Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Inherited inequalities in The Gambia: An analysis of income and literacy opportunity

Domenico Moramarco

domenico.moramarco@uniba.it

University of Bari Aldo Moro

Giorgia Zotti

University of Bari Aldo Moro

Research Article

Keywords:

Posted Date: May 19th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9573936/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Inherited inequalities in The Gambia: An analysis of income and literacy opportunity.

Domenico Moramarco*and Giorgia Zotti†

Abstract

This paper studies inherited inequalities in The Gambia in two dimensions, income and literacy, using data from the 2015 Integrated Household Survey. We interpret inequality of opportunity as a measure of constrained social mobility: the extent to which current outcomes remain predictable from inherited circumstances and parental background. We show that a substantial share of income inequality is explained by circumstances beyond individual control, with birthplace and ethnicity emerging as the main drivers. By contrast, literacy outcomes display a more favorable pattern, with inherited inequalities declining among younger generations.

1 Introduction

Low- and middle-income countries often combine high levels of inequality with strong persistence in socioeconomic status across generations. Yet we still know relatively little about how inequality is transmitted from one generation to the next in these settings. A major obstacle is the scarcity of administrative records and data linking parents and adult children. As a result, conventional measures of intergenerational mobility are difficult to estimate precisely in many low-income countries.

*Department of economics and Finance, University of Bari. Email: domenico.moramarco@uniba.it

†Department of economics and Finance, University of Bari. Email: giorgia.zotti@uniba.it

This paper studies inherited inequality in The Gambia. Building on recent work that connects intergenerational mobility and inequality of opportunity within a common framework (Ferreira & Brunori, 2024; Ferreira *et al.*, 2025b; Brunori *et al.*, 2026). We interpret the inequality of opportunity as a measure of constrained social mobility. The central idea is simple: in an fair society, outcomes such as income, education, or literacy should not be strongly predictable from characteristics over which individuals have no control. When birthplace, ethnicity, gender, or parental background systematically predict later-life outcomes, inequality is not merely incidental; it is transmitted across generations.

This perspective links two literatures that have often developed separately. The literature on intergenerational mobility typically measures the association between parents' and children's outcomes, such as income or education. Instead, the literature on inequality of opportunity asks how much of the current inequality can be attributed to predetermined circumstances, such as family background, place of birth, ethnicity, or gender. Although these approaches differ in origin and emphasis, both address the same underlying question: how much do initial conditions shape life chances? In this sense, inequality of opportunity can be read not only as a normative measure of unfair inequality, but also as an empirical measure of limited mobility (Roemer, 1998; Ferreira & Gignoux, 2011; Ferreira *et al.*, 2025b; Brunori *et al.*, 2026). This interpretation is also supported by empirical evidence of the positive correlation between intergenerational persistence and inequality of opportunity (see, among others, Brunori *et al.*, 2013; Ferreira *et al.*, 2025a).

The concept of *inherited inequality* (Ferreira & Brunori, 2024; Ferreira *et al.*, 2025b; Brunori *et al.*, 2026) provides a particularly useful bridge between these approaches. Rather than requiring a complete distinction between effort and responsibility, which is typical of the inequality of opportunity framework (Roemer, 1998; Checchi & Peragine, 2010; Peragine & Ferreira, 2015), inherited inequality focuses on whether current outcomes are statistically independent of inherited circumstances. If not, the share of inequality associated with those circumstances can be interpreted as inherited inequality and the intergenerational persistence of advantages. This formulation is especially valuable in data-scarce settings, where full parent-child panels are unavailable, but household surveys contain information on predetermined characteristics such as birthplace, ethnicity, gender, and parental background.

The Gambian case is important for both substantive and methodological reasons. The country combines persistent structural disadvantages with important signs of so-

cial progress. Although The Gambia has reduced poverty and inequality over time, major challenges remain, including rural deprivation, demographic pressure, and unequal access to economic opportunities. At the same time, educational reforms—such as the scholarship program introduced for girls in 2001—appear to have improved access to schooling and reduced some gender disparities. These features make The Gambia an informative setting in which to study inequality of opportunity as a measure of realized mobility: background conditions may still strongly shape income outcomes, while educational opportunities may have become less unequal across cohorts.

We focus on two dimensions of advantage: income and literacy. Income captures the distribution of economic resources and is central to the study of inequality and mobility. Literacy captures a basic educational achievement that is closely related to long-term welfare and labor-market opportunities, while also allowing us to examine changes across birth cohorts. Studying both outcomes is important because inherited inequality can evolve differently between domains. For literacy, we also performed a cohort analysis to capture changes between generations and evaluate the long-term effects of educational reforms. By contrast, income outcomes are analyzed for the entire sample, as income data reflect short-term economic dynamics rather than cohort-based structural transformations.

Empirically, the article uses data from the 2015 Gambia Integrated Household Survey. We estimate the inequality of opportunity using an ex-ante parametric approach (Ferreira & Gignoux, 2011; Brunori *et al.*, 2013; Ramos & Van de Gaer, 2016), in which income and literacy are predicted from inherited circumstances, and the inequality is measured in the resulting distribution of predicted outcomes. We complement this with a machine-learning approach based on random forests (Brunori & Neidhöfer, 2021; Brunori *et al.*, 2023a,b,c)), which allows for nonlinearities and interactions among circumstances and provides evidence on the relative importance of different inherited characteristics. This combination allows us to estimate not only the extent of inherited inequality, but also the circumstances that drive it.

The results show that inherited circumstances account for a substantial share of income inequality in The Gambia. Birthplace and ethnicity emerge as especially important predictors, suggesting that economic prospects remain strongly conditioned by territorial and social origin. In this sense, social mobility appears constrained in the income dimension: where individuals are born and the groups into which they are born continue to matter for their position in the income distribution.

The evidence for literacy is more encouraging. Although inherited circumstances also shape literacy outcomes, cohort patterns suggest that educational opportunity has become more equal over time. Younger cohorts show weaker inherited literacy gradients and gender gaps are substantially smaller than among older cohorts. These findings are consistent with the view that educational expansion and targeted schooling policies can reduce some forms of inherited disadvantage, even when economic mobility remains limited.

The paper contributes to the literature in three main ways. First, it provides new evidence on inherited inequalities in The Gambia, a country for which empirical research remains limited. Second, it explicitly frames the inequality of opportunity as a measure of social mobility, thus linking the normative literature on fairness with the empirical literature on intergenerational persistence. Although recent work has emphasized the link between inherited inequality and mobility, this perspective has not been adequately explored in country-specific analyzes for low-income African contexts. Third, it combines traditional and machine-learning methods to identify not only the level of opportunity inequality, but also its main drivers across different outcome dimensions. In doing so, the paper helps to clarify whether observed inequality in The Gambia is better understood as fair inequality, linked to individual effort, or as inherited inequality, linked to social origin.

The remainder of the paper is organized as follows. Section 2 presents a brief historical background on The Gambia. Section 3 describes and further motivates our methodology. Section 4 describes the data and presents the empirical results. Section 5 concludes.

2 The Gambia context

Known as the *smiling coast of Africa*, The Gambia boasts a distinctive geographic location, almost entirely enveloped by Senegal, except for a narrow coastal strip facing the Atlantic Ocean (see Figure 1). This country has an eventful geopolitical history, strongly influenced by the legacy of the British Empire.



Figure 1: The Gambia map

The history of Gambia has been profoundly influenced by the British colonial government (Hughes & Perfect, 2006; Perfect, 2008; Gray, 2015). Gaining independence in 1965, the Country transitioned into a republic in 1970, with Dawda Jawara assuming the presidency. In 1981, a coup by socialist factions led Jawara to seek military support from Senegal. This collaboration birthed the Confederation of Senegambia in 1982, aiming to merge the economies and currencies of the two nations. However, by 1989, Gambia exited the confederation, once again embracing its status as an independent republic, albeit one marked by political vulnerability. This fragility led to a second coup in 1994, resulting in the overthrow of Jawara's government. The coup stifled opposition activities and installed Lieutenant Yahya Jammeh as the new head of state, marking another pivotal moment in Gambia's political landscape.

The Gambia stands out in Sub-Saharan Africa (SSA) for its unique socio-demographic characteristics, which differentiate it from other countries. Indeed, unlike other countries, several World Bank reports over the years have shown how The Gambia is one the most effective countries in fighting poverty and inequality. Fosu & Gafa (2020) showed this for the period between 1998 and 2003. Subsequently, Mungai & Okiya (2019) demonstrated how, between 2010 and 2015, The Gambia experienced a general decrease in poverty levels, accompanied by an increase in poor individuals, especially in rural areas. This phenomenon could be blamed on the country's high fertility rate.

A more mixed conclusion comes from Carrasco Nunez *et al.* (2022), which described The Gambia's economic growth over the past decades as limited despite

implementing fiscal policies to reduce inequality and poverty. In this period, the poverty rate has increased by 5.3%, while the Gini index has decreased compared to previous years. This finding is also confirmed by the World Bank estimates of the Gini index,[1] which went from 48.5% in 1998 to 38.8% in 2020.

Since independence in 1965, The Gambia has experienced economic growth, but remains one of the countries with the lowest levels of education (Foltz & Gajigo, 2012). The Gambian education system provides primary, middle, and upper-secondary education. However, only those who pass the WASSCE exam, administered by the West African Examination Council (WAEC), can enter the university.

In 2001, the country launched a pioneering initiative: a scholarship program dedicated exclusively to girls, developed by the Gambian government with support from the World Bank, UNICEF, and the International Monetary Fund, aiming to reduce the gender gap in education and increase secondary school enrollment. Before 2001, tuition fees were compulsory from middle school, but with the start of reform, female students were exempted from paying fees throughout middle and high school. Although implemented throughout the country, Banjul, the most urbanized and capital region, was excluded. The rural Upper River and Central River areas saw first the implementation of the program, which was later extended to the Lower River and North Bank West Coast.

This initiative has produced remarkable results both demographically, significantly increasing access to female education, reducing the gender gap, and improving educational outcomes (Blimpo *et al.*, 2011, 2019). This intervention is part of a broader policy context for school inclusion that began in the 1990s, when 150 countries, including 20 African countries and The Gambia, joined the United Nations Education for All initiative. This project aimed to reduce or eliminate school fees, which were identified as the main obstacle to access to education, especially primary education (Petrosino *et al.*, 2012; Krishnaratne *et al.*, 2013; Murnane & Ganimian, 2014). As a result, decreasing the cost of access helped reduce the education gap, increasing the share of educated population.

[1] https://data.worldbank.org/indicator/SI.POV.GINI?locations=GM-1W

3 Methodology

3.1 Theoretical framework

This paper aims at analyzing intergenerational persistence of advantages by applying the inequality of opportunity framework in a context where the direct intergenrational link between income levels is not possible.
The foundations of the inequality of opportunity approach lie in the broader literature on justice and responsibility-sensitive equality. Important contributions by Rawls (1971), Sen *et al.* (1980), Dworkin (1981), and Cohen (1989) shaped the normative debate, while Roemer (1993, 1998) provided one of the most influential models to measure and define inequality of opportunity. This framework has subsequently been formalized in empirical work on the measurement of inequality of opportunity (see Ramos & Van de Gaer, 2016, for a survey).

Recent contributions (Ferreira & Brunori, 2024; Ferreira *et al.*, 2025b; Brunori *et al.*, 2026) further show that many measures of mobility and inequality of opportunity can be interpreted within a common framework, in which inherited inequality exists whenever predetermined circumstances help explain the distribution of current outcomes.

Intuitively, both intergenerational persistence and inequality of opportunity measure the extent to which current outcomes are predictable from inherited circumstances. In the standard ex-ante framework, this involves estimating the conditional expectation $\mathbb{E}[y \mid C]$, where y denotes income in the current generation, and C is a vector of inherited characteristics or circumstances. Conventional estimates of intergenerational mobility can be seen as a special case in which C contains only parental income. Inequality of opportunity measures, on the contrary, allows C to include a broader set of characteristics for which individuals cannot be held responsible. Although parental income is rarely observed directly, these measures often include parents' occupation and education, which are strong predictors of it.

Following this intuition, we assess the intergenerational persistence of economic advantages by estimating inequality of opportunity in a model that includes parents' occupation and education as circumstances.

We analyze two relevant outcomes, namely income and literacy, implementing two standard estimation techniques: the so-called parametric approach (Ferreira &

Gignoux, 2011) and the more recent data-driven methods (Brunori *et al.*, 2023c). In addition, regarding literacy analysis in The Gambia, a cohort analysis is conducted to gain a deeper understanding of the evolution of educational attainment in 2015. This cohort analysis involves dividing the population into four separate cohorts, following the history of The Gambia. This approach allows us to understand changes in education levels across several generations, providing a longitudinal perspective on access to and quality of education in the context. It also gives a chance to observe whether policies in favor of girls attending secondary schools have worked. We do not conduct a cohort analysis for income because income levels change with age and short-term economic conditions, making comparisons between generations less reliable. Income inequality is more sensitive to the economic context, and this can influence long-term between birth cohorts in a cross-section.

The next two sections describe in more detail the parametric and data-driven approaches to estimate inequality of opportunity.

3.2 Estimation techniques

The canonical model for studying and defining the inequality of opportunity (Roemer, 1993; Fleurbaey, 1994; Van De Gaer, 1995; Bossert, 1995; Peragine, 1998; Peragine & Ferreira, 2015; Brunori, 2016) considers a population of individuals, each endowed with certain personal characteristics that include circumstances and effort. Circumstances refer to exogenous variables specific to the individual, such as family background characteristics (parents' occupation and education, ethnicity, religion, gender, and place of birth) that are inherited by the individual and for which he should not be held responsible. The circumstances are described as a vector, C, which belongs to a finite set Ω. On the other hand, the effort e considers all factors for which the individual is fully responsible. It is a one-dimensional scalar variable that belongs to a finite set Θ. For each individual, there is an outcome y that is an objective measure of the individual's advantage, such as income, consumption, education, or health. The outcome is generated by a function $g : \Omega \times \Theta \to \mathbb{R}_+$, obtained from the combinations of effort and circumstances: $y = g(C, e)$. In this model, the individual's opportunities, which are not directly observable, can be deduced by observing the joint distributions of circumstances, effort, and outcome.

In a simplified version of this model, the outcome-generating process can be represented by a matrix Y of size $(n \times m)$, where n is the number of *types* - roup of people sharing the same circumstances regardless of the effort exerted - and m is the

number of effort levels. The cells within the matrix are the combinations of circumstances and effort levels that determine an outcome.[2]

Measurement of the inequality of opportunity involves a two-step procedure (Checchi & Peragine, 2010; Ferreira & Gignoux, 2011). In the first step, the original distribution of outcomes, Y, is transformed into a counterfactual distribution, $\widetilde{Y}$, in which all inequality is inequality of opportunity. The second step involves the application of an inequality index to the counterfactual distribution.

In the most prominent approach, also called ex-ante, inequality of opportunity is measured as inequality between types. In other words, the counterfactual distribution $\widetilde{Y}$ is nothing more that a matrix where the elements within each row are replaced by the average of the row. It is then immediate to see that inequality in $\widetilde{Y}$ corresponds to inequality in the distribution of conditional expect incomes $\mathbb{E}[y \mid C]$.

Therefore, the first step fo estimating inequality of opportunity consists in estimating $\mathbb{E}[y \mid C]$. In this paper we use two standard approaches for estimating conditional outcomes: the parametric approach by Ferreira & Gignoux (2011), and the data-driven method based on conditional inference trees and forests Brunori *et al.* (2023c).[3] In what follows, we describe the two methods.

3.2.1 The parametric approach

The ex-ante version of the parametric approach to measure the inequality of opportunity involves estimating the conditional incomes via an OLS regression:

$$\tilde{y}_i = \alpha C_i + \beta e_i + u \tag{1}$$

where $\tilde{x}_i$ is the counterfactual income of individual i, C_i their circumstances, e_i their effort and u is a random component that captures variation due to factors not directly observed.

[2]Usually, a frequency matrix P is associated with the matrix X. Each cell has a value $p_{i,j}$ that informs the proportion of the population that has the exact outcome value obtained from the combination of circumstances and effort levels.

[3]See Brunori *et al.* (2019) and Moramarco *et al.* (2024) for deep discussion of the comparison between standard and data-driven approaches to estimate inequality of opportunity.

First proposed by Bourguignon *et al.* (2007), this method is the most widely employed in the economic literature on the inequality of opportunity. Since circumstances may partially determine effort, the data generating process can be rewritten as $y = g(C, e(C))$. In this sense, the variability attributable to the circumstance vector can be estimated by OLS using the following system of equations:

$$e_i = HC_i + v_i \tag{2}$$

$$\tilde{y}_i = \alpha C_i + \beta(HC_i + v_i) + u_I \tag{3}$$

So that, if

$$\Psi = (\alpha + \beta H) \quad \text{and} \quad \epsilon_i = v_i\beta + u_i$$

then, we can define a reduced form equation that can be estimate using standard OLS techniques:

$$\tilde{y}_i = \Psi C_i + \epsilon_i \tag{4}$$

As many circumstances are likely unobservable, the error terms will not be orthogonal to the regressors, and the coefficient estimates will be biased. However, if one is only interested in identifying inequality of opportunity and does not consider the causal link between circumstances and the outcome, the reduced form can be used.

In the case of a binary outcome, such as literacy, the distribution of the outcomes is determined through logistic regression rather than ordinary least squares (OLS). This technique facilitates the estimation of the probability that an individual will or will not achieve a given outcome based on the available opportunities. After obtaining the probability of being literate, the Gini index is applied.

After computing opportunity inequality, that is, inequality in the outcomes predicted using equation (4), a decomposition analysis can be employed to discern the impact of various circumstances on individual opportunities. In the empirical application, we do this via the well-known Shapley Value decomposition.

3.2.2 The data-driven approach

As shown by Brunori *et al.* (2019) and Moramarco *et al.* (2024), most approaches to estimate the inequality of opportunity, including the parametric one, suffer from upward and downward biases. In general, downward biases are due to missing circumstances,[4] while upward biases appear when there are too many circumstances

[4]This was already observed by Ferreira & Gignoux (2011), more recently, Moramarco *et al.* (2024) showed that this is not the only cause of downward biases.

for the available sample size.

Solving this trade-off ultimately requires a procedure to choose the right number of circumstances on which to condition $\tilde{y}$. Following Brunori *et al.* (2023c), the most recent estimates of the inequality of opportunity (Ferreira *et al.*, 2026) have been produced using machine-learning methods to select the appropriate number and codification of circumstances to estimate the equation (4).

Conditional Inference Trees (CITs), developed by Hothorn *et al.* (2006), have recently been adopted for the estimation of the inequality of opportunity (see Brunori *et al.*, 2023c, for example). CITs are supervised machine-learning algorithms that aim at predicting a dependent variable by partitioning the space of regressors (circumstances). In doing so, they define groups of non-overlapping observations (types) that are homogeneous in terms of the realization of a subsample of the observable predictors. Individual predictions are made taking the average outcome within the groups.

The CIT algorithm takes as input a list of circumstances and a parameter α. At each iteration, the algorithm performs a series of independence tests to identify the circumstance variable that is most associated with the outcome. Formally, for all circumstance variables, it tests the null hypothesis that conditional and unconditional income distributions are statistically indistinguishable.[5] If no p-value is lower than α, then the algorithm stops. Otherwise, the algorithm selects the circumstance variable associated with lowest p-value and performs a binary split of the population according to the possible values of this variable.

The choice of how to split the population is made after performing a series of permutation tests, which in our case corresponds to non-parametric versions of the difference in mean t-test. The split performed is again the one with the lowest p-value. Within each resulting sub-population, the algorithm starts from the independence test phase to attempt further splits. A stopping rule is triggered when the degree of association between the dependent variable and all controls is not significant enough to reject the null hypothesis of independence. In that case, no further split is performed, which prevents the model from growing too deep and overfitting the data.

The result of this algorithm is a partition of the population into types, according

[5]The resulting p-values are Bonferroni-adjusted, which is necessary because multiple hypotheses are tested in this step.

to a subset of all available circumstances. Then, assigning to each individual the average outcome of the type to which they belong, we obtain a data-driven version of $\tilde{y}_i$ in which we can apply standard inequality measures to obtain an estimate of the inequality of opportunity.

The primary limitation of tree algorithms lies in their strong dependence on various factors, such as the choice of the significance level α. To address this issue, an endogenous alpha level is often obtained by applying K-fold cross-validation (Salas-Rojo & Rodríguez (2022)).

As Friedman *et al.* (2009) emphasized, the data structure represents another crucial aspect of these algorithms, encompassing the number of input variables considered, their potential correlation and their distribution. Consequently, predictions derived directly from trees can exhibit significant sensitivity to changes in the data structure. To overcome this limitation, most of the literature considers inequality of opportunity estimates obtained using Conditional Inference Forests.

In brief, a random forest draws different subsamples of the original data and runs CIT on each one. Under the appropriate aggregation procedures, this process smooths the sample dependency and generates robust estimates. For more details, see Brunori *et al.* (2023c) and Ferreira *et al.* (2026).

4 Data description

We used data from the 2015 Integrated Household Survey, carried out by the Gambia Bureau of Statistics (GBOS) between April 2015 and March 2016.[6] The survey collects information from each district of the country and is representative of the population of The Gambia. The analysis sample includes 28,738 individuals over 18 years of age with information about income and circumstances.

The data set provides a comprehensive set of information on socioeconomic circumstances and outcomes. The variables are cleaned and treated as in Ferreira *et al.* (2026), which provides the most recent worldwide estimates of the inequality of opportunity. Our outcome variables are household equivalized disposable income,

[6]The data set was obtained from the World Bank website: https://microdata.worldbank.org/index.php/catalog/3323

adjusted for the age of the household head, and literacy. Literacy is a binary indicator equal to one if the individual is able to read and write in the official language of the country, English, and zero otherwise.

As circumstance variables, we consider gender, region of birth, ethnicity, parental education, and occupation. Tables 5, 6, and 7 in the Appendix offer a detailed description of these aspects. It is evident from the data that male individuals outnumber females. A significant portion of the sample resides in the Kerewan region in the northwest part of The Gambia. In terms of ethnicity, more than 35.2% of the sample identify themselves as Mandinka or Jahanka. When delving into the parental background, it becomes clear that a substantial majority of the population lacks formal education and is predominantly engaged in agriculture. This analysis sheds light on the multifaceted nature of opportunity disparities within the sample.

The outcome variables are detailed in Tables 8 and 9 in the Appendix. The analysis of literacy rates across different cohorts, as presented in Table 9, delineates a significant transformation in the literacy landscape, particularly among women, over successive periods. The sample is segmented into four distinct cohorts, delineating the pre-1965, 1966-1981, 1982-1990, and post-1991 periods. Notably, the number of non-literate individuals demonstrates a decreasing trend, from 4,847 individuals in the earliest cohort to 3,774 in the most recent cohort. In contrast, the count of literate individuals inversely ascends from 1,618 to 3,893 over the same temporal frame. This shift is particularly pronounced in the post-1991 cohort, where the literacy rate among women grows to 46.03%, compared to a modest 9.90% in the cohort preceding 1965. The rising trend in literacy, especially among women, from a mere 9.90% to an impressive 46.03% post-1991 not only marks a pivotal shift, but also reflects the efficacy of educational policies and social changes aimed at enhancing female literacy.

The construction of the cohorts follows the history of The Gambia. The period before 1965 was under British colonial rule. The period between 1966 and 1981 is the first post-colonial period and the early years of the republic, culminating in the 1981 coup. Between 1982 and 1990, the Senegambia Confederation was created and then dissolved, reflecting attempts at economic and political integration with Senegal. Finally, the years after 1991 were the second period of The Gambia's independence. The latter is characterized by a continued political fragility. Indeed, during these years, there was a second coup in 1994.

4.1 Results

Table 1 shows the main results of our analysis on equivalised household disposable income, using the parametric approach. The Gini index, a measure of overall inequality, is 0.53, indicating a significant level of income disparity within the sample. The absolute inequality of opportunity (measured via the Gini index of the conditional income distribution) is 0.1965. As shown by the relative inequality of opportunity measure, this counts for 36.85% of total income inequality.

Table 1: Inequality of opportunity in equivalised household disposable income: parametric estimation

	Gini index
Overall inequality	0.53
Absolute inequality of opportunity	0.1965
Relative inequality of opportunity	0.3685
Circumstance	**Shapley value**
Birthplace	51.79%
Father occupation	7.34%
Father education	2.18%
Sex	0.51%
Mother occupation	8.20%
Mother education	0.98%
Ethnicity	29.00%

The Shapley Decomposition part of the table offers an insightful breakdown of the factors contributing to inequality of opportunity. Birthplace emerges as the most influential factor, accounting for 51.79% of the inequality, highlighting the profound impact of geographic origin on income opportunities. Ethnicity contributes 29% and significantly shapes income opportunities, reflecting the societal divisions and preferences that affect economic outcomesTo[7]. The occupations of both parents also play a significant role, with the Father's occupation and mother's occupation contributing 7.34% and 8.20%, respectively, underscoring the influence of familiar socioeconomic

[7]assess the correlation between *Ethnicity* and *Birthplace*, a Chi-square test was conducted. The results suggest the rejection of the null hypothesis, and corroborated this result, the Cramér's V statistic—a measure quantifying the strength of association between the variables—yielded a value of 0.35, indicating a moderate level of association.

status. The education of the father and the mother further contributes 2.18% and 0.98%, respectively, indicating that parental education is less influential than the occupation. Interestingly, Gender contributes minimally, at 0.51%. This may, however, be the consequence of using household instead of individual level variables.

Table 2 reports our estimates of the absolute inequality of opportunity using data-driven approaches. Our estimates confirm the ability of Random Forest to avoid upward biases. Indeed, the absolute inequality of opportunity is slightly higher when measured using the parametric approach than by machine-learning.

Table 2: Inequality of opportunity in equivalised household disposable income: machine-learning estimation

	Gini index
Conditional Inference Tree	0.232
Random Forest	0.191
Circumstance	**Shapley value**
Birthplace	35.79%
Father occupation	8.68%
Father education	4.40%
Sex	2.31%
Mother occupation	22.41%
Mother education	1.85%
Ethnicity	24.56%

Consistent with the findings of the parametric approach, in the Shapley decomposition based on the machine-learning estimation, Birthplace is identified as the most significant determinant responsible for 35.79% of the inequality. This underscores the critical influence of geographical origin on economic outcomes. Ethnicity also plays an significant role, contributing 24.56% to inequality, reflecting the long-lasting effects of racial and ethnic backgrounds on income opportunities. Following closely, the mother's occupation accounts for 22.41% of the disparities, highlighting the impact of maternal employment status on household income levels. In line with previous analyzes, although impactful, Father's occupation and education exert a comparatively moderate influence (8.68% and 4.40%, respectively). This suggests that while parental background is significant, factors such as the mother's occupation,

along with the individual's Ethnicity and Birthplace, are more critical determinants of income levels. Gender, maintaining its position from previous analyzes, has the minimal impact among the factors examined, contributing only 2.31%.

When comparing the two Shapley decompositions of the parametric and data-driven approaches, the two show some differences in the role of parental background. Both confirm that birthplace and ethnicity explain most income inequality, but the weight of parental education and occupation changes between methods. The parametric model gives more importance to the father's occupation, in line with the traditional dominance of male income sources. The results of machine-learning highlight the growing role of mother's occupation and education, suggesting that maternal work and schooling increasingly influence household opportunities. In general, both sets of results point to the continuing importance of family history in shaping economic outcomes in The Gambia.

When analyzing the inequality of opportunity in literacy, we implement a cohort analysis. Table 3 offers a comprehensive summary of our results based on parametric estimation with a Probit model. This analysis stratifies the sample into cohorts based on birth years: before 1965, 1966–1981, 1982–1990, and after 1991, alongside the entire sample.

The probability of being literate varies significantly across these cohorts, with a clear trend of improvement over time. Indeed, it starts at 25% for those born before 1965 and increases to 51% for individuals born after 1991. Absolute inequality of opportunity, measured by the Gini index, reveals a decline in literacy-based inequality as we move from older to newer cohorts, with the index dropping from 0.4370 in the earliest cohort to 0.1994 in the latest. This suggests a positive trend toward reducing educational disparities over time.

Table 3: Inequality of opportunity in literacy: parametric estimation

	All sample	**Cohort before 1965**	**Cohort 1966–1981**	**Cohort 1982–1990**	**Cohort after 1991**
Probability of being literate	37%	25%	31%	39%	51%
Absolute inequality of opportunity – Gini index	0.2869	0.4370	0.3769	0.3061	0.1994
Shapley decomposition					
Birthplace	27.64%	28.42%	22.08%	30.35%	43.31%
Father occupation	13.30%	3.26%	7.62%	10.18%	13.94%
Father education	9.15%	1.04%	2.84%	6.47%	12.15%
Sex	34.25%	51.18%	54.64%	38.88%	7.14%
Mother occupation	7.70%	2.37%	4.62%	5.95%	8.79%
Mother education	3.76%	0.16%	1.24%	3.43%	4.05%
Ethnicity	4.20%	13.58%	6.97%	4.75%	10.62%

In Shapley, gender emerges as a significant factor, especially in earlier cohorts, with its contribution peaking at 54.64% for the 1966-1981 cohort before dramatically dropping to 7.14% in the post-1991 cohort. This dramatic change highlights the significant steps made towards gender equality in literacy over the years. Birthplace also plays a crucial role, with its influence increasing in the most recent cohort, indicating a growing disparity in literacy based on geographic origin. Parental occupation and education consistently contribute to the inequality of opportunity, although their impacts vary between cohorts. For example, the role of the father's education grows substantially in the latest cohort, suggesting the increasing importance of parental education over time in determining literacy outcomes. Mother's occupation and education, while less influential than sex or birthplace, still significantly impact literacy opportunities. Ethnicity's impact on literacy inequality of opportunity fluctuates across cohorts, with a notable increase in the earliest and latest cohorts, pointing to persistent ethnic disparities in educational outcomes.

Table 4 provides the results of the analysis based on machine-learning estimation. The Gini index shows a clear trend of decreasing inequality of opportunity, from 0.425 in the first cohort to 0.196 in the latest. This reduction indicates significant progress in literacy between generations, with the most substantial decrease observed in the post-1991 cohort, suggesting effective interventions or improvements in educational policies over time.

Table 4: Inequality of opportunity in literacy: machine-learning estimation

	All sample	**Cohort before 1965**	**Cohort 1966–1981**	**Cohort 1982–1990**	**Cohort after 1991**
Types	18	10	10	12	13
Gini index	0.272	0.425	0.352	0.270	0.196
Shapley decomposition					
Birthplace	26.37%	31.71%	25.63%	30.02%	29.99%
Father occupation	19.50%	2.20%	11.93%	18.81%	22.27%
Father education	6.73%	0.00%	0.00%	4.75%	8.60%
Sex	20.75%	45.20%	42.35%	23.02%	6.72%
Mother occupation	14.86%	1.97%	9.46%	14.19%	15.10%
Mother education	2.22%	0.00%	0.00%	0.00%	2.92%
Ethnicity	9.65%	19.38%	11.76%	9.82%	14.37%

Across cohorts, the results based on machine-learning analysis confirm that birthplace remains a persistent source of inequality in literacy, showing that where a person is born continues to shape access to education. However, the sharp decline in the influence of gender points to the success of policies aimed at expanding girls' schooling and reducing the gender gap, such as the 2001 fee-exemption program. The growing role of parental education and occupation in recent cohorts suggests that, as access to basic education has become more equal, family background now matters more for the quality of schooling and the likelihood of reaching higher levels. Overall, these findings indicate that education reforms have improved access, although regional and social disparities still constrain full equality of opportunity.

Comparing the Shapley decompositions from the parametric and machine-learning approaches supports this interpretation. Both methods confirm that birthplace remains the primary driver of literacy inequality, but they differ in how they capture the role of parents. The parametric results give more weight to the father's occupation and education, reflecting the traditional influence of paternal resources on schooling. The machine-learning estimates instead highlight the increasing relevance of the mother's occupation and education, suggesting that women's work and educational experience now play a larger role in shaping children's literacy outcomes. This pattern is consistent with the broader social changes in The Gambia following

the expansion of female education and greater participation of women in economic life.

5 Conclusion

This paper has examined inherited inequalities in The Gambia across two central outcome dimensions, income and literacy, through the lens of the inequality of opportunity framework. Interpreting this framework as a measure of constrained social mobility allowed us to capture the extent to which individual outcomes remain predictable from inherited circumstances.
The analysis combined parametric and machine-learning approaches - using nationally representative data from the 2015 Integrated Household Survey. By doing so, it contributed new evidence from a low-income African context where direct measures of intergenerational mobility are rarely available and where issues of fairness and social mobility are increasingly central to the development debate.

The results demonstrate that inherited circumstances account for a substantial part of income inequality in The Gambia. Birthplace and ethnicity emerge as the strongest determinants, revealing that geography and social origin continue to define individual economic prospects. These findings suggest that opportunity is still largely inherited rather than earned and that social mobility in the income dimension remains limited. Although the overall level of inequality has declined in recent years, our estimates show that roughly one-third of the current income inequality is related to elements beyond individual control. This aligns The Gambia with broader patterns observed in Sub-Saharan Africa, where structural barriers such as rural disadvantage, ethnic stratification, and uneven access to economic opportunities still constrain mobility and reinforce existing divides.

However, literacy findings point to a more dynamic and hopeful trajectory. The cohort analysis reveals tangible progress: younger generations, particularly women, display much lower levels of inherited inequality than their predecessors. The rapid decline in the role of gender among those born after 1991 reflects the success of specific education policies—notably the 2001 scholarship and fee-exemption program for girls—as well as the broader expansion of female education supported by international initiatives such as Education for All. These reforms have reduced gender gaps and improved access to primary and secondary education throughout the country. However, the persistence of birthplace and parental background effects indicates that genuine equality of opportunity has not been fully achieved. Rural and disadvan-

taged regions continue to experience lower literacy rates, showing that while access has expanded, quality and continuity in education remain uneven across space and social groups.

A comparison between the parametric and machine-learning Shapley decompositions further enriches these results. Both methods confirm the central role of birthplace in shaping literacy outcomes, but differ in how they capture the influence of family history. Parametric estimates give greater weight to the occupation and education of the father, consistent with the traditional market and social structures of labor - in which paternal resources predominantly shape the chances of the children. In contrast, the machine-learning model reveals a stronger impact of mother's occupation and education, particularly in younger cohorts. This difference likely reflects recent social transformations: women's participation in education and economic life has expanded, allowing maternal characteristics to increasingly affect the educational outcomes of the next generation. The increasing importance of mothers' education also signals a virtuous cycle in which improvements in women's schooling today translate into better opportunities for their children tomorrow.

Taken together, the evidence supports a two-fold interpretation. First, The Gambia has made significant progress towards a more equitable distribution of educational opportunities, with clear signs that policy interventions can reduce inherited disparities. Second, structural inequalities related to geography, ethnicity, and parental background remain powerful determinants of individual prospects, particularly in the economic sphere. These inequalities are not only distributive but institutional: they reflect how historical, ethnic, and regional divisions have persisted in shaping access to markets, infrastructure, and public services.

From a policy point of view, the results highlight several priorities. Reducing the role of birthplace and ethnicity in determining outcomes requires sustained investment in regional educational infrastructure, teacher quality, and social protection programs that can equalize conditions across regions. Expanding adult and parental education, especially among women, can amplify the intergenerational effects observed in literacy results. In addition, connecting education reforms with employment policies is crucial to ensure that improved schooling actually translates into upward mobility and income gains. Strengthening the link between education, skills, and labor-market demand would help shift the focus from basic literacy achievement to long-term empowerment through work -. In conclusion, the experience of The Gambia illustrates both the challenges and the possibilities of confronting inher-

ited inequality in low-income settings. On the one hand, income inequality remains deeply rooted in contextual and social origins, signaling limited mobility and deep opportunity gaps. However, progress in literacy—particularly the narrowing of gender disparities—demonstrates that targeted policies and sustained political commitment can weaken the transmission of disadvantage between generations. Inherited inequality, therefore, is not a fixed condition, but a symptom of unequal opportunity structures that can be reformed through inclusive development strategies. Continued efforts to make education not only accessible but equitable and of high quality in all regions will be essential to transform current progress into lasting social mobility and a more just future for all Gambians.

References

Blimpo, Moussa P, Evans, David K, & Lahire, Nathalie. 2011. School-based management and educational outcomes: Lessons from a randomized field experiment. *Unpublished manuscript*, **7**.

Blimpo, Moussa P, Gajigo, Ousman, & Pugatch, Todd. 2019. Financial constraints and girls' secondary education: Evidence from school fee elimination in the Gambia. *The World Bank Economic Review*, **33**(1), 185–208.

Bossert, Walter. 1995. Redistribution mechanisms based on individual characteristics. *Mathematical Social Sciences*, **29**(1), 1–17.

Bourguignon, François, Ferreira, Francisco HG, & Walton, Michael. 2007. Equity, efficiency and inequality traps: A research agenda. *The Journal of Economic Inequality*, **5**, 235–256.

Brunori, Paolo. 2016. How to measure inequality of opportunity: a hands-on guide.

Brunori, Paolo, & Neidhöfer, Guido. 2021. The evolution of inequality of opportunity in Germany: A machine learning approach. *Review of Income and Wealth*, **67**(4), 900–927.

Brunori, Paolo, Ferreira, Francisco HG, & Peragine, Vito. 2013. Inequality of opportunity, income inequality, and economic mobility: Some international comparisons. *Pages 85–115 of: Getting development right: Structural transformation, inclusion, and sustainability in the post-crisis era*. Springer.

Brunori, Paolo, Peragine, Vito, & Serlenga, Laura. 2019. Upward and downward bias when measuring inequality of opportunity. *Social Choice and Welfare*, **52**, 635–661.

Brunori, Paolo, Ferreira, Francisco HG, & Neidhöfer, Guido. 2023a. *Inequality of opportunity and intergenerational persistence in Latin America*. WIDER Working Paper.

Brunori, Paolo, Ferreira, Francisco HG, & Salas-Rojo, Pedro. 2023b. Inherited inequality: a general framework and an application to South Africa.

Brunori, Paolo, Hufe, Paul, & Mahler, Daniel. 2023c. The roots of inequality: Estimating inequality of opportunity from regression trees and forests. *The Scandinavian Journal of Economics*, **125**(4), 900–932.

Brunori, Paolo, Ferreira, Francisco HG, & Salas-Rojo, Pedro. 2026. Inherited inequality and the distribution of opportunities inthe United States, China, India,and South Africa.

Carrasco Nunez, Haydeeliz, Jawara, Hamidou, & Meyer, Moritz. 2022. The Effects of Fiscal Policy on Inequality and Poverty in The Gambia.

Checchi, Daniele, & Peragine, Vito. 2010. Inequality of opportunity in Italy. *The Journal of Economic Inequality*, **8**, 429–450.

Cohen, Gerald A. 1989. On the currency of egalitarian justice. *Ethics*, **99**(4), 906–944.

Dworkin, Ronald. 1981. What is equality? Part 1: Equality of welfare. *Philosophy & public affairs*, 185–246.

Ferreira, Francisco HG, & Brunori, Paolo. 2024. Inherited inequality, meritocracy, and the purpose of economic growth.

Ferreira, Francisco HG, & Gignoux, Jérémie. 2011. The measurement of inequality of opportunity: Theory and an application to Latin America. *Review of income and wealth*, **57**(4), 622–657.

Ferreira, Francisco HG, Moramarco, Domenico, & Peragine, Vito. 2025a. Economic development and inequality of opportunity: Kuznets meets the Great Gatsby? *Structural Change and Economic Dynamics*.

Ferreira, Francisco HG, Brunori, Paolo, Neidhöfer, Guido, Salas-Rojo, Pedro, & Sirugue, Louis. 2025b. Inherited Inequality in Latin America.

Ferreira, Francisco HG, Peragine, Vito, Brunori, Paolo, Salas Rojo, Pedro, Moramarco, Domenico, Barajas Prieto, Luis, Barbieri, Teresa, Daza Baez, Nancy, Datt, Gaurav, de Sandi, Vito, *et al.* 2026. Global estimates of opportunity and mobility: a database.

Fleurbaey, Marc. 1994. On fair compensation. *Theory and decision*, **36**, 277–307.

Foltz, Jeremy D, & Gajigo, Ousman. 2012. Assessing the Returns to Education in the Gambia. *Journal of African Economies*, **21**(4), 580–608.

Fosu, Augustin Kwasi, & Gafa, Dede Woade. 2020. Progress on poverty in Africa: How have growth and inequality mattered? *African Review of Economics and Finance*, **12**(1), 61–101.

Friedman, Jerome, Tibshirani, Robert, & Hastie, Trevor. 2009. *The elements of statistical learning: Data mining, inference, and prediction*. Springer-Verlag New York New York.

Gray, John Milner. 2015. *A History of the Gambia*. Cambridge University Press.

Hothorn, Torsten, Hornik, Kurt, & Zeileis, Achim. 2006. Unbiased recursive partitioning: A conditional inference framework. *Journal of Computational and Graphical statistics*, **15**(3), 651–674.

Hughes, Arnold, & Perfect, David. 2006. *A political history of The Gambia, 1816-1994*. University of Rochester Press.

Krishnaratne, Shari, White, Howard, *et al.* 2013. *Quality education for all children? What works in education in developing countries*. Tech. rept. International Initiative for Impact Evaluation (3ie).

Moramarco, Domenico, Brunori, Paolo, & Salas-Rojo, Pedro. 2024. Biases in inequality of opportunity estimates: measures and solutions.

Mungai, Rose, & Okiya, Stephen. 2019. Measuring Inequality of Opportunities in The Gambia.

Murnane, Richard J, & Ganimian, Alejandro. 2014. Improving educational outcomes in developing countries: Lessons from rigorous impact evaluations. *NBER working paper*.

Peragine, Vito, & Ferreira, Francisco. 2015. Equality of opportunity: Theory and evidence. *World Bank Policy Research Paper*, **7217**.

Peragine, Vitorocco. 1998. *ResponsiMility, Compensation and Income DistriMutions*.

Perfect, David. 2008. Politics and society in The Gambia since independence. *History Compass*, **6**(2), 426–438.

Petrosino, Anthony, Morgan, Claire, Fronius, Trevor A, Tanner-Smith, Emily E, & Boruch, Robert F. 2012. Interventions in developing nations for improving primary and secondary school enrollment of children: A systematic review. *Campbell Systematic Reviews*, **8**(1), i–192.

Ramos, Xavier, & Van de Gaer, Dirk. 2016. Approaches to inequality of opportunity: Principles, measures and evidence. *Journal of Economic Surveys*, **30**(5), 855–883.

Rawls, John. 1971. Atheory of justice. *Cambridge (Mass.)*.

Roemer, John E. 1993. A pragmatic theory of responsibility for the egalitarian planner. *Philosophy & Public Affairs*, 146–166.

Roemer, John E. 1998. Equality of Opportunity.

Salas-Rojo, Pedro, & Rodríguez, Juan Gabriel. 2022. Inheritances and wealth inequality: a machine learning approach. *The Journal of Economic Inequality*, **20**(1), 27–51.

Sen, Amartya, *et al.* 1980. Equality of what?

Van De Gaer, DFG. 1995. Equality of opportunity and investment in human capital.

Appendix

Table 5: Socio-demographics - The Gambia 2015 (28738 obs.)

Gender	
Female	14163 (49.3%)
Male	14575 (50.7%)
Region of birth	
Banjul	254 (0.9%)
Basse	5490 (19.1%)
Brikama	4486 (15.6%)
Janjangbureh	4451 (15.5%)
Kanifing	190 (0.7%)
Kerewan	6038 (21.0%)
Kuntaur	4135 (14.4%)
Mansakonko	3694 (12.9%)
Ethnicity	
Bambara	273 (1.0%)
Creole/Aku Marabout	28 (0.1%)
Fula/Tukulur/Lorobo	8370 (29.1%)
Jola/Karoninka	2041 (7.1%)
Mandika/Jahanka	10114 (35.2%)
Manjago	167 (0.6%)
Other	74 (0.3%)
Sarahulleh	2341 (8.1%)
Serer	395 (1.4%)
Wolof	4935 (17.2%)

Table 6: Parental Education - The Gambia 2015 (28738 obs.)

Education levels	**Mother Edu**	**Father Edu**
No school	28239 (98.3%)	27756 (96.6%)
Primary school	239 (0.8%)	185 (0.6%)
Junior secondary school	101 (0.4%)	175 (0.6%)
Upper secondary school	112 (0.4%)	409 (1.4%)
Vocational	34 (0.1%)	86 (0.3%)
University	13 (0.0%)	127 (0.4%)

Table 7: Parental Occupation - The Gambia 2015 (28738 obs.)

Occupation categories	**Mother Occ.**	**Father Occ.**
Accommodation, food service and homely activities	918 (3.2%)	43 (0.1%)
Administrative and service activities	18 (0.1%)	115 (0.4%)
Agriculture, forestry and fishing	26666 (92.8%)	25533 (88.8%)
Arts, entertainment and recreation	13 (0.0%)	20 (0.1%)
Construction, mining and energy supply	1 (0.0%)	292 (1.0%)
Education	97 (0.3%)	746 (2.6%)
Manufacturing	83 (0.3%)	546 (1.9%)
Professional activities (IC, financial, scientific, technical, social)	10 (0.0%)	41 (0.1%)
Public administration and defense	13 (0.0%)	273 (1.0%)
Transportation and storage	6 (0.0%)	299 (1.0%)
Wholesale and retail trader, repair of motor vehicles and motorcycles	913 (3.2%)	830 (2.9%)

Table 8: Equivalized Household Total Income, PPP 2010 - The Gambia 2015

	Min.	Median	Mean	Max.	St.Dev.
Income	0.189	77.39	128.78	6798.28	186.40

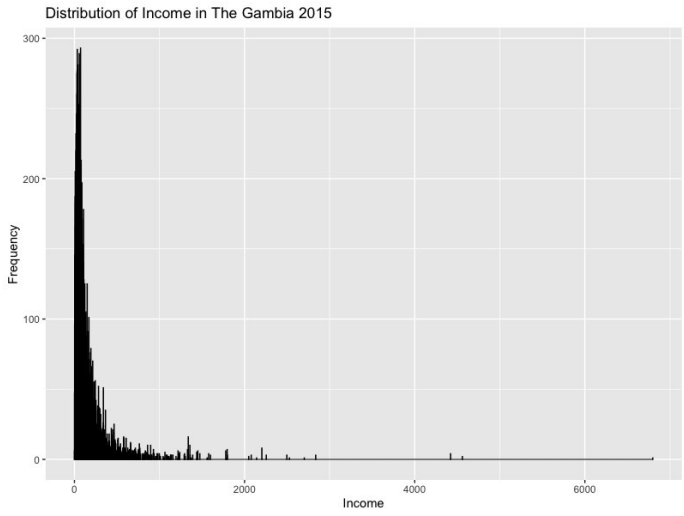


Figure 2: Income distribution in The Gambia 2015

Table 9: Literacy across Cohorts

	All sample	**Cohort before 1965**	**Cohort 1966-1981**	**Cohort 1982-1990**	**Cohort after 1991**
Non Literate	18077	4847	4922	4534	3774
Literate	10587	1618	2210	2866	3893
Total obs.	28664	6465	7132	7400	7667
% literate women	25.70%	9.90%	14.97%	26.06%	46.03%