Economics Letters 137 (2015) 102–105

Contents lists available at ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Lower bounds on inequality of opportunity and measurement error

Carlos Felipe Balcázar *

The World Bank, 1818 H St. NW, Washington DC, United States

HIGHLIGHTS

- Lower bound estimates of inequality of opportunity can have substantial measurement error.
- Lower bound estimates of inequality of opportunity are not likely to be comparable across samples.
- Cross-country comparisons of lower bound estimates of inequality of opportunity can be misleading.

ARTICLE INFO

Article history:
Received 30 March 2015
Received in revised form
17 October 2015
Accepted 20 October 2015
Available online 24 October 2015

JEL classification:
D63
D39
O50

Keywords:
Inequality of opportunity
Lower bounds
Measurement error

ABSTRACT

I show that lower bound estimates of inequality of opportunity can have substantial measurement error, and that measurement error can vary considerably across samples. As a consequence, the traditional cross-country comparisons researchers make can be misleading.

1. Introduction

A fundamental concept in the literature on Inequality of Opportunity (IOp) is that inequalities that derive from circumstances beyond the control of individuals are morally objectionable; society should hold individuals responsible only for the level of effort they exert in comparison to the level that others exert (Roemer, 1998). Therefore, to measure IOp is to measure the amount of inequality that comes from circumstances. However, in practice this is hard to do: First, it is virtually impossible to account for all characteristics that constitute individuals' circumstances. Second, the level of effort is usually unobservable. Here I focus on exploring the former problem.

Barros et al. (2010), Ferreira and Gignoux (2011) and Luongo (2011) show that estimates of IOp based on an incomplete list of circumstances are lower bound estimates of true IOp, even if we make working assumptions about the distribution of an underlying effort variable. The problem is that a lower bound estimate only picks up part of the observed inequality, hence any alternative number which exceeds it would be equally valid (Kanbur and Wagstaff, 2015). This has two implications: First, any lower bound estimate of IOp could increase by adding one non-trivial circumstance; that is, lower bounds can have measurement error. Second, different samples could have the same (different) true value of IOp but have different (the same) estimated values conditional on a set of observable circumstances. Thus, lower bound estimates on different samples may have different levels of measurement error, rendering them incomparable. As a consequence, the traditional cross-country comparisons we make can be misleading.

Using lower bound estimates may lead researchers to infer the share of the variation not attributable to observable circumstances to the dimension of responsibility. But even if we observe all circumstances we may still misestimate IOp. Luck may seem as a fair source of inequality if it is even-handed, but that is not always the case. There are exogenous random factors that affect individuals' outcomes that we should treat as circumstances (e.g., innate talent). However, in practice, we cannot separate effort-related luck from that owing to exogenous factors (Ramos and van de Gaer, 2015). As a result, we may evaluate the opportunity sets as a

* Tel.: +1 202 4581497.
E-mail address: cbalcazarsalazar@worldbank.org.

http://dx.doi.org/10.1016/j.econlet.2015.10.026
0165-1765/

muddle among differences in circumstances and efforts (Pignataro, 2012).

Recently, Niehues and Peichl (2014) attempt to address the problem of partial observability of circumstances by using panel data for Germany and the United States. They use a fixed-effects regression to estimate the amount of inequality that comes from time-invariant circumstances showing that "lower bound estimates substantially underestimate IOp [...]". Moreover, they show that by including luck IOp estimates are higher. However, it is important to note that they do not (cannot) separate effort-related luck from that attributable to circumstances.

In this paper I contribute with further evidence on the magnitude of measurement error derived from using lower bounds. Nonetheless, I focus on exploring an outcome for which effort plays no role: *height-for-age* for children 0–2 years old. This outcome allows me to assume that luck only derives from exogenous factors; hence, it permits me to identify the size of measurement error coming from the empirical impossibility of including all circumstances in the analysis—which by construction is attributable to unobservable circumstances and luck. Also, I show that measurement error varies across countries; this implies that making the traditional cross-country comparisons using lower bound estimates can be misleading.

2. Analytical framework

On the basis of Pignataro (2012), consider a set of N individuals where individual outcomes (y) depend only on circumstances (C) and luck (l); there is no effort. C belongs to a finite set $\Omega = \{C_1, \dots, C_t, \dots, C_T\}$ where $t \in \{1, \dots, T\}$ denotes the type; hence, individuals of each type t have identical circumstances. $l \in \Theta$ is an exogenous scalar random component. And $y = f(C, l)$ where $f : \Omega \times \Theta \to \Re_+$.

Let us assume we observe, only partially, the set of circumstances: $\Omega' = \{C_1, \dots, C_k, \dots, C_K\}$, for each observable type k, where $k \in \{1, \dots, K\}$ and $K < T$. And consider two outcome distributions: First, a *smoothed* distribution, which we obtain by replacing each individual outcome y_i with the outcome mean of the observable type to which he/she belongs, μ_k. I denote this distribution Y_B because it accounts for between-type variation. Second, a *standardized* distribution, which we obtain by replacing each individual outcome y_i with $\frac{\mu}{\mu_k} y_i$, where μ denotes the outcome mean. I denote this distribution Y_W because it accounts for within-type variation.[1]

Traditionally, measuring IOp consists of determining the extent to which the distribution of outcomes Y differs from Y_B. Thus, following Bourguignon et al. (2007) we can compute IOp by comparing Y to Y_B through a scalar measure $I : \{Y\} \to \mathbb{R}_+$ such that

$$\Theta = 1 - \frac{I(Y_B)}{I(Y)}.$$

Nonetheless, to estimate Θ using Y_B or Y_W – so that $I(Y_B) = I(Y) - I(Y_W)$ – leads to different evaluations for almost all inequality indexes.

Foster and Shneyerov (2000) introduce the *path-independent decomposable* class of additive inequality measures for which the two alternatives yield the same results, which reduce to the mean-log-deviation (MLD).[2] Hence, using the MLD

$$I(Y) = I(Y_W) + I(Y_B).$$

Note that by using Ω', the partition of the population we obtain is coarser than the one we would obtain from using Ω. This causes an underestimation of the level of inequality. This is easy to see since if $I(Y_W) > 0$, then $I(Y_B) < I(Y)$. Thus, $I(Y_B)$ represents a *lower bound* estimate of IOp.

3. An empirical exercise

The data comes from 27 Demographic Household Surveys (DHS) circa 2008. I focus on kids 0–2 years of age because we can confidently assume that they have not exerted efforts.[3] Also, the advantage I explore is height-for-age, which for example has been related to higher cognitive test scores and higher labor earnings (Case and Paxson, 2008).

Although the DHS provides information on a broad number of circumstances, some of these contain many missing values, compromising the representativeness of the samples. Therefore, I include variables that do not compromise the representativeness of the data: age of the toddler (in months), gender (dummy), birth order (numerical), age of the mother (in years), height of the mother (in centimeters), educational attainment of the mother (in years), the DHS' index of wealth and location (urban or rural).[4]

Now, since toddlers' height increases in variance with age and varies by sex, I use the World Health Organization's growth charts to create a standardized measure for height (y), which corresponds to the equivalent height the child would have had if (s)he were a 24 months old female (Pradhan et al., 2003). More formally

$$y = F^{-1}_{\bar{a},\bar{g}}(F_{a,g}(h)),$$

where F is the distribution function of heights in the reference population for the age and sex group of an individual of age a and gender g; h is the actual height of that individual; $\bar{a} = 24$ months; and $\bar{g} = female$.

3.1. Results

Given that y follows a normal distribution (WHO, 2006), I estimate Y_B for each country by means of a linear regression[5]:

$$y_i = \alpha + \beta C_i^o + \varepsilon_i,$$

where y_i is the observed standardized height for individual i; C_i^o is the vector of observable circumstances; and ε_i is an idiosyncratic error. Note that the distribution of predicted values of this regression ($\widehat{Y_B}$), corresponds to the distribution we would obtain by replacing individuals' outcomes with the average outcome of their respective type. Hence $MLD(\widehat{Y_B})$ corresponds to a lower bound estimate of IOp.

Provided that the MLD is path independent, we can compute

$$IR(\widehat{Y_W}) = \left[1 - \frac{MLD(\widehat{Y_B})}{MLD(Y)}\right] \times 100,$$

with $IR(\widehat{Y_W}) \in [0, 100]$. $IR(\widehat{Y_W})$ provides us with a comparable cross-country estimate of the extent of measurement error that arises from the impossibility of including all observable circumstances.

[1] Note that Y_W considers those inequalities caused by unobservable circumstances and (exogenous) luck.

[2] An alternative to the MLD is the Atkinson inequality index, it satisfies multiplicative decomposability and path-independence (Donni et al., 2014). However, I do not use it due to space restrictions.

[3] Toddlers start relating their own actions to their surrounding environment after 24 months of age (Rochat, 1998).

[4] My full sample consists of 67,985 children. After dropping biological implausible values for height following the WHO (2006) guidelines, and dropping missing values in my variables of interest, I lose 6% of my full sample. Information on sample sizes at https://sites.google.com/site/cfbalcazars/misc.

[5] Using a non-parametric approach would lead to biased estimations of type-means due to small cell-sizes.