Dipartimento di Scienze economiche e metodi matematici

Southern Europe Research in Economic Studies

Upward and downward bias when measuring inequality of opportunity

Paolo Brunori, Vito Peragine, Laura Serlenga

SERIES Working Papers n. 05/2016

Upward and downward bias when measuring inequality of opportunity

Paolo Brunori[*], Vito Peragine[†], Laura Serlenga[‡]

September 14, 2016[§]

Abstract

We show that, when measuring inequality of opportunity with survey data, scholars incur two types of biases. A well-known downward-bias, due to partial observability of circumstances that affect individual outcome, and an upward bias, which depends on the econometric method used and the quality of the available data. We suggest a simple criterion to balance between the two sources of bias based on cross validation. An empirical application, based on 26 European countries, shows the usefulness of our method.

Keywords: inequality of opportunity, model selection, variance-bias trade-off.
JEL: C52, D3, D63

[*]University of Bari & Life Course Centre, corresponding author: Dipartimento di Scienze Economiche e Metodi Matematici, Largo Santa Scolastica 53, 70124, Bari. paolo.brunori@uniba.it

[†]University of Bari.

[‡]University of Bari.

[§]This paper was presented at the worksop *Equality of opportunity and social mobility. Towards an international database*, 10-11 June 2016, University of Bari. We are grateful to all participants for their comments and suggestions.

1 Introduction

The measurement of inequality of opportunity (IOp) is a growing topic in economics, in the last two decades the number of empirical contributions to this literature has exploded (Ferreira and Peragine, 2015; Roemer and Trannoy, 2015). The vast majority of this contributions are based on the approach proposed by Roemer (1998) to measure IOp. This method consists in a two-step procedure: first, starting from a distribution of outcome (typically income or consumption), a counterfactual distribution is derived, which reproduces only unfair inequality (i.e inequality due to circumstances beyond individual control) and does not reflect any inequality arising from choice and effort of individuals; second, the inequality is measured in this counterfactual distribution.

The empirical literature has extensively used two approaches to estimate the counterfactual distribution based on survey data: a parametric and a non-parametric approach. One of the main drawbacks of both approaches is that, unless all circumstances beyond individual control that affect outcome are observable, they produce biased estimate of IOp. While the magnitude of this bias may be impossible to identify (Bourguignon et al., 2013), under few assumptions it can be shown that the sign of the bias is negative (Roemer, 1998; Ferreira and Gignoux, 2011; Luongo, 2011). This explains why IOp estimates are generally interpreted as lower-bound estimates of the real one. Some authors have challenged the usefulness of those lower bound measures (Kanbur and Wagstaff, 2015; Balcazar, 2015). In particular, Balcazar (2015) and Ibarra et al. (2015) have suggested that the downward bias in estimated IOp may lead to a substantial underestimation of the real level of IOp. Moreover, Ibarra et al. (2015) run a number of simulation showing that not observing the top 5 percent of the income distribution produces downward biased IOp estimates.

Typically, authors address this problem by using - if available - rich data sources containing an increasing number of circumstances (Biörklund et al., 2012). In this paper, we show that attempts to reduce the downward bias by increasing the number of circumstances, might increase the variance of the estimated counterfactual distribution and, in turn, give rise to an upward bias. The magnitude of the distorsion depends on the quality of the data used and the flexibility of the model implemented. Even if in some cases the bias can be substantial, we notice that this aspect has been surprisingly neglected in the empirical literature the IOp measurement.

In what follows we suggest a method to select the best econometric specification in order to balance both sourse biases. The method is based on cross validation and can be easily implemented with survey data. In order to show the usefulness of our approach we implement the proposed method to estimate IOp in 26 European countries.

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

2 Downward and upward biased IOp

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

$$y = g(C, e)$$

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

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

2.1 Counterfactuals estimation

Ex-ante IOp can be estimated either by a parametric or a non parametric approach. Checchi and Peragine (2010) propose to estimate $\tilde{Y}$ non-parametrically following the typical two stages: (i) after partitioning the sample into types on the basis of all observable circumstances, they choose the arithmetic mean of type t, denoted by μ_t, as the value v_t of type t, hence they estimate $\mu_1, ..., \mu_T$; (ii) for each individual i belonging to type t they define $\tilde{y}_i = \hat{\mu}_t$ - where is the sample estimate for μ_t - and measure inequality in $\tilde{Y}$. The reliability of those estimates requires a sufficient number of observations in each type and, in practice, this might represent a severe constraint as individuals are, most likely, not uniformly distributed across types. In order to overcome this drawback, scholars tend to limit the number of circumstances in the definition of types or to aggregate types in broader categories: for example, districts of birth are aggregated in macro-region, parental occupations only distinguishes white from blue collar, and so on.

Bourguignon et al. (2007) propose to measure *ex-ante* IOp estimating $\tilde{Y}$ parametrically as the prediction of the following reduced form regression

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

where c_{ij_k} identifies each category by means of a dummy variable and χ_{j_k} is the corresponding coefficient. The parametric approach does not directly measure types' mean but captures the variability explained by the circumstances by a linear approximation of their effect on outcome. In particular, parametric estimation has been proposed as a good alternative to the nonparametric one when few observations are available (Ferreira and Gignoux, 2011; Ibarra et al., 2015).

However, we notice that the two methods coincide when the counterfactual is obtained by the prediction of a regression model where y is regressed on all possible interactions among the circumstances[1]. In this case each regressor captures the effect of belonging to one of all the possible circumstances combination, which is the effect of belonging to a given type:

$$y_i = \sum_{t=1}^{T} \beta_t \pi_{it} + u_i \tag{2}$$

where π_{it} are T binary variables obtained by interacting all categories.

In all other cases, the corresponding IOp measures might be very different, and - by construction - the parametric approach (1) will explain less inequality than the nonparametric (2). Moreover, while the linear specification might be too restrictive, the choice to include the full number of combinations among categories might lead to a large variance of the estimated counterfactual distribution.

[1]Note that this does not apply to continuous variables which are however very rarely used to measure circumstances in empirical contributions.

2.2 Variance-bias trade-off in estimating IOp

It has been shown that, if the "true" set of circumstances is not fully observable, the estimated IOp will be lower than the real IOp. This result follows from the assumption of orthogonality between circumstances and effort (see on this Roemer, 1998) and explains why IOp measures are generally interpreted as lower-bound estimates of IOp.

Typically authors attempt to solve this problem by using rich dataset containing an increasing number of circumstances (Bjorklund et al., 2012). Recently, Niehues and Peichl (2014) endorse an extreme perspective and, by exploiting longitudinal datasets, they measure IOp including individual fixed effects among circumstances beyond individual control.

However, when using survey data, whenever one makes an effort to reduce the downward bias by increasing the number of circumstances, she obtains a counterfactual distribution based on a finer partition in types where, by construction, each type contains less observations. This strategy might lead to both (i) higher between-group inequality and (ii) a larger sample variance when estimating the effect of C on y.

Surprisingly, the empirical literature on the estimates of IOp has so far neglected this second implication[2]. We face the classical variance-bias trade-off: if we are willing to reduce the downward bias, we have to accept higher uncertainty on the shape counterfactual distribution.

Following similar reasoning, it is important to notice that, when measuring inequality, higher variance of the estimated distribution implies an upward bias. This result is easily shown applying what Chakravarty and Eichhron (1994) proved for the case of estimating inequality when the variable of interest is measured with an error. The same result can be applied here: instead of the classical measurement error discussed in Chakravarty and Eichhron (1994), we consider a variable - the type mean - which is estimated with a higher sample variance, the finer the partition in types is.[3]

This discussion clarifies that, when estimating IOp, we should be worried about two sources of distortion that bias our measure in opposite directions: partial observability and sample variance of the counterfactual distribution. The following section proposes a simple method for choosing the best model, i.e. the best way to exploit information contained in survey data, balancing the two biases.

[2]Brunori et al. (2015) working with Sub-Saharan African surveys have only noticed that the use of very detailed circumstances such as 'village of birth' tends to dramatically increase estimated IOp.

[3]A formal proof is available upon request.

3 Model selection for measuring IOp

In this section we propose a method to select the most suitable model among: the simple linear model (1), which provides the lowest extreme IOp estimates; a flexible model which includes the full number of combinations among categories (2) and leads to the highest value of the IOp estimates; and all the intermediate specifications which only include subset of categories' combinations.

In a statistical learning framework, we evaluate the variance bias trade-off in terms of model predictions. On one hand, a more flexible model reduces the typical downward bias in IOp measurement but increases the prediction variance leading to upward (IOp) bias. On the other hand, a more restricted model reduces the variance and hence the upward bias, but suffers of omitted variable bias, the typical downward bias well known in the literature. In what follows, we propose to exploit the property of Mean Square Error (MSE) and choose the best model conditioned to available information by means of Cross Validation (CV). In a regression setting the MSE is defined as:

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

where y is the dependent variable, x are the regressors, and $i = 1, ..., n$ are the observations. For given out of sample observations y_0 and x_0, the MSE can always be decomposed in variance of $\hat{f}(x_0)$, square bias of $\hat{f}(x_0)$ and variance of the error term such as

$$E\left(y_0 - \hat{f}(x_0)\right)^2 = Var(\hat{f}(x_0)) + \left[Bias\left(\hat{f}(x_0)\right)\right]^2 + Var(\epsilon)$$

where $\hat{f}(x_0)$ are the predictions. Since the variance of the error cannot be reduced, it turns out that in order to minimize the MSE, we need to minimize both the bias and the variance. A comparison among different specifications is performed by CV. In a CV procedure, the sample is randomly divided into k equal-sized parts. Leaving out part k (test sample), the model is fitted to the other $k - 1$ parts (training sample) whereas out of sample predictions are obtained for the left-out k^{th} part.[4] For each specification, the average of the k MSEs is stored and the best specification is selected by minimizing it. This simple CV procedure is the widely acknowledged criterion that we propose to select the best specification among a number of possible alternatives: model (1), (2) and all the alternative specifications obtained interacting only a subset of circumstances.

[4]Cross-validation compared with AIC, BIC and adjusted R^2 provides a direct estimate of the error. Overfitted models will have high R^2 values, but will perform poorly in predicting out-of-sample cases. CV is also useful to choose among alternative non linear specifications together with non nested models.

4 An empirical illustration

We use the EUSILC 2011 dataset on 26 countries: Austria (AT), Belgium (BE), Bulgaria (BG), Czech Republic (CZ), Denmark (DK), Estonia (EE), Germany (DE), Great Britain (UK), Greece (GR), Finland (FI), France (FR), Hungary (HU), Italy (IT), Latvia (LV), Lithuania (LT), Luxemburg (LU), the Netherlands (NL), Norway (NO), Poland (PL), Portugal (PT), Romania (RO), Spain (ES), Slovakia (SK), Slovenia (SI), Sweden (SE) and Switzerland (CH). Following Checchi et al. (2016), we restrict the sample to individuals aged between 30 and 60 who are either working full or part-time, unemployed or fulfilling domestic tasks and care responsibilities. Our outcome variable is disposable income. We consider gender, country of origin and family background as circumstances affecting individual incomes irrespective of individual responsibility. All variables are categorical and we consider them in both a parsimonious and a broad categories' partition. In the most parsimonious case, four binary variables are included in the regressions: gender (male/female), country of origin (native/foreign), parental education (low/ high) and parental occupation (elementary/not elementary). In the broadest partition, we consider the following division: gender (male/female), country of origin (native/ EU foreign/ non EU foreign), mother and father occupation (coded in 10 values each)[5] and parental education (coded in 4 values each)[6]. Table 1 shows the descriptive statistics of the most parsimonious case while Figure 1 shows the Gini IOp measures of three cases: (i) the linear most parsimonious case (*low*), widely adopted by the literature, with no interactions and categories coded as binary variables; (ii) the fully interacted model where the categories are defined in the widest case and are fully interacted (*up*); (iii) an intermediate measure computed considering the best model selected by the CV method (*best*).

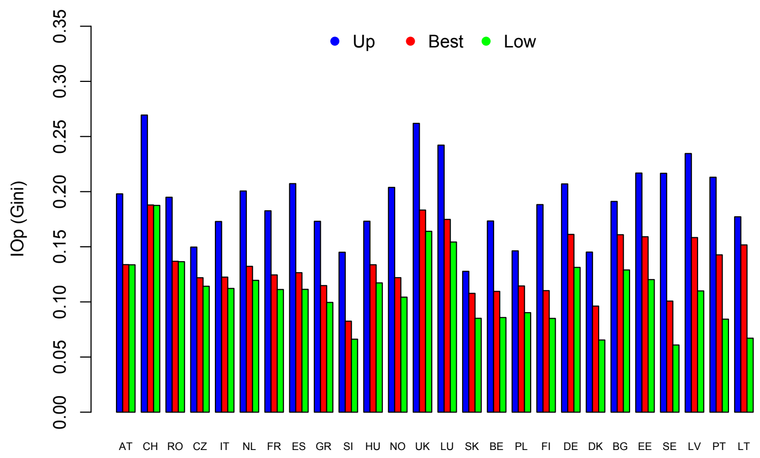
The three alternative measures clearly differ among each other, in some cases (left) the best model and the linear model coincide. In other cases (right) the best model is far from the linear specification and close to the most flexible one. An immediate implication is that the position achieved in the countries' ranking clearly depends on the model specification chosen by the researcher. If we compare our *best* measure with alternative estimations obtained in other studies which use the same EUSILC data like Brzezinski (2015) and Checchi et al (2016), we notice that the final assessment might largely differ, see Figure 2.[7]

[5]ISCO-08: Armed forces occupations; Managers; Professionals; Technicians and associate professionals; Clerical support workers; Service and sales workers; Skilled agricultural, forestry and fish; Craft and related trades workers; Plant and machine operators; Elementary occupations.

[6]Could neither read nor write in; Low level (pre-primary, primary education); Medium level (upper secondary education); High level (first stage of tertiary education).

[7]Both Brzezinski (2015) and Checchi et al (2016) consider a linear specification (with no interaction term). Brzezinski (2015) estimates a reduced-form OLS regressions of equivalized personal income on circumstances whereas Checchi et al (2016) estimates an heckman selection model of disposal individual income on circum-

Figure 1: IOp in 26 European countries under different model specifications



Source: EU-SILC, 2011 Note: The Figure shows each country's IOp measure with three alternative methods: (i) the linear most parsimonious case (low), (ii) the fully interacted model (up); (iii) the best model selected (best).

This evidence suggests that it is extremely important to introduce a widely accepted statistical criteria to select the best fitted model among many possible specifications.

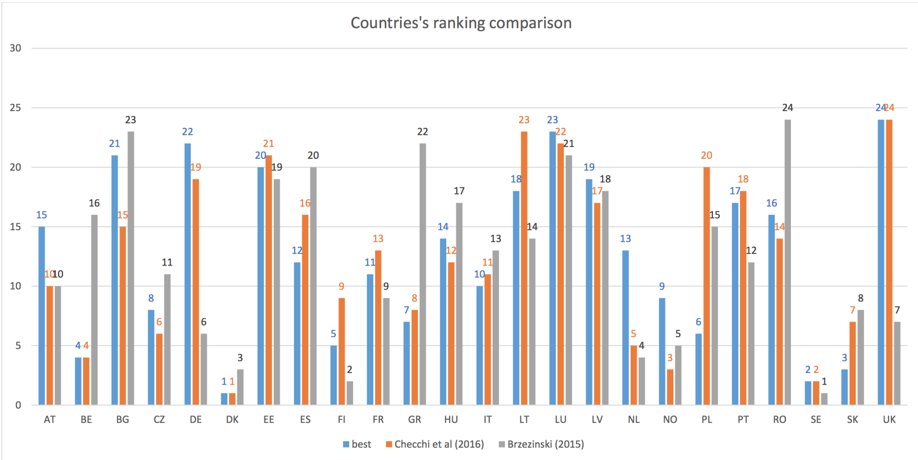
5 Conclusions

Scholars are well aware that the IOp estimates are mostly downward biased. The bias is a consequence of the partial observability of circumstances beyond individual control that do affect individual outcome. However, since IOp is measured as inequality in a counterfactual distribution of unfair inequality, a second possible source of bias might be related to the sample variance of the estimated counterfactual distribution. In this paper, we discuss this further source of bias, which has been surprisingly neglected by the empirical literature on IOp measure, and show that it implies an upward bias. We stress that the empirical specification used to estimate IOp, might largely influence its measure and suggest that, when choosing among alternative models, scholars should opt for the best balance between the two sources of bias.

We interpret this problem as a typical variance-bias trade-off and propose to adopt a simple

stances.

Figure 2: IOp in 26 European countries under different model specifications



Source: EU-SILC, 2011 Note: Note: The Figure shows each country's raking position derived by three alternative IOp measures: our best selected model, Checchi et al (2016) and Brzezinski (2015).

CV method to find the best solution. Finally, we show the empirical relevance of our intuition and implement the proposed method to measure IOp in 26 European countries. The proposed exercise clarifies that the choice of the model specification affects the estimated IOp and shows the importance to have a widely accepted criterion to identify the best possible specification.

References

- Balcazar, C. (2015), "Lower bounds on inequality of opportunity and measurement error", *Economics Letters*, 137: 102-105.

- Biörklund, A., Jäntti A. and Roemer J. (2012), "Equality of Opportunity and the Distribution of Long-Run Income in Sweden", *Social Choice and Welfare*, 39: 675-696.

- Bourguignon, Ferreira F., and Menendez M., (2007), "Inequality of Opportunity in Brazil", *Review of Income Wealth*, 53: 585-618.

- Bourguignon, Ferreira F., and Menendez M., (2013), "Inequality of Opportunity in Brazil: A Corrigendum", *Review of Income Wealth*, 59: 551-555.

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

- Brunori P., Palmisano, F., Peragine, V. (2015), "Inequality of opportunity in Sub Saharan Africa", World Bank Policy Research Working Papers 7782.

- Brunori, P., Ferreira F., Peragine, V. (2013), "Inequality of opportunity, income inequality and mobility: some international comparisons", in *Getting Development Right: Structural Transformation, Inclusion and Sustainability in the Post-Crisis Era* edited by E. Paus. Palgrave Macmillan.

- Checchi, D., and Peragine V. (2010), "Inequality of Opportunity in Italy", *Journal of Economic Inequality*, 8: 429-450.

- Checchi, D., Peragine V., Serlenga L., (2016), "Inequality of Opportunity in Europe: Is There a Role for Institutions?" , in Cappellari L., Polachek S., Tatsiramos K. (ed.) *Inequality: Causes and Consequences* , Research in Labor Economics, Volume 43, Emerald.

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

- Ferreira F., and V. Peragine. (2015). "Equality of Opportunity: Theory and Evidence", in *Oxford Handbook of Well-Being and Public Policy* edited by M. Adler and M. Fleurbaey. Oxford University Press.

- Kanbur, R., Wagstaff, A. (2015), "How useful is inequality of opportunity as a policy construct?", *CEPR Discussion Paper* DP10508.

- Ibarra L., Martinez C., Adan L. (2015), "Exploring the sources of downward bias in measuring inequality of opportunity." Policy Research working paper; no. WPS 7458. Washington.

- Luongo P. (2011),"The Implication of Partial Observability of Circumstances on the Measurement of Inequality of Opportunity", in Rodriguez, J. eds. , *Research on Economic Inequality*, 19: 23-49.

- Niehues, J., and Peichl A. (2014), "Upper bounds of inequality of opportunity: theory and evidence for Germany and the US", *Social Choice & Welfare*, 43: 63-79.

- Ramos, X., Van de gaer, D. (2015), "Measurement of Inequality of Opportunity Based on Counterfactuals", *Paper presented at the 6th ECINEQ meeting.* University of Luxembourg, 2015.

- Roemer, J. (1998), "Equality of Opportunity", Cambridge, MA: Harvard University Press.

- Roemer J., and Trannoy A ., (2015), "Equality of Opportunity", In *Handbook of Income Distribution* Edited by Anthony B. Atkinson and Francois Bourguignon. vol. 2. Elsevier.

Appendix

A large number of observable circumstances implies upward biased IOp estimate

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

A finer partition of the population and, therefore, smaller sample size leads to a larger distortion of the sample mean. Also, considering the variance bias-trade off, when estimating a group mean we get higher sample variance the smaller is the sample size of the group. Hence, in the case of between group inequality, we expect the distortion of the counterfactual distribution to increase with the number of groups in which we have partitioned the population especially if the variance in the groups is high. An implication of this might be that the more we exploit information contained in the data, the more we will upward bias our between-group inequality measure. More in details, if μ_t is the type mean when the number of observations within types is small, we expect a biased estimates of sample mean, such that $\tilde{\mu}_t = \mu_t + \eta$ where $\tilde{\mu}_t$ is the estimated type mean, μ_t is the "true" parameter and η is the standard error of $\tilde{\mu}_t$, i.e. $\frac{\sigma}{\sqrt{N_t}}$. Simulations prove that the error component leads to a positive distortion and by construction converges to zero as $N_t \to \infty$. Following Chakravarty and Eichhron (1994) we can easily prove that between inequality derived by a larger partition of the population is an overestimation of that derived by smaller (and more representative) ones.

Assuming that U is strictly concave by Jensen's inequality, we have

$$E\left(U\left(\tilde{\mu}_t|\mu_t\right)\right) < U\left(E\left(\tilde{\mu}_t|\mu_t\right)\right)$$

given that

$$\begin{aligned} \tilde{\mu}_t &= \mu_t + \eta \qquad (3) \\ &= E\left(\tilde{\mu}_t|\mu_t\right) + \overbrace{\left(\tilde{\mu}_t - E\left(\tilde{\mu}_t|\mu_t\right)\right)}^{\eta} \\ \tilde{\mu}_t - \eta &= E\left(\tilde{\mu}_t|\mu_t\right) \text{ from (3) } \tilde{\mu}_t - \eta = \mu_t \end{aligned}$$

and

$$U\left(E\left(\tilde{\mu}_t|\mu_t\right)\right) = U\left(\mu_t\right) \tag{4}$$

Then

$$E\left(U\left(\tilde{\mu}_t|\mu_t\right)\right) < U\left(\mu_t\right)$$

Taking expectation of both sides of (4) with respect μ_t, we get

$$E\left(U\left(\tilde{\mu}_t\right)\right) < U\left(E\left(\mu_t\right)\right) \tag{5}$$

Given that $\tilde{\mu}_t$ and μ_t asymptotically - as $N_t \to \infty$ - have the same mean and U is strictly concave.

Therefore, given the circumstances observed, IOp estimates are an upward biased estimate of the real between-type inequality. The bias is monotonically increasing with the number of observed circumstances and is monotonically decreasing with the sample size.

Table 1: Descriptive statistics

	AT	BE	BG	CH	CZ	DE	DK	EE	GR	ES	FI	FR	HU
obs	6097	5892	6989	7433	8538	12342	5781	5233	5990	15174	9550	10859	13067
median disposable income	19946	20700	2454	43648	7214	17608	33506	5113	12902	13800	27613	19658	4322
female	0.53	0.52	0.51	0.54	0.52	0.54	0.53	0.52	0.52	0.52	0.5	0.52	0.54
age (years)	45.36	45.03	46.19	45.73	45.56	46.5	46.66	45.77	45.26	45.07	46.52	45.7	46
foreign	0.16	0.17	0.01	0.25	0.03	0.06	0.07	0.13	0.10	0.10	0.04	0.10	0.01
parental occupation: elementary	0.15	0.09	0.21	0.09	0.13	0.06	0.05	0.17	0.09	0.19	0.08	0.27	0.26
parental education: low	0.37	0.51	0.51	0.24	0.60	0.11	0.10	0.34	0.78	0.83	0.49	0.76	0.61
	IT	LT	LU	LV	NL	NO	PL	PT	RO	SE	SI	SK	UK
obs	3648	20652	5296	6632	4654	11179	4927	15238	5755	7699	6469	12926	6712
median disposable income	17972	17944	3791	31163	20946	31864	45198	4844	7899	2092	24082	11102	14876
female	0.51	0.52	0.54	0.52	0.52	0.53	0.51	0.52	0.53	0.52	0.52	0.51	0.53
age (years)	45.67	45.32	47.58	44.57	46.52	46.22	45.5	45.96	46.03	46.18	45.67	45.58	45.62
foreign	0.09	0.07	0.52	0.13	0.06	0.09	0.01	0.08	0.01	0.14	0.12	0.01	0.12
parental occupation: elementary	0.15	0.37	0.13	0.27	0.04	0.06	0.16	0.19	0.13	0.01	0.11	0.26	0.16
parental education: low	0.76	0.58	0.51	0.41	0.38	0.23	0.49	0.93	0.85	0.34	0.67	0.35	0.55

Source: EUSILC (2011)