luck instead of to inequality of opportunity. Ramos and Van de gaer (2012) discuss this approach in detail and provide additional reasons why it yields lower-bound estimates of inequality of opportunity.

In this article, we describe the user-written command `iop`, which implements several recently proposed methods to estimate ex-ante inequality of opportunity. `iop` can estimate inequality of opportunity for both continuous and dichotomous variables. Moreover, by dichotomizing ordered variables at every possible level and applying the methods for dichotomous variables, `iop` can also handle ordered variables. We focus on two methods proposed by Ferreira and Gignoux (2011) and Ferreira and Gignoux (2014) for continuous variables. For dichotomous outcome variables, we implement the method proposed by Paes de Barros, de Carvalho, and Franco (2007) and a translation invariant version of it suggested by Soloaga and Wendelspiess Chávez Juárez (2013b). Focusing on these methods is justified by their practical use in recent empirical applications and their ability to be used to estimate other methods. For instance, by including dummy variables for each type[2] and applying the method proposed by Ferreira and Gignoux (2011), we get the results proposed by Checchi and Peragine (2010).

1. Fleurbaey and Peragine (2013) discuss the differences in detail and provide conditions in which the two approaches are incompatible.
2. A "type" is defined by a combination of circumstances. Thus all circumstances are identical within a type.

In addition to the point estimates of inequality of opportunity, we propose and implement two decomposition methods. First, total inequality of opportunity can be decomposed according to different circumstances by using the Shapley decomposition. The purpose of this decomposition is to understand which circumstances drive inequality of opportunity. This decomposition allows the user to understand how much all circumstances affect inequality and how much each circumstance contributes to total inequality of opportunity. Second, `iop` proposes an Oaxaca-type decomposition of the difference between two groups in a composition and a coefficient effect. This decomposition is used to analyze differences in the level of inequality of opportunity between two geographical units or between the same unit at different times. The Oaxaca decomposition identifies what part of the observed differences is due to differences in the distributions of circumstances and what part is due to differences in the impact of circumstances on the outcome variable.

In this article, we first introduce the regression approach to the measurement of inequality of opportunity in section 2. We then present the command `iop` in section 3 and include examples using the Programme for International Student Assessment (PISA) data in section 4. In the conclusion, we address some limitations and issues of the command and provide an outlook on future developments.

2 Methods

2.1 The regression approach

There are different approaches to assess inequality of opportunity. The regression approach comprises many approaches, which all assess ex-ante inequality of opportunity. To discuss this family of methods, we first introduce some notation. Let y be the outcome variable of interest and $\boldsymbol{C}$ be a matrix of circumstances beyond the control of the individual. The core element of these methods is to relate the outcome to the vector of circumstances. In general, we can describe this by the expected conditional outcome

$$\widehat{y} = E(y \,|\, \boldsymbol{C}) \tag{1}$$

which can be estimated in different ways according to the research question and the dependent variable. For instance, Paes de Barros, de Carvalho, and Franco (2007) have a binary outcome variable (for example, access to schooling) and use a logit or probit model to estimate (1). Ferreira and Gignoux (2011) use income as a dependent variable and estimate the same equation with an ordinary-least squares (OLS) regression and with nonparametric methods by averaging over types.[3] Checchi and Peragine (2010) also estimate inequality of opportunity for income and perform a similar analysis but use only nonparametric estimation techniques to assess (1). Finally, Ferreira and Gignoux (2014) use linear regression for test scores.

3. "Types" is defined as a group of individuals sharing the same circumstances.

Independently of the way (1) is estimated, inequality of opportunity is then computed using a common inequality measure $I(\cdot)$ applied to $\widehat{y}$:

$$\theta_a = I(\widehat{y})$$

The idea behind this is simple. All variation in the vector $\widehat{y}$ is exclusively due to circumstances; hence, it refers to inequality of opportunity. The best choice of the appropriate inequality measure depends on the scope of the analysis and on the dependent variable. Paes de Barros, de Carvalho, and Franco (2007) use the dissimilarity index, Ferreira and Gignoux (2011) use the mean logarithmic deviation, and Ferreira and Gignoux (2014) use the variance. Dividing the absolute inequality measure by the same metric $I()$ applied to the actual outcome y gives a relative measure of inequality of opportunity:

$$\theta_r = \frac{I(\widehat{y})}{I(y)}$$

This last step is possible only when the inequality measure $I(\cdot)$ is equally defined for $\widehat{y}$ and y. For example, this is not the case when the actual outcome is binary and $\widehat{y}$ is the estimated probability.

The choice of the appropriate inequality measure $I(\cdot)$ is crucial and depends mainly on the outcome variable. Table 1 provides an overview of different measures proposed in the literature and implemented in the command `iop`.

Table 1. Different methods to estimate ex-ante inequality of opportunity

	Ferreira and Gignoux (2011)	Ferreira and Gignoux (2014)	Paes de Barros, de Carvalho, and Franco (2007)	Soloaga and Wendelspiess Chávez Juárez (2013b)
Variable type	Continuous, with inherent scale	Continuous, with arbitrary mean and dispersion	Dichotomous and ordered	Dichotomous and ordered
Example	Income	PISA score	Access to schooling	Access to schooling
Method to estimate $E(y\|C)$	OLS	OLS	Probit or logit	Probit
Inequality measure $I(y)$	Mean log deviation: $\frac{1}{N}\sum_{i=1}^{N}\ln(\frac{\mu}{y_i})$	Variance: $\frac{1}{N}\sum_{i=1}^{N}(y_i-\overline{y})^2$	Dissimilarity index: $\frac{1}{N\overline{y}}\sum_{i=1}^{N}\|y_i-\overline{y}\|$	Modified dissimilarity index: $\frac{2}{N}\sum_{i=1}^{N}\|y_i-\overline{y}\|$
Absolute measure θ_a	Yes	No	Yes	Yes
Relative measure θ_r	Yes	Yes	No	No
Translation invariant	No	Yes	No	Yes
Scale invariant	Yes	Yes	Yes	No
Abbreviation used in `iop`	`fg1a` or `fg1r`	`fg2r`	`pdb`	`ws`

The methods presented in table 1 are for continuous and dichotomous variables. However, by dichotomizing ordered variables at each possible level, one can also apply the latter two methods to ordered variables. The methods are mainly different in terms of properties. For the continuous case, Ferreira and Gignoux (2011) use a method that is particularly well suited for variables such as income, which has an inherent scale. For example, income is naturally defined from zero to infinity. However, sometimes, the continuous variable has no such natural points. For instance, student test scores can be translated and rescaled without losing the sense of the variable. Here the method proposed by Ferreira and Gignoux (2014) should be preferred because their measure is both translation and scale invariant, while the former method is only scale invariant.

With respect to dichotomous variables, two methods are proposed: one method ensures scale invariance, and the other method ensures translation invariance. Paes de Barros, de Carvalho, and Franco (2007) use a logit or probit model to estimate the conditional probability and to apply the dissimilarity index. This method ensures scale invariance of the inequality of opportunity measure, but it is sensitive to translation. It is used, for instance, to compute the Human Opportunity Index introduced by the World Bank and explained in Paes de Barros et al. (2009).[4] Soloaga and Wendelspiess Chávez Juárez (2013b) apply a variation of this method that focuses on translation invariance of the measure instead of on scale invariance.

We focus on these four references for two reasons. First, these methods have been used the most in recent empirical work. Second, these methods are members of a larger family of methods, and they allow the user to also estimate some related approaches. For instance, by creating dummies for each type and using them as circumstances, we get the nonparametric estimator proposed by Checchi and Peragine (2010). Moreover, `iop` can handle other non- or semiparametric methods, such as splines.

2.2 Decompositions of the inequality of opportunity measure

The regression approach provides us with a point estimate of absolute or relative inequality of opportunity. However, to fully understand the phenomenon of inequality of opportunity and its evolution, one may want to further decompose the measure. There are two interesting decompositions. First, we can decompose inequality of opportunity in a given country into its sources by estimating the relative importance of each circumstance. This decomposition is based on the Shapley value. Second, we can decompose the difference in inequality of opportunity between two populations. For example, the different populations can refer to different countries or to the same country in two points of time, by gender, etc. This Oaxaca decomposition allows us to distinguish which part of the difference is due to different distributions of circumstances and which part is due to differences in how the circumstances affect the outcome. We will now discuss the two decomposition methods in more detail.

4. To estimate the Human Opportunity Index, Stata users can download the command `hoi` (Azevedo et al. 2010).

The Shapley decomposition

The measure of total inequality of opportunity can be divided into its components, attributing a part of total inequality to each circumstance. We will use the Shapley decomposition. To compute the Shapley decomposition, we first estimate the inequality measure for all possible permutations of the circumstance variables. We then compute the average marginal effect of each circumstance variable on the measure of inequality of opportunity. This procedure is computationally intensive because 2^K (K = number of circumstances) must be computed. However, the Shapley decomposition has substantial advantages over other decomposition methods. First, the decomposition is order independent, and second, the different components equal the total value.

As a note of caution, Ferreira and Gignoux (2014) argue that such decomposition should not be seen as causal and can give only an idea of the relative importance. This is because most circumstances are highly correlated, so the coefficients might suffer from multicollinearity. This multicollinearity is a problem for the decomposition but not for the point estimates of inequality of opportunity.

Group decomposition in the spirit of Oaxaca

A second decomposition of inequality of opportunity that might be of interest is the decomposition in subgroups, for instance, women and men. To have inequality of opportunity, two conditions must be satisfied: people have to differ in circumstances (the composition effect), and these circumstances must affect the outcome. Therefore, differences in inequality of opportunity can be based on differences in the circumstances (composition effect) or on differences in the impact of circumstances on the outcome (association effect). We propose this decomposition by computing the inequality of opportunity for each group individually and then by computing counterfactual inequality of opportunity measures. For instance, this is done by computing the level of inequality of opportunity of women by using the returns to circumstances (the estimated regression coefficients) of men. All differences between the true value for women and this counterfactual measure are attributable to differences in the circumstances (composition effects). Table 2 shows the four possible inequality measures that can be computed for the two groups.

Table 2. Group decomposition

	Coefficients	
Distribution	Men	Women
Men	$I\left(X_{\male}\widehat{\beta}_{\male}\right)$	$I\left(X_{\male}\widehat{\beta}_{\female}\right)$
Women	$I\left(X_{\female}\widehat{\beta}_{\male}\right)$	$I\left(X_{\female}\widehat{\beta}_{\female}\right)$

On the diagonal (upper left to lower right), we have the actual inequality of opportunity estimates for both genders. The upper-right value is the counterfactual estimate using the coefficients of women and the composition of circumstances of men. The lower-left value is based on men's coefficients and women's circumstances. This decomposition approach has been used, for instance, by Contreras et al. (2012) in a study on inequality of opportunity in Chile. Besides comparing two groups, this method also allows the researcher to analyze the differences in one group for two different points in time. Understanding where the differences between two groups or two periods are coming from is crucial for policy design.

3 The iop command

3.1 Syntax

`iop` *depvar* [*indepvars*] [*if*] [*in*] [*weight*] [`, detail shapley(`*stat*`) sgroup(`*str*`)`
`oaxaca(`*groupvar stat*`) type(d|o|c) logit bootstrap(`*int*`)`]

where *depvar* is the outcome variable (for example, income or access to education), and *indepvars* are the circumstance variables as defined in section 2. *stat* refers to the measure of inequality of opportunity that should be decomposed, and *groupvar* is a categorical variable containing the definition of subgroups of the sample (for example, gender dummy).

`fweight`s and `iweight`s are allowed; see [U] **11.1.6 weight** for details.

Note that the first version of `iop` had a different syntax and estimated only the method proposed by Paes de Barros, de Carvalho, and Franco (2007).[5] The old syntax is still working to ensure backward compatibility.[6] Nevertheless, we encourage all users to switch to the new syntax because it offers more convenient analyses.

3.2 Description

The command `iop` implements the four methods presented in table 1 and performs the two decomposition methods presented in section 2.2. First, we can compute a decomposition in the relative contribution of each circumstance by using the idea of the Shapley decomposition (Shorrocks 1982). Second, we can compute a decomposition for subpopulations defined in variable *groupvar* by using the Oaxaca–Blinder decomposition (Oaxaca 1973; Blinder 1973).

5. See Soloaga and Wendelspiess Chávez Juárez (2013a) for details.
6. `iop` automatically recognizes which syntax the user requests and adapts the analysis. When one uses the old syntax, a warning is displayed.

The point estimates of inequality of opportunity

The algorithm used by `iop` is very simple and is based on existing Stata commands. For binary variables, `iop` first estimates a probit model of the outcome variable on the set of circumstances. For continuous variables, it performs an OLS estimation. For ordered variables, `iop` estimates the probit model on each possible definition of the dichotomous variable, meaning that it creates a new dummy variable for each level of the ordered variable.

Once the algorithm estimates the regression, either probit or OLS, it computes the predicted values and applies the corresponding inequality measure. This provides a point estimate of inequality of opportunity. To get the relative measure, it further divides the values by the same inequality measure (for example, by the mean log deviation) of the original outcome variable.[7]

3.3 Options

`iop` has seven options to adapt the analysis to the researcher's needs. Three options are used to activate and adapt the decomposition methods, one option is used to activate the bootstrap standard errors, and the remaining options allow the user to change from the probit to the logit model, to display more details, and to correct the type of dependent variable if it was guessed incorrectly.

`detail` makes the underlying regressions (OLS, probit, or logit) visible. By default, these regressions are not displayed.

`shapley(`*stat*`)` estimates the relative importance of each circumstance variable. The argument tells `iop` which statistics to decompose. The possible values depend on the type of the variable:

Type	Possible arguments
Continuous	`fg1a`, `fg1r`, and `fg2r`
Dummy/Ordered	`pdb` or `ws`

The Shapley decomposition becomes very computationally intensive when the number of circumstances increases. Therefore, it is advisable to use this option with only a few circumstance variables.

`sgroup(`*str*`)` allows the user to group some circumstance variables when computing the Shapley value and to reduce the number of computations required. Grouping variables makes sense particularly when the variables are directly related (for example, father's and mother's education) or when they are inseparable (age and age

7. The relative measure is available for only continuous outcome variables, because the inequality measure is equally defined for the actual and the conditional outcome. For binary variables, the actual outcome is dichotomous, while the conditional outcome (probability) is continuous. This makes it impossible to compute the relative inequality of opportunity measure in a sound way.

squared). To define the groups, the user has to indicate the variable names and separate the groups by a comma. For instance, assume we have the 4 variables `x1`, `x2`, `z1`, and `z2` and would like to group the `x` and the `z` variables. To do this, we indicate `sgroups(x1 x2, z1 z2)`. In this case, the computation of the Shapley value requires $2^2 = 4$ instead of $2^4 = 16$ estimations. Note that the grouping of variables affects the computation of the Shapley value but does not affect the estimation of inequality of opportunity.

`oaxaca(`*groupvar stat*`)` activates the Oaxaca-type decomposition. The option takes two string arguments. The argument *groupvar* indicates the variable that contains the groups, and the second argument indicates which statistics must be decomposed. The group variable must be numeric and can contain value labels that are used in the display to make the output more readable. The decomposition works for only the absolute measures of inequality of opportunity (`fg1a`, `pdb`, `ws`). For the relative measures, such decomposition does not make sense, because the difference might also be due to the total amount of inequality. By correcting for that, we would be back to the absolute measure. For ordered variables, the option `oaxaca()` is not implemented, because it would yield an unmanageable amount of decompositions. In this case, a certain threshold should be chosen to dichotomize the ordered variable, and the decomposition should be used for only this threshold.

`type(d|o|c)` specifies the variable type. This option is optional because `iop` tries to figure out the type of the dependent variable on its own. If `iop` fails to identify the type, you can specify it with this option. The possible values are `d` (dummy variables), `o` (ordered variables), and `c` (continuous variables).

`logit` changes the model from the default probit to a logit model. This option is relevant only for dichotomous and ordered variables.

`bootstrap(`*int*`)` allows the user to add bootstrap standard errors to the point estimates. The argument *int* corresponds to the number of replications the user wants to estimate. Obtaining the bootstrap standard errors can be computationally intensive, so we suggest the users start with a relatively small number of replications.

3.4 Stored results

`iop` stores the following in `r()`:

Scalars

`r(pdb)`	`pdb` measure
`r(ws)`	`ws` measure
`r(fg1a)`	`fg1a` measure
`r(fg1r)`	`fg1r` measure
`r(fg2r)`	`fg2r` measure
`r(pdbSD)`	bootstrap std. error of `pdb`
`r(wsSD)`	bootstrap std. error of `ws`
`r(fg1aSD)`	bootstrap std. error of `fg1a`
`r(fg1rSD)`	bootstrap std. error of `fg1r`
`r(fg2rSD)`	bootstrap std. error of `fg2r`
`r(bootN)`	number of bootstrap replications

Matrices

`r(iop)`	matrix with all inequality measures
`r(oaxaca)`	matrix of Oaxaca-type decomposition

The exact number of elements `iop` stores depends on the analysis performed. With respect to the scalars, it returns only computed values, thus it does not provide any empty scalars. For example, it provides the scalars with the bootstrap standard errors only when it uses the bootstrap method. For the matrices, the `r(iop)` is always given, while the matrix `r(oaxaca)` is provided only if such an analysis is performed.

4 Examples

In this section, we present some examples using the 2006 PISA data.[8] In a first example, we estimate the level of inequality of opportunity for a specific country, and we perform the Shapley decomposition to identify the main drivers. In a second example, we compare different countries and use the Oaxaca-type decomposition to figure out the origin of the differences.

Example 1: Analyzing inequality of opportunity in PISA scores

First, we estimate the level of inequality of opportunity for Germany by using the test scores in mathematics as dependent variables and a set of family characteristics as circumstances. Among these explanatory variables, we have the occupation status of the father, parental education, the number of books at home, and a dummy for immigrants. To estimate inequality of opportunity, we indicate the dependent variable, and then we indicate the set of circumstances and specify the `if` qualifier to limit the analysis to Germany (`if cnt=="DEU"`). To complete the simple estimation, we ask `iop` to decompose the statistic `fg2r` by circumstances by using the Shapley decomposition (`shapley(fg2r)`).

8. Available at http://pisa2006.acer.edu.au/downloads.php.

For the Shapley decomposition, we define four variable groups: mother's and father's education (grouped together); a dummy for immigrants; the number of books at home; and the three indicators for the occupation of the father are grouped together.

The output of iop is given as follows:

```
. local circumstances="misced fisced immig books fcat1 fcat2 fcat3"
. iop pv1math `circumstances´ if cnt=="DEU", bootstrap(100) shapley(fg2r)
> sgroups(misced fisced,immig,books,fcat1 fcat2 fcat3)
I assume the variable to be: continuous
If this is not correct, use option type
Bootstrapping...done!
---------------------------------------------------------------------
    Inequality of opportunity in pv1math
---------------------------------------------------------------------
Method                                 |     Absolute   Relative
---------------------------------------------------------------------
Ferreira-Gignoux (with scale)          |     0.004109   0.225217
  Bootstrap std. err.                  |  (  0.000258)  ( 0.000258)
Ferreira-Gignoux (without scale)       |  not defined   0.241480
  Bootstrap std. err.                  |                ( 0.012628)
---------------------------------------------------------------------
Observations:              3967
Bootstrap replications:     100
Decomposition (Shapley method)
------------------------------
Variable            Value      In percentage
---------------------------------------------
Group 1             0.014590       6.04%
Group 2             0.034399      14.25%
Group 3             0.136462      56.51%
Group 4             0.056029      23.20%
---------------------------------------------
TOTAL               0.241480      100.00%
---------------------------------------------
The groups are defined as follows:
Group 1: misced fisced
Group 2: immig
Group 3: books
Group 4: fcat1 fcat2 fcat3
```

At the beginning of the output, iop indicates the type of variable that was assumed. In this case, iop correctly identified a continuous variable. If the variable type is detected incorrectly, it can be overwritten using the option type(). The main estimation of inequality of opportunity is presented in the first panel. In this case, the panel presents all three possible estimates for continuous variables. Because the PISA score has no inherent scale, we recommend using the second line (without a scale), which is the method proposed by Ferreira and Gignoux (2014).

The value 0.241 tells us that about one quarter of all heterogeneity in the PISA scores is due to observed circumstances. This means that about one quarter of total inequality can be considered to be ethically offensive and is not due to students' different levels of effort or to luck. The bootstrap standard errors below the point estimates are based on 100 replications and are about 1.2%, which is relatively small. Recall that this is a lower-bound estimate for the reasons outlined in the introduction and discussed in depth by Ramos and Van de gaer (2012).

The second panel provides results for the Shapley decomposition of the estimated inequality of opportunity. The results are presented by level and as percentages of total inequality of opportunity. In our example, the number of books at home accounts for more than half of total inequality of opportunity, while parental education categories (`fisced` and `misced`) do not account for much. A father's job categories (`fcat1`–`fcat3`) account together for about 23% of total inequality of opportunity, and immigration status accounts for about 14%. Note, however, that Ferreira and Gignoux (2014) argue that this decomposition must be used with caution. Highly correlated circumstances might lead to biased coefficients, which is not directly a problem for the estimation of θ_{IOP}. However, it might be problematic for the decomposition in relative contributions of circumstances.

Example 2: Dichotomous outcome and Oaxaca-type decomposition

In the second example, we use the same data but include Canada and the United States in the analysis. Instead of using the PISA score as we did above, we now use a binary indicator, taking the value of 1 for students that have achieved 500 points or more on the PISA test and 0 otherwise.[9] Moreover, we use the option `oaxaca(country pdb)` to decompose the measure proposed by Paes de Barros, de Carvalho, and Franco (2007) in the spirit of an Oaxaca–Blinder decomposition. The variable `country` is a categorical variable (numerical) with value labels.

```
. iop math500 misced fisced immig books fcat1 fcat2 fcat3, oaxaca(country pdb)
I assume the variable to be: dichotomous (dummy)
If this is not correct, use option type
```

Inequality of opportunity in *math500*		
Method	Absolute	Relative
PdB (Dissimilarity index)	.115558	not defined
ws (adapted DI)	.266169	not defined

```
Observations:              27832

Oaxaca-like decomposition

Group variable:    country
Statistic:         pdb
```

	Coefficients of		
Distribution	CAN	GER	USA
CAN	0.09212	0.13256	0.17369
GER	0.10727	0.15270	0.20991
USA	0.10433	0.15184	0.19557

9. Five hundred is the average PISA test score of all countries. We perform this dichotomization exclusively for illustrative purposes to show how `iop` handles binary variables.

The output produced by `iop` starts again by guessing the type of dependent variable and then provides the general analysis of all the countries together. The point estimates are 0.116 and 0.266 for the two methods, respectively. The adapted dissimilarity index (`ws`) is defined on the interval of 0 to 1, so the value of 0.266 suggests that a rather large amount of inequality is due to circumstances.

This general analysis is followed by the Oaxaca-type decomposition presented in matrix form. On the diagonal, we have the estimate for each country, where Canada displays the lowest level of inequality of opportunity, followed by Germany and the United States. The remaining values are counter-factual estimates, where the column refers to the estimated coefficients, and the rows refer to the composition. For instance, the value in the first column, `CAN`, and the last row, `USA`, would be the level of inequality of opportunity with the distribution of circumstances of the United States and the estimated coefficients of Canada. The value lies much closer to the original value of Canada, which suggests that most of the difference is due to differences in the link between circumstances and outcome, while very little of the difference is due to a different structure of circumstances.

Example 3: Ordered variables

A final example presents the output for ordered variables. For simplicity and for illustrative purposes, we use the same variables as before. Instead of dichotomizing the scores as we did for example 2, we create an ordered variable with 4 categories. The categories are

$$\texttt{mathORD} = \begin{cases} 1 & \text{if score} \leqslant 400 \\ 2 & \text{if } 400 < \text{score} \leqslant 500 \\ 3 & \text{if } 500 < \text{score} \leqslant 600 \\ 4 & \text{if } 600 < \text{score} \end{cases}$$

Additionally, to change the variable, we indicate to `iop` that we want to use logit instead of probit. The output is as follows:

```
. iop mathORD misced fisced immig books fcat1 fcat2 fcat3, logit
Note: logit was used instead of probit
I assume the variable to be: ordered
If this is not correct, use option type
----------------------------------------------------------
    Inequality of opportunity in mathORD
----------------------------------------------------------
Threshold            |    PdB        ws
----------------------------------------------------------
mathORD < 2          |  0.029191   0.105503
mathORD < 3          |  0.115895   0.267051
mathORD < 4          |  0.249776   0.173206
----------------------------------------------------------
Only absolute estimates are reported
Observations:     27832
```

Before showing the estimates, `iop` informs the user that it used logit instead of probit and that it detected an ordered variable. The actual output of results is very much like the output of results for dummy variables, the difference being that there are two estimates for every possible threshold of the ordered variable. In this respect, the first line provides the estimate for inequality of opportunity in the probability of having at least 400 points in the score. The second and the third lines are for at least 500 and at least 600 points, respectively. Note that the second line is comparable—but, because of the change from probit to logit, not identical—with example 2. The example shows that using the scale invariant measure (`PdB`), we find the highest values for the highest threshold, while the translation invariant measure indicates the highest level of inequality for the threshold in the middle.[10]

5 Concluding remarks and limitations

In this article, we described the user-written command `iop`, which estimates several methods to assess ex-ante inequality of opportunity. In addition to the point estimates, `iop` proposes two decompositions. The first decomposition allows the researcher to identify the relative importance of the included circumstances using the Shapley value. The second decomposition allows the researcher to better understand differences in inequality of opportunity between groups (for example, between countries or regions within a country) using an Oaxaca-type decomposition. The main goal of the command `iop` is to offer interested researchers an easy-to-use command that allows them to estimate inequality of opportunity with different methods. The choice of the implemented methods was driven by the recent use of these methods in empirical applications. We are confident that `iop` supports most of the commonly used methods.

In this respect, we would also like to highlight some limitations of the current version of the command. First, `iop` estimates only ex-ante inequality of opportunity. It would be difficult to combine the alternative ex-post methods in one command, because they require different data and have substantially different approaches. Second, `iop` supports parametric estimates by using OLS for continuous variables and by using probit and logit models for dichotomous variables. However, a nonparametric approach based on type averages can be estimated by using type dummy variables in the parametric regression. Finally, `iop` omits analytical standard errors of the estimators and limits itself to bootstrap standard errors for the point estimates. There are no bootstrap standard errors included for the decompositions, because their statistical properties are unclear.

We plan to further develop `iop` in accordance with the propositions of estimators for inequality of opportunity. We are always happy to receive comments and suggestions for future developments.

10. A discussion on the conceptual differences between the two measures can be found in Soloaga and Wendelspiess Chávez Juárez 2013b.

6 References

Azevedo, J. P., S. Franco, E. Rubiano, and A. Hoyos. 2010. hoi: Stata module to compute Human Opportunity Index. http://ideas.repec.org/c/boc/bocode/s457191.html.

Blinder, A. S. 1973. Wage discrimination: Reduced form and structural estimates. *Journal of Human Resources* 8: 436–455.

Checchi, D., and V. Peragine. 2010. Inequality of opportunity in Italy. *Journal of Economic Inequality* 8: 429–450.

Contreras, D., O. Larrañaga, E. Puentes, and T. Rau. 2012. The evolution of opportunities for children in Chile, 1990–2006. *Cepal Review* 106: 107–124.

Ferreira, F. H. G., and J. Gignoux. 2011. The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of Income and Wealth* 57: 622–657.

———. 2014. The measurement of educational inequality: Achievement and opportunity. *World Bank Economic Review* 28: 210–246.

Fleurbaey, M., and V. Peragine. 2013. *Ex ante* versus *ex post* equality of opportunity. *Economica* 80: 118–130.

Oaxaca, R. 1973. Male–female wage differentials in urban labor markets. *International Economic Review* 14: 693–709.

Paes de Barros, R., M. de Carvalho, and S. Franco. 2007. Preliminary notes on the measurement of socially-determined inequality of opportunity when the outcome is discrete. Working paper.

Paes de Barros, R., F. H. G. Ferreira, J. R. Molinas Vega, and J. Saavedra Chanduvi. 2009. *Measuring Inequality of Opportunity in Latin America and the Caribbean*. Washington, DC: The World Bank.

Ramos, X., and D. Van de gaer. 2012. Empirical approaches to inequality of opportunity: Principles, measures, and evidence. Working paper, ECINEQ. http://ftp.iza.org/dp6672.pdf.

Roemer, J. E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.

Shorrocks, A. F. 1982. Inequality decomposition by factor components. *Econometrica* 50: 193–211.

Soloaga, I., and F. Wendelspiess Chávez Juárez. 2013a. iop–Estimar desigualdad de oportunidades cuando el indicador es binario. In *Aplicaciones en Economía y Ciencias Sociales con Stata*, ed. A. M. Velázquez. College Station, TX: Stata Press.

———. 2013b. Scale vs. translation invariant measures of inequality of opportunity when the outcome is binary. http://iariw.org/c2013brazil.php.

About the authors

Florian Wendelspiess Chávez Juárez is a postdoctoral scholar at the University of Geneva, Switzerland.

Isidro Soloaga is professor of economics at the Universidad Iberoamericana in Mexico City.