Lower and Upper Bound Estimates of Inequality of Opportunity for Emerging Economies*

Paul Hufe, Andreas Peichl & Daniel Weishaar

August 16, 2021

Abstract

Equality of opportunity is an important normative ideal of distributive justice. In spite of its wide acceptance and economic relevance, standard estimation approaches suffer from data limitations that can lead to both downward and upward biased estimates of inequality of opportunity. These shortcomings may be particularly pronounced for emerging economies in which comprehensive household survey data of sufficient sample size is often unavailable. In this paper, we assess the extent of upward and downward bias in inequality of opportunity estimates for a set of twelve emerging economies. Our findings suggest strongly downward biased estimates of inequality of opportunity in these countries. To the contrary, there is little scope for upward bias. By bounding inequality of opportunity from above, we address recent critiques that worry about the prevalence of downward biased estimates and the ensuing possibility to downplay the normative significance of inequality.

JEL-Codes: D31; D63; I32
Keywords: Equality of Opportunity; Inequality; Emerging Economies

*__Hufe__: ifo Munich, LMU Munich, IZA (paul.hufe@econ.lmu.de); **Peichl** (corresponding author): ifo Munich, LMU Munich, IZA and CESifo (peichl@econ.lmu.de), Postal Address: ifo Munich, Poschingerstraße 5, 81679 Munich, Germany; **Weishaar**: LMU Munich (daniel.weishaar@econ.lmu.de). We thank John E. Roemer (editor) and two anonymous referees for many insightful comments on earlier drafts. This paper benefited from discussions with Paolo Brunori, Daniel Mahler, and Valentin Lang. Furthermore, we are grateful to seminar and conference audiences in Canazei, Luxembourg and Munich. We gratefully acknowledge funding from Deutsche Forschungsgemeinschaft (DFG) through NORFACE project "IMCHILD: The impact of childhood circumstances on individual outcomes over the life-course" (PE 1675/5-1).

1 Introduction

Equality of Opportunity (EOp) is an ideal of distributive justice that garners wide-spread public support and is plausibly related to macro-economic indicators of development (Aiyar and Ebeke, 2019; Alesina et al., 2018; Cappelen et al., 2007; Ferreira et al., 2018; Marrero and Rodríguez, 2013). However, limitations in the underlying data sources lead to both upward and downward biased estimates of inequality of opportunity (IOp). Both biases are potentially large in emerging countries where the data quality is arguably worse than in industrialized economies. However, it is not clear ex ante which of the two biases prevails and whether IOp estimates rather tend to be downward or upward biased. In this paper, we address this uncertainty by constructing lower bound (LB) and upper bound (UB) estimates of IOp for twelve emerging economies and compare them to estimates from the conventional approach.

EOp distinguishes ethically justifiable (fair) inequalities from unjustifiable (unfair) inequalities using the concepts of circumstances and effort.[1] Circumstances are defined as all factors that are not under the control of the individual—for instance, the biological sex, the parental background and the birthplace. To the contrary, working hours and educational decisions are under the (partial) control of individuals and are therefore characterized as efforts. Opportunity egalitarians consider inequalities based on exogenous circumstances as unfair, while inequalities resulting from effort exertion are deemed fair sources of inequality (among others Arneson, 1989; Cohen, 1989).

This distinction is not only relevant from a normative perspective but provides important insights for the patterns and drivers of economic development (Ferreira et al., 2018; Marrero and Rodríguez, 2013; Neidhöfer et al., 2018; Peragine et al., 2014). For instance, a leveled playing field fosters human capital accumulation by providing incentives for skill acquisition (Mejía and St-Pierre, 2008). Furthermore, circumstance-based variation in life outcomes reflects horizontal inequality and segregation, both of which are important drivers of social tensions and conflict (Rohner, 2011).

What we call the "standard approach" (S) towards IOp estimation in this paper, constructs a counterfactual distribution of life outcomes from a linear prediction using all circumstance information observable by the econometrician. In line with the opportunity-egalitarian doctrine, inequality in this counterfactual distribution is considered "unfair" since it only varies with immutable circumstance characteristics. Due to limitations in the underlying data sources, this conventional method can lead to both upward and downward biased empirical measurements of IOp. First, due to the partial observability of circumstances, standard IOp estimates tend to be *downward* biased (Balcázar, 2015; Hufe et al., 2017). The downward bias may be particularly pronounced in countries that lack household surveys combining information on the outcome of interest with rich information on individual characteristics. Most emerging economies fall

[1] Among others, this dichotomy is formalized in Fleurbaey (1995) and Roemer (1998).

into this category. Second, if the ratio between the number of parameters to be estimated and the available degrees of freedom is large, the ensuing noise in the parameter estimates will artificially inflate the measured impact of observed circumstances on individual life outcomes (Brunori et al., 2019b). Emerging economies may again be particularly susceptible to such *upward* bias in standard IOp estimates since the sample sizes of available household surveys tend to be comparatively small. Ex ante it is unclear which of the two biases prevails for the group of emerging economies. As a consequence, policy makers that rely on standard estimates may over- or underestimate the true degree of IOp and enact policy measures without considering the uncertainty around such estimates (Kanbur and Wagstaff, 2016).

In this paper, we address the uncertainty around empirical IOp estimates by drawing on longitudinal household surveys from twelve emerging economies which enable us to estimate both LB and UB measures of IOp. First, we calculate LB measures of IOp by estimating the impact of *observable circumstances* on incomes with a cross-validated lasso procedure. Assessing statistical models by out-of-sample cross-validation disciplines the process of model selection and therefore prevents overfitting the circumstance parameters to the estimation sample. As a consequence, the relevant circumstance parameters are estimated with less noise which in turn cushions upward biases in IOp measures.

Second, we leverage the panel dimension of the data to calculate UB estimates based on the individual fixed effect (FE) estimator proposed in Niehues and Peichl (2014). By their most common definition, circumstance characteristics are time-constant but partly unobservable by the econometrician. Individual FEs capture the full set of *unobservable circumstances* and therefore yield the maximum amount of outcome variation that can be explained by circumstances. However, individual FEs also capture time-constant effort variables and therefore may overstate the extent of unequal opportunities. Hence, they yield an upper bound of the true IOp estimate.

Our results can be summarized as follows. In emerging economies the standard approach of estimating inequality of opportunity produces results that closely align with the lower bound. In theory, the restricted data infrastructures of many emerging economies could lead to either upward biased (small sample sizes) or downward biased (little circumstance information) estimates. In practice, the latter concern clearly dominates the former in our sample. With respect to individual (equivalized household) incomes, the average difference between the standard estimate and the lower bound estimate is 5.7 (5.0) percentage points (pp). To the contrary, the average distance between the standard estimate and the upper bound estimate is 22.8pp (28.5pp).

These results from emerging economies contrast recent evidence for European countries. For example, Brunori et al. (2018) show for a set of European countries that standard estimates may be upward biased by up to 300%. This contrast emphasizes that the particularities of data environments are crucial for an assessment of the relative importance of upward and downward

biases. Second, the large distance between the standard estimate and the upper bound estimate in emerging economies emphasizes the concern of providing misleading reference points to policymakers who could use downward-biased estimates of IOp to downplay the moral significance of inequality (Kanbur and Wagstaff, 2016). In the absence of data innovations, providing reasonable bounds on inequality of opportunity may be the only way to address such concerns. Our paper is the first to conduct such a bounding exercise for a set of emerging economies with broad geographical coverage and thereby contributes to the growing literature on EOp in these countries.[2]

The remainder of this paper is organized as follows. In section 2 we formalize the EOp concept and outline the corresponding estimation strategies for its LB and UB measures. After introducing the data sources in section 3, we present results and robustness analyses for both LB and UB estimates in section 4. Section 5 concludes the paper.

2 Conceptual Framework

Important life outcomes such as income and consumption are determined by an extensive vector of personal characteristics that can be subsumed by a binary classification into circumstances and efforts. Those characteristics that are completely beyond the realm of individual control are called circumstances. To the contrary, those characteristics that are at least partially controlled by individuals are called efforts. The more the distribution of outcomes depends on circumstances, the stronger the violation of the opportunity-egalitarian ideal and the higher the measure of inequality of opportunity.

Consider a finite population indexed by $i \in \{1, ..., N\}$.[3] Each individual is characterized by the tuple $\{y_{it}, \mathbf{C_i}, \mathbf{E_{it}}\}$. y_{it} constitutes the period-specific outcome of interest, $\mathbf{C_i}$ the vector of time-invariant circumstances, and $\mathbf{E_{it}}$ period-specific effort. Life outcomes are a function of circumstances and efforts:

$$y_{it} = f(\mathbf{C_i}, \mathbf{E_{it}}(\mathbf{C_i})).^{4} \tag{1}$$

[2]See Alesina et al. (2019) and Brunori et al. (2019a) for work on Africa, Ferreira and Gignoux (2011) for work on Latin America, as well as Andreoli and Fusco (2019) and Brock et al. (2016) for comparative work including Eastern Europe and Central Asia.

[3]We follow the notational conventions established in Ferreira and Gignoux (2011).

[4]Note that the current literature largely abstracts from time-variant circumstance characteristics. This abstraction can be rationalized by the blurry distinction between time-variant factors beyond individual control and individual efforts. For example, consider local economic shocks or local outburst of conflict as potential embodiments of time-variant circumstances. Their effect could be confounded by individual migration decisions which are at least partially under individual control. However, as we outline below, our normative framework accounts for the effect of such factors to the extent that they are correlated with time-constant factors such as the region of birth.

Note that we allow circumstances to have a direct and an indirect impact on the outcome of interest. For example, certain groups may be excluded from offices and positions based on outright discrimination (direct impact). However, such discrimination may also lead to adjustments in individual effort exertion since the imposed circumstance constraints alter the individual optimization calculus (indirect impact). Whether the correlation between circumstances and efforts contributes to the fair or the unfair part of inequality is widely debated (Jusot et al., 2013). In this paper we follow Roemer (1998) who proposes that outcome differences due to a correlation between circumstances and effort constitute a violation of EOp.[5]

The literature on EOp further distinguishes the ex-ante from the ex-post approach (Ramos and Van de gaer, 2016). While the ex-ante approach requires that there are no differences in life outcomes across circumstance types, the ex-post approach demands that individuals exerting the same effort enjoy the same level of advantage. In this paper we focus on the ex-ante approach. That is, we use $\mathbf{C_i}$ to construct a partition of disjunct types $\Pi = \{T_1, ..., T_P\}$ such that all members of a type are homogeneous in circumstances. The average outcome of type k is denoted by μ_t^k. EOp is achieved if type-means in period t are equalized across types, i.e. if $\mu_t^k = \mu_t^l \ \forall \ l, k \mid T_k, T_l \in \Pi$.

Computing inequality in a counterfactual distribution $M_t = (\mu_{1t}^1, ..., \mu_{it}^k, ..., \mu_{Nt}^P)$, in which each individual i of type k is assigned its corresponding type outcome μ_t^k yields a scalar measure of IOp. It decreases with Pigou-Dalton transfers between circumstance types but is invariant to such transfers within circumstance types. Inequality in the counterfactual distribution of type-means can thus be considered unfair as it only depends on disparities due to immutable circumstance characteristics.

Standard Estimation (S). The standard approach towards IOp measurement (Bourguignon et al., 2007; Ferreira and Gignoux, 2011) constructs an estimate for the counterfactual distribution of type means in a two-step procedure. First, for the year of interest t we estimate:

$$\ln y_{it} = \alpha + \boldsymbol{\beta} * \mathbf{C_i} + \epsilon_{it}. \tag{2}$$

Note that this specification accounts for both the direct and the indirect effect of circumstances since the correlation between $\mathbf{C_i}$ and $\mathbf{E_{it}}$ is implicitly captured by $\boldsymbol{\beta}$. Second, we use the vector of estimated parameters $\hat{\boldsymbol{\beta}}$ to parametrically construct an estimate for the distribution of type

[5]This normative assumption is adopted by much of the empirical literature on IOp but can be easily relaxed, see Jusot et al. (2013) and Niehues and Peichl (2014). We refrain from doing so in our empirical application since restricting samples on availability of effort information would further reduce the number of observations.

means $\tilde{M}_t^S = (\tilde{\mu}_{1t}^S, ..., \tilde{\mu}_{it}^S, ..., \tilde{\mu}_{Nt}^S)$:

$$\tilde{\mu}_{it}^S = \exp\left\{\hat{\alpha} + \hat{\boldsymbol{\beta}} * \mathbf{C_i} + \frac{\sigma^2}{2}\right\}.^{6} \tag{3}$$

Lower Bound Estimation (LB). Conceptually, Ferreira and Gignoux (2011) show that the outlined standard estimate of IOp is a LB of its true value if the circumstance vector $\mathbf{C_i}$ contains only a subset of all relevant circumstances. Empirically, however, this lower bound measure may be upward biased due to sampling variance in the distribution of type means (Brunori et al., 2019b). With decreasing sample size and increasing size of the circumstance set, the available degrees of freedom to estimate $\boldsymbol{\beta}$ shrink. The ensuing noise in $\hat{\boldsymbol{\beta}}$ artificially inflates the variance in the distribution of estimated type means $\tilde{M}_t^S$, which in turn leads to upward biased lower bound measures of IOp.

The literature has proposed different methods to address the upward bias in IOp estimates. Using the European Union Survey on Income and Living Conditions (EU-SILC), Brunori et al. (2019b) select models by 5-fold cross validation. Thereby, the authors pre-specify a large variety of potential models which differ in circumstance characteristics and their interactions. After estimating these models on random folds of the data, the algorithm chooses the model which minimizes the average out of sample mean squared error.[7] An alternative approach to model selection are conditional inference trees and forests (Brunori et al., 2018). The regression tree method recursively splits the data according to the circumstance variables which have the strongest association with the outcome of interest while regression forest provide average estimates over multiple regression trees applied to random subsets of the data.

In this work we calculate lower bound estimates based on two different cross-validated lasso estimations that select the relevant circumstances to maximize the out-of-sample prediction accuracy of the model. Lasso estimations have two advantages in comparison to previous methods. First, one does not have to pre-specify the models to be evaluated by cross-validation—the preferred method in Brunori et al. (2019b). Second, they are less computationally expensive than random forests—the preferred method in Brunori et al. (2018). In Figure B.1, we use EU-SILC data to validate the lasso methodology against the findings of Brunori et al. (2019b) and Brunori et al. (2018). Both lasso estimates align very closely with the alternative estimation procedures. The implied Pearson correlation coefficients are 0.90/0.87 in comparison to the findings of Brunori et al. (2019b), and 0.91/0.89 in comparison to the findings of Brunori et al. (2018). All correlation coefficients are not statistically different from one at the 5% significance

[6] $\frac{\sigma^2}{2}$ represents the residual variance that corrects for differences in the marginal impact of circumstances due to the log-transformation (Blackburn, 2007).

[7] Intuitively, k-fold cross-validation works as follows. The sample is divided into k-folds. Under each specification, the model parameters are estimated on $k-1$ folds and the ensuing predictions are benchmarked against the data points in the kth fold. Repeating this procedure k times, one chooses the model that delivers the lowest average mean-squared prediction error across the k iterations.

level.

In both estimation approaches, we first estimate

$$\underset{\beta}{\operatorname{argmin}} \sum_i \underbrace{(\ln y_{it} - \alpha^{lasso} - \sum_j \beta_j^{lasso} * C_{ij})^2}_{(1)} + \underbrace{\sum_j \lambda |\beta_j^{lasso}|}_{(2)}. \tag{4}$$

Part (1) of equation 4 is a perfect mirror of the OLS algorithm used to estimate equation 2. Part (2) however introduces a penalization term that varies with the absolute value of the estimated coefficient $\hat{\beta}_j^{lasso}$. The larger (smaller) the penalization term λ, the more (less) parsimonious the model and the lower the variance (bias) in the predictions based on the parameter vector $\hat{\boldsymbol{\beta}}^{lasso}$. We choose the optimal parameterization of λ by means of 5-fold cross validation.[8]

The first lower bound estimate (LB1) uses the resulting vector $\hat{\boldsymbol{\beta}}^{lasso}$ to construct the counterfactual distribution $\tilde{M}_t^{LB1} = (\tilde{\mu}_{1t}^{LB1}, ..., \tilde{\mu}_{it}^{LB1}, ..., \tilde{\mu}_{Nt}^{LB1})$:

$$\tilde{\mu}_{it}^{LB1} = \exp\left\{\hat{\alpha}^{lasso} + \hat{\boldsymbol{\beta}}^{lasso} * \mathbf{C_i} + \frac{\sigma^2}{2}\right\}. \tag{5}$$

The second lower bound estimate (LB2) implements a post-OLS lasso estimation (Hastie et al., 2013). We only retain the subset $\mathbf{C^r} \subseteq \mathbf{C}$, i.e. those circumstances whose coefficients were not shrunk to zero in equation 4. Then, we estimate $\hat{\boldsymbol{\beta}}^{Post-lasso}$ by running an OLS regression on the restricted set of circumstances:

$$\ln y_{it} = \alpha^{Post-lasso} + \boldsymbol{\beta}^{Post-lasso} * \mathbf{C_i^r} + \epsilon_{it}. \tag{6}$$

We use $\hat{\boldsymbol{\beta}}^{Post-lasso}$ to construct the counterfactual distribution $\tilde{M}_t^{LB2} = (\tilde{\mu}_{1t}^{LB2}, ..., \tilde{\mu}_{it}^{LB2}, ..., \tilde{\mu}_{Nt}^{LB2})$:

$$\tilde{\mu}_{it}^{LB2} = \exp\left\{\hat{\alpha}^{Post-lasso} + \hat{\boldsymbol{\beta}}^{Post-lasso} * \mathbf{C_i^r} + \frac{\sigma^2}{2}\right\}. \tag{7}$$

Note that LB1 and LB2 are just different estimates of the same parameter vector. The choice between these two estimation methods is not straightforward. On the one hand, Belloni and Chernozhukov (2013) argue that the post-lasso may have a superior prediction accuracy than the standard lasso approach. On the other hand, the methodological validation based on EU-SILC reveals that the standard lasso approach tends to align more closely with the results in Brunori et al. (2018, 2019b) (Figure B.1). In our empirical application, we refer to standard lasso

[8] The general idea of cross-validation is explained in footnote 7. In the case of lasso estimations, its implementation is as follows: We re-estimate equation 4 for different values of λ on each of the five folds. Ultimately, we choose λ that on average minimizes the mean-squared prediction error across the five folds. The mean-squared prediction error is a standard measure of prediction accuracy (Hastie et al., 2013) and the appropriate target statistic to trade-off upward and downward bias in inequality of opportunity estimates (Brunori et al., 2019b). In Table A.3 we show the chosen values of λ for each country in our sample.

as our baseline LB estimate. However, we show that our main conclusions are insensitive to this choice.[9]

Upper Bound Estimation (UB). Since S and LB are based on the subset of observable circumstances only, the resulting IOp estimates may be downward biased. Following Niehues and Peichl (2014) we therefore construct UBs of IOp using an individual fixed effects (FE) estimator. Assuming circumstances to be time-invariant, individual FEs capture the full set of $\mathbf{C_i}$ even though not all circumstances are observable by the econometrician. A counterfactual distribution of type means constructed from individual FEs thus captures the upper ceiling of outcome variation that can be attributed to the impact of circumstances. In particular, the smoothed distribution of the UB is constructed as follows.

First, using observations from all periods $v \neq t$, we estimate the individual FE c_i while accounting for common year-specific shocks u_v:[10]

$$\ln y_{iv} = c_i + u_v + \epsilon_{iv}. \tag{8}$$

Second, we regress the individual outcome in period t on the estimated individual FE:

$$\ln y_{it} = \Psi * \hat{c}_i + \epsilon_{it}. \tag{9}$$

Third, we use the vector of parameters $\hat{\Psi}$ to construct the counterfactual distribution $\tilde{M}_t^{UB} = (\tilde{\mu}_{1t}^{UB}, ..., \tilde{\mu}_{it}^{UB}, ..., \tilde{\mu}_{Nt}^{UB})$:

$$\tilde{\mu}_{it}^{UB} = \exp\left\{\hat{\Psi} * \hat{c}_i + \frac{\sigma^2}{2}\right\}. \tag{10}$$

Note that this estimator would yield the true estimate of IOp if c_i captured time-invariant circumstances only. However, the individual FE may also absorb time-invariant effort exertion (e.g. long-term motivation, ambition) leading to an UB interpretation of this IOp estimate.

Inequality Measurement. We follow existing IOp literature and summarize the information in counterfactual distributions $\tilde{M}_t^S$, $\tilde{M}_t^{LB1}$, $\tilde{M}_t^{LB2}$, and $\tilde{M}_t^{UB}$ by the mean log deviation (MLD) and the Gini coefficient. The MLD is part of the generalized entropy class of inequality measures satisfying symmetry, the Pigou–Dalton transfer principle, scale invariance, population

[9] The post-lasso approach will yield results that are more in line with standard estimations based on OLS. This is the case since standard lasso retains parameter estimates that are shrunk by penalization. To the contrary—and analogous to OLS—post-lasso re-estimates these parameters without penalization.

[10] Accounting for year-specific shocks is necessary since the panel data used to estimate the fixed effect are unbalanced. In case of a balanced panel, the individual fixed effect would be completely orthogonal to the year-specific shock, i.e. one could abstract from u_v.

replication, as well as additive and path-independent subgroup decomposability (Foster and Shneyerov, 2000; Shorrocks, 1980). However, the MLD is very sensitive to low incomes many of which are smoothed out when constructing counterfactual distributions. Therefore, Brunori et al. (2019a) argue in favor of using the Gini index in spite of its imperfect subgroup decomposability.[11] For both inequality measures, we provide relative measures of IOp that relate the MLD (Gini) of the counterfactual distributions $\tilde{M}_t^S$, $\tilde{M}_t^{LB1}$, $\tilde{M}_t^{LB2}$ and $\tilde{M}_t^{UB}$ to the actual outcome distribution Y_t. The latter measures can be interpreted as the share of total inequality that is explained by circumstances and thus violates the opportunity-egalitarian ideal.

3 Data

We estimate IOp in income and consumption expenditure for twelve emerging economies in different geographical areas of the world ranging from Africa (Ethiopia, Malawi, South Africa, Tanzania), Central and South America (Argentina, Chile, Mexico, Peru), Europe and Central Asia (Russia), to East and South-East Asia (China, Indonesia, Thailand). The country selection is guided by the availability of household panel data with i) information on relevant circumstance variables, and ii) a sufficient number of observations in the longitudinal dimension.[12] Table A.1 provides an overview of the underlying data sources.

We consider three outcomes of interest. First, we calculate IOp in individual income—before or after taxes and transfers depending on data availability. Second, we account for resource sharing at the household level and calculate IOp in equivalized household income. Accounting for resource sharing at the household level is particularly relevant in emerging economies since female participation in formal labor markets tends to be low (Cubas, 2016). Third, to derive a more direct measure of IOp in material well-being, we also consider equivalized household consumption expenditures. Household income and consumption expenditure are deflated by the modified OECD equivalence scale.

Throughout the paper, we restrict ourselves to within-country comparisons. Table A.1 documents many differences across the underlying data sources. These include differences in the reference period, the income and consumption expenditure aggregates, the detail of available circumstance characteristics, as well as the sampled populations. For example, while the data for Mexico avails net income information until 2004, the data for Thailand provides gross income figures until 2016. The Ethiopian panel provides a rather parsimonious set of circumstances for a rural fraction of the population, whereas the Russian panel provides a rich set of circumstances for a nationally representative sample of households. We therefore refrain from

[11] Technically, the Gini coefficient nevertheless yields conservative IOp estimates as the residual in the Gini decomposition does contain elements of between-group inequality (Brunori et al., 2019a).

[12] For countries, in which multiple panel data sets are available, we use the data set with the highest number of waves.

cross-country comparisons but focus our discussion on intra-country comparisons between the different estimation approaches.

To ensure the consistency of these intra-country comparisons, we only retain those units of observation for which we observe (i) all circumstance variables, and (ii) positive outcomes in all available outcome dimensions for at least three periods of observation. We further restrict our samples to individuals aged 25-55.[13]

Table 1 displays relevant summary statistics for the estimation of S, LB, and UB by country.

4 Results

Figure 1 displays bounds of relative IOp, i.e. the percentage of total inequality that can be explained by exogenous circumstances.[14] Standard estimates (S) indicate IOp based on all observable circumstances available in the particular country data set. Lower bound estimates (LB) also use the full set of observable circumstances but account for potential upward biases through lasso estimation in which irrelevant circumstance parameters are shrunk to zero.[15] Upper bound estimates (UB) account for unobservable circumstances through the FE estimation procedure outlined in section 2.

Individual Income. Panel (a) shows the results for individual income. The standard IOp estimate (S) for individual income ranges from 9.3% (Argentina) to 30.6% (Peru, South Africa). Accounting for sampling variation and the ensuing potential for upward biases in S provides only minor reductions in IOp. According to LB, between 6% (China) and 25.9% (Peru) of outcome inequality must be considered unfair. The average difference between S and LB estimates amounts to 5.7pp.[16] When using the post-lasso OLS procedure, the average difference is even smaller and equals 0.5pp. These results suggest that the standard estimation approach (S) is largely uncompromised by overfitting circumstance parameters to the available data. Instead—and in line with the theoretical reasoning of Ferreira and Gignoux (2011)—the standard approach indeed recovers estimates close to the lower bound (LB) estimate in all countries under consideration. Note that this result stands in contrast to recent evidence for European countries suggesting that the standard approach overestimates lower bound IOp by up to 300% (Brunori et al., 2018, 2019b). This difference is reconciled by the quality of the underlying data

[13] In Appendix Table A.2 we show how samples change as we sequentially impose these data restrictions.

[14] Point estimates for absolute IOp, relative IOp, as well as total inequality are disclosed in Table A.3.

[15] As highlighted above: Unless otherwise indicated the LB estimate refers to the standard lasso estimation.

[16] This cross-country average conceals heterogeneity. In particular, the lower the sample size relative to the number of estimated circumstance parameters, the larger the difference between S and LB. See Table A.3 where we list the ratio of sample size and estimated parameters by country.

Table 1: Circumstance Information by Country

	Standard Estimate (S) and Lower Bound (LB)			Upper Bound (UB)				
	N	Circumstances	Parameter	N (FE)	Start	End	Min. Years	Avg. Years
Argentina	3,019	gender, year of birth, place of birth	8	6,038	2013	2014	3	3.00
Chile	2,808	gender, year of birth, place of birth, education of father/mother, ethnicity, labor force status of father/mother, chronic disease	37	8,424	2006	2008	4	4.00
China	243	gender, year of birth, ethnicity, urbanity of birthplace	22	717	1988	2010	3	3.95
Ethiopia	660	gender, year of birth, education of father/mother, ethnicity, religion	69	2,433	1994	2004	3	4.69
Indonesia	786	gender, year of birth, education of father/mother, ethnicity, religion, language	29	2,036	1992	2006	3	3.59
Malawi	362	gender, year of birth, education of father/mother, religion	16	995	2004	2008	3	3.67
Mexico	3,050	gender, year of birth, language	5	16,552	1999	2004	5	6.43
Peru	2,193	gender, year of birth, place of birth, language, chronic disease	37	5,878	1998	2011	3	3.68
Russia	1,181	gender, year of birth, place of birth, urbanity of birthplace, education of father/mother, labor force status of father/mother, height	54	10,816	1994	2016	5	10.16
South Africa	670	gender, year of birth, place of birth, education of father/mother, ethnicity	48	2,331	2008	2015	4	4.48
Tanzania	221	gender, year of birth, place of birth, ethnicity, religion	18	819	1991	2004	3	4.71
Thailand	465	gender, year of birth, education of father/mother, wealth of parents, family plot size	15	6,338	1997	2016	3	14.63

Notes: Column 2 displays the number of observations in the year of interest t. Column 3 lists country-specific circumstances used to estimate standard (S) and lower bound (LB) measures of inequality of opportunity. Column 4 shows the total number of parameters associated with country-specific circumstances. Columns 4-7 describe the longitudinal distribution of data points to estimate upper bound (UB) measures of inequality of opportunity. The displayed minimum and average number of years include the year of interest t plus years v that are used in the fixed effects estimation.
Source: Own calculations based on data described in Table A.1.

sources. While the richness of the European data confers the opportunity to overfit the circumstance information to the data, the sparsity of circumstance information in the household surveys under consideration prevents upward biases in the standard estimate (S).

The lower bound estimator selects the circumstance parameters with the highest out-of-sample prediction accuracy. In Table A.4, we show for each outcome of interest, which of the circumstance variables and categories are chosen by the lasso estimator in a particular country. Across all countries, gender plays a prominent role reflecting concerns about gender inequality in the context of emerging and developing economies (Jayachandran, 2015). However, it is important to note that the selection of particular variables by lasso only indicate a predictive correlation and does not necessarily imply a causal relationship. For instance, even though both maternal and paternal education could causally affect the income of individuals, a high correlation between fathers' and mothers' education might lead the lasso to choose only one of the two circumstance characteristics.

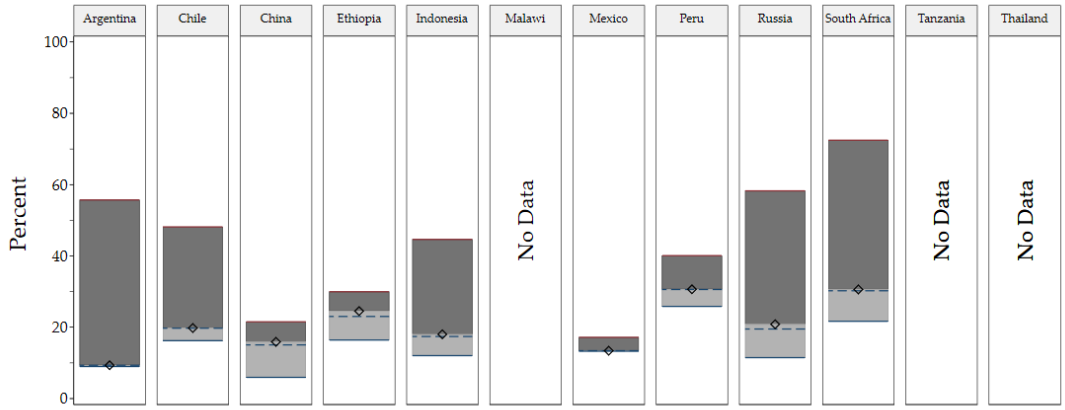
While sparse circumstance information limits the scope for upward biases, it may lead to downward biases due to the neglection of circumstances that are unobserved by the econometrician. Therefore, we take account of *unobservable* circumstances by means of the fixed effect estimation outlined in section 2. The UB estimates of IOp vary between 17.2% (Mexico) and 72.5% (South Africa). On average, UB exceeds S by 22.8pp. It therefore yields a significant upward correction of IOp in comparison to S and LB, respectively. The difference between UB and S is broadly comparable to the respective gap in developed economies (Niehues and Peichl, 2014). As such, our results reflect recent concerns that downward biased IOp estimates based on observable circumstance characteristics provide misleading reference points as regards the normative significance of inequality (Kanbur and Wagstaff, 2016).[17]

Household Income. Panel (b) of Figure 1 displays analogous IOp estimates for equivalized household income. In contrast to the results on individual income, we thereby account for resource sharing at the household level and heterogeneity in household compositions. Estimates for S (LB) decrease for the vast majority of countries and now lie between 1.2% in Argentina (0%, China) and 35.9% in South Africa (24.7%, South Africa). This decrease follows from the assumption of resource sharing at the household level that largely nullifies gender-based differences in incomes. Hence, the average difference between S and LB remains at a very low level of 5.0pp. Again, using the alternative post-lasso OLS estimation strategy decreases this difference to 1.3pp. To the contrary, the UB estimates are largely comparable to their individual income analogues. According to UB, IOp ranges between 8.6% (Mexico) and 73.9% (South

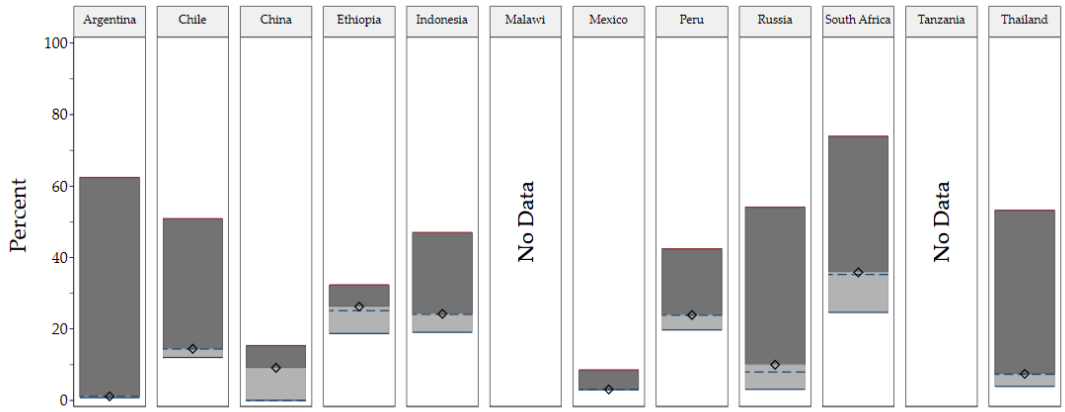
[17] Due to differences in the underlying data, we refrain from comparing our results to other IOp estimates in the relevant countries: See for example, Brock et al. (2016), Brunori et al. (2019a), Ferreira et al. (2018), Ferreira and Gignoux (2011), Golley et al. (2019), Piraino (2015), Song and Zhou (2019), Wendelspiess Chávez Juárez (2015), and Zhang and Eriksson (2010). These differences pertain to reference periods, the considered outcomes of interest, the detail of available circumstance characteristics, sample selection criteria, estimation methods, as well as inequality indices. However, we provide detailed information on these studies in Table C.1.

Figure 1: Bounds of Inequality of Opportunity

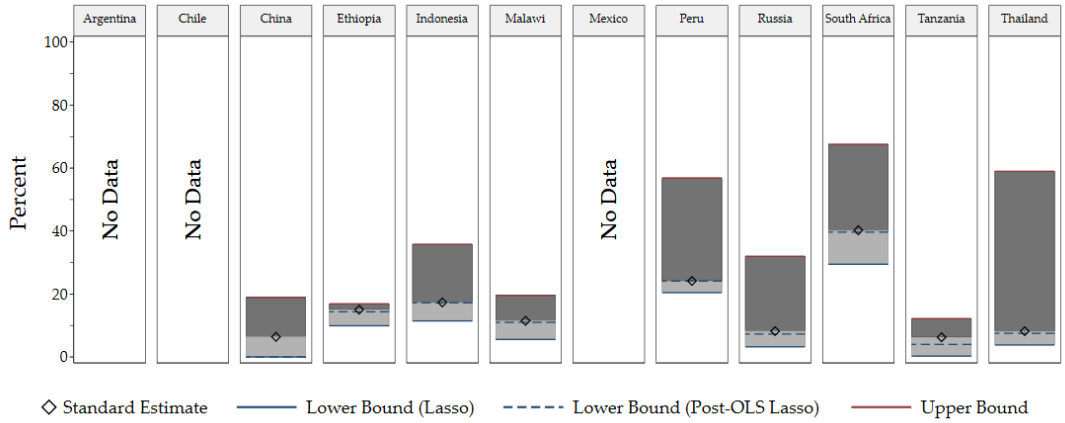
(a) Individual Income



(b) Equivalized Household Income



(c) Equivalized Household Expenditure



Notes: The figure shows estimates of relative IOp for individual incomes (Panel (a)), equivalized household incomes (Panel (b)) and equivalized household expenditures (Panel (c)) based on the MLD. Standard estimates (S) use the full set of country-specific circumstances disclosed in Table 1. Lower bound (LB) estimates use the full set of country-specific circumstances disclosed in Table 1 but estimate the relevant parameters by means of a lasso estimation to account for sampling variance. Upper bound (UB) estimates are based on predictions from individual fixed effects. *Source:* Own calculations based on data described in Table A.1.

Africa). As a consequence, the average difference between S and UB increases from 22.8pp to a level of 28.5pp when considering household instead of individual incomes. Our general conclusion, however, remains intact: In the context of the developing economies under consideration, the standard estimation approach recovers an estimate close to LB. However, its large distance to UB suggests severe underestimations due to the influence of unobservable circumstances.

Household Expenditure. In Panel (c), we show IOp estimates for equivalized household expenditure. There are different explanations for potential deviations of IOp in household expenditure and household income. First, if households smooth consumption its distribution is less unequal than the distribution of income. Additionally, assuming transitory fluctuations to be more strongly reflected in the outcome distribution Y_t than the smoothed distribution $\tilde{M}_t$, we would expect relative IOp in consumption expenditures to be *higher* than in income.[18] In fact, this is the pattern observed by Ferreira and Gignoux (2011) when comparing IOp in income and consumption for five Latin-American countries. Second, even if households smooth consumption, expenditures for consumption items, especially durables, can be lumpy (Meyer and Sullivan, 2017). This tendency is amplified by the fact that reference periods for expenditure reporting are oftentimes shorter (e.g. weekly, monthly, quarterly) in order to allow survey respondents to recall their expenditures in different categories. Again, assuming transitory fluctuations to be more strongly reflected in the outcome distribution Y_t than the smoothed distribution $\tilde{M}_t$, we would expect relative IOp in consumption expenditures to be *lower* than in income. Which of the two tendencies dominates is an empirical question and varies with the mode of data collection in the different countries. In our country sample the second channel tends to dominate. Compared to relative IOp in household income, IOp in household expenditure is on average 2.5pp (S), 1.6pp (LB), and 4.5pp (UB) lower. However, there is heterogeneity across countries. According to the standard estimate, relative IOp for household expenditure is higher than IOp for income in Peru, South Africa, and Thailand. The reverse is true for China, Ethiopia, Indonesia, and Russia.

Estimates for S (LB) with respect to consumption expenditure lie between 6.3% in Tanzania (0%, China) and 40.3% in South Africa (29.5%, South Africa). According to UB, IOp ranges between 12.2% (Tanzania) and 67.6% (South Africa). As a consequence, the average difference between S and LB (UB) amounts to 5.9pp (20.2pp). These findings support our conclusion that the standard estimation approach recovers an estimate close to LB.

Sensitivity Analysis. We conduct four sensitivity checks in which we probe the robustness of our conclusions to alternative specification choices.

[18] A similar line of thought can be found in Bourguignon et al. (2007) who argue that the presence of transitory fluctuations in the residual tends to bias IOp estimates downward.

MLD vs. Gini coefficient. The majority of empirical IOp estimations draw on the MLD due to its path-independent decomposability property. In the context of IOp measurment, this property allows for a perfect decomposition into circumstance-based unfair inequality and effort-based fair inequality. However, as noted by Brunori et al. (2019a) the MLD's senstivity to low income values leads to low relative measures of IOp.

Hence, we replicate our analysis based on the Gini coefficient and show the results in Figure 2. Indeed, relative IOp based on the Gini is larger than suggested by the MLD. For individual incomes, the standard estimate on average increases by 30pp and now lies between 34.1% (Argentina) and 68.1% (Peru). The corresponding UB on average increases by 26pp and ranges from 43.5% (Mexico) to 89.8% (South Africa). The LB on average increases by 27.8pp and lies between 28.7% (China) and 62.3% (Peru). The pattern is very similar for equivalized household income and expenditure (see Table A.3).

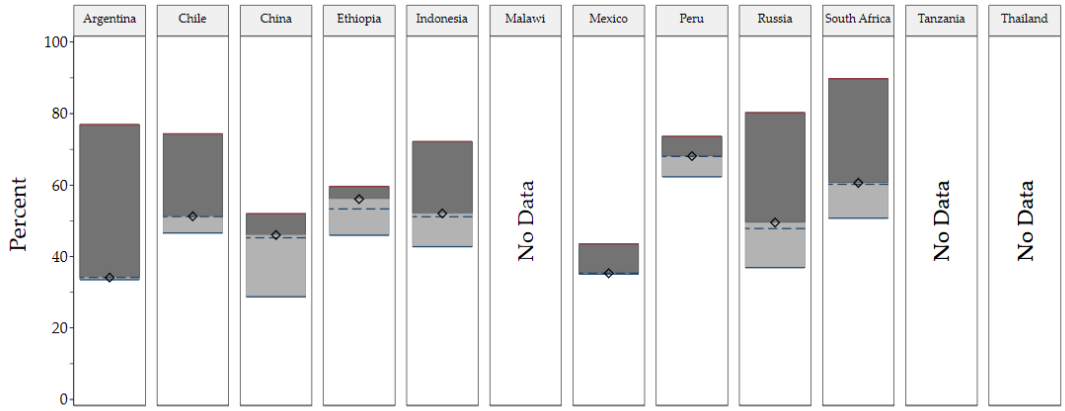
These results indicate that the attenuating effect implied by the tail sensitivity of the MLD largely outweighs the attenuating effect implied by the imperfect decomposability of the Gini coefficient. Furthermore, although using the Gini coefficient widens the gap between S and LB, the difference between UB and S is still larger for the majority of outcomes and countries in our sample. This observation confirms that independent of the inequality measure, the potential for downward biased IOp estimates is much larger than the potential to overestimate IOp in emerging economies.

Circumstance Availability. The differences between S and LB (UB) may vary with the size of the invoked circumstance set. To test the relevance of this concern in our sample, we re-estimate S and LB while restricting ourselves to a harmonized set of circumstances that is available in all countries under consideration. The internationally comparable circumstance set includes gender and year of birth. In Panel (a) of Figure 3 we plot the difference between S and UB (LB) according to the harmonized circumstance specification (y-axis) against the analogous differences in our baseline estimates (x-axis). The closer data points align with the 45 degree line, the more similar the results between the baseline and the alternative specification.

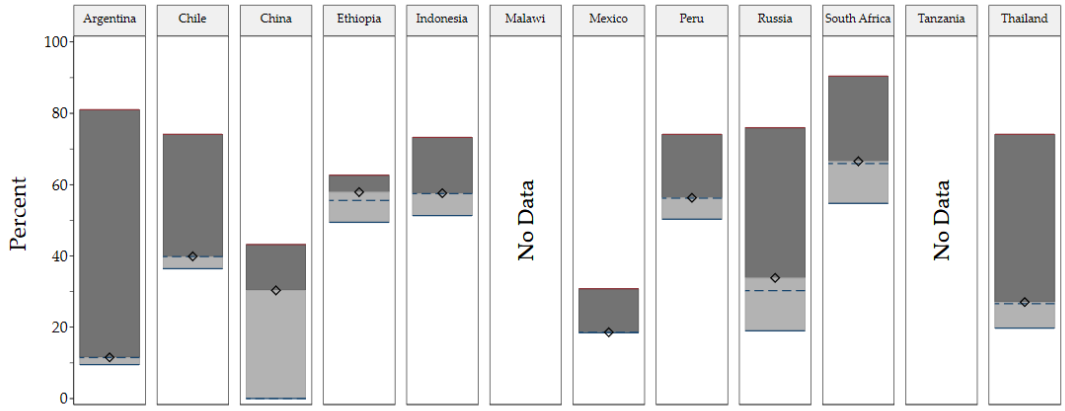
Restricting the circumstance set mechanically attenuates S but leaves UB unaltered. It is therefore unsurprising that the difference between S and UB increases for all countries under consideration. The reverse holds true for the difference between S and LB. In fact, the restriction of the circumstance set leads to a zero difference between S and LB for the majority of the country cases. These results therefore confirm our main conclusion: The more parsimonious the circumstance set, the stronger the correspondence between S and LB and the higher the downward bias. Unfortunately, we cannot run the reverse test by increasing the number of circumstances. Therefore, we cannot provide a direct assessment of the precise conditions under which S and LB come adrift.

Figure 2: Bounds of Inequality of Opportunity, Gini coefficient

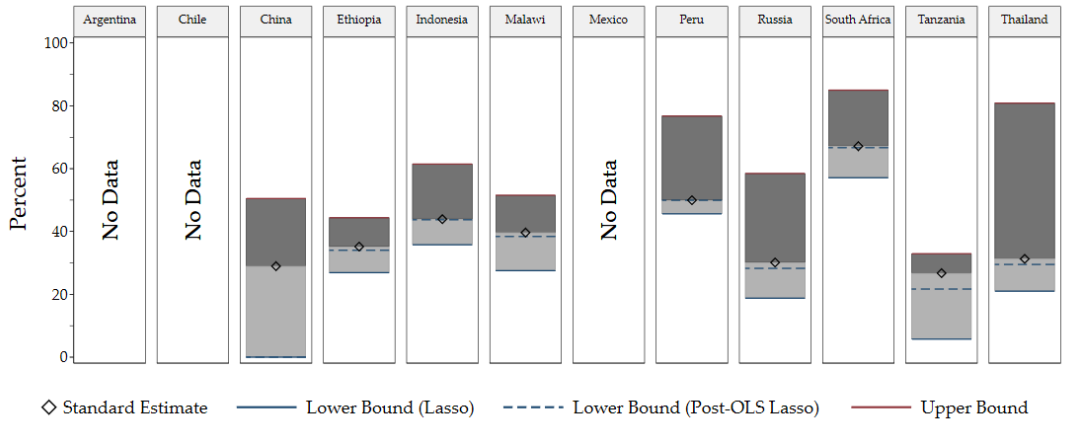
(a) Individual Income



(b) Equivalized Household Income



(c) Equivalized Household Expenditure

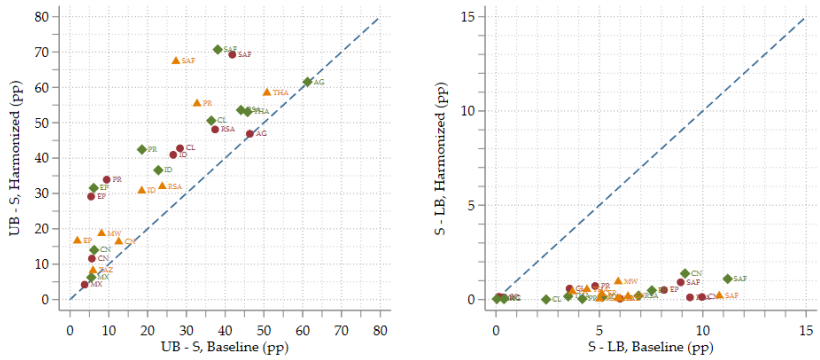


Notes: The figure shows estimates of relative IOp for individual incomes (Panel (a)), equivalized household incomes (Panel (b)) and equivalized household expenditures (Panel (c)) based on the Gini coefficient. Standard estimates (S) use the full set of country-specific circumstances disclosed in Table 1. Lower bound (LB) estimates use the full set of country-specific circumstances disclosed in Table 1 but estimate the relevant parameters by means of a lasso estimation to account for sampling variance. Upper bound (UB) estimates are based on predictions from individual fixed effects.
Source: Own calculations based on data described in Table A.1.

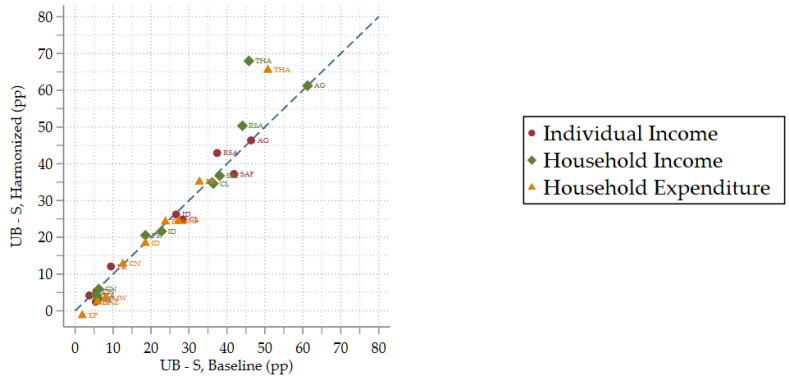
Figure 3: Sensitivity Checks

Difference UB - S	Difference S - LB

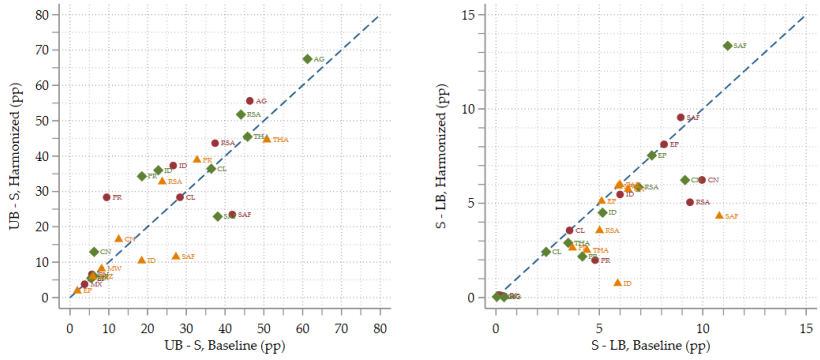
(a) Harmonized Circumstance Set



(b) Harmonized Number of Periods



(c) Harmonized Year of Interest



Notes: The figure shows the robustness of our results according to three variations. In Panel (a) we harmonize the set of circumstances. In Panel (b) we harmonize the number of periods used to calculate UB. In Panel (c) we harmonize the year of interest for the calculation of IOp according to the scheme outlined in Table A.5. In all figures, the x-axis shows the percentage point (pp) difference between the standard estimate (S) and the lower bound (LB) (upper bound (UB)) according to our baseline specification. The y-axis provides analogous statistics after the respective harmonization.
Source: Own calculations based on data described in Table A.1.

Number of Periods. The difference between S and UB may differ with the number of periods used to construct the individual FEs. In the baseline we set a minimum threshold for the number of periods used to calculate the fixed effect. However, in spite of implementing this minimum threshold the de facto number of observations used for the construction of the individual FEs is not bounded from above and therefore varies across countries (Table 1). To test the relevance of this concern, we construct UB estimates in which we restrict the sample to the three most recent observations for each individual in each country. In Panel (b) of Figure 3 we plot the differences between S and UB according to this harmonized specification (y-axis) against the analogous differences according to our baseline estimates (x-axis). The closer data points align with the 45 degree line, the more similar the results between the baseline and the alternative specification.

We find that all data points with respect to the difference between S and UB closely align to the 45 degree line. This pattern suggest that even short panels deliver reliable indicators for UB inequality of opportunity. Note that the panel length impinges upon the UB estimate only. Therefore, all differences between S and LB remain unaffected by this harmonization.

Year of Interest. Our results may be sensitive to alternations in the time period of interest. In our baseline analysis we focus on the most recent available data years covering a range from 2009 to 2017. Therefore, we replicate our analysis for the country-specific wave in closest proximity to 2009.[19] In Panel (c) of Figure 3 we plot the differences between S and UB (LB) according to this harmonized specification (y-axis) against the analogous differences according to our baseline estimates (x-axis). The closer the data points align with the 45 degree line, the more similar the results between the baseline and the alternative specification.

Given that a society's opportunity structure is shaped by long-run institutional features, one would expect these differences to be small. Indeed, we find that the data points for the difference between S and UB closely group around the 45 degree line. A similar conclusion holds for the difference between S and LB although the dispersion around the 45 degree line is somewhat larger.

5 Conclusion

Measures of IOp are of considerable policy relevance since they reflect widely-held principles of distributive justice and plausibly correlate with measures of economic development. In spite of their interest, point estimates of IOp are surrounded by severe uncertainty since they can be both upward and downward biased. Due to poorer data infrastructures with smaller

[19] Table A.5 shows the country-specific year chosen for this sensitivity check.

sample sizes and less information on circumstance characteristics, IOp estimates in emerging economies may be particularly susceptible to both biases and it is unclear which of the two biases prevails.

We show that downward bias clearly dominates in the context of emerging economies. On the one hand, sparsely populated circumstance sets restrict the scope for overfitting circumstance information to the data. As a consequence, standard estimates of IOp strongly correspond to their lower bound analogues. This result stands in contrast to recent evidence from countries with richer data environments. On the other hand, the sparsity of observable circumstance information leads to large differences between standard estimates of IOp and their upper bound analogues. The extent of these differences is largely comparable to more developed countries and ranges between 20pp and 30pp.

While we provide reasonable bounds for IOp in these countries, substantial differences between lower and upper bound IOp remain. Our results therefore tie in with recent concerns that downward biased IOp estimates could misguide judgments on the normative significance of inequality. In the future, such gaps may be closed as better data sets become available. However, until such innovations materialize, bounding the range of potential estimates remains a viable way to limit the scope for downplaying the normative significance of inequality in the countries of interest.

References

AIYAR, S. and C. EBEKE (2019). "Inequality of Opportunity, Inequality of Income and Economic Growth". *IMF Working Paper Series* 19/34.

ALESINA, A., S. HOHMANN, S. MICHALOPOULOS, and E. PAPAIOANNOU (2019). "Intergenerational Mobility in Africa". *mimeo*.

ALESINA, A., S. STANTCHEVA, and E. TESO (2018). "Intergenerational Mobility and Preferences for Redistribution". *American Economic Review* 108 (2), pp. 521–554.

ANDREOLI, F. and A. FUSCO (2019). "Robust Cross-Country Analysis of Inequality of Opportunity". *Economics Letters* 182, pp. 86–89.

ARNESON, R. J. (1989). "Equality and Equal Opportunity for Welfare". *Philosophical Studies* 56 (1), pp. 77–93.

BALCÁZAR, C. F. (2015). "Lower Bounds on Inequality of Opportunity and Measurement Error". *Economics Letters* 137, pp. 102–105.

BELLONI, A. and V. CHERNOZHUKOV (May 2013). "Least Squares after Model Selection in High-Dimensional Sparse Models". *Bernoulli* 19 (2), pp. 521–547.

BLACKBURN, M. L. (2007). "Estimating Wage Differentials without Logarithms". *Labour Economics* 14 (1), pp. 73–98.

BOURGUIGNON, F., F. H. G. FERREIRA, and M. MENÉNDEZ (2007). "Inequality of Opportunity in Brazil". *Review of Income and Wealth* 53 (4), pp. 585–618.

BROCK, J. M., V. PERAGINE, and S. TONINI (2016). "Inequality of Opportunity". *Transition for all: Equal opportunities in an unequal world*. EBRD. Chap. 3, pp. 81–98.

BRUNORI, P., P. HUFE, and D. MAHLER (2018). "The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees". *ifo Working Papers* 252.

BRUNORI, P., F. PALMISANO, and V. PERAGINE (2019a). "Inequality of Opportunity in Sub-Saharan Africa". *Applied Economics* 51 (60), pp. 6428–6458.

BRUNORI, P., V. PERAGINE, and L. SERLENGA (2019b). "Upward and Downward Bias When Measuring Inequality of Opportunity". *Social Choice and Welfare* 52 (4), pp. 635–661.

CAPPELEN, A. W., A. D. HOLE, E. Ø. SØRENSEN, and B. TUNGODDEN (2007). "The Pluralism of Fairness Ideals: An Experimental Approach". *American Economic Review* 97 (3), pp. 818–827.

COHEN, G. A. (1989). "On the Currency of Egalitarian Justice". *Ethics* 99 (4), pp. 906–944.

CUBAS, G. (2016). "Distortions, Infrastructure, and Female Labor Supply in Developing Countries". *European Economic Review* 87, pp. 194–215.

FERREIRA, F. H. G., C. LAKNER, M. A. LUGO, and B. ÖZLER (2018). "Inequality of Opportunity and Economic Growth: How Much Can Cross-Country Regressions Really Tell Us?" *Review of Income and Wealth* 64 (4), pp. 800–827.

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 (4), pp. 622–657.

FLEURBAEY, M. (1995). "Three Solutions for the Compensation Problem". *Journal of Economic Theory* 65 (2), pp. 505–521.

FOSTER, J. E. and A. A. SHNEYEROV (2000). "Path Independent Inequality Measures". *Journal of Economic Theory* 91 (2), pp. 199–222.

GOLLEY, J., Y. ZHOU, and M. WANG (2019). "Inequality of Opportunity in China's Labor Earnings: The Gender Dimension". *China & World Economy* 27 (1), pp. 28–50.

HASTIE, T., R. TIBSHIRANI, and J. FRIEDMAN (2013). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Heidelberg: Springer.

HUFE, P., A. PEICHL, J. E. ROEMER, and M. UNGERER (2017). "Inequality of Income Acquisition: The Role of Childhood Circumstances". *Social Choice and Welfare* 143 (3-4), pp. 499–544.

JAYACHANDRAN, S. (2015). "The Roots of Gender Inequality in Developing Countries". *Annual Review of Economics* 7 (1), pp. 63–88.

JUSOT, F., S. TUBEUF, and A. TRANNOY (2013). "Circumstances and Efforts: How Important Is Their Correlation for the Measurement of Inequality of Opportunity in Health?" *Health Economics* 22 (12), pp. 1470–1495.

KANBUR, R. and A. WAGSTAFF (2016). "How Useful Is Inequality of Opportunity as a Policy Construct?" *Inequality and Growth: Patterns and Policy*. Ed. by K. BASU and J. E. STIGLITZ. Vol. 1. London: Palgrave Macmillan UK. Chap. 4, pp. 131–150.

MARRERO, G. A. and J. G. RODRÍGUEZ (2013). "Inequality of Opportunity and Growth". *Journal of Development Economics* 104, pp. 107–122.

MEJÍA, D. and M. ST-PIERRE (2008). "Unequal Opportunities and Human Capital Formation". *Journal of Development Economics* 86 (2), pp. 395–413.

MEYER, B. D. and J. X. SULLIVAN (2017). "Consumption and Income Inequality in the U.S. since the 1960s". *National Bureau of Economic Research Working Paper Series* 23655.

NEIDHÖFER, G., J. SERRANO, and L. GASPARINI (2018). "Educational Inequality and Intergenerational Mobility in Latin America: A New Database". *Journal of Development Economics* 134, pp. 329–349.

NIEHUES, J. and A. PEICHL (2014). "Upper Bounds of Inequality of Opportunity: Theory and Evidence for Germany and the US". *Social Choice and Welfare* 43 (1), pp. 73–99.

PERAGINE, V., F. PALMISANO, and P. BRUNORI (2014). "Economic Growth and Equality of Opportunity". *World Bank Economic Review* 28 (2), pp. 247–281.

PIRAINO, P. (2015). "Intergenerational Earnings Mobility and Equality of Opportunity in South Africa". *World Development* 67 (1), pp. 396–405.

RAMOS, X. and D. VAN DE GAER (2016). "Empirical Approaches to Inequality of Opportunity: Principles, Measures, and Evidence". *Journal of Economic Surveys* 30 (5), pp. 855–883.

ROEMER, J. E. (1998). *Equality of Opportunity*. Cambridge: Harvard University Press.

ROHNER, D. (2011). "Reputation, Group Structure and Social Tensions". *Journal of Development Economics* 96 (2), pp. 188–199.

SHORROCKS, A. F. (1980). "The Class of Additively Decomposable Inequality Measures". *Econometrica* 48 (3), pp. 613–625.

SONG, Y. and G. ZHOU (2019). "Inequality of Opportunity and Household Education Expenditures: Evidence from Panel Data in China". *China Economic Review* 55 (1), pp. 85–98.

WENDELSPIESS CHÁVEZ JUÁREZ, F. (2015). "Measuring Inequality of Opportunity with Latent Variables". *Journal of Human Development and Capabilities* 16 (1), pp. 106–121.

ZHANG, Y. and T. ERIKSSON (2010). "Inequality of Opportunity and Income Inequality in Nine Chinese Provinces, 1989–2006". *China Economic Review* 21 (4), pp. 607–616.

A Additional Tables

Table A.1: Data Overview

Country	Country Abbr.	Panel Period	Waves	Data	Source	Individual	Household		Sample Weights Available
						Income	Income	Expenditure	
Argentina	AG	2003-2015	12	Encuesta Permanente de Hogares	Instituto Nacional de Estadística y Censos (INDEC)	Net	Net	-	Yes
Chile	CL	2006-2009	4	Encuesta Panel CASEN	Ministerio de Desarrollo Social, Chile	Net	Net	-	Yes
China	CN	1988-2014	10	China Health and Nutrition Survey	Carolina Population Center at the University of North Carolina at Chapel Hill and National Institute for Nutrition and Health (NINH) at the Chinese Center for Disease Control and Prevention (CCDC)	Net (Labor)	Net (Labor)	Yes	No
Ethiopia	EP	1994-2009	6	Ethiopia Rural Household Survey	International Food Policy Research Institute (IFPRI), Washington DC	Net	Net	Yes	No
Indonesia	ID	1992-2013	5	Indonesian Family Life Survey (IFLS)	RAND Social and Economic Well-Being	Net (Labor)	Net (Labor)	Yes	No
Malawi	MW	1998-2010	4	Malawi Longitudinal Study of Family and Health	Population Studies Center at the University of Pennsylvania and College of Medicine at the University of Malawi and Invest in Knowledge (IKI) in Zomba, Malawi	-	-	Yes	No
Mexico	MX	1999-2009	7	Encuesta Evaluation de los Hogares (ENCEL)	International Food Policy Research Institute (IFPRI), Washington DC	Net	Net	-	No
Peru	PR	1998-2011	14	Encuesta Nacional de Hogares, Condiciones de Vida y Pobreza	Instituto Nacional de Estadística e Informatica	Gross	Gross	Yes	Yes
Russia	RSA	1994-2017	22	Russia Longitudinal Monitoring Survey (RLMS)	National Research University "Higher School of Economics", OOO "Demoscope", Carolina Population Center, University of North Carolina at Chapel Hill and the Institute of Sociology of the Federal Center of Theoretical and Applied Sociology of the Russian Academy of Sciences	Net (Labor)	Net (Labor)	Yes	Yes
South Africa	SAF	2008-2017	5	National Income Dynamics Study (NIDS)	Southern Africa Labour and Development Research Unit (SALDRU), University of Cape Town	Net (Labor)	Net (Labor)	Yes	Yes
Tanzania	TAZ	1991-2010	6	Kagera Health and Development Survey	Economic Development Initiatives	-	-	Yes	No
Thailand	THA	1997-2017	21	Townsend Thai Data	The Townsend Thai Project	-	Gross	Yes	No

Table A.2: Sample Selection

Country	Outcome	Full Sample	+Age 25-55	+Circumstance Availability	+Outcome Availability (Recent Year)	+Outcome Availability (Longitudinal)
Argentina	Equiv. HH Income	N = 93,473 ARP 73,704	N = 37,190 ARP 80,235	N = 37,181 ARP 80,240	N = 31,577 ARP 83,832	N = 3,019 ARP 78,103
Chile	Equiv. HH Income	N = 21,087 CLP 2,632,864	N = 8,483 CLP 2,736,782	N = 6,307 CLP 2,782,129	N = 4,918 CLP 2,925,172	N = 2,808 CLP 3,093,657
China	Equiv. HH Income	N = 10,434 CNY 34,858	N = 5,579 CNY 40,624	N = 5,047 CNY 40,573	N = 1,118 CNY 32,430	N = 243 CNY 35,796
Ethiopia	Equiv. HH Income	N = 6,982 ETB 18,253	N = 2,117 ETB 18,465	N = 764 ETB 17,169	N = 742 ETB 17,645	N = 660 ETB 18,570
Indonesia	Equiv. HH Income	N = 31,035 IDR 13,231,805	N = 14,382 IDR 14,516,044	N = 5,426 IDR 18,097,790	N = 3,620 IDR 20,517,590	N = 786 IDR 26,036,236
Malawi	Equiv. HH Expenditure	N = 3,397 MWK 23,053	N = 2,243 MWK 24,301	N = 440 MWK 27,512	N = 440 MWK 27,512	N = 362 MWK 30,052
Mexico	Equiv. HH Income	N = 30,789 MXP 21,039	N = 11,075 MXP 20,169	N = 9,443 MXP 20,279	N = 6,047 MXP 22,390	N = 3,050 MXP 21,123
Peru	Equiv. HH Income	N = 71,758 PEN 6,189	N = 26,710 PEN 7,314	N = 26,095 PEN 7,375	N = 15,280 PEN 9,075	N = 2,193 PEN 9,013
Russia	Equiv. HH Income	N = 15,201 RUB 287,663	N = 7,597 RUB 309,332	N = 1,738 RUB 309,794	N = 1,383 RUB 328,866	N = 1,181 RUB 323,750
South Africa	Equiv. HH Income	N = 21,306 ZAR 40,550	N = 8,602 ZAR 49,357	N = 3,187 ZAR 65,522	N = 2,402 ZAR 74,904	N = 670 ZAR 71,921
Tanzania	Equiv. HH Expenditure	N = 4,289 TZS 486,370	N = 2,681 TZS 504,302	N = 936 TZS 532,841	N = 936 TZS 532,841	N = 221 TZS 464,709
Thailand	Equiv. HH Income	N = 3,649 THB 156,869	N = 1,439 THB 180,727	N = 473 THB 150,957	N = 473 THB 150,957	N = 465 THB 150,089

Notes: The table shows how the step-wise sample selection procedure changes the number of observations and the mean outcome variable denoted in local currency. The sequence is as follows: full sample (column 3), age restriction (column 4), full circumstance availability (column 5), observability of outcome variables in the year of interest (column 6), observability of outcome variables in longitudinal dimension (column 7).
Source: Own calculations based on the panel survey data described in Table A.1.

Table A.3: Absolute and Relative Inequality of Opportunity, Baseline Specification

					Parameters					Inequality		MLD, Absolute				MLD, Relative (%)				Gini, Absolute				Gini, Relative (%)			
	Country	Year	N	N (FE)	P	P^S	N/P^S	P^{LB}	λ^*	MLD	Gini	S	LB 1	LB 2	UB	S	LB 1	LB 2	UB	S	LB 1	LB 2	UB	S	LB 1	LB 2	UB
Individual Income	Argentina	2015	3,019	6,038	8	7	431.29	7	0.003	0.285	0.386	0.027	0.026	0.027	0.159	9.33	8.99	9.33	55.70	0.132	0.129	0.132	0.297	34.13	33.50	34.13	76.97
	Chile	2009	2,808	8,424	37	26	108.00	16	0.015	0.395	0.430	0.078	0.064	0.078	0.190	19.79	16.24	19.74	48.16	0.220	0.200	0.220	0.319	51.25	46.58	51.17	74.36
	China	2014	243	717	22	14	17.36	7	0.070	0.530	0.471	0.084	0.032	0.080	0.114	15.91	5.95	15.05	21.54	0.217	0.135	0.213	0.245	46.01	28.73	45.25	52.03
	Ethiopia	2009	660	2,433	69	33	20.00	12	0.056	0.873	0.622	0.214	0.143	0.201	0.262	24.54	16.42	23.04	29.97	0.349	0.286	0.331	0.371	56.08	45.95	53.33	59.64
	Indonesia	2013	786	2,036	29	20	39.30	9	0.036	0.533	0.478	0.096	0.064	0.093	0.238	18.04	12.04	17.41	44.65	0.249	0.204	0.244	0.345	52.07	42.75	51.13	72.20
	Mexico	2009	3,050	16,552	5	4	762.50	4	0.001	0.186	0.311	0.025	0.025	0.025	0.032	13.42	13.28	13.42	17.18	0.110	0.109	0.110	0.135	35.33	35.09	35.33	43.54
	Peru	2010	2,193	5,878	37	33	66.45	26	0.017	0.695	0.489	0.213	0.180	0.213	0.279	30.64	25.85	30.59	40.10	0.333	0.305	0.333	0.361	68.11	62.31	68.01	73.66
	Russia	2017	1,181	10,816	54	46	25.67	14	0.033	0.234	0.358	0.049	0.027	0.046	0.137	20.85	11.47	19.52	58.27	0.177	0.132	0.171	0.287	49.52	36.89	47.84	80.30
	South Africa	2017	670	2,331	48	34	19.71	20	0.025	0.418	0.471	0.128	0.091	0.126	0.303	30.62	21.68	30.24	72.46	0.286	0.239	0.284	0.423	60.62	50.71	60.22	89.78
Household Income	Argentina	2015	3,019	6,038	8	7	431.29	6	0.007	0.236	0.365	0.003	0.002	0.003	0.147	1.17	0.79	1.16	62.42	0.042	0.035	0.042	0.296	11.55	9.51	11.49	81.01
	Chile	2009	2,808	8,424	37	26	108.00	18	0.009	0.261	0.385	0.038	0.031	0.038	0.133	14.45	12.03	14.41	50.88	0.154	0.140	0.153	0.285	39.88	36.42	39.83	74.11
	China	2014	243	717	22	14	17.36	1	0.185	0.517	0.480	0.047	0.000	0.000	0.079	9.14	0.00	0.00	15.36	0.146	0.000	0.000	0.208	30.36	0.00	0.00	43.21
	Ethiopia	2009	660	2,433	69	33	20.00	13	0.050	0.953	0.640	0.250	0.178	0.240	0.308	26.26	18.73	25.14	32.36	0.370	0.316	0.355	0.401	57.92	49.43	55.56	62.67
	Indonesia	2013	786	2,036	29	20	39.30	12	0.022	0.482	0.473	0.117	0.092	0.116	0.226	24.23	19.08	24.09	47.00	0.273	0.243	0.272	0.346	57.61	51.32	57.48	73.24
	Mexico	2009	3,050	16,552	5	4	762.50	4	0.000	0.198	0.331	0.006	0.006	0.006	0.017	3.07	3.03	3.07	8.56	0.062	0.061	0.062	0.102	18.59	18.46	18.59	30.81
	Peru	2010	2,193	5,878	37	33	66.45	25	0.015	0.561	0.470	0.134	0.111	0.134	0.238	23.91	19.74	23.86	42.45	0.265	0.237	0.264	0.348	56.32	50.31	56.21	74.09
	Russia	2017	1,181	10,816	54	46	25.67	11	0.031	0.194	0.330	0.019	0.006	0.015	0.105	10.02	3.12	7.96	54.10	0.112	0.063	0.100	0.251	33.85	19.00	30.27	75.96
	South Africa	2017	670	2,331	48	34	19.71	20	0.033	0.475	0.495	0.170	0.117	0.167	0.351	35.85	24.65	35.26	73.94	0.329	0.271	0.326	0.447	66.53	54.74	65.92	90.42
	Thailand	2017	465	6,338	15	12	38.75	5	0.034	0.289	0.419	0.021	0.011	0.021	0.154	7.43	3.94	7.32	53.22	0.113	0.083	0.112	0.311	27.05	19.75	26.62	74.12
Household Expenditure	China	2014	243	717	22	14	17.36	1	0.255	1.427	0.766	0.091	0.000	0.000	0.270	6.38	0.00	0.00	18.95	0.222	0.000	0.000	0.387	28.96	0.00	0.00	50.50
	Ethiopia	2009	660	2,433	69	33	20.00	14	0.029	0.613	0.557	0.092	0.061	0.088	0.104	15.02	9.91	14.35	16.90	0.196	0.150	0.190	0.247	35.20	26.90	34.03	44.39
	Indonesia	2013	786	2,036	29	20	39.30	12	0.028	0.398	0.480	0.069	0.046	0.068	0.143	17.33	11.45	17.21	35.81	0.211	0.172	0.210	0.295	43.93	35.78	43.75	61.47
	Malawi	2010	362	965	16	12	30.17	7	0.101	1.213	0.720	0.139	0.067	0.133	0.238	11.47	5.56	10.99	19.60	0.285	0.198	0.276	0.371	39.65	27.54	38.41	51.55
	Peru	2010	2,193	5,878	37	33	66.45	27	0.007	0.187	0.331	0.045	0.038	0.045	0.106	24.13	20.43	24.12	56.88	0.166	0.151	0.166	0.254	49.96	45.62	49.94	76.75
	Russia	2017	1,181	10,816	54	46	25.67	16	0.033	0.472	0.518	0.039	0.015	0.034	0.151	8.22	3.21	7.28	32.00	0.156	0.097	0.146	0.303	30.13	18.78	28.27	58.48
	South Africa	2017	670	2,331	48	34	19.71	21	0.030	0.503	0.531	0.203	0.148	0.200	0.340	40.26	29.46	39.66	67.57	0.357	0.303	0.354	0.452	67.18	57.11	66.69	85.01
	Tanzania	2010	221	819	18	15	14.73	3	0.114	0.586	0.569	0.037	0.002	0.023	0.071	6.27	0.28	4.00	12.20	0.152	0.033	0.123	0.187	26.74	5.74	21.68	32.94
	Thailand	2017	465	6,338	15	12	38.75	4	0.065	0.764	0.620	0.063	0.029	0.057	0.451	8.19	3.80	7.51	58.98	0.194	0.130	0.183	0.501	31.29	21.01	29.51	80.87

Notes: The table shows baseline IOp estimation results for individual income, household income, and household expenditure in terms of the MLD and the Gini coefficient. Column 2 shows year t for which IOp is estimated. Column 3 (4) display the number of observations in years v for the fixed effect estimation. Columns 5-9 provide information on the total number of circumstance categories (P), the number of estimated parameters under the standard approach (P^S), the ratio between number of observations and estimated parameters, the number of parameters selected by the lasso estimation (P^{LB}), and the selected value of lambda in the cross validation (λ^*). Columns 10 and 11 display total inequality in terms of the MLD and the Gini coefficient. The remaining columns display baseline absolute and relative IOp measures. Standard estimates (S) use the full set of country-specific circumstances disclosed in Table 1. Lower bound (LB) estimates use the full set of country-specific circumstances disclosed in Table 1 but estimate the relevant parameters by means of a (post-OLS) lasso estimation to account for sampling variance. Upper bound (UB) estimates are based on predictions from individual fixed effects. LB 1 refers to the standard lasso. LB 2 refers to the post-OLS lasso.
Source: Own calculations based on the panel survey data described in Table A.1.

Table A.4: All (Lasso-Selected) Parameter Categories

Country	Parameters selected by lasso: Individual Income (●), Household Income (◆), and Household Expenditure (▲)
Argentina	**gender**●◆, **birthyear**●◆, **birthplace** (current place of residence●, different place than current residence◆, other province●◆, neighboring country●, other country●◆)
Chile	**gender**●, **birthyear**●◆, **father education** (no schooling●◆, primary, secondary●◆, tertiary●◆), **mother education** (no schooling●◆, primary, secondary●◆, tertiary●◆), **birthplace** (national, foreign), **ethnicity** (not member of any indigeneous population●◆, Aymara●, Rapa Nui, Quechua, Mapuche◆, Atacameno●◆, Coya, Kawaskar, Yagan, Diaguita●◆), **chronic disease**●◆ (yes/no), **labor force status of father** (not working●, employer◆, self-employed◆, employed◆, domestic worker, armed forces◆), **labor force status of mother** (not working●, employer, self-employed, employed, domestic worker, armed forces◆)
China	**gender**●, **birthyear**●, **ethnicity** (Han●, Mongolian, Hui, Tibetan, Vaguer, Miao, Yi, Zhuang, Buyi, Korean, Man, Dong, Yao, Tujia, other●), **birthplace urbanity** (city●, suburban, county capital city, village●)
Ethiopia	**gender**▲, **birthyear**▲, **father education** (no schooling●◆, some nursery school, 1st grade, 2nd grade, 3rd grade▲, 4th grade, 5th grade, 6th grade▲, 7th grade, 8th grade, 9th grade, 10th grade, 11th grade, 12th grade, uncompleted non-university higher education, completed non university higher education, university education, adult literacy program◆, other literacy program, parochia education, koranic education, other), **mother education** (no schooling, some nursery school●◆, 1st grade, 2nd grade, 3rd grade, 4th grade, 5th grade, 6th grade, 7th grade, 8th grade, 9th grade, 10th grade, 11th grade, 12th grade, uncompleted non-university higher education, completed non university higher education, university education, adult literacy program●, other literacy program, parochia education, koranic education, other), **ethnicity** (Amhara●▲, Oromo●◆▲, Tigrai●◆▲, Adere, Afar, Gurage●◆, Somali, other, Gedeo●◆▲, Gamo, Kembata▲, Wolaita●◆▲, Hadiya◆, Saho●◆▲), **religion** (None◆, Orthodox, Catholic▲, Muslim, Other Christian, Protestant, Traditional●◆▲, other)
Indonesia	**gender**◆▲, **birthyear**●◆▲, **father education** (no schooling, elementary schooling●◆▲, junior high, senior high●◆, junior college/college/university▲, other◆) **mother education** (no schooling, elementary schooling, junior high, senior high●◆▲, junior college/college/university, other), **ethnicity** (Jawa●◆▲, Sunda●◆, Bali●◆▲, Minang, Betawi, other▲), **religion** (Islam▲, Catholic◆▲, Protestant, Hindu, Buddha, Konghucu), **foreign language**●◆▲
Malawi	**gender**▲, **birthyear**, **father education** (no schooling▲, primary schooling, more than primary schooling, other), **mother education** (no schooling▲, primary schooling▲, more than primary schooling, other), **religion** (Catholic, Protestant▲, Revival, Moslem▲, Traditional, other)
Mexico	**gender**●◆ , **birthyear**●◆, **indigeneous**●◆
Peru	**gender**●◆▲, **birthyear**●◆▲, **birthplace** (Amazonas●◆▲, Áncash●◆, Apurímac●◆, Arequipa●◆▲, Ayacucho●◆▲, Cajamarca●◆▲, Callao●◆▲, Cusco▲, Huancavelica●◆, Huánuco●◆▲, Ica●◆, Junín●▲, La Libertad●◆▲, Lambayeque●◆▲, Lima●◆▲, Loreto●◆▲, Madre de Dios●◆▲, Moquegua◆▲, Pasco, Piura●◆▲, Puno, San Martin, Tacna▲, Tumbes▲, Ucayali●▲, Other county▲), **chronic disease**●◆▲, **language** (Quechua●◆▲, Aymara, other native language●◆▲, Spanish●◆▲, foreign language, deaf-dumb●◆▲)

Continued on next page

Table A.4: All (Lasso-Selected) Parameter Categories (cont.)

Country	Parameters selected by lasso: Individual Income (●), Household Income (◆), and Household Expenditure (▲)
Russia	**gender**$^{●}$, **birthyear**, **father education** (without education/illiterate, elementary school/incomplete secondary school$^{●▲}$, professional courses, vocational training without secondary education, vocational training with secondary education, secondary education$^{▲}$, technical community college$^{●▲}$, institute/university/academy$^{◆▲}$, post-graduate course, academic degree$^{●}$), **mother education** (without education/illiterate, elementary school/incomplete secondary school$^{●◆▲}$, professional courses$^{▲}$, vocational training without secondary education, vocational training with secondary education, secondary education, technical community college, institute/university/academy$^{●◆▲}$, post-graduate course, academic degree), **birthplace** (Russia, Ukraine, Belorussia$^{●▲}$, Azerbaizhan, Kazakhstan$^{◆}$, Uzbekistan, other country$^{▲}$), **father occupation** (armed forces$^{◆}$, legislators/senior officials/managers$^{●}$, professionals$^{▲}$, technicians/associate professionals, clerks, service workers/shop market sales work, skilled agricultural and fishery worker$^{▲}$, craft and related trade workers, plant and machine operators/assemblers, elementary occupations), **mother occupation** (armed forces$^{●▲}$, legislators/senior officials/managers, professionals, technicians/associate professionals$^{◆▲}$, clerks, service workers/shop market sales work, skilled agricultural and fishery worker, craft and related trade workers, plant and machine operators/assemblers, elementary occupations$^{●◆}$), **birthplace urbanity** (city$^{●◆▲}$, urban-type settlement, village/Derevnia/Kishlak/Aul$^{●◆▲}$), **height**$^{●◆}$
South Africa	**gender**$^{●◆▲}$, **birthyear**$^{●◆▲}$, **father education** (Grade R/0, Grade 1$^{●◆▲}$, Grade 2, Grade 3$^{●◆▲}$, Grade 4, Grade 5$^{▲}$, Grade 6$^{●}$, Grade 7, Grade 8$^{◆▲}$, Grade 9$^{●◆}$, Grade 10$^{◆▲}$, Grade 11$^{●▲}$, Grade 12$^{●◆▲}$, other, no schooling$^{●◆▲}$, National Certificate Vocational 2, National Certificate Vocational 4, NTC 1, NTC 2, NTC 3), **mother education** (Grade R/0, Grade 1, Grade 2$^{●}$, Grade 3$^{●◆}$, Grade 4, Grade 5$^{●◆▲}$, Grade 6$^{▲}$, Grade 7$^{◆}$, Grade 8$^{●◆▲}$, Grade 9, Grade 10$^{●◆▲}$, Grade 11$^{▲}$, Grade 12$^{●◆▲}$, other, no schooling$^{●◆▲}$, National Certificate Vocational 2, National Certificate Vocational 4, NTC 3), **foreign birthplace**$^{▲}$, **ethnicity** (African, Coloured$^{●◆}$, Asian/Indian$^{●◆▲}$, White$^{●◆▲}$, other)
Tanzania	**gender**$^{▲}$, **birthyear**$^{▲}$, **birthplace** (non-foreign/foreign), **ethnicity** (Mhaya, Mnyambo, Mhangaza, Msubi, Kishubi, Mzinza, other), **religion** (Musilim, Catholic, Protestant, Other Christian, Traditional, other)
Thailand	**gender**, **birthyear**, **father education** (no education, less than P4, P4, more than P4), **mother education** (no education$^{◆▲}$, less than P4, P4, more than P4), **wealth of parents** (among the poorest households in the village$^{◆}$, around the middle in terms of wealth$^{◆▲}$, among the rich households in the village), **land size of parents**$^{◆▲}$

Notes: The table shows the circumstance categories used to calculate standard estimates (S). Circumstance variables are denoted in boldface; their respective categories are listed in parentheses. Superscripts indicate variables chosen by the lasso procedure for the lower bounds (LB) of three different outcome variables: individual income (●), household income (◆), household consumption expenditure (▲).
Source: Own calculations based on data described in Table 1.

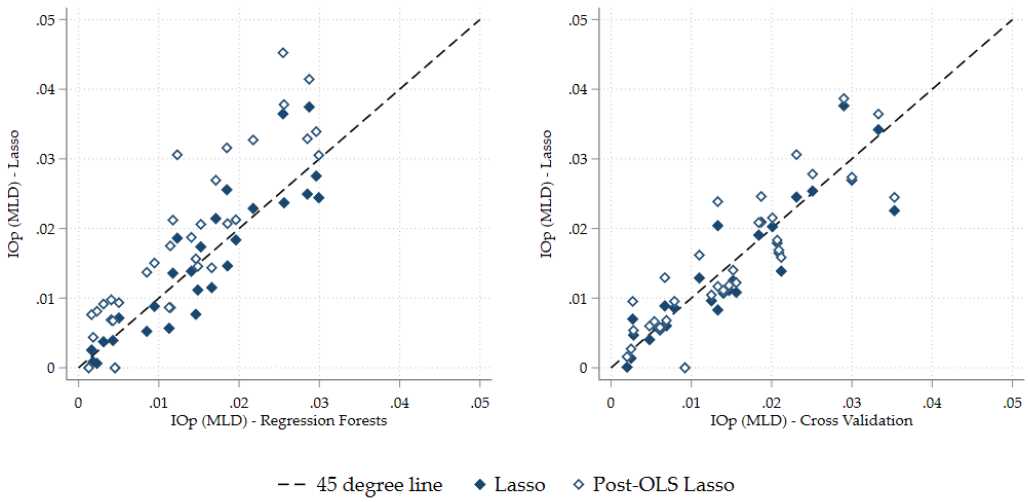
Table A.5: Year of Interest, Baseline and Harmonized

	Year Baseline	Year Harmonized	Difference in Years
Argentina	2015	2009	6
Chile	2009	2009	0
China	2014	2010	4
Ethiopia	2009	2009	0
Indonesia	2013	2006	7
Malawi	2010	2010	0
Mexico	2009	2009	0
Peru	2010	2009	1
Russia	2017	2009	8
South Africa	2017	2008	9
Tanzania	2010	2010	0
Thailand	2017	2009	8

Notes: The table shows the country-specific baseline year, the harmonized year, and their difference for the sensitivity check concerning the year of interest t. See section 4 for details.
Source: Own calculations based on data described in Table A.1.

B Additional Figures

Figure B.1: Methodological Validation



Notes: The figure shows how lower bound methodologies from the literature compare to the lasso estimation in this paper. The left panel compares the lasso procedure with the regression forest estimates from Brunori et al. (2018) while the right panel compares the lasso procedure with the cross-validated model selection approach in Brunori et al. (2019b). Filled diamonds refer to the standard lasso. White diamonds refer to post-OLS lasso.
Source: Own calculations based on EU-SILC, Brunori et al. (2018), Brunori et al. (2019b).

C Existing Studies

Table C.1: Overview Existing Studies

Country	Year	Study / Source	Data Source	Sample Size	Sample Restrictions	Circumstances	Outcome	Method	Index	Abs.	Rel. (%)
Argentina	2014	EqualChances.org	Encuesta Nacional sobre la Estructura Social	N.A.	Working age respondents	Education of Parents, Occupation of Parents, Origin (Race, Ethnic Origin, Area of Birth)	Equivalized Household Disposable Income	Parametric	Gini	0.159	41.99
Chile	2005	Ferreira et al. (2018)	Socio-Economic Database for Latin America and the Caribbean	N.A.	N.A.	Gender, Ethnicity, language, region of residence	Net Household Income	Parametric	MLD	0.460	5.30
Chile	2009	EqualChances.org	Encuesta de Caracterización Socioeconómica Nacional	N.A.	Working age respondents	Education of Parents, Occupation of Parents, Origin (Race, Ethnic Origin, Area of Birth)	Equivalized Household Disposable Income	Parametric	Gini	0.230	49.94
China	2010	Golley et al. (2019)	Survey of Women's Social Status in China	15,974	Age 24-65	Gender, Education of Father, Occupation of Father, Hukou status at birth, Region, Age	Indididual Labor Earnings	Parametric	MLD	0.140	25.00
China	2010-2014	Song and Zhou (2019)	China Family Panel Studies	5,892	Households with children at school or positive education expenditures	Gender, Education of Father/Mother, Hukou Status at 3 years old	Individual Income	Parametric	Theil	0.069	21.70
China	1989-2006	Zhang and Eriksson (2010)	China Health and Nutrition Survey	1,287	Age 20-50, parental background information through longitudinal matching available	Gender, Age, Birthplace, Education of Parents, Employment of Parents, Income of Parents	Individual Income	Parametric	Gini	N.A.	63.00
Ethiopia	2005	Ferreira et al. (2018)	Demographic and Health Survey	N.A.	N.A.	Region of birth, Religion, Mother tongue	Wealth Index	Parametric	Variance	6.160	7.70
Indonesia	2005	Ferreira et al. (2018)	Demographic and Health Survey	N.A.	N.A.	Religion	Wealth Index	Parametric	Variance	2.450	2.10
Malawi	2010-2011	Brunori et al. (2019a)	Third Integrated Household Survey	30,137	N.A.	Sex, Birthplace, Parental Education	Household Income	Tree	Gini	0.235	49.64
Mexico	2006	Wendelspiess Chávez Juárez (2015)	Mexican Social Mobility Survey	5,277	Age 25-64	Gender, Education of Father / Mother, Parents owned House, Socio-Economic Situation at age 14, Indigeneous Group	Individual Income	Parametric	Variance	N.A.	18.30
Mexico	2009	EqualChances.org	Mexican Family Life Survey	N.A.	Working age respondents	Education of Parents, Occupation of Parents, Origin (Race, Ethnic Origin, Area of Birth)	Equivalized Household Disposable Income	Parametric	Gini	0.135	23.80
Peru	2001	Ferreira and Gignoux (2011)	Encuesta Nacional de Hogares	13,621	Age 30-49, household heads or spouses	Education of Father/Mother, Ethnicity, Region of Birth, Gender	Household per Capita Income	Non-Parametric	MLD	0.163	29.30
Russia	2006-2016	Brock et al. (2016)	Life in Transition Survey	N.A.	N.A.	Gender, Birthplace Urbanity, Ethnicity, Education of Father/Mother, Membership of Communist Party	Self-Reported Income over the last twelve months	Parametric	Gini	0.130	34.50
South Africa	2008-2012	Piraino (2015)	National Income Dynamics Study	2,587	Age 20-44, only male respondents	Race, Education of Father	Individual Gross Income	Non-Parametric	MLD	N.A.	23.70
South Africa	2014	EqualChances.org	National Income Dynamics Study	N.A.	Working age respondents	Education of Parents, Occupation of Parents, Origin (Race, Ethnic Origin, Area of Birth)	Equivalized Household Disposable Income	Parametric	Gini	0.337	57.66
Tanzania	2010-2011	Brunori et al. (2019a)	World Bank, National Panel Survey	20,569	N.A.	Sex, Birthplace, Parental Education	Household Income	Tree	Gini	0.177	44.71

Notes: The table provides information on previous published IOp studies covering the countries in our sample. Information about methodological details and IOp estimates always refer to the preferred estimate in the respective study.