The Stata Journal (*yyyy*) *vv*, Number *ii*, pp. 1–14

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. It is therefore the ethically offensive part of inequality. Several estimation procedures were proposed over the last years. `iop` is a comprehensive and easy-to-use command which implements many of them. It handles continuous, dichotomous and ordered variables. Additionally to the point estimates, `iop` provides bootstrap standard errors and two decomposition methods.

Keywords: st0001, iop, inequality of opportunity, dissimilarity index, mean log deviation, decomposition

1 Introduction

The concept of inequality of opportunity received much attention over the last decade in development economics. 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 non-offensive inequality. Effort should lead to different outcomes, thus inequality due to different effort levels might be desirable. In contrast, the ethically offensive part of inequality is the part that is due to circumstances beyond the control of individuals. Circumstances are all 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 take a school exam as example. If students get different grades because they studied more or less, we consider the inequality in the grade as something desirable. If the differences in grades were only due to the family background and not to different effort level, the same inequality would be considered as ethically offensive. The goal is therefore to split total inequality into its ethically offensive and non-offensive 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 equality of opportunity is achieved when all people making the same degree of effort achieve the

 st0001

same outcome, independently of their circumstances. While the two approaches seem to differ only marginally at first sight, there are important differences making them incompatible in some cases[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 as compared to the *ex-post* approach. For both approaches the main challenge is that both effort and luck are generally not observable and therefore it is difficult to distinguish them empirically. While the *ex-post* approach requires at least an estimate of effort, it is possible to estimate *ex-ante* inequality of opportunity without it. This is likely to be the main reason why empirical applications focus mostly on *ex-ante* inequality of opportunity, given that the estimation of effort requires very strong assumptions. We follow the empirical applications and focus on *ex-ante* inequality of opportunity.

Several methods to assess *ex-ante* inequality of opportunity have been proposed over the last years. The regression approach became a very popular approach 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 non-parametric regression methods. The intuition is that in a world of equal opportunities the circumstances should not matter and therefore the regression should have a very low fit. In case of finding an effect of circumstances on 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. The main reason is that the part of inequality due to unobserved circumstances might be wrongly attributed to effort and luck instead of 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 Stata command `iop`, which implements several recently proposed methods to estimate *ex-ante* inequality of opportunity. `iop` is able to 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 (2013) for continuous variables. For dichotomous outcome variables we implement the method proposed by Paes de Barros et al. (2007) and a translation invariant version of it suggested by Wendelspiess Chávez Juárez and Soloaga (2013). The focus on these methods is justified by their practical use in recent empirical applications and the ability to use them also 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).

In addition to the point estimates of inequality of opportunity, two decomposition meth-

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, within a type all circumstances are identical.

ods are proposed and implemented.
First, total inequality of opportunity can be decomposed according to the different circumstances using the Shapley decomposition. The idea of this decomposition is to understand which circumstances are driving inequality of opportunity. It therefore allows the user to understand not only how much all circumstances affect inequality, but which circumstance contribute by how much 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. The idea of this decomposition is to analyze differences in the level of inequality of opportunity between two geographical units or the same unit in different times. The Oaxaca decomposition identifies what part of the observed differences is due to the fact that the distributions of circumstances are different and what part to the fact that the impact of the circumstances on the outcome variable is different.

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 along with examples using the 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

As mentioned in the introduction, there are different approaches to assess inequality of opportunity. The regression approach comprises a large number of approaches, which all assess *ex-ante* inequality of opportunity. To discuss the general idea of this family of methods, we first introduce some notation. Let y be the outcome variable of interest and C 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 a general way, we can describe this by the expected conditional outcome

$$\widehat{y} = E[y|C] \tag{1}$$

which can be estimated in different ways according to the research question and the dependent variable. For instance, Paes de Barros et al. (2007) has a binary outcome variable (e.g. access to schooling) and use a logit or probit model to estimate equation (1). Ferreira and Gignoux (2011) use income as dependent variable and estimate the same equation with an OLS regression and also with non-parametric 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 non-parametric estimation techniques to asses equation (1). Finally, Ferreira and Gignoux (2013) use linear regression for test scores.

Independently of the way equation (1) is estimated, inequality of opportunity is then

3. A *type* is defined as a group of individuals sharing the same circumstances

computed using a common inequality measure $I(.)$ applied to $\widehat{y}$:

$$\theta_a = I(\widehat{y}) \tag{2}$$

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 the dependent variable. Paes de Barros et al. (2007) use the dissimilarity index, Ferreira and Gignoux (2011) the mean logarithmic deviation and Ferreira and Gignoux (2013) the variance. Besides this absolute inequality measure, we can divide it by the same metric applied to the actual outcome to get a relative measure of inequality of opportunity:

$$\theta_r = \frac{I(\widehat{y})}{I(y)} \tag{3}$$

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

The choice of the appropriate inequality measure $I(.)$ 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` .

The methods presented in Table 1 are for continuous and dichotomous variables. However by dichotomizing ordered variables at each possible level, the latter two methods can also be applied to ordered variables. The differences of the methods are mainly in terms of properties. For the continuous case, Ferreira and Gignoux (2011) is particularly well suited for variables such as income, which have an inherent scale. Income for instance is naturally defined from zero to infinity. In some cases 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. In this case, the method proposed by Ferreira and Gignoux (2013) should be preferred, as their measure is both translation and scale invariant, while the former is only scale invariant.

With respect to dichotomous variables two methods are proposed, one assuring scale invariance and the other translation invariance. Paes de Barros et al. (2007) uses a logit or probit model to estimate the conditional probability and applies then 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 for the computation of the Human Opportunity Index (HOI) introduced by the World Bank and explained in Paes de Barros et al. (2009)[4]. Wendelspiess Chávez Juárez and Soloaga (2013) is simply a variation of this method which focuses on translation invariance of the measure instead of scale invariance.

We focus on these four references for several reasons. First, these methods seem to be the most used in recent empirical work. Second, these methods are members of a larger family of methods and allow the user also to estimate some related approaches.

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

	Ferreira and Gignoux (2011)	Ferreira and Gignoux (2013)	Paes de Barros et al. (2007)	Wendelspiess Chávez Juárez and Soloaga (2013)
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: $\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{y})^2$	Dissimilarity index: $\frac{1}{N\bar{y}}\sum_{i=1}^{N}\|y_i-\bar{y}\|$	Modified dissimilarity index: $\frac{2}{N}\sum_{i=1}^{N}\|y_i-\bar{y}\|$
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 / fg1r`	`fg2r`	`pdb`	`ws`

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

For instance, by creating dummies for each type and using them as circumstances, we get the non-parametric estimator proposed by Checchi and Peragine (2010). Moreover, `iop` can also handle other non- or semi-parametric 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, in order to fully understand the phenomenon of inequality of opportunity and its evolution, one might want to decompose the measure further. 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 idea of the Shapley value. Second, we can decompose the difference in inequality of opportunity between two populations. The different populations can for example refer to different countries, the same country in two points of time, gender, etc. This Oaxaca-type 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 the effect of these circumstances on the outcome.

We will now discuss the two decomposition methods in some more details.

The Shapley-decomposition

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

As a note of caution, Ferreira and Gignoux (2013) argue that such decomposition should not be seen as causal and can only give an idea of the relative importance. The reason is that most circumstances are highly correlated and therefore the coefficients might suffer the problem of multicollinearity. This multicollinearity is only 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 particular interest is the decomposition in subgroups, for instance women and men. In order to have inequality of opportunity two conditions must be satisfied. First, people have to differ in circumstances (the composition effect) and these circumstances must have an effect on the outcome. Differences in inequality of opportunity can therefore be based on differences in the circumstances (composition effect) and/or 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 in a second step by computing counterfactual inequality of opportunity measures. This is done, for instance, by computing the level of inequality of opportunity of women using the returns to circumstances (the estimated regression coefficients) of men. All difference between the true value for women and this counterfactual measure would then be attributable to differences in the circumstances (composition effects). Table 2 shows the 4 possible inequality measure that can be computed for two groups.

	Coefficients	
Distribution	Men	Women
Men	$I(X_{♂}\hat{\beta}_{♂})$	$I(X_{♂}\hat{\beta}_{♀})$
Women	$I(X_{♀}\hat{\beta}_{♂})$	$I(X_{♂}\hat{\beta}_{♂})$

Table 2: Group decomposition

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. Finally, 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 allows the researcher also to analyze the differences for the same 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 (e.g. 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 (e.g. gender dummy).

`fweights` and `iweights` 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 et al. (2007)[5]. The old syntax is still working to ensure backward compatibility[6]. Nevertheless, we encourage all users to switch to the new syntax, as 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, a decomposition in the relative contribution of each circumstance can be computed, using the idea of the Shapley decomposition (Shorrocks 1982). The second decomposition can be made for subpopulations defined in variable *groupvar* and follows the idea of the Oaxaca-Blinder decomposition (Oaxaca 1973; Blinder 1973).

The point estimates of inequality of opportunity

The algorithm used by `iop` is very simple and based on existing Stata commands. For binary variables, `iop` first estimates a probit model of the outcome variable on the set

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

of circumstances. For the case of continuous variables an OLS estimation is performed. For ordered variables, the probit model is estimated on each possible definition of the dichotomous variable, meaning that for each level of the ordered variable a new dummy variable is created.

Once the regression, either probit or ordinary least squares, is estimated, the predicted values are computed and the corresponding inequality measure applied. This provides a point estimate of inequality of opportunity. To get the relative measure, the value is further divided by the same inequality measure (e.g. mean log deviation) of the original outcome variable[7].

3.3 Options

`iop` has seven options to adapt the analysis to the needs of the researcher. Three options are used to activate and adapt the decomposition methods, one to activate the bootstrap standard errors and the remaining allow the user to change from the probit to the logit model, to display more details and to correct a wrong guessing of the type of dependent variable.

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

`shapley(`*stat*`)` is the option needed to estimate the relative importance of each circumstance variable. The argument is used to tell `iop` which statistics has to be decomposed. 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, it is therefore advisable to use this option only with a few circumstance variables.

`sgroup`(*str*) allows the user to group some circumstance variables in the computation of the Shapley-value and consequently to reduce the number of computations required. Grouping variables makes particularly sense when they are directly related (e.g. father's and mother's education) or if they are simply not separable (age and age 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 4 variables x1, x2, z1 and z2 and would like to group the x and the z variables. To do this, simply indicate `sgroups(x1 x2, z1 z2)`. In this case the computation of the Shapley-value requires only $2^2 = 4$ instead of

7. The relative measure is only available for continuous outcome variables, as 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.

$2^4 = 16$ estimations. Note that the grouping of variables affects only the computation of the Shapley-value and not the estimation of inequality of opportunity.

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

`type(d|o|c)` tells `iop` which type of variable the outcome variable is. This option is normally not needed, as `iop` detects the type automatically. In case `iop` does not identify the type correctly, this option overwrites `iop`'s guess.

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

`bootstrap`(*int*) allows the user to add bootstrap standard errors to the point estimates. The argument *int* correspond to the number of replications the user wants to estimate. Note that obtaining the bootstrap standard errors can be computationally intensive, we therefore suggest the user to start with a relatively small number of replications.

3.4 Returned results

`iop` is a `r-class` command and returns all computed statistics as scalars and matrices:

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` returns depends on the performed analysis. With respect to the scalars, only computed values are returned, thus no empty scalars are provided. For example, the scalars with the bootstrap standard errors are only provided when the bootstrap method was used. For the matrices, the `r(iop)` is always given, while the matrix `r(oaxaca)` is only provided if such an analysis was performed.

4 Examples

In this section we present some examples using the 2006 PISA data[8]. In a first step, we estimate simply the level of inequality of opportunity for a specific country and 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 where the differences are coming from.

Example 1: Analyzing inequality of opportunity in PISA scores

First, we estimate the level of inequality of opportunity for Germany, using the test scores in mathematics as dependent variable 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 simply indicate first the dependent variable, followed by the set of circumstances and the if-condition to limit the analysis to Germany (`if cnt=="DEU"`). Additionally to the simple estimation, we ask `iop` to decompose the statistic `fg2r` by circumstances using the Shapley-decomposition (`shapley(fg2r)`). For the Shapley-decomposition, we define 4 groups of variables. Father's and mother's education on the one hand and the three indicators for the occupation of the father on the other 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", boot(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
--------------------------------------------
```

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

```
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 very beginning of the output, `iop` indicates the type of variable that was assumed. In this case, `iop` correctly identified a continuous variable. In case of a wrong detection, this can be overwritten using the option `type`. In the first panel, the main estimation of inequality of opportunity is presented. In this case, all three possible estimates for continuous variables are displayed. Since the PISA score has no inherent scale, it is advisable to use the second line (without scale), which is the method proposed by Ferreira and Gignoux (2013).

The value of 0.241 tells us that about one quarter of all heterogeneity in the PISA scores are due to observed circumstances. Put differently, about one quarter of total inequality can be considered to be ethically offensive and is not due to different effort levels of students. 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 estimates for the reasons outlined in the introduction and discussed in depth by Ramos and Van de gaer (2012).

The second panel of results is the Shapley-decomposition of the estimated inequality of opportunity. The result is presented in level and as percentage 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 (`fisced` and `misced`) do not account for much. Father's job categories (`fcat1-fcat3`) account together for about 23% of total inequality of opportunity and the immigration status for about 14%. Note, however, that Ferreira and Gignoux (2013) argue that the use of this decomposition has to be done with caution. Highly correlated circumstances might lead to biased coefficients, which is not directly a problem for the estimation of θ_{IOP}. However, for the decomposition in relative contributions of circumstances it might be problematic.

Example 2: Dichotomous outcome and oaxaca-type decomposition

In the second example we use the same data but include Canada and the United States to the analysis. Instead of the PISA score as above, we now use a binary indicator taking the value of 1 for students having achieved 500 points or more at the PISA test and 0 otherwise[9]. Moreover, we use the option `oaxaca(country pdb)` to decompose the measure proposed by Paes de Barros et al. (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
```

The output produced by `iop` starts again by indicating the guess on the type of dependent variable, followed by the general analysis of all 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 zero to one, thus the value of 0.266 suggest a rather large amount of inequality that 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 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, suggesting that the largest part of the difference is due to differences in the link between circumstances and outcome, while only very little is due to a different structure of circumstances.

9. 500 is the average of all countries. We perform this dichotomization exclusively for illustrative purpose in order to show how `iop` handles binary variables.

Example 3: Ordered variables

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

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

Additionally to the change of the variable, we indicate `iop` that we prefer logit instead of probit. The output is the following:

```
. 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 logit was used instead of probit and that it detected an ordered variable. The actual output of results is very much like the one for dummy variables, with the difference 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 line are for at least 500 and at least 600 points respectively. Note that the second line is comparable - but due to the change from probit to logit not identical - to 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].

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

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 allows the researcher to identify the relative importance of the included circumstances using the Shapley value. The second decomposition allows him or her to better understand differences in inequality of opportunity between groups (e.g. 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 allowing them to estimate inequality of opportunity with different methods. The choice of the implemented methods was essentially driven by the recent use of these methods in empirical applications. We are therefore 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 a same command, as they require different data and have a substantially different approach. Second, `iop` supports only parametric estimates using ordinary least squares for continuous and probit and logit models for dichotomous variables. A non-parametric approach based on type-averages can, however, also be estimated by the use of type-dummy variables in the parametric regression. Finally, `iop` does not include analytical standard errors of the estimators and limits itself to bootstrap standard errors for the point estimates. For the decompositions there are no bootstrap standard errors included because their statistical properties are unclear.

As a concluding remark we would highlight that we plan to further develop `iop` in accordance with the propositions of estimators for inequality of opportunity. In this respect, we are always happy to receive comments and suggestions for future developments.

6 References

Azevedo, J. P., S. Franco, E. Rubiano, and A. Hoyos. 2010. HOI: Stata module to compute Human Opportunity Index. Statistical Software Components, Boston College Department of Economics. http://ideas.repec.org/c/boc/bocode/s457191.html.

Blinder, A. S. 1973. Wage Discrimination: Reduced Form and Structural Estimates. *The 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: pp.107–124.

Ferreira, F. H., and J. Gignoux. 2011. The Measurement of Inequality of Opportunity: Theory and an Application to Latin America. *The Review of Income and Wealth* 57(4): pp. 622–657.

———. 2013. The Measurement of Educational Inequality: Achievement and Opportunity. Forthcoming in The World Bank Economic Review. Advance Access published February 20, 2013. doi: 10.1093/wber/lht004.

Fleurbaey, M., and V. Peragine. 2013. Ex ante versus ex post equality of opportunity. *Economica* 80(317): 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. Mimeo.

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

Ramos, X., and D. Van de gaer. 2012. Empirical approaches to inequality of opportunity: Principles, measures, and evidence. ECINEQ Working Paper 2012-259.

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

Shorrocks, A. 1982. Decomposition by Factor Components. *Econometrica* 50(1): pp. 193–211.

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

Wendelspiess Chávez Juárez, F., and I. Soloaga. 2013. Scale vs. Translation Invariant Measures of Inequality of Opportunity when the Outcome is Binary. http://ssrn.com/abstract=2226822.

About the authors

Florian Wendelspiess Chávez Juárez is PhD-student in econometrics at the University of Geneva, Switzerland.

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