

Inequality of opportunity in sub-Saharan Africa

Paolo Brunori^a, Flaviana Palmisano^b and Vitorocco Peragine^c

^aUniversity of Florence & University of Bari, Florence, Italy; ^bSapienza University of Rome, Rome, Italy; ^cUniversity of Bari, Bari, Italy

ABSTRACT

Inequality of opportunity is defined as the difference in individuals' outcome systematically correlated with morally irrelevant pre-determined circumstances, such as ethnicity, socio-economic background, area of birth. This definition has been extensively studied by economists on the assumption that, in addition to being normatively undesirable, it can be related to low potentials for growth. However, empirical estimations of inequality of opportunity require accessing rich data sources, rarely available in poorer countries. In this paper, we exploit 13 consumption household surveys to evaluate inequality of opportunity in 10 Sub-Saharan African countries. According to our results, the portion of total inequality that can be attributed to exogenous circumstances is between 40% and 56% for the generality of countries. Our estimates are significantly higher than what has been found by previous studies. We detect a positive association between total consumption inequality and inequality of opportunity, and we study the different sources of unequal opportunities. The place of birth and the education of the father appear to exert the most relevant role in shaping inequality of opportunity in the region.

KEYWORDS

Consumption inequality; equality of opportunity; sub-Saharan Africa; circumstances

1. Introduction

Sub-Saharan Africa (SSA) countries are especially known for their high levels of economic inequality and poverty (see, for instance, Moradi and Baten 2005; Thorbecke 2013). However, the specific features of these inequalities remain largely understudied. Yet the understanding of the different sources of inequality is a necessary step toward the implementation of policies that may foster a sustained and 'shared' growth in these countries. There is, in fact, a rooted consensus on the argument that not all inequalities are the same: in particular, it has been convincingly argued that the degree of the inequality caused by differences at birth (such as gender, ethnicity, or parental background) or, more generally, by factors beyond the individual control may be related to low growth, more so than other effort-based inequalities (see World Bank 2006; Bradbury and Triest 2016; Marrero, Rodriguez, and Van der Weide 2016; Marrero and Rodríguez 2013). The idea is that when exogenous circumstances play a strong role in determining the individual outcome, there is a sub-optimal allocation of resources and a lower potential for growth. To put it differently,

the existence of inequality traps, which systematically exclude some groups of the population from participation in the economic activity, is harmful to growth because they discourage effort and investments by individuals, provoke a loss of productive potential, and contribute to social and institutional instability. The arguments above suggest that analysing the specific 'horizontal' dimensions of inequality is particularly important in both developing and underdeveloped countries.

One way to assess these kinds of inequalities is to endorse the Equality of Opportunity (EOp) approach (see Roemer 1998; Fleurbaey 2008), which provides a model to distinguish between that part of inequality caused by exogenous circumstances outside the individual responsibility, considered to be objectionable and therefore deserving a compensatory intervention, and the part of inequality generated by individual choices and effort, which is on the contrary considered to be unproblematic and not to be eliminated. The EOp theory has spurred a huge amount of theoretical and empirical works focussing on the measurement of inequality of opportunity (see the recent surveys by Ferreira and Peragine 2016; Ramos and Van de Gaer 2015; Roemer and

Trannoy 2016). However, most of the literature has been concerned with inequality of opportunity (IOp) in Western developed countries, with only a small set of studies dedicated to developing countries.¹ One reason for this is that measuring opportunity inequality is not an easy task: its informational requirements are quite high if compared to the standard measurement of income or consumption inequality.² Hence, as argued above, such analysis would be particularly needed in developing countries.

This paper is a contribution in this direction as it is the first attempt to evaluate inequality of opportunity in a large set of SSA countries. In particular, we contribute to the understanding of economic inequality in 10 SSA countries (i) by estimating a lower-bound for the portion of consumption inequality, which can be attributed to inequality of opportunities, (ii) by identifying the most disadvantaged groups of the population in each country, (iii) by evaluating the relative importance of different circumstances in describing inequality of opportunity. This analysis can help understanding the social and economic mechanisms that describe inequalities and can help identifying priorities in anti-poverty policies in different countries.

Our analysis is made possible through the availability of large-sample surveys built upon a common methodology and providing information on the socio-economic background of adult individuals. We use a set of 13 surveys that were implemented during a period ranging from 2000 to 2013 and covering the following countries: Comoros, Ghana, Guinea, Madagascar, Malawi, Niger, Nigeria (two waves), Rwanda, Tanzania (two waves) and Uganda (two waves).

Our estimates uncover a dramatic picture. Total consumption inequality is remarkable in all the samples considered, although quite variable across them: the Gini index ranges from 0.53 for Comoros to 0.31 for Niger, but in general the Gini index is around 0.4 in all countries considered. The entire region of SSA is confirmed as one of the most unequal regions in the world. As far as

inequality of opportunity is concerned, our estimates witness that the impact of exogenous circumstances is noticeable in every country, although this impact is quite variable across them: the portion of total inequality which can be attributed to (the observable) exogenous circumstances is between 40% and 56% for the generality of countries considered. This is a striking result, particularly if one considers that the computed measures are only lower bound estimates of the inequality of opportunity level in each country. We also look at the association between total consumption inequality and inequality of opportunity: although some re-rankings do exist, the data show a positive relationship between the two kinds of inequalities consistently with what the literature has found for Western countries. Although the analysis does not allow a causal interpretation of the result, we identify the relative ability of circumstances in predicting the outcome. Such analysis shows that the ‘sources’ of unequal opportunities also differ across countries. Comoros, Ghana, Guinea and Niger, for example, show a large impact of birthplace. Father education is instead the most important circumstance in Madagascar, Malawi, Rwanda and Tanzania. It is mother education for Nigeria and ethnicity in Uganda. By contrast, because our outcome of interest is household income, sex is the circumstance playing the weakest role in all countries considered.

Our results differ substantially from the only previous contribution that has focussed on inequality of opportunity in SSA³ Cogneau and Mesplé-Soms (2008) analysed five SSA countries (Ivory Coast, Ghana, Guinea, Madagascar and Uganda) by using data collected between 1985 and 1994. They use a very coarse set of circumstances (parental background) and, in fact, their results show a much lower level of inequality of opportunity: with some variation between countries, their estimates show that the portion of inequality attributed to exogenous circumstances is between 10% and 20%. Unlike Cogneau and Mesplé-Soms (2008), we extend the analysis to

¹In particular, only two contributions exist in the literature, namely Cogneau and Mesplé-Soms (2008) and Piraino (2015) that propose an analysis of inequality of opportunity for African countries.

²See on this Hassine (2011) and Singh (2011).

³See also Brunori, Palmisano, and Peragine (2018b) for a study of IOp and growth in Uganda.

a larger set of countries, and we consider a bigger set of circumstances for each country; moreover, we provide a more data-extended and methodologically intensive analysis.

The paper is organized as follows. Section 2 describes the measurement methodology. Section 3 describes the data and Section 4 presents the results of our analysis for the periods and countries considered. Section 5 provides a summary of the current findings and concludes with suggestions for further research on IOP in SSA countries.

II. Methodology

A model of equality of opportunity

The canonical model of EOp assumes that the outcome of an individual, y , is entirely determined by two classes of variables: circumstances and effort (see Roemer 1998; Van de Gaer 1993; Peragine 2002). For simplicity, we refer here to the individual outcome as ‘income’, but any other interpretation of outcome, such as consumption, is possible. Circumstances are denoted by c and belong to a finite set Ω : examples are gender, age, ethnicity, region of birth, or parental background. These are factors beyond an individual’s control but nonetheless exogenously affect income. Effort is denoted by e and belongs to the set Θ , and it may be treated either as a continuous or a discrete variable. This is a factor that endogenously affects the individual income since it is the result of one’s own choices. The different forms of luck that may affect the individual income can be classified either as circumstances or as responsibility characteristics.⁴

Individual income can then be expressed as follows:

$$y = g(c, e) \quad (1)$$

The production function $g : \Omega \times \Theta \rightarrow \mathbb{R}_+$ is assumed to be monotonic in e , while circumstances and effort are assumed to be orthogonal.⁵ This is a reduced form model in which neither the

opportunities themselves nor the individual decision process to exert a given level of effort are explicitly modelled. The model builds on the argument that (non-observable) individual opportunities can be inferred by observing joint distributions of circumstances, effort and income, which fully characterize a population of individuals. For simplicity, let us treat effort, as well as each element of the vector of circumstances, as discrete variables. This would allow the population to be partitioned in two ways: into types in which all individuals share the same circumstances and into tranches in which everyone shares the same degree of effort.

Roughly speaking, the source of unfairness in this model is given by the effect that circumstance variables have on individual outcomes. To measure this effect a two-step procedure can be adopted: first, the actual distribution is transformed into a counterfactual distribution that reflects only and fully the unfair inequality, while all the fair inequality is removed. In the second step, a measure of inequality is applied to this counterfactual distribution. The construction of the counterfactual distribution should reflect two distinct and independent principles: the reward principle, which is concerned with the apportion of outcome to effort and, in some of its formulations, requires to respect the outcome inequalities due to effort; the compensation principle, according to which all outcome inequalities due to exogenous circumstances are unfair and should be compensated for by society. In particular, in the analysis developed in this paper, we adopt the ex-ante approach of the compensation principle, according to which there is equality of opportunity if the set of opportunities is the same for all individuals, regardless of their circumstances. Hence in the ex-ante version, the compensation principle requires reducing the inequality between these opportunity sets. In the model introduced above, the income distribution of a given type is interpreted as the opportunity set of all individuals with the same set of circumstances. Hence, the

⁴Although in Sub-Saharan African countries individual outcomes are particularly sensitive to random shocks (climate conditions, conflicts, international prices of raw materials, etc.), in this paper we do not focus specifically on how luck affects different individuals. From a normative point of view, this means that we consider the effect of luck to be unproblematic as long as it is orthogonal to circumstances. See Lefranc and Trannoy (2017) and Villar (2017) for alternative proposals to measure the influence of luck.

⁵This assumption is motivated by the theoretical argument that it would be hardly sustainable to hold people responsible for the factor e in a situation in which it was dependent on exogenous characteristics.

focus is on the inequality between-types: the counterfactual distribution should eliminate the inequality within the types (reward) and reflect the inequality between the types (ex-ante compensation). Let us underline here a dual interpretation of the types in the EOp model: on one hand, the type is a component of a model that, starting from a multivariate distribution of income and circumstances, allows us to obtain a distribution of (the value of) opportunity sets enjoyed by each individual in the population. On the other hand, given the nature of the circumstances typically observed and used in empirical applications, the partition into types may be of interest *per se*: they can often identify well-defined socio-economic groups, possibly deserving special attention by policymakers.⁶

In addition to normative considerations, the choice of which methodology to adopt reflects data availability. The database we use in this paper does not contain a satisfactory measure of effort. Therefore, we focus on the ex-ante approach and we use the between-types inequality measure, which was proposed, among others, by Peragine (2002), Checchi and Peragine (2010), Ferreira and Gignoux (2011), Brunori, Peragine, and Serlenga (2018c). It relies on a counterfactual distribution, Y_s , which is obtained by replacing each individual's income by the average income of the type an individual belongs to, independently of the level of effort exerted.⁷

Estimation

The smoothing transformation, intended to remove all inequality within types, can be performed by using either a non-parametric or a parametric method. The non-parametric approach directly implements Roemer's theory: it identifies types as groups of individuals with identical circumstances beyond individual control (e.g. same parental background, same area of birth, same ethnicity). This method is frequently used in empirical contributions, but its implementation is severely constrained by the availability of data.

The larger the number of observable circumstances, the finer becomes the partition into types. With a granular partition of the population, the size of each type decreases, bringing about a decline in the precision of the estimates of the type mean, consequently giving rise to an upward bias in the estimation of IOp (Brunori, Peragine, and Serlenga 2018c).

The parametric approach estimates a Mincerian equation to explain outcome as a log-linear combination of circumstances with an ordinary least square regression (Ferreira and Gignoux 2011). The use of a parametric approach imposes a precise functional form linking circumstances and outcome and making the model more parsimonious in terms of degree of freedom. The cost of using an OLS estimation is an arbitrary assumption about the data generating process. In practice, the set of circumstances used in empirical analyses typically includes parental education, parental occupation, area of birth and ethnicity. The categories of such variables do not have cardinal meaning, and, to make the model operational, each one needs to be transformed into a dummy variable. If no cardinal circumstance is observable, the OLS estimation brings to the estimation of a shift in the regression intercept associated with each category of every circumstance, for instance, having white-collar parents or being a first-generation immigrant. This implies a severe restriction in the construction of the counterfactual distribution because it imposes a fixed effect for each circumstance. For example, it could be the case that being a first-generation immigrant has a completely different meaning depending on whether one's parents are university professors or construction workers. In a parametric approach, this effect is defined to be the same. To take into account the interaction between circumstances, one needs to interact dummies. However, once all dummies are interacted, one intercept is estimated for each type, and our OLS estimate becomes equivalent to the non-parametric approach. Thus, on one side, the motivations for

⁶Alternatively, the literature has developed the ex-post approach, according to which there is equality of opportunity if and only if all those who exert the same effort end up with the same outcome. The compensation principle, in the ex-post version, is thus defined with respect to individuals with the same effort but different outcomes. Opportunity inequality within this approach is measured as inequality within the tranches.

⁷The use of the average income of the type for the smoothing transformation is justified, from a normative point of view, in light of the utilitarian reward principle, according to which society should express full neutrality with respect to inequalities due to the effort.

the use of a parametric approach appear to be clear if cardinal measures of circumstances are available (such as parental income). They are less convincing if all circumstances can only be modelled through dummies.⁸

In this paper, we follow the non-parametric approach proposed by Brunori, Hufe, and Mahler (2018a). This approach obtains the partition in types by estimating a conditional inference regression tree. Introduced by Morgan and Sonquist (1963) and popularized by Breiman et al. (1984), regression trees are algorithms that aim at predicting out of sample an outcome based on a number of covariates. This is done partitioning the space of the regressors in non-overlapping regions, called terminal nodes or leaves.

An example is reported in Figures 1 and 2. That tree partitions a two-dimensional space (circumstance 1 and circumstance 2) to predict a binary outcome (black or white).⁹ In Figure 2 the two splitting points (circumstance 1 at 2.5 and circumstance 2 at 15), and the three terminal nodes are easily identified.

Once the tree is estimated, one can use the average outcome in each terminal node to predict the individual outcome. A very deep tree, in which each terminal node contains a few observations, will very closely fit the data in the sample. However, such a tree would be useless for the

purpose of predicting out of sample, it is clearly an overfitted model that would produce very different prediction if it was estimated on a different sample, that is, it is affected by a very large sampling variance.

There are many ways to construct trees for out of sample prediction and avoid overfitting. Conditional inference trees use a sequence of statistical tests as a rule to grow the tree. The algorithm proposed by Hothorn et al. (2006) proceeds as follows:

- (1) set a confidence level $(1 - \alpha)$, typically 0.99;
- (2) test the null hypothesis of independence between any of the covariates (circumstances) and the response (outcome). Stop if this hypothesis cannot be rejected ($p - value \geq \alpha$). Otherwise, select the circumstance with the strongest association to the outcome (lowest $p - value$);
- (3) for every possible ways of dividing the selected circumstance into two regions, that is for every possible value that could be used as splitting point, test the difference between the average outcome of two resulting regions. Select the splitting point associated with the smallest probability that, under the assumption that the null hypothesis is true, the sample mean difference

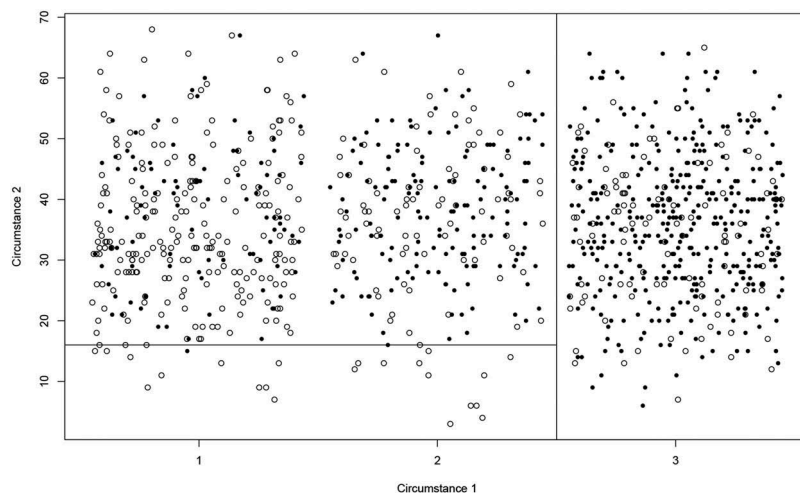


Figure 1. Classification tree: a binary outcome and two controls.

Note: Black dots indicate a failure, white dots indicate a positive outcome.

⁸See (Bourguignon, Ferreira, and Menéndez 2007; Ferreira and Gignoux 2011; Ferreira, Gignoux, and Aran 2011) for more details on the parametric approach.

⁹Note that regression trees used to predict a binary outcome are generally called 'classification trees'.

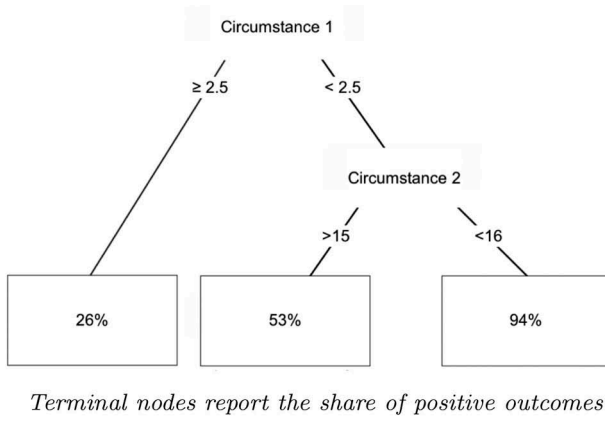


Figure 2. Classification tree.

between the two resulting subgroups would be greater than or equal to the observed (lowest $p - value$);

(4) recursively repeat steps 2 and 3.

The use of conditional inference trees to obtain the partition of types has a number of advantages. First, the choice of which circumstance to use is no longer arbitrary. One can select a very large set of circumstances and the algorithm will use only the covariates that have greater explanatory power. Second, the model specification is no longer arbitrary: how circumstances interact in determining the outcome is driven by the attempt of the algorithm to explain outcome's variability. Third, but not less important, opportunity trees tell us a story about the structure of opportunity which is immediate to understand even without formal statistical training.

Regression trees are not immune from limitations. First, they can be misleading when two or more covariates are highly correlated. In fact, when one of the two correlated covariates has been used to determine a split, it is unlikely that the second will play any role in the tree, although it may be nearly as correlated as the first one with the dependent variable. Second, they assume non-linearity in the data generating process.

We accommodate both issues by complementing our estimates obtained with trees with estimates obtained with conditional inference forests. Loosely speaking a random forest is a collection of hundreds of trees. Each tree is estimated using a subsample of the original observations and, at each splitting point, a subsample of the original controls.

With forests, the problem of correlated regressors disappears since selecting a subsample of controls, when the number of trees is sufficiently large, all controls are allowed to play a role in determining the structure of some trees. In addition, obtaining predictions by averaging across hundreds of trees has the effect of weakening the non-linearity assumption that holds for every single tree.

Finally, excluding some of the variables from the tree provides an immediate method to evaluate relative variable importance. This is typically quantified as the drop in explained variability due to the exclusion of each control.

In what follows, we will use conditional inference trees to describe the partition in types suggested by the data. We will then estimate the counterfactual distribution both directly using trees and by averaging individuals' predictions using conditional inference forests.

IOp index

Inequality of opportunity is measured by applying an inequality index (I) to the counterfactual distribution estimated using the above methodology (Y_s). Doing so one can obtain an absolute IOp index and a relative IOp index, respectively, denoted as follows:

$$IOp^{abs} = I(Y_s) \text{ and } IOp^{rel} = \frac{I(Y_s)}{I(Y)} \quad (2)$$

In most empirical analyses, the measure of overall inequality and consequently inequality of opportunity used is the Mean Logarithmic Deviation (MLD). In this paper, the focus is on the Gini coefficient, as we explain below.

There are three arguments in favour of the Gini coefficient and against the use of the MLD. The first argument is based on the evidence that the MLD is very sensitive to extreme values, much more than the Gini coefficient. As explained before, IOp is measured by applying an index of inequality to a smoothed distribution that eliminates all the inequality due to effort. This smoothed distribution, by definition, does not contain extreme values, as they are removed by the smoothing process. Now, the high sensitivity

of the MLD to extreme values implies that the reduction of inequality by going from the original to the smoothed distribution will be much higher for the MLD than for the Gini coefficient. That is to say – *ceteris paribus* – *rel-IOp* as measured by the MLD will be much lower than *rel-IOp* as measured by the Gini coefficient. Symmetrically, the MLD is insensitive to small levels of inequality that typically characterize between-group inequality with sufficiently large groups. Using MLD we would, therefore, obtain estimates levelled toward zero. This would limit our ability to appreciate the between-country difference in *IOp*. In order to show the importance of the difference between the two measures, consider Figure 3 plotting the MLD (red-dashed line) and the Gini coefficient (black solid line) for 1000 log normal distributions with mean zero and different levels of inequality. The Gini coefficient increases linearly with the increase in the standard deviation, whereas the convexity of the MLD curve reveals a low sensitivity of this index with respect to low levels of inequality and high sensitivity with respect to extreme values. The figure also helps understand why empirical estimates systematically find higher *IOp* in terms of Gini coefficient than in terms of MLD. When the standard deviation is above 1, the Gini coefficient is smaller than the MLD. However, a value of the standard deviation above 1 for a log-normal distribution indicates a level of inequality that corresponds to a Gini coefficient above 0.50. This inequality is, therefore, higher than what is on average found empirically for

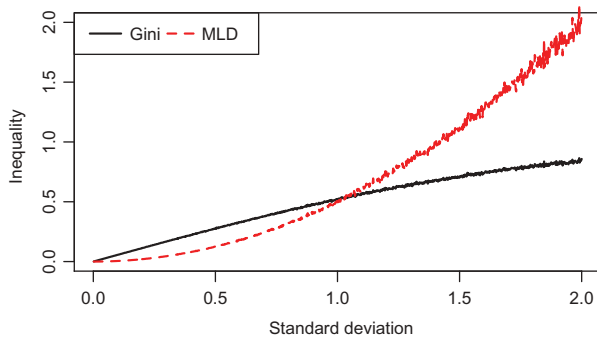


Figure 3. Sensitivity to inequality: Gini vs MLD.

Source: Simulation log-normal distributions of 10,000 observations.

distributions of income or consumption, which are by construction weakly higher than *IOp*.

The second argument is based on the observation that the use of the MLD is generally motivated by its decomposability property, because it is perfectly decomposable in between- and within-types inequality. However, one can question if exact decomposability of inequality in a between and a within component is really a desirable property in the light of the fact that, for instance, the Lorenz partial ordering does not decompose exactly in a between and within component. Furthermore, this result relies on a restricted definition of the concept of decomposability, that is aggregativity,¹⁰ a powerful practical concept but lacking strong normative support (see Bourguignon 1979). Last, differently from the Gini coefficient whose values belong to a well-defined interval [0,1], the MLD is not bounded above, making the interpretation of the results less straightforward.

For these reasons, in this paper, we follow Aaberge, Mogstad, and Peragine (2011) and use the Gini index, which has well-known desirable characteristics, although it is not perfectly decomposable in between- and within-types inequality whenever the type income distributions overlap.¹¹

Therefore, in general:

$$\text{Gini}(Y) = \text{Gini}(Y_{\text{within}}) + \text{Gini}(Y_s) + K \quad (3)$$

K is a residual greater than zero when there is overlapping between the types' distributions. K measures the part of inequality that is jointly determined by effort and circumstances, but that cannot be disentangled into effect of effort and effect of circumstances.

We will focus on the two *IOp* indexes:

$$\begin{aligned} IOp^{abs} &= \text{Gini}(Y_s) \text{ and } IOp^{rel} \\ &= \frac{\text{Gini}(Y_s)}{\text{Gini}(Y_{\text{within}}) + \text{Gini}(Y_s) + K} \end{aligned} \quad (4)$$

Note that as far as the relative *IOp* index is concerned, K is part of the denominator but not part of the numerator. This makes IOp^{rel} more conservative than if the normalization was obtained with a perfectly decomposable inequality measure.

¹⁰This property implies that there is no need to know the exact distribution of incomes within the subgroups of a population to compute the inequality measure of that population, only the inequality measures of the subgroups and their aggregate characteristics are necessary.

¹¹See Lambert and Aronson (1993) for an insightful discussion on the Gini decomposition.

Relative circumstance importance

Finally, each circumstance may play a different role in the determination of IOp. However, if one or more relevant circumstances are not observable, and if we cannot exclude that they are correlated with observable circumstances, an exact causal identification of the relative role of each circumstance is impossible. However, a description of the relative role of observable circumstances may be of some interest (Ferreira and Gignoux 2011).

Coherently with our methodological framework, this analysis can be performed by means of random forests, that are obtained drawing several bootstrap samples from the original data and fitting a regression tree to each sample. At each split, a random subsample of the set of predictors to select from is used. This allows covariates – that would have not played a role in a single tree based on all circumstances – to enter the prediction function. Even though their selection can be inefficient for the prediction purpose of a single tree, their use can reveal informative interactions with other covariates that may turn out useful to improve the predictive performance of the forest (Strobl et al. 2008).

Using random forests, it is possible to quantify the relative importance of each circumstance by permutation, that is, by randomly permuting the values of a given circumstance across observations in the sample. Doing so, the original association of the permuted circumstance with the outcome is broken. This implies that, when the tree obtained after permutation is used to predict observation not included in the bootstrap sample (out-of-bag), the predictive accuracy decreases. The difference in prediction accuracy before and after the permutation, averaged over all trees, is a measure of the importance of each variable that we interpret as a measure of the relative importance of each circumstance in determining the overall outcome.

III. Data

Our analysis is based on the following surveys:

- *Enquête Intégrale auprès des Ménages* (EIM) for Comoros (year 2004), carried out by the

Statistical Office of the Ministry of Land Planning and Settlement;

- *Ghana Living Standards Survey* (GLSS) for Ghana (year 2012-2013), carried out by the Ghana Statistical Service – National Data Archive (GSS);

- *Enquête Intégrée de Base pour l'Évaluation de la Pauvreté* (EIBEP) for Guinea (year 2002-2003), carried out by the National Directorate of Statistics (Ministry of Economics and Finance)

- *Enquête Périodique auprès des Ménages* (EPM) for Madagascar (year 2005), carried out by the National Institute of Statistics (INSTAT);

- *Third Integrated Household Survey* (IHS3) for Malawi (year 2010-2011), carried out by the National Statistical Office of Malawi;

- *National Survey on Household Living Conditions and Agriculture* (ECVM) for Niger (year 2011), carried out by the National Institute of Statistics of Niger;

- *General Household Survey* (GHS) for Nigeria (years 2010-2011 and 2012-2013), carried out by the National Bureau of Statistics of Nigeria;

- *Enquête Intégrale sur les Conditions de Vie des Ménages* (EICV) for Rwanda (year 2000), carried out by National Institute of Statistics of Rwanda (NISR);

- *National Panel Survey* (NPS) for Tanzania (years 2008-2009 and 2010-2011), carried out by the National Bureau of Statistics of Tanzania;

- *Uganda National Panel Survey* (UNPS) for Uganda (years 2009-2010 and 2010-2011), carried out by the Uganda Bureau of Statistics.

All surveys are listed in [Table 1](#) with the year they refer to, their original sample size, and a link to the documentation. All surveys are representative at a national level and cover both urban and rural areas. Our analysis is based on a sub-sample of the original data obtained by considering only individuals aged 15 years or more for whom information about circumstances beyond individual control are available.¹² The outcome considered is per capita consumption, which encompasses consumption for both food and non-food goods, that is, we assume a proportional intra-household distribution of consumption and zero economies of scales in consumption. Although we use different surveys, the results

¹²All individuals with missing information on the circumstances are dropped from the analysis (see [Appendix I](#)).

Table 1. Data sources.

Country	Survey	Year	Sample size	Documentation
Comoros	EIM	2004	18,373	IHSN
Ghana	GLSS	2012–2013	39,826	GSS
Guinea	EIBEP	2002–2003	25,319	INSG
Madagascar	EPM	2005	30,271	INSTAT
Malawi	IHS3	2010–2011	30,137	World Bank
Niger	ECVM	2011	12,118	World Bank
Nigeria	GHS	2010–2011	14,916	World Bank
Nigeria	GHS	2012–2013	14,560	World Bank
Rwanda	EICV	2000	17,69	INSR
Tanzania	NPS	2008–2009	9175	World Bank
Tanzania	NPS	2010–2011	11,394	World Bank
Uganda	UNPS	2009–2010	8268	World Bank
Uganda	UNPS	2010–2011	7509	World Bank

are comparable across countries since the consumption variable has been adjusted for inflation and translated into 2011 purchasing power parity (PPP) international dollars (World Bank 2015).

A fundamental step in the measurement of inequality of opportunity is the identification of the vector of circumstances. This is a normative choice, subject to the constraint of data availability. Our data contain information on a small set of basic circumstances, but nonetheless of prominent importance. For each country, in fact, we can observe a subset of the following: sex, birthplace, parental education and occupation, ethnicity. (see Table 2 for details).

As for the specific circumstances, there is a broad consensus regarding the magnitude of gender inequality and its effects on the economic development of a country (see, among others, Nordman, Robilliard, and Roubaud 2011; Berik, Rodgers, and Seguno 2009). This is particularly salient for SSA where notwithstanding efforts to reduce gender differences in well-being, inequality is still prevalent in numerous domains. However, because in this analysis the outcome considered is household consumption, our estimates will systematically underestimate the role of gender in shaping opportunities.

Table 2. Circumstances observed by country.

Country	Circumstances				
	Sex	Birthplace	Parental education	Parental occupation	Ethnicity
Comoros	✓	✓	✓	✓	
Ghana	✓	✓	✓		✓
Guinea	✓	✓	✓	✓	
Madagascar	✓	✓	✓		✓
Malawi	✓	✓	✓		
Niger	✓	✓			✓
Nigeria	✓		✓	✓	
Rwanda	✓	✓	✓	✓	
Tanzania	✓	✓	✓		
Uganda	✓	✓			✓

Parental education and occupation are widely used in the empirical literature on IOp that has dealt with richer countries. A vast amount of evidence has been produced on the effect of socio-economic background on children's outcomes during adulthood. This literature is however traditionally Western-centric and has rarely concentrated on SSA countries. Nevertheless, there is also evidence supporting the argument that parental education and occupation act as circumstances on individual outcome in the specific SSA context. For instance, it has been shown that, in these countries, the nutritional status of a child is strongly correlated to parental occupation with obvious, although indirect, consequences on his outcome in the future (Madise, Matthews, and Margetts 1999). Parental education, instead, has been shown to be an important factor in determining whether or not a child is currently attending school; whereas, school improvements in parental education have been shown to increase the schooling of children, which, in addition to improving their health and reducing the status of extreme poverty, has direct effects on the outcome prospects of these children (see, among others, Lassibille and Tan 2005; Schultz 2004).

The Sub-Saharan region contains many of the most ethnically diverse countries on Earth. Not surprisingly, ethnicity and birthplace are variables of paramount importance in SSA, a region where ethnic divisions play the most important role in how the governance of society works. Kimenyi (2006) explains how the most common form of corruption entails the distribution of rewards, jobs, contracts and promotions, on the basis of ethnicity. Even today, SSA countries face impressive challenges to peace and stability and have fallen prey to continuous armed ethnic conflicts, which arrest or even reverse the growth and development process of this specific part of the African continent. Between 1946 and 2002, not less than 1.37 million battle-related deaths occurred in 47 civil wars in SSA (Lacina and Gleditsch 2005). In 2011, for instance, SSA had 91 instances of this type of conflicts, compared to the 89 of 2010 (see Brautigam and Knack 2014; De Ree and Nillesen 2009). Moreover, previous studies have shown that high levels of ethnic diversity are strongly linked to high informal market premiums, poor financial development, low provision of infrastructure and low levels of education. Ethnicity

has a strong influence on inequality in Africa where ethnic fractionalization has given rise to a political economy of unequal subsidies and discrimination (Easterly and Levine 1997; Milanovic 2003). The area is also characterized by regional disparities in access to opportunities. Hence, it appears natural to treat ethnicity and birthplace as circumstances in the context of our analysis.

It is important to note that cross-country comparisons of IOP must be interpreted while bearing in mind the heterogeneity of available information. In particular, the subset of circumstances used may vary across countries, as different surveys usually collect different information on circumstances. The higher is the number of observable circumstances the larger is the probability that the algorithm detects significant splitting points. Moreover, the sample size varies across countries. *Ceteris paribus*, larger sample sizes will also tend to produce deeper trees. And a deeper tree tends to produce larger between-group inequality. In Appendix III we discuss this issue, and we show how IOP index can be corrected to take into account heterogeneity in the number of terminal nodes.

Opportunity trees

All partition in types are reported in Appendix II. We comment here in detail, for a didactic

purpose, only the Nigerian opportunity tree for 2012–2013 reported in Figure 4, and the tree of Comoros and Guinea.

In Nigeria, we have five observable circumstances: sex, mother and father education, mother and father occupation by industry. Based on this information we can partition the population into nine types. The first splitting point, based on father education, determines two subtrees: to the right, we have the two really advantaged types made of individuals whose father had a high level of education. Independently of the other circumstances highly educated father is reported by a minority of respondents (the population shares displayed in the terminal node are 1.7% and 2.2%, respectively). Both types have an expected outcome largely above the national average: 60% and 229% above. The other subtree contains poorer types. Among respondents with a father with low level of education, the second most statistically relevant circumstance is father occupation. With the majority of respondents reporting a father working in agriculture and fishing, mining, electricity and utilities, services, or had no work. This

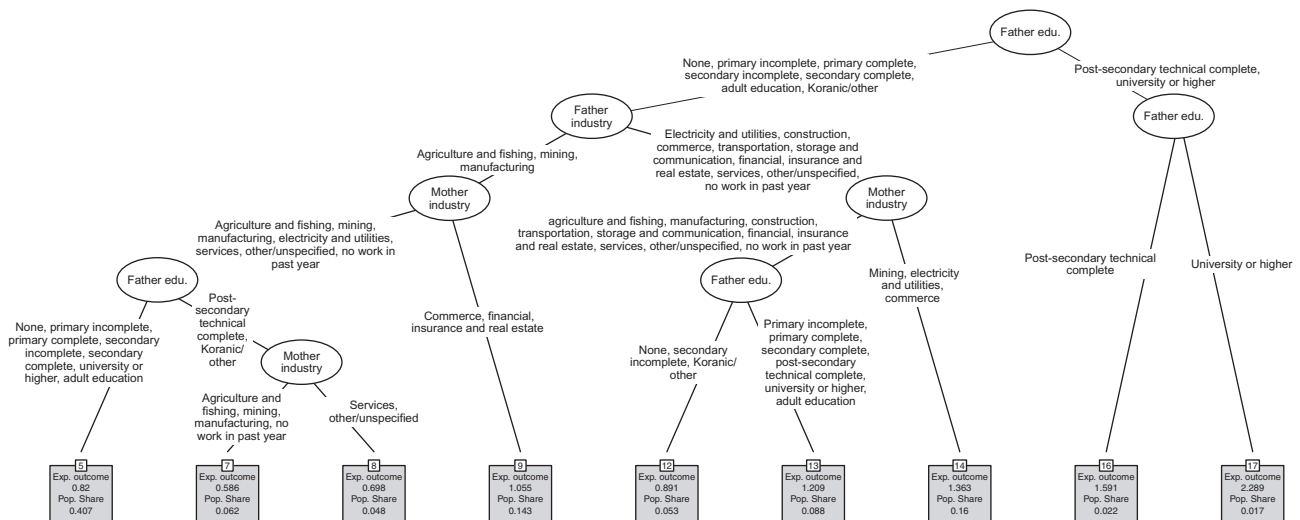


Figure 4. Opportunity structure: Nigeria 2012. Source:

Authors' calculation based on surveys listed in Table 1

type (number 5) contains 40% of the sample and have an expected outcome 18% lower than the national average. However, the worst-off type is type number 7 with an expected outcome 42% lower than the average. The structure of the tree shows how circumstances' interaction correlates with per capita disposable consumption. Type 14, for example, is made of individuals reporting a poorly educated father but both father and mother not working in agriculture and fishing, then the two sectors characterized by lowest productivity and salaries. Parental industry seems to more than counterbalance the low level of education of the father: the expected outcome for type 14 is 36% higher than the national average.

Figures A7–A18 in Appendix II contain the partition in types obtained estimating conditional inference trees for each country using the entire sample of adult respondents with non-missing information about circumstances beyond individual control. For instance, looking at the opportunity profile reported in Figure A7, we find that the population in Comoros can be partitioned into four types, and this partition is mostly determined by birth location and father occupation, although for this country we also observe sex and parental occupation.¹³ The most advantaged type (denoted with number 7) is composed of people born in countries other than African countries or in the cities of Domoni, Sima, Dimani, Hambou, Hamanvou, independently of the values of the other circumstances (about the 20% of the population). For this type, the value of the opportunity set (proxied here by the mean per capita consumption of the individuals in this type) is about 64% higher than the mean per capita consumption in the population. In the definition of the other types, father education enters into play. It should be highlighted here that what matters are the distinction between individuals with father that has at least primary education and individuals with father that has below primary education. This is a feature that characterizes the opportunity profiles of all the other countries as well. In Comoros, then, the expected outcome of the most disadvantaged type (denoted with number 4) is about 30%

below that of the expected outcome in the population.

As an alternative example, consider the opportunity profile of Guinea reported in Figure A9. As opposed to the previous example, in the case of Guinea, all circumstances are relevant for the partitioning of the populations into types: they are 19. In this opportunity tree, birth location appears in seven nodes, implying that this is an important circumstance. In this country, the most advantaged type (denoted with number 37) enjoys an expected outcome that is about 3.46 times the average outcome in the population. It is worth noticing that father education is fundamental in determining such economic advantage. In fact, consider type 36, which differs with respect to the best type only in the education of father, its expected outcome is less than one-third of the expected outcome of the best type. Last, the worst type in Guinea (denoted with number 32) is benefiting of an expected outcome that is about 40% below the average outcome in the population.¹⁴

IV. Inequality of opportunity

An immediate way of estimating IOp in the samples consists of directly measuring the between-type Gini considering the expected outcome and population share of terminal nodes. This approach produces the estimates for IOp^{abs} and IOp^{rel} reported in the fifth and sixth column of Table 3. The seventh column reports the IOp^{rel} after it has been corrected in order to take into account the number of types. This correction is called ELMO because the correction method was proposed by Elbers, Lanjouw, Mistiaen and Ozler in 2008 and is fully discussed in Appendix III. Finally, the last two columns report estimates obtained with conditional inference forests that are our preferred estimates, as discussed above. All inequality estimates are computed by using the Gini coefficient.

Total inequality is remarkable in all countries, although quite variable across them: the Gini coefficient ranges from 0.53 for Comoros to 0.31 for Niger, but in general, the Gini is around 0.40. The entire region of SSA is confirmed as one of the

¹³One reason for the exclusion of parental occupation could be the high number of missing values as reported in Table A5 in Appendix I.

¹⁴For the sake of brevity, we only describe in the main text these few cases.

Table 3. Total consumption inequality and inequality of opportunity estimates.

Country	Mean per capita consumpt.	Types	Inequality	IOp^{obs} tree	IOp^{rel} tree	IOp^{rel} tree (ELMO)	IOp^{obs} forest	IOp^{rel} forest
Comoros 2004	2936.88 (2755.37–3126.42)	4	0.5302 (0.5101–0.5500)	0.1753 (0.1308–0.2323)	0.3306	0.3807	0.2468	0.4654
Ghana 2012–2013	1795.58 (1767.32–1819.67)	24	0.4156 (0.4114–0.4198)	0.2070 (0.2140–0.2276)	0.4981	0.5011	0.2341	0.5634
Guinea 2002–2003	1024.04 (982.14–1073.22)	19	0.4260 (0.4094–0.4430)	0.1680 (0.1431–0.2300)	0.3944	0.4500	0.1714	0.4024
Madagascar 25	411.60 (401.22–418.19)	17	0.3710 (0.3623–0.3802)	0.1535 (0.1506–0.1736)	0.4136	0.4224	0.1690	0.4556
Malawi 2010–2011	853.50 (835.41–873.32)	17	0.4734 (0.4671–0.4806)	0.2350 (0.2373–0.2595)	0.4964	0.5018	0.2378	0.5023
Niger 2011	1070.69 (1052.12–1099.29)	20	0.3128 (0.3076–0.3185)	0.1086 (0.0961–0.1161)	0.3471	0.3521	0.1263	0.4036
Nigeria 2010–2011	1296.54 (1270.76–1326.23)	17	0.3890 (0.3842–0.3955)	0.1669 (0.1643–0.1833)	0.4290	0.4385	0.1694	0.4355
Nigeria 2012–2013	1603.90 (1562.14–1642.39)	9	0.3893 (0.3780–0.4003)	0.1524 (0.1434–0.1804)	0.3916	0.4114	0.1774	0.4557
Rwanda 2000	633.57 (616.09–649.42)	11	0.4598 (0.4479–0.4696)	0.2100 (0.1938–0.2234)	0.4351	0.4477	0.2112	0.4594
Tanzania 2008–2009	1127.70 (1105.22–1157.98)	13	0.3917 (0.3843–0.3996)	0.1708 (0.1691–0.1895)	0.4361	0.4423	0.1944	0.4963
Tanzania 2010–2011	1107.97 (1083.94–1139.23)	19	0.3955 (0.3896–0.3907)	0.1768 (0.1670–0.1986)	0.4471	0.4517	0.1993	0.5040
Uganda 2009–2010	1142.95 (1094.78–1188.59)	12	0.4524 (0.4409–0.4660)	0.2158 (0.2105–0.2481)	0.4770	0.4992	0.2201	0.4865
Uganda 2010–2011	1040.03 (986.65–1085.19)	10	0.4743 (0.4557–0.4920)	0.2415 (0.2192–0.2781)	0.5091	0.5333	0.2454	0.5174

most unequal regions in the world. For the three countries for which observations for more than 1 year are available (Nigeria, Tanzania and Uganda) the results bear witness to an increase in inequality; hence, the recent dynamics, where available, show a regressive pattern.

As far as the inequality of opportunity is concerned, we focus on inequality of opportunity measured using random forest. Note, however, that as shown in Figure 5, with the exception of Comoros the two measures – IOp computed using the trees technique and IOp computed using the forests technique – are strongly correlated. Forests are obtained with 200 conditional inference trees. Because the number of observable circumstances is low, for each tree, $k - 1$ circumstances are used at each splitting point (where k is the number of observable circumstances). Each tree is based on a subsample of observations of half the original sample size. The share of inequality that can be attributed to different exogenous factors is extremely high and variable across all countries: it ranges between 40% for Guinea and 56% for Ghana and is more generally between 40% and 50% for the other SSA countries. In other words, according to the observed circumstances, almost half of the observed inequalities in consumption can be attributed to exogenous factors, that is, to inequality of opportunity. This is a striking result, particularly if one considers that the computed measures are only

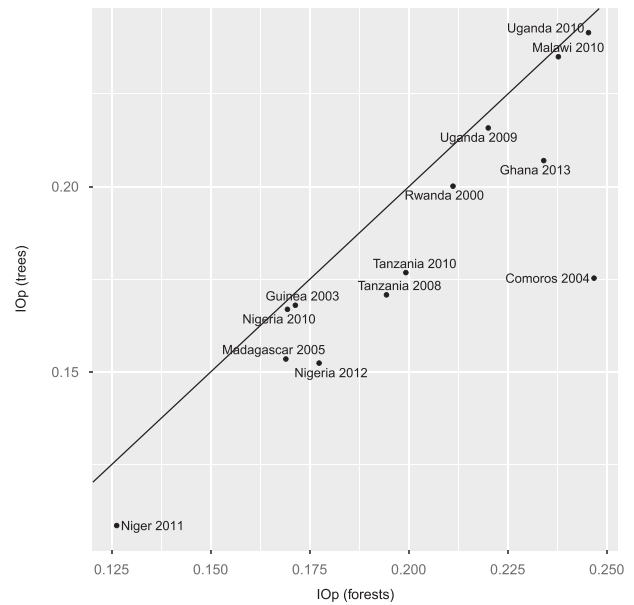


Figure 5.: Inequality of opportunity: trees vs. forests.

Source: Authors' calculation based on surveys listed in Table 1

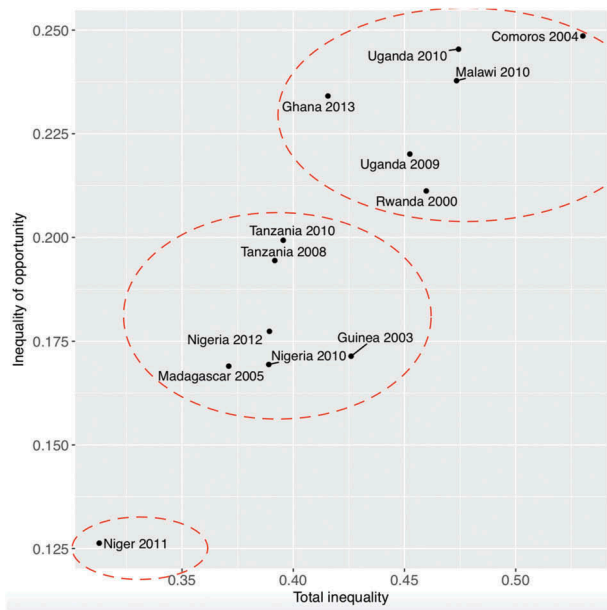


Figure 6. Total inequality and inequality of opportunity.

Source: Authors' calculation based on surveys listed in [Table 1](#)

lower bound estimates of the inequality of opportunity level obtained with a rather limited number of observable circumstances. Note also that, for the three countries for which we have two observations in time, the trend differs: in Nigeria the share has increased from 43% in 2010–2011 to 45% in 2012–2013, in Tanzania it remained stable from 2008–2009 and 2010–2011, in Uganda it sharply increased from 48% to 51% in only 1 year.

It is interesting to look at the association between total consumption inequality and (absolute) opportunity inequality computed using random forests as depicted in [Figure 6](#). This figure could be interpreted as a generalization of the so-called 'Great Gatsby' curve ([Corak 2013](#)), showing a negative relationship between income inequality and social mobility.¹⁵ Our results show that countries with higher consumption inequality are also characterized by a higher level (portion) of inequality of opportunity, although there is also some re-ranking between countries. Notable here are the cases of Niger and Comoros: the former having lowest values of both total inequality and inequality of opportunity, much lower than the other countries; the latter performing worst

according to both total inequality and inequality of opportunity but with values that are similar to other countries (namely, Uganda and Malawi). Interestingly, Comoros is also the country that, according to the sample used in this paper, has the highest level of average per capita consumption, that is about 2937 PPP \$. Whereas, Niger is ranked in the bottom part of the cross-country consumption distribution with an average per capita consumption of about 1071 PPP \$.

Overall, we observe three clusters. The first cluster is made of countries characterized by high level of overall inequality and high level of IOp (Comoros, Ghana, Malawi, Rwanda, Uganda). The second is composed of countries with (relatively) moderate levels of overall inequality and IOp (Guinea, Madagascar, Nigeria, Tanzania). The last cluster encompasses only one country with low levels for both kinds of inequality (Niger, as discussed above).

All countries in the first and second cluster are characterized by impressive processes of economic growth (about 7% per year). However, these robust economic growths have not been pro-poor, inclusive or egalitarian. For instance, in Uganda, inequality between regions tripled from 1992–1993 to 2009–2010. Inequality increased also within regions.¹⁶ As reported by UN, while in 2004, the richest 10% of Malawians consumed 22 times more than the poorest 10%, in 2011 the richest 10% was spending 34 times more than the poorest 10%. Similarly, Ghana has experienced increasing economic growth of over 7% per year on average since 2005. However, this country is categorized as having one of the fastest increasing inequality levels in Africa.

Madagascar, which belongs to the second cluster, is an exception. It has been characterized by a deterioration of the economic situation of people in the upper quintile of the income distribution, which contributed to reduce inequality. This explains why, despite very low incomes and high poverty incidence, Madagascar remains among the most inequality virtuous countries as compared to those treated in this paper and, in general, middle of the range of

¹⁵It must be noted that the original version of the Great Gatsby Curve actually says that present inequality reduces future intergenerational mobility so some temporal gap is needed.

¹⁶See Brunori, Palmisano, and Peragine (2018b) for more details.

¹⁷The ranking of countries is robust to the choice of the inequality index, whether Gini or MLD, although the use of MLD reports lower relative of both inequality of opportunity in level and relatively to overall inequality.

values for sub-Saharan countries and worldwide (see World Bank 2014).¹⁷

By contrast, Niger, representing the third cluster, shows the lowest level of both overall inequality and IOp as compared to the other countries considered in this paper. This result is confirmed in Beegle et al. (2016), although it might be argued that the lowest level of IOp can be in part due to the small number of circumstances observed (ethnicity, sex, birthplace).

Detecting the contribution of the specific circumstances

In this section, we analyse the degree of association between each circumstance and the level of individual consumption to shed light on the relative importance of the different circumstances in determining IOp. This analysis does not identify the causal effect of each circumstance on IOp; unobservable determinants of the individual outcome are likely to be correlated with the observable circumstances preventing a causal identification (see Ferreira and Gignoux 2011 for a discussion). Nevertheless, the description of the different degrees of association may help provide an interpretative framework for our estimates of IOp across the Sub-Saharan African countries considered. In doing so, we go beyond the distinction between IOp and other inequalities, and we decompose IOp by source. The same level of IOp can have a rather distinct meaning depending on the relative importance of different circumstances. A country in which ethnic inequalities have a prominent role may appear different from a country in which the main channel of transmission of wealth is parental education or occupation.

Table 4 contains the relative importance associated with each observable circumstance, obtained applying the permutation procedure to the random forest of each county, as described in section 2.4. Forests are obtained with 200 trees, using half of the observations and excluding one regressor at each splitting point.

The value of the coefficients in the tables varies from 0 to 1. A value equal to 1 means that the circumstance has maximum importance; the lower is the value of the coefficient, the less important will be that circumstance in generating inequality of opportunity.

According to our results, the place of birth and the education of the father are the most influential circumstances in SSA: each of them explains the largest share of IOp in four countries. In particular, birth location observed for all countries with the exception of Nigeria, is the most influential circumstance in Comoros, Ghana, Guinea and Niger. It is the second most important in Tanzania and Uganda. Indeed, these countries are characterized by a neat regional division. For instance, in Niger, the northern regions are extremely poor and characterized by wasteland and absence of any kind of industry, as opposed to the centre and southern regions. Similarly, most of the inequalities in Ghana can be mainly explained by regional inequalities that are mostly due to the continuation in the post-colonial era of the colonial policy of investing in regions with exportable products and providing supporting infrastructure only in such regions (see Ananim, Mariwah, and Sebu 2012). This circumstance is comparatively less relevant in Madagascar and Malawi.

Father education, observed in all countries with the exception of Niger and Uganda, is the most relevant source of inequality in Madagascar, Malawi, Rwanda and Tanzania, whereas in Comoros father education appears to be a weak factor of influence. Interestingly the first three are the poorest countries in the sample, with Tanzania ranked in the middle of the countries mean per capita consumption distribution. More in general, parental education is an important driver of IOp because it helps improving the health and nutritional behaviour of the most disadvantaged segments of the population more than proportionally. Indeed, mother education arises to be the most relevant circumstance in Nigeria; it is also quite significant in Malawi, Madagascar and Tanzania.

Ethnicity helps to explain the greatest part of IOp in Uganda. It also remains an important determinant of IOp for Ghana, Madagascar and Niger. The ethnicity impact on IOp usually acts through the underprovision or biased allocation of public goods (especially of education as compared to electricity or water) and prevalence of patronage goods because of the heavy influence of ethnic loyalties on policies (Baldwin and Huber 2010).

Although parental occupation never arises to be the most relevant circumstance, it does play a distinguishing role in some countries. In

Table 4. Random forests relative circumstances importance.

Country	Birth location	Father occupation	Father education	Mother occupation	Mother education	Ethnicity	Sex
Comoros 2004	1.0000	0.0991	0.1012	0.1255	0.0068		0.0146
Ghana 2012–2013	1.0000	0.4230	0.7683	0.7705	0.3938	0.7006	0.1816
Guinea 2002–2003	1.0000	0.3967	0.6100	0.9484	0.2011		0.0308
Madagascar 2005	0.3123		1.0000		0.5804	0.7532	0.0000
Malawi 2010–2011	0.3904		1.0000		0.8173		0.0000
Niger 2011	1.0000					0.6190	0.0000
Nigeria 2010–2011		0.7639	0.7533	0.7205	1.0000		0.0000
Nigeria 2012–2013		0.7199	0.7237	0.6654	1.0000		0.0000
Rwanda 2000	0.4765	0.8339	1.0000	0.4316	0.4528		0.0000
Tanzania 2008–2009	0.9616		1.0000		0.5700		0.0000
Tanzania 2010–2011	0.7877		1.0000		0.5976		0.0000
Uganda 2009–2010	0.6265					1.0000	0.0178
Uganda 2010–2011	0.7325					1.0000	0.0074

particular, mother occupation is the second most relevant circumstance in Guinea, Ghana and Comoros, and father occupation is the second most relevant circumstance in Rwanda. It can be noticed, however, that in Comoros these circumstances have a very low explicative power, compared to the other countries in which parental occupation is ranked as the second most important circumstance.

Last, our results show that the incidence of sex is marginal and null in many cases. This result does not come with a surprise as IOP is measured on the household per capita consumption. The only exception is Ghana wht, a relative importance of about 18%. This figure can only be explained with a negative correlation between per capita consumption and share of female household member due, for example, to high divorce rate (Clark and Hamplová 2013).

V. Conclusion

In the last two decades, SSA has enjoyed a period of unprecedented growth. The other side of the coin of this ‘African Renaissance’ is, however, a less successful alleviation of serious distributional issues, since this sustained growth has rarely resulted in a truly inclusive pattern. As we have shown in this paper, inequality is a serious issue in this African sub-continent, much more serious than it is in the other developing countries around the world.

Inequality in SSA countries is generated by many factors. The area of birth and the education level of the parents are, for instance, among the most important factors. Inequality of opportunity, that is, the extent to which these kinds of factors are systematically correlated with valuable

outcomes of individuals in adulthood, contribute to increase overall inequality and violate principles of fairness. Although the empirical literature on IOP measurement has proliferated in the last decades, there are very few contributions that focus on inequality of opportunity in SSA countries. The lack of estimates for this part of the world is mainly due to the lack of reliable data on individual outcome and circumstances.

This paper has utilized 13 household consumption surveys to assess IOP in 10 SSA countries. All information about exogenous factors provided by these surveys have been used. These encompass information on region of birth, parental education and occupation, ethnicity and sex. We have complemented the analysis by estimating the partial effect of each circumstance in determining IOP.

Overall, inequality of opportunity is very high in every country included in this analysis, although this is quite variable across them, and countries with higher total inequality do not always show higher IOP.

In all countries analysed, circumstances beyond individual control such as ethnicity, birthplace and parental background interact in determining individual opportunity in a much more complex way than what we typically observe in Western societies. Moreover, it is important to stress that our measurement approach considers unproblematic any outcome variability not systematically correlated with observable circumstances. This assumption is commonly accepted by the literature in the context of richer countries. However, it may be questioned in highly risky environments in which a large share of inequality can be explained by luck.

Thus, as a focus of future research, the interaction of country-specific circumstances and the

magnitude of unexplained variability should be examined with country-specific and more data-intensive studies to further elucidate the best possible methods for determining IOp in non-Western countries.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

- Aaberge, R., M. Mogstad, and V. Peragine. 2011. "Measuring Long-Term Inequality of Opportunity." *Journal of Public Economics* 95 (3): 193–204.
- Annim, S. K., S. Mariwah, and J. Sebu. 2012. "Spatial Inequality and Household Poverty in Ghana." *Economic Systems* 36 (4): 487–505.
- Baldwin, K., and J. D. Huber. 2010. "Economic versus Cultural Differences: Forms of Ethnic Diversity and Public Goods Provision." *American Political Science Review* 104 (4): 644–662.
- Beegle, K., L. Christiaensen, A. Dabalen, and I. Gaddis. 2016. *Poverty in a Rising Africa*. Washington, DC: World Bank Publication, World Bank.
- Berik, G., Y. Rodgers, and S. Seguino. 2009. "Feminist Economics of Inequality, Development, and Growth." *Feminist Economics* 15 (3): 1–33.
- Bourguignon, F. 1979. "Decomposable Income Inequality Measures." *Econometrica: Journal of the Econometric Society* 47 (4): 901–920.
- Bourguignon, F., F. H. G. Ferreira, and M. Menéndez. 2007. "Inequality of Opportunity in Brazil." *Review of Income and Wealth* 53 (4): 585–18.
- Bradbury, K., and R.K. Triest. 2016. "Inequality of Opportunity and Aggregate Economic Performance." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 2 (2), 178–201.
- Brautigam, D. A., and S. Knack. 2014. "Foreign Aid, Institutions, and Governance in Sub-Saharan Africa." *Economic Development and Cultural Change* 52 (2): 255–285.
- Breiman, L., J. H. Friedman, R. A. Olshen, and C. J. Stone. 1984. *Classification and Regression Trees*. Wadsworth, California: Cole Statistics/Probability Series.
- Brunori, P., F. Palmisano, and V. Peragine. 2018b. "Consumption Dynamics and Inequality of Opportunity with an Application to Uganda." *Review of Development Economics* 22 (2): 632–657.
- Brunori, P., P. Hufe, and D. G. Mahler. 2018a. "The Roots of Inequality: Estimating Inequality of Opportunity from Regression Trees." World Bank Policy Research Working Papers N. 8349.
- Brunori, P., V. Peragine, and L. Serlenga. 2018c. "Upward and Downward Bias When Measuring Inequality of Opportunity". *Social Choice and Welfare* First Online: 23 November 2018, 52 (4): 635–661.
- Checchi, D., and V. Peragine. 2010. "Inequality of Opportunity in Italy." *Journal of Economic Inequality* 8 (4): 429–450.
- Clark, S., and D. Hamplová. 2013. "Single Motherhood and Child Mortality in sub-Saharan Africa: A Life Course Perspective." *Demography* 50 (5): 1521–1549.
- Cogneau, D., and S. Mesplé-Soms. 2008. "Inequality of Opportunity for Income in Five Countries of Africa." In *Research on Economic Inequality*, edited by J. Bishop and B. Zheng, 99–128. Vol. 16. Bingley, UK: Emerald Group Publishing Limited.
- Corak, M. 2013. "Income Inequality, Equality of Opportunity, and Intergenerational Mobility." *The Journal of Economic Perspectives* 27: 79–102.
- De Ree, J., and E. Nillesen. 2009. "Aiding Violence or Peace? the Impact of Foreign Aid on the Risk of Civil Conflict in sub-Saharan Africa." *Journal of Development Economics* 88 (2): 301–313.
- Easterly, W., and R. Levine. 1997. "Africa's Growth Tragedy: Policies and Ethnic Divisions." *The Quarterly Journal of Economics* 112: 1203–1250.
- Elbers, C., P. Lanjouw, J. Mistiaen, and B. Ozler. 2008. "Re-Interpreting Sub-Group Inequality Decompositions." *The Journal of Economic Inequality* 6 (3): 1569–1721.
- Ferreira, F. H. G., and V. Peragine. 2016. "Individual Responsibility and Equality of Opportunity." chapter 25 In *The Oxford Handbook of Well-Being and Public Policy*, edited by M. D. Adler and M. Fleurbaey (pp. 751–777). Oxford: Oxford University Press.
- 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): 622–657.
- Ferreira, F. H. G., J. Gignoux, and M. Aran. 2011. "Measuring Inequality of Opportunity with Imperfect Data: The Case of Turkey." *The Journal of Economic Inequality* 9 (4): 651–680.
- Fleurbaey, M. 2008. *Fairness, Responsibility and Welfare*. Oxford: Oxford University Press.
- Hassine, N. B. 2011. "Inequality of Opportunity in Egypt." *World Bank Economic Review* 26 (2): 265–295.
- Hothorn, T., K. Hornik, and A. Zeileis. 2006. "Unbiased Recursive Partitioning: A Conditional Inference Framework." *Journal of Computational and Graphical Statistics* 15 (3): 651–674.
- Kimenyi, M. S. 2006. "Ethnicity, Governance and the Provision of Public Goods." *Journal of African Economies* 15 (suppl 1): 62–99.
- Lacina, B., and N. P. Gleditsch. 2005. "Monitoring Trends in Global Combat: A New Dataset of Battle Deaths." *European Journal of Population* 21 (2–3): 145–166.
- Lambert, P. J., and J. R. Aronson. 1993. "Inequality Decomposition Analysis and the Gini Coefficient Revisited." *The Economic Journal* 103 (420): 1221–1227.

- Lassibille, G., and J. P. Tan. 2005. "The Returns to Education in Rwanda." *Journal of African Economies* 14 (1): 92–116.
- Lefranc, A., and A. Trannoy. 2017. "Equality of Opportunity, Moral Hazard and the Timing of Luck." IZA working paper 10645.
- Madise, N. J., Z. Matthews, and B. Margetts. 1999. "Heterogeneity of Child Nutritional Status between Households: A Comparison of Six Sub-Saharan African Countries." *Population Studies* 53 (3): 331–343.
- Marrero, G., J. G. Rodriguez, and R. Van der Weide. 2016. "Unequal Opportunity, Unequal Growth." Policy Research Working Paper 7853, World Bank.
- Marrero, G. A., and J. G. Rodriguez. 2013. "Inequality of Opportunity and Growth." *Journal of Development Economics* 104: 107–122.
- Milanovic, B. 2003. "Is Inequality in Africa Really Different?" World Bank Policy Research Working Paper, (3169).
- Moradi, A., and J. Baten. 2005. "Inequality in Sub-Saharan Africa: New Data and New Insights from Anthropometric Estimates." *World Development* 33 (8): 1233–1265.
- Morgan, J. N., and J. A. Sonquist. 1963. "Problems in the Analysis of Survey Data, and a Proposal." *Journal of the American Statistical Association* 58: 415–434.
- Nordman, C. J., A. S. Robilliard, and F. Roubaud. 2011. "Gender and Ethnic Earnings Gaps in Seven West African Cities." *Labour Economics* 18: S132–45.
- Peragine, V. 2002. "Opportunity Egalitarianism and Income Inequality: The Rank-Dependent Approach." *Mathematical Social Sciences* 44: 45–64.
- Piraino, P. 2015. "Intergenerational Earnings Mobility and Equality of Opportunity in South Africa." *World Development* 67: 396–405.
- Ramos, X., and D. Van de Gaer. 2015. "Approaches to Inequality of Opportunity: Principles, Measures and Evidences." *Journal of Economic Surveys* 30 (5): 855–883.
- Roemer, J. E. 1998. *Equality of Opportunity*. Cambridge, MA: Harvard University Press.
- Roemer, J. E., and A. Trannoy. 2016. "Equality of Opportunity: Theory and Measurement." *Journal of Economic Literature* 54 (4): 1288–1332.
- Schultz, P. 2004. "Evidence of Returns to Schooling in Africa from Household Surveys: Monitoring and Restructuring the Market for Education." *Journal of African Economies* 13 (2): 95–148.
- Singh, A. 2011. "Inequality of Opportunity in Earnings and Consumption Expenditure: The Case of Indian Men." *Review of Income and Wealth* 58: 79–106.
- Strobl, C., A. Boulesteix, T. Kneib, T. Augustin, and A. Zeileis. 2008. "Conditional Variable Importance for Random Forests." *BMC Bioinformatics* 9 (307).
- Thorbecke, E. 2013. "The Interrelationship Linking Growth, Inequality and Poverty in Sub-Saharan Africa." *Journal of African Economies* 22 (suppl 1): i15–i48.
- Van de Gaer, D. 1993. "Equality of opportunity and investment in human capital." Ph.D. diss., Katholieke Universiteit Leuven: Belgium.
- Villar, A. 2017. *Lectures on Inequality, Poverty and Welfare*. Switzerland: Springer-Verlag.
- World Bank. 2006. *World Development Report 2006: Equity and Development*. Washington, DC: World Bank.
- World Bank. 2014. *World Development Report 2014: Risk and Opportunity: Managing Risk for Development*. Washington, DC: World Bank.

World Bank. 2015. *Purchasing Power Parities and the Real Size of World Economies*. 2015 International Bank for Reconstruction and Development The

Table A5. Share of missing circumstances.

Country	Birth location	Father occupation	Mother occupation	Father education	Mother education	Ethnicity	Sex
Comoros 2004	5.78%	44.91%	59.54%	12.59%	13.92%	100%	0%
Ghana 2012–2013	0.037%	3.35%	6.69%	4.37%	2.07%	0.92%	0%
Guinea 2002–2003	0.016%	5.45%	16.31%	1.48%	12.48%	100%	0%
Madagascar 2005	0.47%	100%	100%	9.38%	8.02%	0.01%	0.01%
Malawi 2010–2011	0.07%	100%	100%	0.33%	0.32%	100%	0%
Niger 2011	0.17%	100%	100%	100%	100%	0.0082	0%
Nigeria 2010–2011	100%	1.21%	1.82%	1.26%	1.43%	100%	0%
Nigeria 2012–2013	100%	4.32%	4.18%	3.61%	3.75%	100%	0%
Rwanda 2000	1.75%	1.99%	2.01%	8.86%	5.83%	100%	1.75%
Tanzania 2008–2009	0.26%	100%	100%	9.186	5.533	100%	0%
Tanzania 2010–2011	0.65%	100%	100%	9.643	6.112	100%	0%
Uganda 2009–2010	7.78%	100%	100%	100%	100%	8.876	0%
Uganda 2019–2011	3.69%	100%	100%	100%	100%	4.733	0%

World Bank, Washington DC.

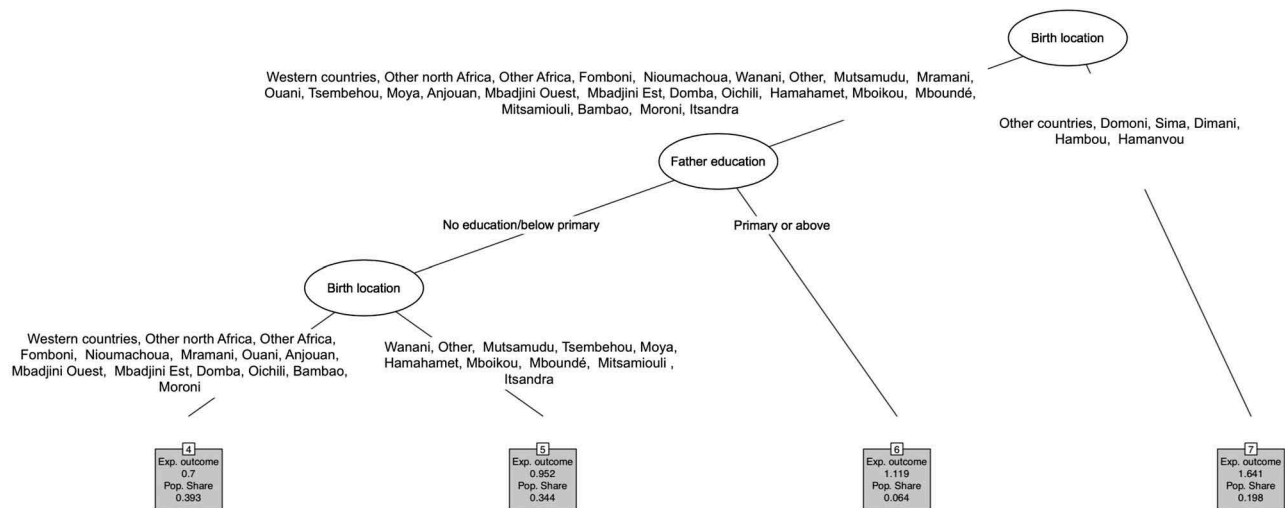


Figure A7. Opportunity structure: Comoros 2004.

Source: Authors' calculation based on surveys listed in Table 1.

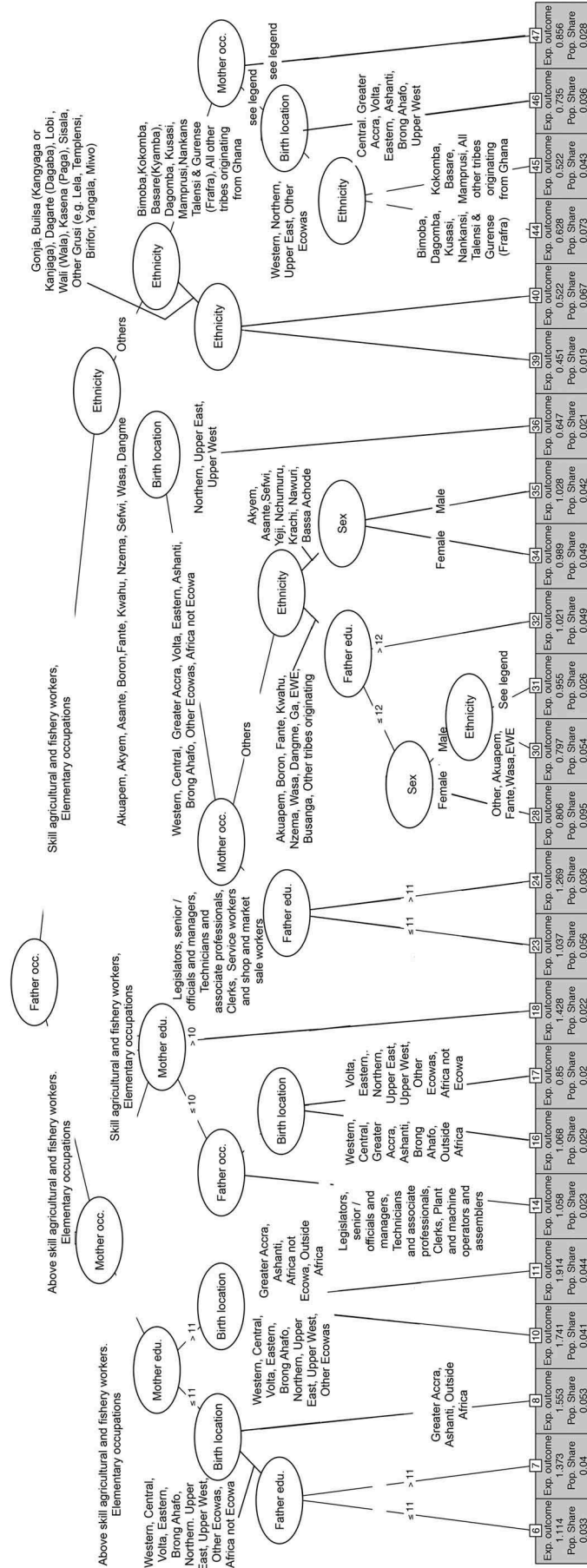
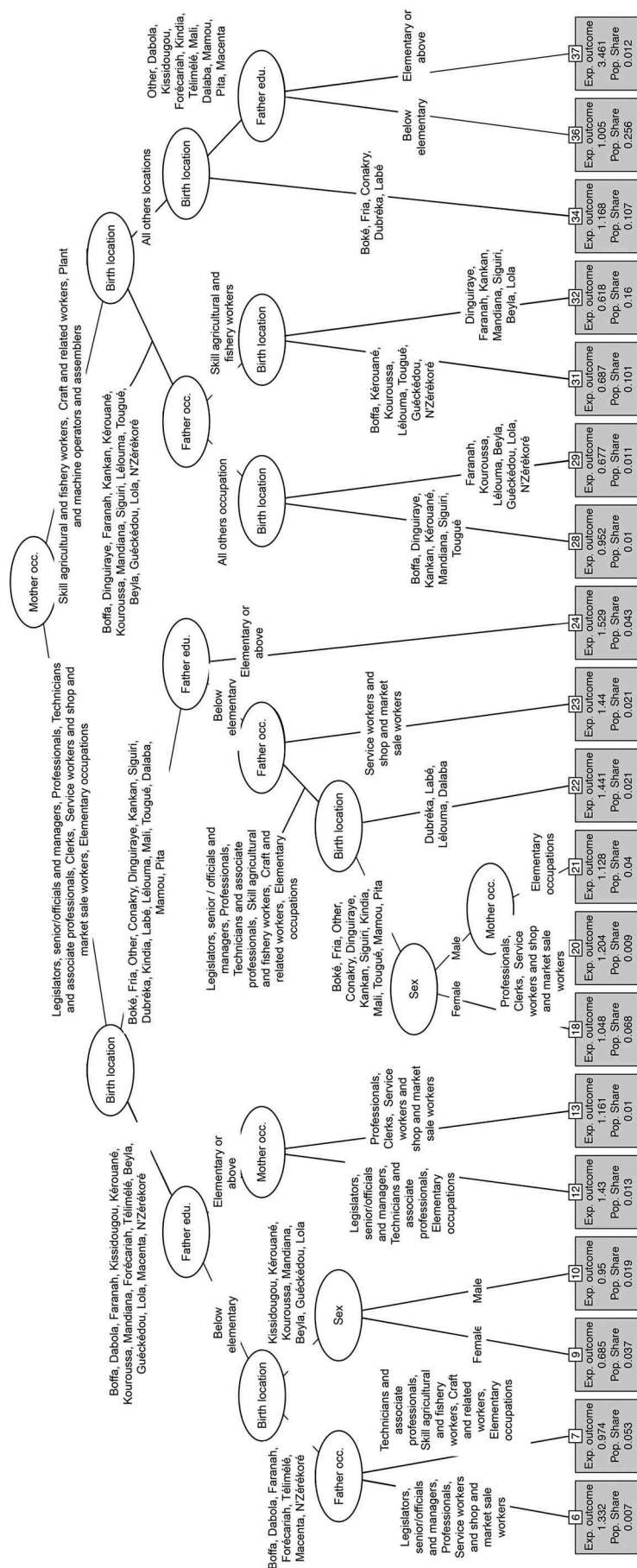
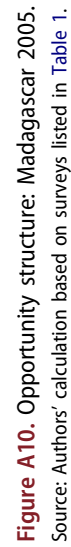


Figure A8. Opportunity structure: Ghana.

Source: Authors' calculation based on surveys listed in Table 1. Node 31 Ethnicity: Boron (Brong) (including Banda), Kwahu, Nzema, Dangme (Ada, Shai, Krobo, Osudoku, Ningol), Ga, Busanga, Other tribes originating from outside Ghana. Node 47 Mother occupation: Professionals, Clerks, Service workers and shop and market sale workers, Plant and machine operators and assemblers.





Source: Authors' calculation based on surveys listed in Table 1.

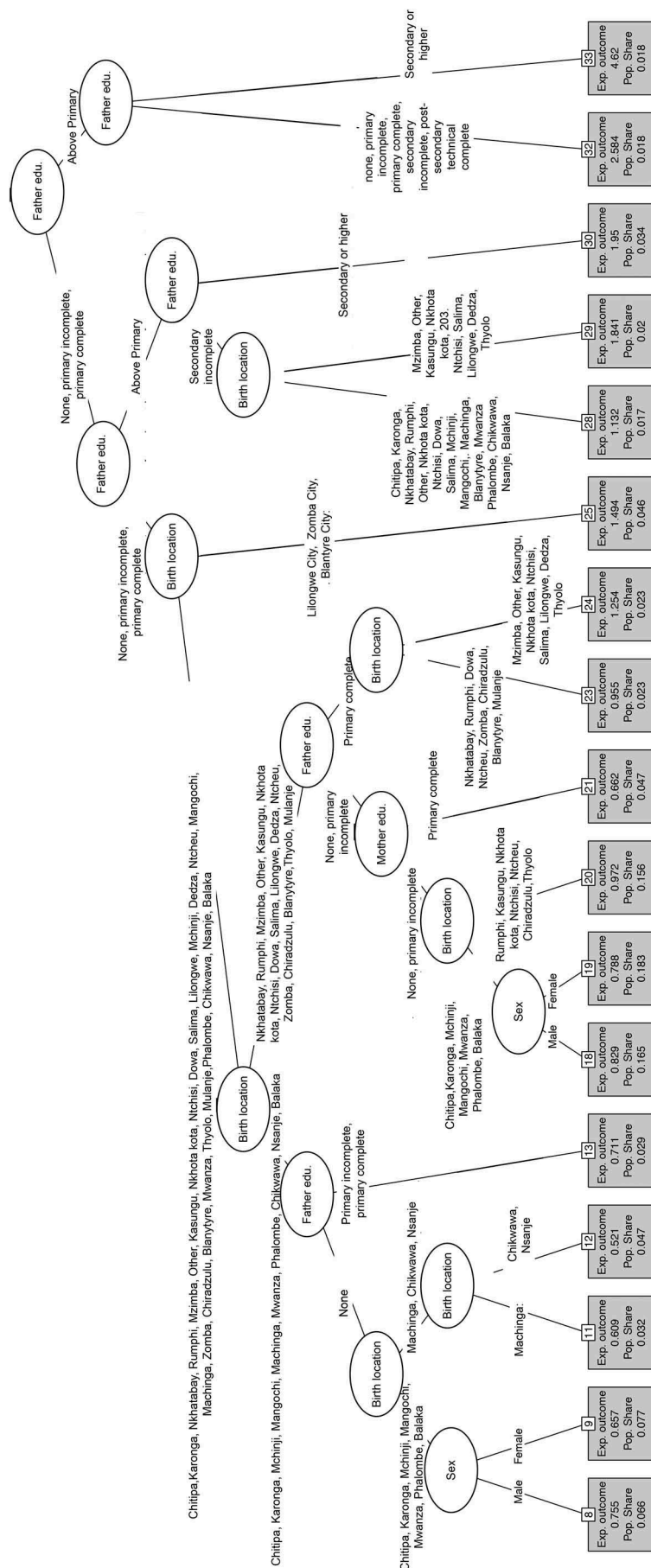
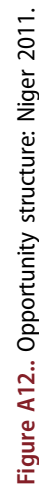


Figure A11. Opportunity structure: Malawi 2010.
Source: Authors' calculation based on surveys listed in Table 1.



Source: Authors' calculation based on surveys listed in Table 1.

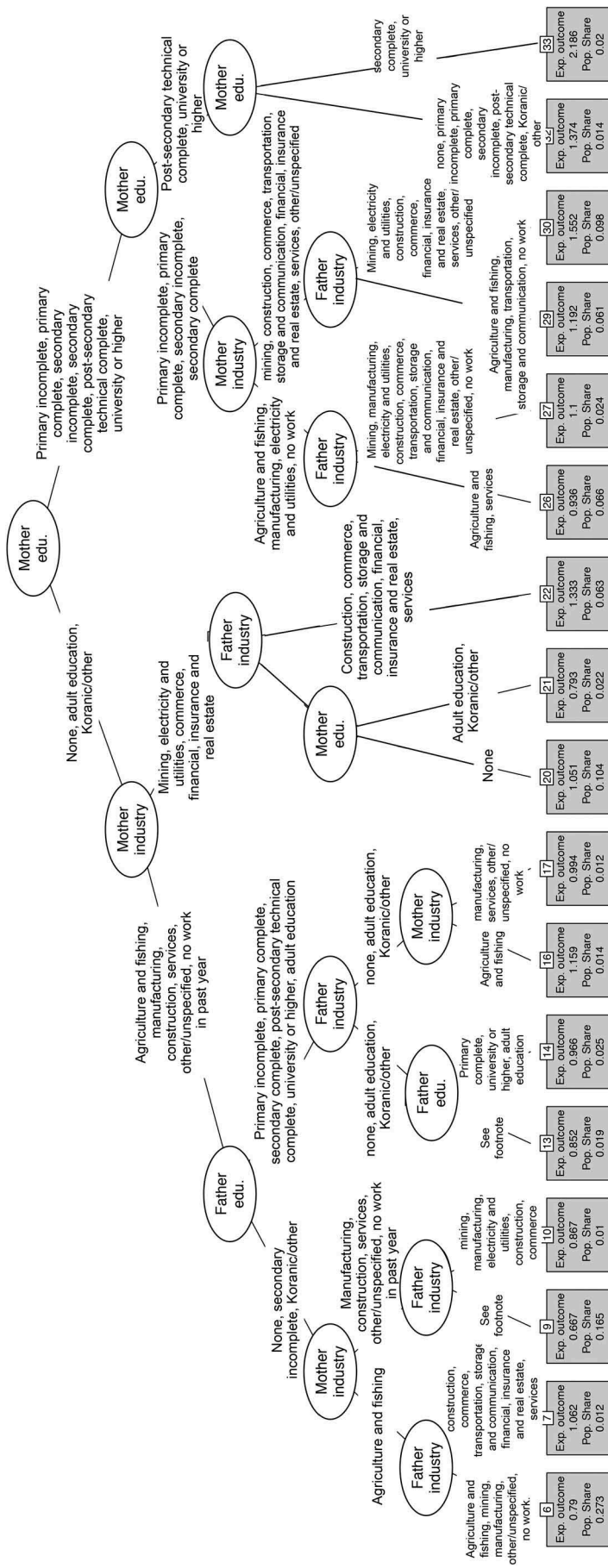


Figure A13. Opportunity structure: Nigeria 2010.
Source: Authors' calculation based on surveys listed in Table 1.

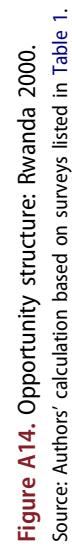


Figure A14. Opportunity structure: Rwanda 2000.

Source: Authors' calculation based on surveys listed in Table 1.

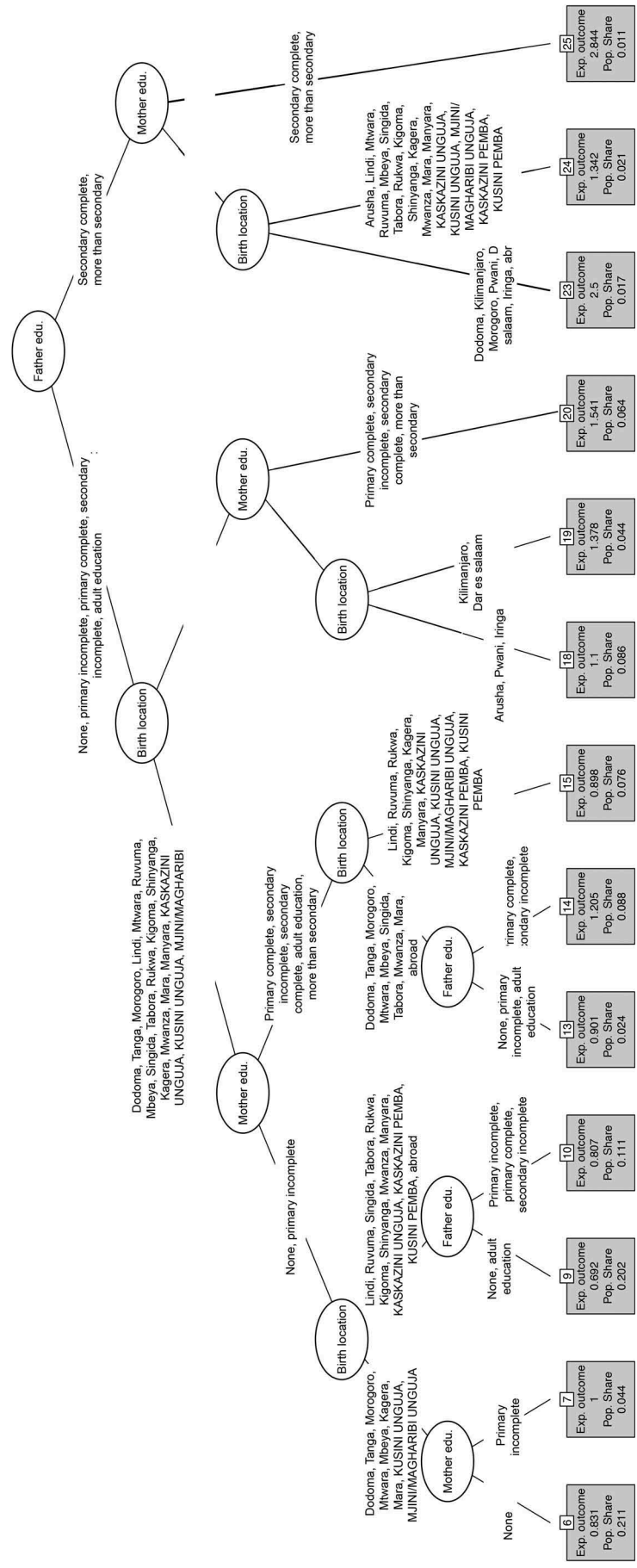


Figure A15. Opportunity structure: Tanzania 2008.
Source: Authors' calculation based on surveys listed in Table 1.

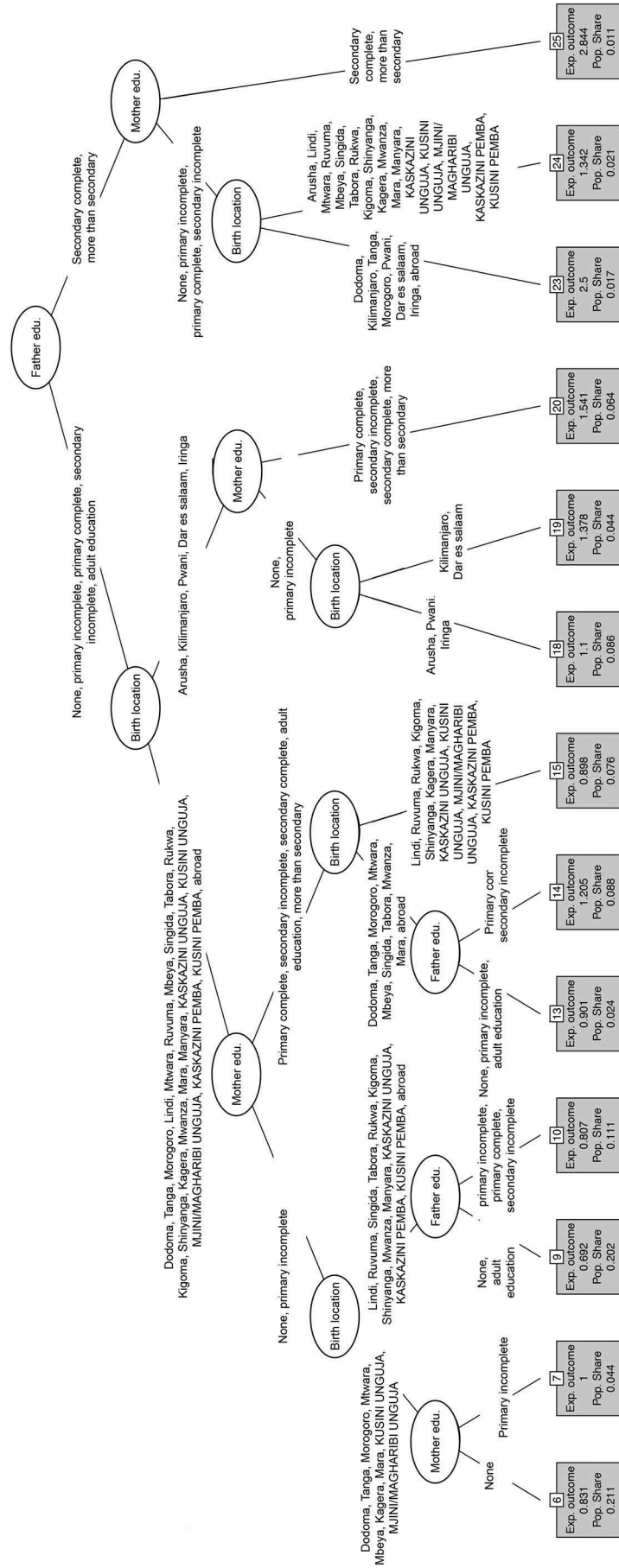


Figure A16. Opportunity structure: Tanzania 2010.
Source: Authors' calculation based on surveys listed in Table 1.

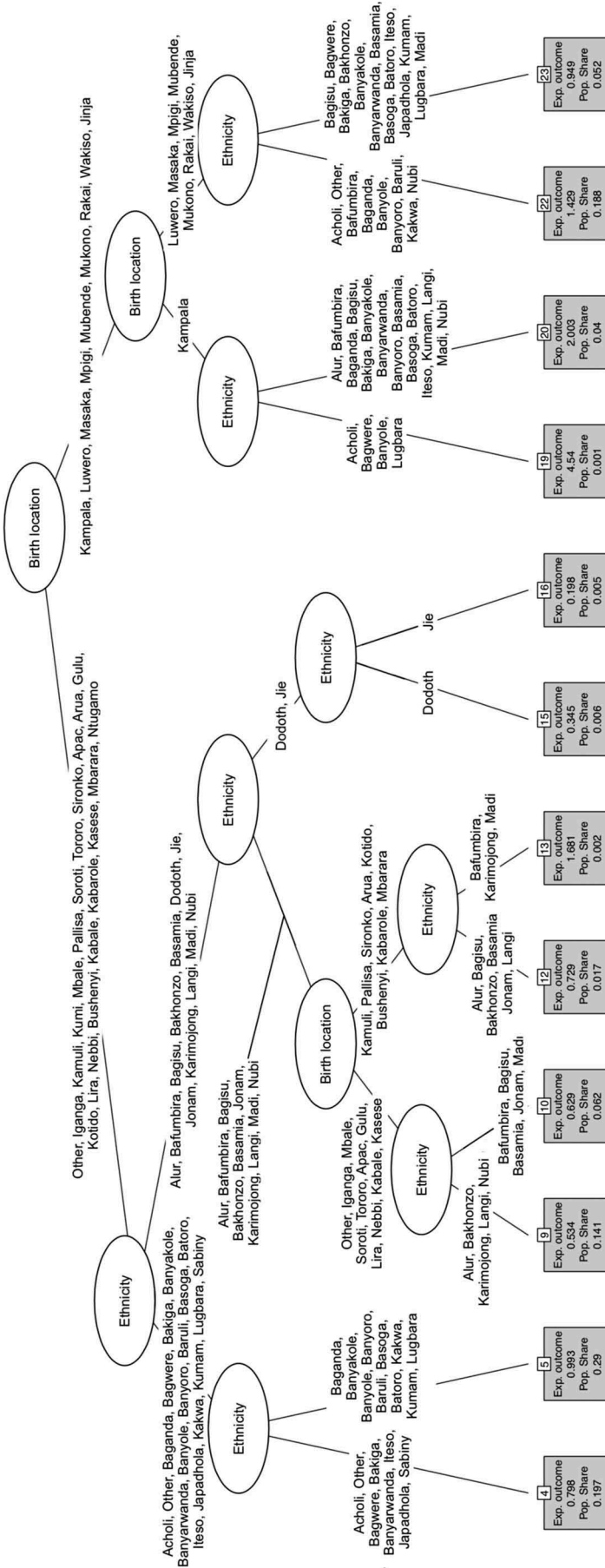
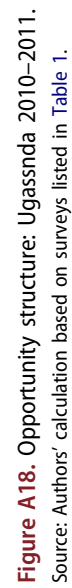


Figure A17. Opportunity structure: Uganda 2009–2010.
Source: Authors' calculation based on surveys listed in Table 1.



Source: Authors' calculation based on surveys listed in Table 1.

Appendix I. Information on missing circumstances

Appendix II. Opportunity profiles

Appendix III. Adjusted IOP

Given a set of selected circumstances defined on the basis of normative grounds and observability constraints, any within-type variation in individual outcome is attributed to personal effort. However, the vector of observed circumstances is likely to be a sub-vector of the theoretical ('true') vector of all possible circumstances that determine a person's outcome. Hence, as in any other empirical analysis of this kind, we face the issue of omitted circumstance variables. This problem is often addressed by the argument that the IOP estimates should be interpreted as lower-bound estimators of the true inequality of opportunity, that is, the inequality that would be captured by observing the full vector of circumstances. It can be shown, in fact, that increasing the number of observed circumstances increases IOP (see Ferreira and Gignoux 2011).

However, this interpretation renders IOP estimates barely comparable across studies, particularly when comparing, for instance, the IOP of a country with a large number of observable circumstances to the IOP of another country with only few observable circumstances. Another important reason that can make the number of types differing across countries is the sample size. Using conditional inference trees, the larger the sample size, the higher is the number of expected terminal nodes in the tree, independently of the level of IOP in the population.

Elbers et al. (2008) discuss this issue in a more general setting concerning any estimate of between-group inequality. They claim that when decomposing total inequality into a between and a within component, the estimate of between-group inequality might be artificially too low because it compares between-group inequalities with the inequality measured in a counterfactual population in which each individual is a group. To overcome this problem they propose an adjusted measure of between-group inequality, which is equivalent to the actual between-group inequality normalized by the maximum possible between-group inequality that could be reached in the population, given the number of groups. The latter is defined as the extent of between-group inequality in a counterfactual distribution (Y_a) obtained by ranking outcomes from the lowest to the richest and then partitioning the distribution in such a way that the groups have the same population share as the actual group. Hence, the adjusted IOP (Adj-IOP) can be expressed as follows:

$$\text{Adj-IOP} = \frac{I(Y_s)}{I(Y_a)} \quad (5)$$

Although the problem they are looking at does not exactly correspond to our problem of partial observability, their solution can be usefully applied to this context. This adjusted measure is appealing as it accounts for the number of types and their relative weights. Adj-IOP solves, at least in part, the problem of comparing IOP estimates based on different number of observable characteristics. Therefore, in the empirical analysis, we propose estimates of both IOP and Adj-IOP. It deserves to be noted that, to the best of our knowledge, this is the first contribution that extends the application of an adjusted measure of inequality between groups to the analysis of inequality of opportunity.

The seventh column of Table 3 reports the adjusted IOP^{rel} . As discussed above, the normalization of inequality with respect to the number of types is particularly relevant in the present context, as we are comparing IOP in countries whose specific consumption distribution is partitioned into a very different number of types.

Figure A20 plots the correction to our IOP estimates that is generated by the computation of the Adj-IOP. This correction terms are measured as the difference between IOP and Adj-IOP as a percentage of IOP and is plotted against the number of types. It is characterised by a clear pattern, approaching zero as the number of types increases.

The figure makes it clear that the adjustment procedure does not add particularly relevant information in our context. The correction is above 10% only for Comoros, a country with an extremely small tree. It is interesting that this correction is somehow already obtained when using random forests: IOP increases by 30%. For all the other countries the correction is never above 5% and close to zero for partitions made of more than 15 types. Hence, the higher the number of types the lower the impact of the adjustment, and this result is rather general. To grasp this drawback consider Figure A20, plotting the difference between total inequality measured using the Gini coefficient, twice the area between the black Lorenz curve and the diagonal, and the maximum between-group Gini coefficient, twice the area between the blue broken line, for three hypothetical group partitions: 1 group, 5 groups, 10 groups. The difference between the two possible denominators of IOP will depend on the shape of the original Lorenz curve; the example clarifies that this difference approaches zero very quickly as the number of types increases. This is because the Lorenz curve of the Y_a distribution can be expressed as follows:

$$L(Y_a) = \frac{\sum_{h \leq j-1} N_h \mu_h + N_j \mu_j p_j}{N \mu} \quad (6)$$

From the expression above it clearly results that as the number of groups increases and approaches the number of individuals in the population, N_h and N_j tend to 1 and μ_h and μ_j tend to the income of individuals ranked, respectively, h

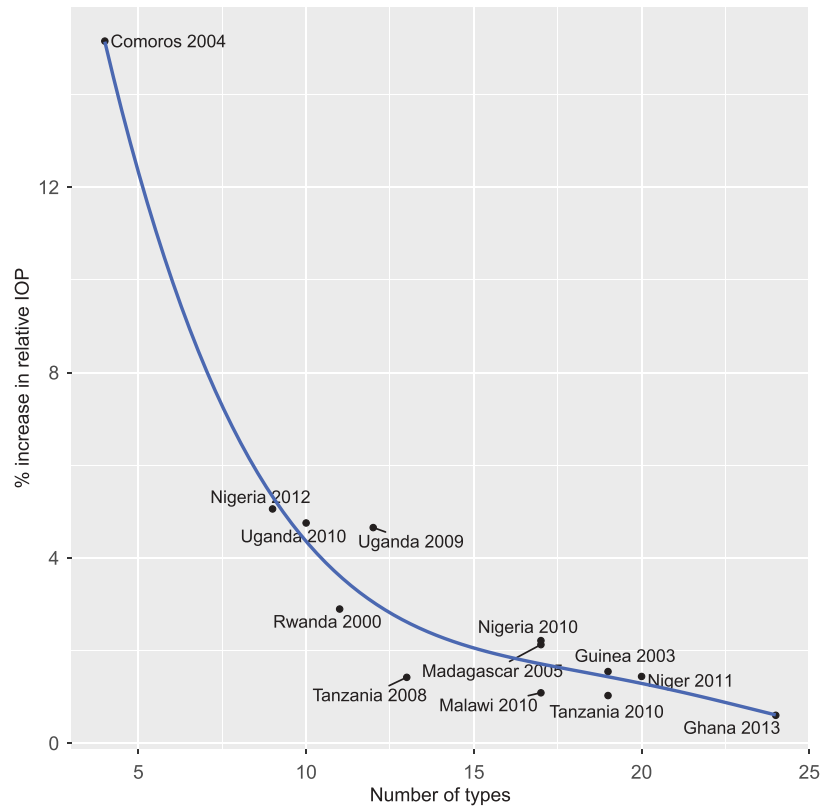


Figure A19. Adj-IOP correction and number of types. The relationship is interpolated with a polynomial of degree 4.
Source: Authors' calculation based on surveys listed in [Table 1](#).

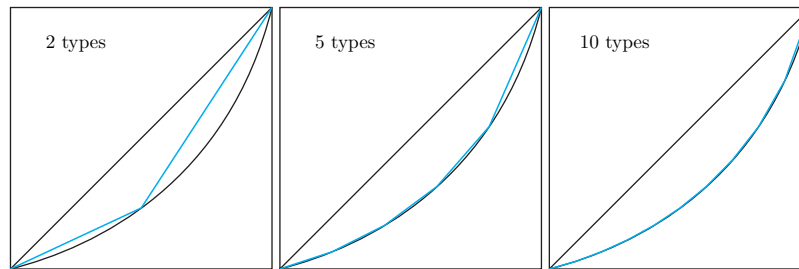


Figure A20. Lorenz curve and maximum between type inequality Lorenz curve.

Note: Lorenz curves for the maximum between-group inequality (light blue) are drawn assuming a population partitioned into equally sized types.

and j . As a result, $L(Y_a)$ would correspond to $L(y) = \frac{\sum_{h=1}^j y_h}{N\mu}$.
Therefore, the adjustment proposed by Elbers et al. (2008)

loses relevance whenever the number of types is in the order of tens.