Dipartimento di Economia e Finanza

Southern Europe Research in Economic Studies

Assessing the extent, the evolution, and the sources of inequality of opportunity in Sierra Leone and The Gambia

Vito Peragine

Giorgia Zotti

SERIES Working Papers n. 06/2024

Assessing the extent, the evolution, and the sources of inequality of opportunity in Sierra Leone and The Gambia

Vito Peragine*, Giorgia Zotti†

Abstract

This paper investigates inequality of opportunity in Sierra Leone and The Gambia, focusing on income and literacy dimensions. Utilizing an ex-ante parametric normative framework combined with regression tree machine learning techniques, the study examines trends in income, consumption, and literacy to elucidate inequality patterns. A detailed Shapley decomposition analysis identifies and quantifies the key factors driving inequality in these domains. A novel aspect of this research is the cohort analysis of literacy, which structures cohorts based on significant historical events in each country. Findings indicate that educational opportunities are more equitably distributed among younger cohorts, signaling a shift toward increased equity. Furthermore, targeted policies such as Sierra Leone's Bunumbu Project and initiatives to reduce school fees for girls in The Gambia have significantly enhanced literacy rates among these cohorts, contributing to a reduction in the urban-rural divide. Despite these improvements, the analysis highlights that gender disparities persist, underscoring the need for continued focus on gender-inclusive policies.

Keywords: Inequality of Opportunity, Literacy and Income Inequality, Educational Policy, Sub-Saharan Africa (SSA)

JEL Codes: I21, I24, I28

Aknowledgement: The authors express their sincere gratitude to Statistics Sierra Leone (Stats SL) and the World Bank for generously providing the data utilized in this study. We are also deeply appreciative of the insightful feedback and constructive suggestions offered by participants at the GEOM Conference and the International Conference on "Poverty, Inequality, and Intergenerational Mobility: Measurement Issues, Empirical Evidence, and Policies," which significantly contributed to refining earlier drafts of this work.

*E-mail address: vitorocco.peragine@uniba.it
Department of Economics and Finance, Università degli Studi di Bari, Largo Abbazia S. Scolastica, 70124 Bari, Italy

†Corresponding author.
E-mail address: giorgia.zotti@uniba.it
Department of Economics and Finance, Università degli Studi di Bari, Largo Abbazia S. Scolastica, 70124 Bari, Italy

1 Introduction

Sub-Saharan Africa (SSA) is recognized not only for its status as the world's poorest region but also for its pronounced inequalities. Indeed, some estimates of the Gini Index, from Aguilar *et al.* (2022), suggest that the average value of the coefficient in the region is equal to 41.5%. The economic inequality that exists in SSA is an obstacle to economic and political development and poverty alleviation.
Moreover, the literature on inequality of opportunity (IOp) in the SSA context offers critical insights into the distribution of wealth and resources. Ferreira *et al.* (2018) contributes to the existing literature by estimating inequality of opportunity in many countries (10 are in SSA) over the period 1980-2005, focusing on wealth as a key outcome variable. Meanwhile, Brunori *et al.* (2019a) extends this research by specifically assessing IOp in 10 SSA countries using data ranging from 2000 to 2013, marking a significant focus for the region. They find that IOp accounts, on average, for 47% of the inequality in consumption. In addition, a report from the World Bank (Bank (2022)) extends the analysis to southern Africa by examining IOp in five countries. However, its analysis does not include parental characteristics. Despite this limitation, the report illustrates the extent of IOp, with figures ranging from 15 to 26%, rising to 46% in South Africa when racial disparities are taken into account. Tetteh-Baah *et al.* (2024) then enlarges the survey to horizontal inequality across Africa, assessing various outcomes by location, ethnicity, gender, and religion without including income or consumption. Atamanov *et al.* (2024) set a new standard by providing insights that enable the evaluation of gender inequality. Their research investigates inequality of opportunity across 18 countries in Sub-Saharan Africa, unveiling that such inequality is significantly more severe and widespread than earlier assessments suggested.

The literature reports a correlation between IOp and economic growth (Marrero & Rodríguez (2013); Bradbury & Triest (2016); Brunori *et al.* (2013)) supporting this relationship[1]. This perspective is particularly relevant for severely poor Sub-Saharan African (SSA) nations, including Sierra Leone, underscoring the need for targeted measures to decrease IOp to promote economic progress.

This paper aims to enrich the existing literature by examining the IOp in terms of income and literacy in Sierra Leone and The Gambia, focusing on the analysis of data collected through Integrated Household Surveys. Two distinct methodologies are employed to assess the IOp. The first adopts a normative approach using a parametric technique, while the second is a data-driven approach applying machine learning techniques. The normative approach encompasses the ex-ante parametric methodology[2], while the machine learning approach consists of a random forest algorithm to measure inequality[3]. This latter method is also applied in the second part of the study to analyze the evolution of the literacy rate through cohort analyses. The cohorts are defined based on significant political events of each country.

[1] Ferreira *et al.* (2018) presents a contrary view, finding no correlation between IOp and growth.

[2] Ferreira & Gignoux (2011); Ramos & Van de Gaer (2016); Brunori *et al.* (2013); Ferreira & Gignoux (2014)

[3] Brunori *et al.* (2018); Brunori & Neidhöfer (2021); Brunori *et al.* (2023a); Brunori *et al.* (2023b)

The analysis of inequality of Opportunity arose in Western societies from the 1970s to the 1990s, with notable contributions from Roemer (1998) and Checchi & Peragine (2010). Its philosophical support, deeply rooted in political philosophy and social justice, was significantly shaped by scholars like Rawls (1971), Dworkin (1981), Sen *et al.* (1980), and Cohen (1989). These pioneers laid initial definitions and methodologies for measuring inequality of Opportunity.
Inequality of Opportunity emerged as a critique of consequentialism (Peragine & Ferreira (2015)), which emphasizes the distribution of outcomes and the processes through which those outcomes are achieved. At its core, inequality of Opportunity reflects the concept of social justice, which aims to ensure equal access to a society's wealth and opportunities for all individuals.
To ensure a just and equitable distribution of opportunities and resources within society, examining the factors influencing individual outcomes is imperative. The attainment of success is shaped by two key elements: circumstances and effort (Roemer (1998)). The former are factors beyond an individual's control, such as family background and natural disasters. On the other hand, effort originates from personal responsibility and deliberate choices made by individuals. People who share the same circumstances can be categorized into types, while those who exert similar levels of effort can be grouped into tranches.

Equality of opportunity hinges upon distinguishing between fair and unfair inequalities, as certain inequalities are considered more ethically acceptable than others (Ferreira & Peragine (2013)). This concept is guided by two fundamental principles: the compensation principle and the reward principle. The former asserts that inequalities arising from diverse circumstances are ethically unfair. Conversely, the reward principle addresses assessing economic inequality from different levels of effort.
Within the inequality of Opportunity, the compensation principle can be understood through two distinct approaches: ex-ante and ex-post (Peragine & Ferreira (2015)). The ex-ante approach endeavors to equalize opportunities before the realization of outcomes. By contrast, the ex-post approach focuses on equalizing outcomes after efforts and results have materialized. Both approaches provide distinct perspectives on addressing the compensation principle within the framework of inequality of Opportunity. Indeed, the ex-ante focuses on equality of Opportunity between sets of individual circumstances, while the ex-post aims at equality in outcomes resulting from exercising the same level of effort.
Parallel to the compensation principle, the reward has two prominent interpretations: the utilitarian and the liberal. According to the latter, individuals should be rewarded based on their efforts. Conversely, the utilitarian interpretation suggests that rewards should be distributed to maximize total well-being.
Studying the intricate link between inequality of Opportunity and education is critical to understanding the dynamics of socioeconomic inequality. Indeed, inequality concerns the disparities in well-being experienced by individuals within a society. These disparities are mainly attributable to variations in income levels, consumption, access to health care, education, and life expectancy. Important scientific evidence points to the strong interconnections between economic inequality, inequality of Opportunity, and educational inequality (Ferreira & Walton (2005); Bourguignon *et al.* (2007)).

Inequality in education is widely recognized as a key factor contributing to the growing income disparity (Petcu (2014)). This understanding is rooted in the neoclassical model of human capital proposed by Schultz (1961) and Becker *et al.* (1964), which suggests that income inequality arises from inequality in investments in human capital, particularly in education. Education, as argued by Weiss (1995), is often considered a signal of productivity, leading to higher wages.

However, despite the theoretical consensus linking educational inequality to income inequality, empirical evidence presents a more complex picture. Research findings on this topic appear to be conflicting. For instance, Keller (2010), using data from the World Inequality Database, finds that among individuals with primary and secondary school education levels, income inequality is lower compared to those with tertiary education levels. Indeed, tertiary education increases income inequality, particularly in developing countries. Similarly, Dao *et al.* (2013) identifies a positive association between educational inequality and income inequality in nineteen developing and emerging countries in sub-Saharan and northern Africa. In contrast, Checchi (2004) and Földvári & van Leeuwen (2011) find no significant correlations between income inequality and educational inequality based on data from Organisation for Economic Co-operation and Development (OECD) countries as well as other regions, including sub-Saharan Africa, North Africa, the Middle East, and South America. Moreover, most studies examining these opportunity disparities have primarily focused on developed countries, particularly OECD countries (Földvári & van Leeuwen (2011)) or Latin American countries (Bourguignon *et al.* (2007)).

Regarding the connection between economic inequality and educational outcomes, recent studies shed light on the close association between disparities in academic achievement and income distribution. Reardon (2011) presents findings suggesting that economic inequality worsens the achievement gap in education in the United States. Anderson *et al.* (2010), utilizing student-level data from the OECD's International Student Assessment for OECD and non-OECD countries, demonstrates that economic inequality hurts mathematics performance. Similar results are observed in the United States by Condron (2011). However, it should be noted that empirical evidence specifically focused on sub-Saharan African countries remains limited (Brunori *et al.* (2019a); Atamanov *et al.* (2024)).

The paper is organized as follows: the next sections provide a brief historical context of Sierra Leone and The Gambia. The subsequent section provides a comprehensive overview of the methodology employed. Subsequently, section 3 describes the datasets implemented for the analysis. In section 4, the results are presented, followed by the conclusions.

1.1 Sierra Leone Context

Sierra Leone is situated in the southern region of West Africa's Sub-Saharan area (Figure 1). It was named by the Portuguese explorer Pedro de Sintra in the 15th century, who was the first European to map and record the harbor of Freetown. The name originally given by the Portuguese, Serra Lyoa, meaning "Lion Mountains" was inspired by the hills encircling the harbor (Crowder (2013)).

Figure 1: Sierra Leone Map

In 1961, Sierra Leone declared independence from the British Empire, ending a period of colonial rule. This new era was soon met with political challenges, as the nation's main political parties, the Sierra Leone People's Party (SLPP) and the All People's Congress (APC), found themselves in stark opposition. In 1967, the APC led by Stevens won the elections, initiating a tumultuous period characterized by economic downturns and political instability. Stevens' politics was a turbulent period, marked by economic and political crises. In the early 1980s, new elections were conducted in a guerrilla atmosphere, which saw Stevens lacking support. Thus, Momoh was elected, whose tenure similarly grappled with economic difficulties and was marred by several coup attempts, reflecting the continuing political volatility.

The early 1990s marked what is often called a difficult decade for Sierra Leone, dominated by the emergence of the Revolutionary United Front (RUF). This rebel group plunged the country into a series of brutal civil conflicts. These battles were not just political skirmishes but escalated into full-blown civil wars, devastating the nation and its people.
A significant turning point came in 2002 when the RUF was finally disarmed. This was partly due to the intervention of the United Nations, which was crucial and gradually inaugurated a period of political stabilization and improved governance. The following years saw a slow but steady recovery of Sierra Leone's political landscape, which, in turn, began to reflect positively on the living conditions of the population. Thanks to international

support and a renewed commitment to peace and democracy, Sierra Leone has embarked on a journey of reconstruction and healing, seeking to overcome the shadows of past conflicts.

After the civil war that struck Sierra Leone in the 1970s, education was deemed crucial for the country's recovery and development. Three years before becoming an independent nation, the government began laying the foundation to address issues of equity and access to education, anticipating the challenges that a universal primary system would face without post-graduation opportunities (Pai (2013)). The Bunumbu Project in 1974[4] had the aim to make education more relevant for rural communities by developing a primary curriculum with a rural orientation and training teachers equipped with modern pedagogical skills. This project aspired to overcome the lack of qualified teachers, obsolete teaching methods, and inadequate educational resources by focusing on practical programs and a closer integration between school life and community life. Despite intentions to diversify secondary education and reduce the mismatch between educational output and labor market demand by increasing agricultural productivity and rural employment, there were obstacles (Lebby & Lutz (1982)). The merger of education and development, although aimed at promoting rural development and combating disparities, led to disillusionment and frustration among community members, some of whom would have preferred a more traditional academic education over the rural-oriented curricula proposed by the Bunumbu Project (Sleight (1964); LEONE *et al.* (1975); Pai (2013)). These reforms, designed to reflect the social and ideological needs of the population, highlighted the importance of innovative educational projects aimed at improving the living, economic, and cultural conditions of the country.

The evolution of educational policies in Sierra Leone has followed a path aimed at supporting access, quality, and relevance of education in response to the country's socioeconomic and cultural needs[5]. Starting with the Constitution of 1991, last amended in 2008, which does not recognize education as a human right. Still, as a fundamental principle of state policy, subsequent laws and reforms have gradually sought to expand access to education and improve its quality. The National Council for Technical, Vocational, and Other Academic Awards Act of 2001 established an independent entity to validate and certify awards in technical, vocational education, and teacher training. The subsequent Education Act of 2004 made basic education compulsory and free, imposing penalties on parents neglecting their children's education and establishing the right to basic education for all citizens. With the Child Rights Act of 2007, every child gained the right to education, with special attention to disabled children, ensuring them special care and educational opportunities. The introduction of the Sierra Leone Teaching Service Commission Act in 2011 aimed to improve the professional status and economic well-being of teachers. The Sierra Leone Education Sector Plan (2007-2015) outlined an overall strategy to build on the post-war education gains, aiming for universal primary education and improvement of its quality, demonstrating a growing commitment towards inclusive and sustainable educational development in the country.

[4] https://thedocs.worldbank.org/en/doc/952781614597519779-0240021979/original/WorldBankGroupArchivesFolder113525I.pdf

[5] Alltheinformationinthisparagraphderivedfrom:https://www.right-to-education.org/sites/right-to-education.org/files/resource-attachments/RTE_country_factsheet_Sierra_Leone_January_2016_0.pdf

Moreover, Sierra Leone faces significant challenges related to gender discrimination, negatively affecting their education. The government implemented policies to improve access to and the quality of female education despite conflicting laws that prohibit child marriage but allow minors under 18 to marry with parental consent. This creates a complex situation that requires further intervention to ensure the protection and education of girls, underlining the need for targeted efforts to address these specific challenges as part of the broader commitment to educational reform and gender equality.

1.2 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 2). This Country has an eventful geopolitical history, strongly influenced by the legacy of the British Empire.

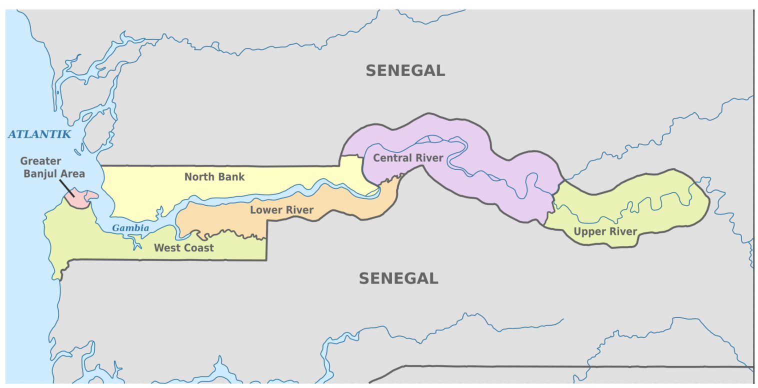


Figure 2: Tha Gambia Map

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 the most prosperous country in combating poverty and inequality. Fosu & Gafa (2020) showed this regarding the period between 1998 and 2003. Subsequently, Mungai & Okiya (2019) demonstrated how, between 2010 and 2015, The Gambia experienced an overall 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 decade as limited despite implementing fiscal policies to reduce inequality and poverty. However, the poverty rate has increased by 5.3%, while the Gini index has decreased compared to previous years. This finding is also confirmed by estimates of the Gini index conducted by the World Bank [6]. It appears to have reduced by 10% from 1998 (48.5%) to 2020 (38.8%). This suggests a decline in economic inequality despite persistent challenges.

Since gaining independence in 1965, The Gambia has experienced modest economic growth yet 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 examination, administered by the West African Examination Council (WAEC), can enter university.

In 2001, it 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), Blimpo *et al.* (2019)). This move 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 percentage of the population educated.

2 Methodology

Two methodologies are employed to assess the inequality of Opportunity in income and literacy in The Gambia, and each is implemented in the theoretical framework of the inequality of Opportunity. The first method involves a normative approach, employing a parametric technique in its ex-ante version. Alternatively, the analysis can be data-driven, leveraging regression trees and random forests to understand patterns and disparities in the data.

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

Both these methodologies are used to analyze both outcome indicators. 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, thus 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.

2.1 The Normative Approach

The canonical model to study and define the inequality of opportunity[7] considers a population of N individuals, each endowed with certain personal characteristics comprising of 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 $C = (c_1, ..c_n) \in \Omega$. On the other hand, the effort, e considers all factors for which the individual is fully responsible. It is a scalar, one-dimensional variable belonging to $(e_1, ..e_m) \in \Theta$. For each i, there is an outcome x which 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 $x = g(C, e)$.

In the model, the individual's opportunities, which are not directly observable, can be deduced by observing the joint distributions of circumstances, effort, and outcome. The outcome-generating process can be represented by a matrix X of size $(n \times m)$, where n is the number of circumstances (rows) and m is the number of effort levels (columns). Each row identifies a *type*, a group of people sharing the same circumstances regardless of the effort exerted. Similarly, each column identifies a *tranche*, a group of individuals who exert the same level of effort regardless of their circumstances. The cells within the matrix, $x_{i,j}$, are the combinations of circumstances and effort levels that determine an outcome.
Usually, a frequency matrix P is associated with the matrix X. Each cell has a value $p_{i,j}$ which informs the proportion of the population having the exact outcome value obtained from the combination of circumstances and effort levels.

The measurement of inequality of Opportunity entails a two-step procedure. In the first step, the original distribution of outcomes, X, is transformed into a counterfactual distribution, $\widetilde{X}$, which fully accounts for the inequality of Opportunity. The second step involves the application of an inequality index to the counterfactual distribution. Notably, Checchi & Peragine (2010) and Ferreira & Gignoux (2011) have proposed a measure of inequality

[7]From Peragine & Ferreira (2015) and Brunori (2016) and based on: Roemer (1993); Van De Gaer (1995); Fleurbaey (1994); Bossert (1995); Peragine (1998).

of Opportunity that involves generating a counterfactual distribution through an ex-ante approach, utilizing both non-parametric and parametric methods.
When using the ex-ante approach to measure inequality of Opportunity, the focus is on inequality between types. Consequently, the counterfactual distribution represents inequality of Opportunity between people sharing the same circumstances, ignoring inequality within each group.

To assess the extent of inequality of Opportunity between different types, it is crucial to identify pertinent circumstances, including gender, ethnicity, socioeconomic background, parental education, or geographical location. The objective is to create groups with similar circumstances capable of influencing opportunities. Types are identified on the basis of ethical and normative considerations. Once the types have been identified, an outcome variable of interest is chosen to assess inequality of Opportunity. This variable might encompass educational attainment, income, health outcomes, or any other relevant metric that reflects the range of opportunities accessible to individuals.
The assessment of inequality between types facilitates an examination of the portion of overall outcome inequality that can be attributed to disparities between distinct groups or types within a population. It offers insights into the extent and nature of inequalities arising from circumstances beyond an individual's control, such as social background or demographic factors. By identifying and quantifying these disparities, policymakers and researchers can develop a deeper understanding of the sources of inequality and devise targeted interventions to reduce disparities and promote equal opportunities for all types or groups within society.

2.1.1 Ex-Ante Parametric Approach

The ex-ante version of the Parametric Approach to measure the inequality of Opportunity involves estimating the outcome-generating function $x_{i,j} = g(c_i, e_j)$ via OLS regression:

$$\tilde{x_{i,j}} = \alpha c_i + \beta e_j + u \tag{1}$$

First proposed by Bourguignon *et al.* (2007), this method is the most widely employed in the economic literature on the inequality of Opportunity. Again, the value of the outcome is generated by c_i, the vector of circumstances, e_j, the vector of effort, and u is a random component that captures variation due to factors not directly observed. Since circumstances partially determine effort, the Equation can be rewritten as $x = g(c_i, e_j(c_i))$, where v is once again a random component. In this direction, the variability attributable to the circumstance vector can be estimated by OLS using the following system of equations:

$$e_j = Hc_i + v_j \tag{2}$$

$$\tilde{x_{i,j}} = \alpha c_i + \beta(Hc_i + v_j) + u \tag{3}$$

$$\text{if} \quad \Psi = (\alpha + \beta H)c_i \quad and \quad e = v_j\beta + u$$

then:

$$\tilde{x_{i,j}} = \Psi c_i + e_j \tag{4}$$

Equation (1) describes the process of reaching the outcome. 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.

Estimating Equation (4) is crucial. The first term represents the direct effect of circumstances on the outcome, and the second is the direct effect of effort on the outcome.

After calculating the contractual distribution, an inequality index is applied on X than on $\tilde{x_{i,j}}$. The value obtained on $\tilde{x_{i,j}}$ represents the amount of absolute inequality due to Opportunity and effort. To obtain the relative inequality, the ratio between the two indices is calculated, which gives the percentage of inequality due to circumstances.

The paper implements the ex-ante version of the parametric approach, in which effort levels are not taken into account but only the circumstances available.

$$\tilde{x_{i,j}} = \Psi c_i \tag{5}$$

In the scenario of a binary outcome, such as literacy, the distribution of 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 opportunities available. After obtaining the probability of being literate, the Gini index is applied.

After computing opportunity inequality using the parametric ex-ante method, a decomposition analysis can be employed to discern the impact of various circumstances on individual opportunities. It is essential to know how much the circumstances considered in the analysis influence individual opportunities.

The inequality of opportunity value obtained through Equation (5) does not expressly inform about the impact of each circumstance. It simply assesses how much inequality in outcomes due to circumstances amounts to. In situations where the need arises to decompose opportunity inequality computed for a continuous outcome, such as consumption, the Shapley decomposition becomes applicable. The Shapley decomposition presents notable advantages compared to alternative decomposition methods. Notably, it is order-independent, and the sum of its constituent components equals the overall value.

Shapley's Value, originally formulated in the context of cooperative game theory, offers a method for fairly distributing the gains (or losses) of collective action among all participants based on their individual contributions. When this concept is applied to the decomposition

of inequality by factor components, it seeks to understand how different sources of the outcome considered (income, literacy) contribute to a society's overall inequality[8]. In detail, applying the Shapley Value to the decomposition of inequality involves examining how the elimination of each source of outcome affects total inequality. Consider different sources of outcome, such as income wages, investments, and social benefits, that contribute in different ways to a country's economic inequality. To understand the specific weight of each source in overall inequality, assume that we remove one source of the outcome at a time and observe how inequality changes (Chantreuil & Trannoy (1997)).
The main problem here is that there is no natural order for removing these sources of income. Thus, the effect of removing one source and the other might be different from the effect of removing them in reverse order (Chantreuil & Trannoy (2011)). Accordingly, Shapley's Value proposes to consider all possible sequences by which outcome sources could be removed and then average the impacts of these removals. In practice, the effect of each individual outcome source on overall inequality is analyzed, taking into account all possible combinations in which income sources could be eliminated, and then an average is calculated.

2.2 The Regression Trees and Forest Approach

Machine learning (ML) is a computational technique in data-driven analysis processes. It utilizes algorithms to extract information, identify patterns, and make statistical decisions with minimal human intervention, thus avoiding common errors made by researchers, such as variable selection, assessing significance, or discretizing continuous variables.

Measuring inequality of opportunity attempts to assess how much circumstances influence inequality in economic or social outcomes among individuals. There are two causes, identified by Brunori *et al.* (2019b), that allocate the underestimation and overestimation of inequality of opportunity. Underestimation of inequality of opportunity stems from the difficulty of observing and measuring all possible circumstances that might influence individuals' outcomes. Many of these circumstances may not be captured by the available data or may be mismeasured, thus leading to a downward assessment of real inequality of opportunity. The second source of bias concerns the sampling variance of the estimated counterfactual distribution. When trying to measure inequality of opportunity, a counterfactual distribution is compared to the original distribution of outcomes to estimate inequality of opportunity. However, because this counterfactual distribution is based on estimates derived from sample data, it is subject to sampling variance. This sampling variance may overestimate inequality of opportunity.
Machine Learning techniques do not exhibit these biases, and they aim to study the relationship between a dependent variable, such as annual income and circumstances, through algorithms.
These algorithms are used to construct conditional inference trees Hothorn *et al.* (2006), followed by forests of trees. Conditional inference trees are executed once on the data, while

[8] Shorrocks (1982); Young (1985); Chantreuil & Trannoy (1997); Chantreuil & Trannoy (1999); Chantreuil & Trannoy (2011)

forests of trees employ a bootstrap approach, repeating the process multiple times to reduce and rectify potential measurement errors.
To fully understand the algorithm's functioning, it is essential to first understand the role of conditional inference trees and then delve into the role of forests of trees.

2.2.1 Conditional Inference Tree

Tree-based algorithms partition datasets into groups based on statistical criteria, enabling sequential and hierarchical decision-making. Subsequently, each observation is assigned the average value of the dependent variable within its respective group. In the specific context of measuring inequality of opportunity, this technique divides the data into partitions based on observable circumstances, creates types within these partitions to further group observations with similar characteristics, and then assigns each observation the average value of its type-specific outcome.
In this case, the selection of trees is based on conditional inference algorithms, which are not limited to partitioning continuous variables exclusively Hothorn *et al.* (2006). These algorithms have already been employed in other studies concerning the measurement of inequality of opportunity (Brunori *et al.* (2019a), Brunori *et al.* (2019b), Brunori & Neidhöfer (2021)).

The algorithm consists of four stages.
(I) The first involves selecting the significance level for a hypothesis test between the outcome variable, w, and each input variable C (circumstance). Meanwhile, let x represent one among the potential values in the continuous variable, and z denote the resulting subsamples.

$$H_0^C = D(w|C) = D(w) \tag{6}$$

Subsequently, the Bonferroni correction is applied to adjust the p-values, reducing the risk of type I errors.

$$p_{adj} = 1 - (1-p)_c \tag{7}$$

(II)The input variable (circumstance) with the lowest adjusted p-value is selected as the possible splitting variable. If the selected p-value is lower than α, then the population is split according to that circumstance.

$$p_{adj} < \alpha \tag{8}$$

(III) Cut points are identified after determining whether a circumstance serves as a splitting variable. Binary variables require straightforward splitting, while continuous variables necessitate evaluating all potential subdivisions.

$$w_z = w_i : C_i < x \tag{9}$$

$$w_{-z} = w_i : C_i \geq x \tag{10}$$

(IV) This process is iterated for subsamples until no input variable significantly correlates with the outcome variable. The depth of the resulting regression tree depends on the chosen significance level. When the null hypothesis of independence cannot be rejected, the algorithm terminates, and the tree is constructed. Finally, the algorithm assigns each observation its expected wealth conditioned on the membership type.

The primary limitation of tree algorithms lies in their strong dependence on various factors, such as the choice of significance level. To address this issue, an endogenous alpha level is often utilized through the application of 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 considered input variables, their potential correlation, and their distribution. Consequently, predictions derived directly from trees can exhibit significant sensitivity to changes in the data structure. While tree methods typically demonstrate favorable performance within the sample, legitimate concerns arise regarding their validity beyond the sample. Therefore, to ensure the robustness of results, it is necessary to complement the analysis with a more comprehensive technique.

2.2.2 Forest Approach

Conditional inference Forests represent an advanced methodology in statistical analysis, evolving from Conditioned inference trees. This concept relies on bootstrapping, a resampling technique, to generate multiple estimates and evaluate variations in the results through repeated tests of the independence hypothesis on different trees. The key idea is to create multiple decision trees and observe how the conclusions vary among them to determine the consistency of the results.

The methodological approach of Conditional inference Forests begins with a random selection of circumstances. This is essential to minimize the risk of introducing variables that are not actually independent or that could lead to bias in the analysis. Next, the construction of a large number of inference trees, which can range from 100, 200, or even 500, ensures the robustness of the analysis. This process exploits different samples or sets of variables for each tree, thus ensuring that the results are robust and not affected by arbitrary selection of a particular sample or set of variables. The final step involves the application of an endogenously adjusted alpha value, which is critical in establishing the statistical significance of the results. This alpha value helps confirm that the inferences drawn from the analysis are valid and reliable, making the entire process a key pillar for empirical research that aims to generate robust and defensible statistical conclusions.

Disparities between the constructed trees' results are analyzed and resolved through the law of large numbers, producing a sample distribution that closely approximates the true distribution of the data. This process, as illustrated by Schlosser *et al.* (2019), provides a comprehensive and reliable view of the analyzed data.

Brunori & Neidhöfer (2021) explore the factors influencing the construction of these forests, emphasizing the importance of the random selection of circumstances and the construction of large numbers of trees to ensure accurate and meaningful analyses. This methodological approach offers a powerful tool for the analysis of complex data, allowing the identification of significant patterns and relationships within the data with a high degree of reliability.

	Sierra Leone	The Gambia
Dataset	SL Integrated Household Survey 2018	Gambia Integrated Household Survey 2015
Outcomes	Consumption, Literacy	Income, Literacy
Circumstances	Birthplace (14 regions) Parental Education (8 levels) Parental Occupation (14 categories) Gender (F, M) Religion (11 clusters)	Birthplace (8 regions) Parental Education (6 levels) Parental Occupation (11 categories) Gender (F, M) Ethnicity (10 clusters)
Obs.	10947	28738
Literates	3905	10587

Table 1: Dataset characteristics comparison between Sierra Leone and The Gambia.

3 Data description

The datasets implemented in this paper are sourced from (i) the World Bank website[9], namely the Gambia - Integrated Household Survey 2015, carried out by Gambia Bureau of Statistics (GBOS), and (ii) from the Sierra Leone Integrated Household Survey (SLIHS) conducted by the National Statistics Institute of Sierra Leone 2018. Each survey collects information from each district and is nationally representative. Both analysis samples include individuals over 18 with information about income and circumstances.

These datasets provide a comprehensive range of information (Table 1) encompassing various socioeconomic circumstances, including Gender, Ethnicity, region of birth, parental education, occupation, total household income, relationship to the head of the household, age, household size, personal education, occupation, weights, and literacy.
Some variables are implemented to fill in the missing data. For example, if parental education information is missing, household information is imputed. In the case of the *household size* variable, it plays an important role in establishing an equivalence scale for calculating annual family income. Following the recommendation by Cowell & Van Kerm (2015), working with households of different sizes requires using equivalence scales derived from the same dataset, even if there may be ambiguity regarding their construction. *Weights* to ensure that the sample is representative of the population.

In the analyses, the variables of interest are the household's reported annual income and literacy status. The latter is coded as a binary variable in both datasets, where it assumes a value of 1 if the individual can read and write in the official language of the country and 0 if they cannot. In Sierra Leone, consumption is used instead of income.
Considering income (consumption), each household's value is divided by the square root of its size, a technique endorsed by Buhmann *et al.* (1988). This adjustment ensures that the income figures are comparable across households of varying sizes. Further improvement of

[9]https://microdata.worldbank.org/index.php/catalog/3323

Sierra Leone	Overall Sample	Colonial Period (before 1967)	Independence (1968-1982)	Momoh Government (1982-1992)	Political Fragility (after 1992)
Total Obs.	10939	2694	3205	2489	2551
% Literate	35.7%	17.89%	27.3%	37%	63.78%
The Gambia	**Overall Sample**	**Colonial Period** (before 1965)	**Independence** (1966-1981)	**SenGambia Confederation** (1982-1990)	**Political Fragility** (after 1991)
Total Obs.	28664	6465	7132	7400	7667
% Literate	36.93%	25%	31%	38.7%	50.8%

Table 2: Literacy rates in Sierra Leone and The Gambia

the consumption data is achieved by applying the CPI_PPP 2017 index for price level adjustments and conducting a regression analysis to account for the age of the household head.

In the analysis of IOp, the circumstances considered are the following: Gender, region of birth, Ethnicity, parental education, and occupation. In Sierra Leone, the religion variable is used as a proxy for ethnicity. Indeed, as defined by Smith (1978) there is a correlation between the two dimensions.
Tables 6 to 8 in the Appendix offer a detailed description of the circumstances in Sierra Leone and The Gambia. It is evident from the data that male individuals outnumber females. When delving into 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 analysis of literacy rates across different cohorts, as presented in Table 2, delineates a significant transformation in the literacy landscape, particularly among women, over successive periods. The sample is segmented into four distinct cohorts according to the historical events that each country experienced. Notably, the number of non-literate individuals demonstrates a decreasing trend from the earliest cohort to the most recent cohort. In contrast, the count of literate individuals inversely ascends over the same temporal frame.

4 Results

4.1 Inequality of Opportunity in Income/Consumption

In this section, the empirical findings from the ex-ante parametric approach and the machine learning techniques focus on the IOp in income and consumption.

Table 3 provides a quantitative assessment of the overall income inequality through the lens of the Gini index. Meanwhile, Table 4 shows the core results of the income distribution analysis by applying (i) the Ex-Ante version of the Parametric method in the Normative Approach framework and (ii) the results of the machine learning techniques. Finally, Table 5 shows the Shapley Decomposition, where the determinants of IOp in income are identified.

The regression applied to build the counterfactual distribution in the Ex-Ante Parametric Normative Approach is the following:

$$\tilde{x}_{i,j} = \alpha+\beta_1\cdot\text{Gender}_i+\beta_2\cdot\text{Ethnicity}_i+\beta_3\cdot\text{Birthplace}_i+\beta_4\cdot\text{Mother\_edu}_i+\beta_5\cdot\text{Father\_edu}_i+\beta_6\cdot\text{Mother\_occ}_i+\beta_7\cdot\text{Father\_occ}_i \tag{11}$$

Indeed, the Gini index stands at 34% in Sierra Leone and approaches 55% in The Gambia, highlighting notable income disparities within the sample. The Normative Approach analysis shows that inequality attributable to differing opportunities accounts for 55% of the overall inequality in Sierra Leone and 36% in The Gambia. Additionally, findings from random forest techniques corroborate these results, indicating that factors beyond individual control significantly contribute to income inequality.
This analysis demonstrates the derived Gini index values from these models, highlighting Random Forest's ability to avoid the problem of overfitting. Indeed, absolute IOp is slightly higher when measured by the parametric approach than by machine learning.
Figures 5 and 6 in the Appendix illustrate the Conditional Regression Trees resulting from the machine learning technique, which reveals the presence of 17 and 19 distinct nodes, respectively, for Sierra Leone and The Gambia. Each node represents a specific portion of the population, characterized by sharing the same circumstances, and identifies its correlated average income.

	Sierra Leone	**The Gambia**
Overall Gini Index	0.3397	0.5442

Table 3: Gini Index Comparison between Sierra Leone and The Gambia

	Ex-Ante Parametric Approach			**Regression Tree Approach**	
	Sierra Leone	**The Gambia**		**Sierra Leone**	**The Gambia**
IOp Gini	0.1853	0.1965	**Ctree Gini**	0.173	0.232
Relative IOp	54.55%	36.10%	**Random Forest Gini**	0.156	0.191

Table 4: Comparison of Gini Indices using Different Approaches for Sierra Leone and The Gambia

The Shapley decomposition (Table 5 and Figure 3) further explains the relative importance of various sociodemographic factors in accounting for the observed income disparities. For both countries, Birthplace is identified as the most significant determinant responsible for inequality. This underscores the critical influence of geographical origin on economic outcomes. In The Gambia, Ethnicity also plays a significant role, contributing 24.56% to the inequality, thereby reflecting the long-lasting effects of racial and ethnic backgrounds on income opportunities. This reflects the societal divisions and preferences that affect economic outcomes[10]. While the role of religion is more marginal in Sierra Leone.

[10]To assess the correlation between *Ethnicity* and *Birthplace*, a Chi-square test was conducted. The

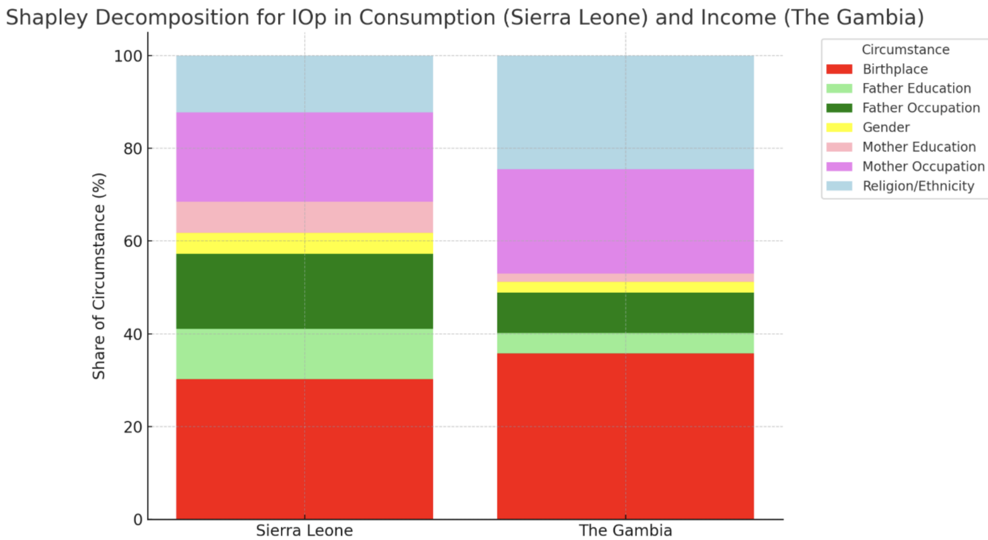


Figure 3: Shapley Decomposition for Sierra Leone and The Gambia - IOp in Consumption/Income

Following closely, again, for both countries, the mother's occupation accounts for the disparities, emphasizing the impact of maternal employment status on household income levels. On the other hand, Father's occupation and education exert a comparatively moderate influence. 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 has the minimal impact among the examined factors, contributing only 4.54% and 2.31%. This could indicate a smaller extent of direct gender-based income inequality within the analyzed sample, highlighting the interplay of sociodemographic factors in shaping economic disparities.

Circumstance	**Sierra Leone**	**The Gambia**
Birthplace	30.19%	35.79%
Religion/Ethnicity	12.30%	24.56%
Father Occupation	16.17%	8.68%
Father Education	10.86%	4.40%
Gender	4.54%	2.31%
Mother Education	6.66%	1.85%
Mother Occupation	19.28%	22.41%

Table 5: Shapley Decomposition for Sierra Leone and The Gambia - IOp in Consumption/Income

results suggest the rejection of the null hypothesis, underscored by an exceptionally low p-value. Corroborating 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.

4.2 Inequality of Opportunity in Literacy

The cohort analysis of inequality of opportunity in literacy is presented in this section.

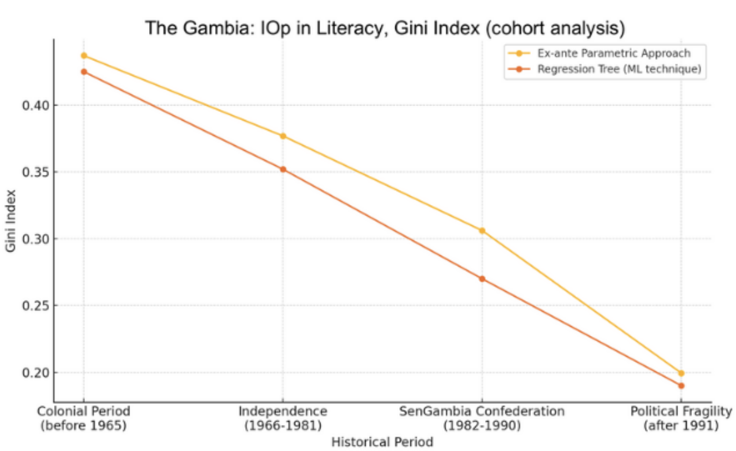
4.2.1 The evolution of literacy rates in The Gambia

For the analysis of The Gambia, Table 6 provides a comprehensive examination of inequality of opportunity (IOp) in literacy across various cohorts. The analysis categorizes the sample into cohorts, according to historical events, based on birth years: before 1965, 1966-1981, 1982-1990, and after 1991, in addition to assessing the entire sample.
The IOp in literacy is measured implementing theParametric Approach with a Probit model for an ex-ante perspective and the Regression Tree Technique for the Machine Learning Approach. The probability of being literate shows significant variation across these cohorts, demonstrating a clear trend of improvement over time. Specifically, the literacy rate increases from 25% for those born before 1965 to 51% for individuals born after 1991. The Absolute IOp, measured by the Gini index, indicates a decline in literacy-based inequality over time, with the index dropping from 0.4370 in the earliest cohort to 0.1994 in the latest. This trend suggests a positive movement toward reducing educational disparities over time. The findings are corroborated by results obtained through Machine Learning techniques, which also show a clear trend of decreasing inequality. The Gini index calculated via these techniques decreases from 0.425 in the earliest cohort to 0.196 in the latest. This reduction signifies substantial progress in literacy across generations, with the most significant decrease observed in the post-1991 cohort, implying effective interventions or improvements in educational policies over time. The Appendix includes Tree charts for each cohort, identifying groups with similar opportunity sets.
Figure 4 presents a graph on the left that illustrates the trends of the Gini indices computed using the two different methodologies across cohorts.

The Shapley decomposition (Table 6 and right part of Figure 4) offers insight into the contributors to the inequality of opportunity in literacy. Birthplace consistently influences all cohorts, indicating the lasting impact of geographic origin on literacy outcomes. They were followed by the gender variable, which stands out for its drastically decreasing influence from 45.20% in the earliest cohort to 6.72% in the latest, reflecting substantial strides toward gender equality in literacy. The importance of fathers' occupations and education varies significantly, with a marked increase in their contributions in more recent cohorts, highlighting changing socioeconomic dynamics. The role of the mother's occupation remains relatively stable, emphasizing the consistent impact of maternal employment on literacy opportunities. Interestingly, mother's education, which has zero contribution in the middle cohorts, sees an increase in the latest cohort, suggesting evolving perceptions of maternal education's role in recent years. Ethnicity's contribution also varies, peaking in the earliest cohort and remaining a significant factor throughout, indicating persistent ethnic disparities in literacy access and outcomes.

Table 6: The Gambia: IOp in Literacy - Cohort analysis

	All sample	**Cohort before 1965**	**Cohort 1966-1981**	**Cohort 1982-1990**	**Cohort after 1991**
Probability to be literate	36.93%	25%	31%	38.7%	50.8%
Gini index Parametric	0.2869	0.4370	0.3769	0.3061	0.1994
Gini Index ML	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	19.5%	2.20%	11.93%	18.81%	22.27%
Father Education	6.73%	0%	0%	4.75%	8.6%
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%	0%	0%	2.92%
Ethnicity	9.65%	19.38%	11.76%	9.82%	14.37%



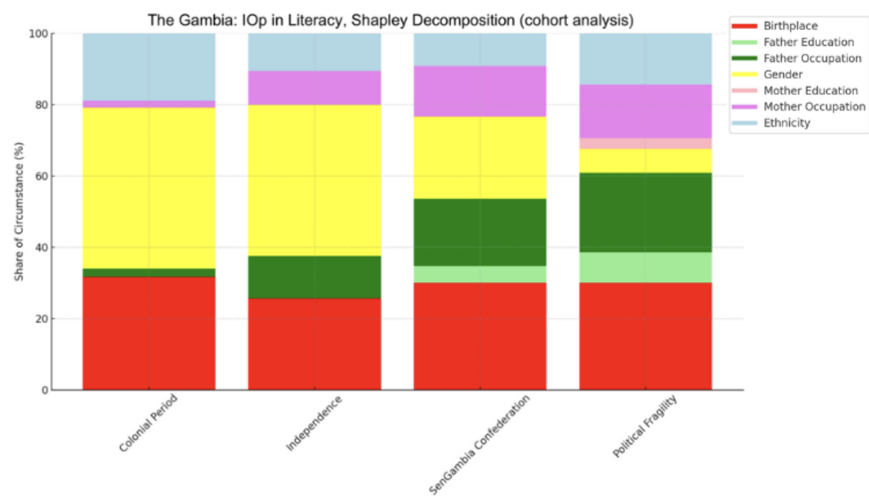


Figure 4: The Gambia: on the left the trend of the IOp Gini Indeces across cohorts; on the right the Shapley decomposition per each cohort

The dataset under examination reveals varied yet relatively comparable sizes across the cohorts, with the second cohort emerging as the most populous. A noteworthy observation pertains to the evolving dynamics of the variables across these cohorts, suggesting a diminishing impact of ethnicity on younger generations compared to their older counterparts. This

shift indicates a potential generational change in the social determinants of educational and economic outcomes.
Furthermore, a critical analysis of the gender variable unveils a significant transformation in its influence across generations. Contemporary educational reforms, which have been extensively documented in existing literature, underscore the pivotal role of policies facilitating free access to secondary education for females. These reforms have markedly contributed to elevating the educational attainment among women, thereby substantially narrowing the gender gap in literacy and education levels. Such trends not only reflect the direct impact of policy interventions on gender parity in education but also highlight the broader societal shifts towards greater gender equality. This evolution underscores the effectiveness of targeted educational policies in mitigating traditional barriers to female education. It suggests a progressive realignment of gender roles within the socio-economic fabric of the population.

4.2.2 The evolution of literacy rates in Sierra Leone

The cohorts delineated in this analysis are built following the historical events that characterized Sierra Leone as a nation. The oldest cohort, preceding 1967, reflects the literacy landscape in the final period of colonialism, marking the final stages of British rule. Sierra Leone's independence in 1967 indicated a new era. As the country transitioned under the Stevens government between 1968 and 1982, which included a state of emergency, the literacy levels demonstrated modest improvements due to the introduction of the Bunumbu Project. It targeted improvements in the quality of primary education within rural areas of Sierra Leone. The project's emphasis on community development and practical skills training suggests that regional educational programs can have a profound influence on literacy outcomes.

The cohort from 1983 to 1991, corresponding to the one-party government under President Momoh, reveals a policy context reflecting a literacy landscape that shows incremental progress. Indeed, despite adversity, this cohort reveals a surprisingly large increase in literacy, potentially indicating the resilience of educational policies aimed at the universalization of education. This process began in the early 1950s and continued into the 1990s (Pai (2013)).
The 1992 cohort represents a period of economic challenges and political instability, including several coups. In 2004, the Education Act made primary and middle school completion a national goal. The data may suggest that government and informal community-based educational initiatives have paid off by increasing literacy levels[11].
Taken together, the data across cohorts shapes the interplay between governance, political stability, and educational policy in shaping literacy outcomes. The increasing trend in literacy across cohorts, particularly the leap in the 1992 cohort, underscores the enduring spirit of Sierra Leone's commitment to education amidst formidable challenges.

In analyzing the inequalities of opportunity in literacy, the cohort analysis, presented in Table 7, offers insightful findings about the dynamics across different generations. A notable decline in the Gini index, computed with both the normative and datadriven approaches,

[11] $https : //www.right - to - education.org/sites/right - to - education.org/files/resource - attachments/RTE_country_factsheet_Sierra_Leone_january_2016_0.pdf$

suggests a substantial reduction in literacy inequality over time. This trend is underscored by the Shapley decomposition (Table 7 and Figure 5), which attributes inequality to various circumstances.

The highest contribution to inequality consistently comes from the region of birth, accentuating the pivotal role of geography in literacy attainment. This confirms how important the implementation of the Bunumbu Project, which was targeted to reduce the educational gap between rural and urban areas, is.
The influence of Sex on inequality has fluctuated significantly, peaking in the cohort of 1968-1982, which may reflect the patriarchal society in Sierra Leone (McFerson (2012)). Moreover, Parental influence, as observed through occupation and education, remains a steady determinant of literacy, although its impact appears to be diminishing in more recent cohorts. A sharp decrease in the Gini index within the 1992 cohort aligns with educational reforms and increased access to literacy resources, suggesting effective policies could be bridging the gap. However, the persistent presence of these determinants in the Shapley decomposition highlights the enduring challenge of fully mitigating the impact of socioeconomic and demographic factors on literacy opportunities.

Table 7: Sierra Leone: IOp in Literacy - Cohort analysis

	All sample	**Cohort before 1967**	**Cohort 1968-1982**	**Cohort 1983-1991**	**Cohort after 1992**
Probability to be literate	35.7%	17.89%	27.3%	37%	63.78%
Gini index Parametric	0.397	0.549	0.5029	0.416	0.2178
Gini Index ML	0.357	0.449	0.430	0.357	0.196
Shapley decomposition					
Birthplace	19.03%	22.47%	16.64%	19.46%	21.09%
Father Occupation	15.73%	15.57%	13.71%	12.7%	14.19%
Father Education	15.79%	5.80%	11.41%	13.18%	16%
Sex	15.92%	25.65%	28.10%	24.36%	17.68%
Mother Occupation	15.79%	10.54%	15.17%	16.17%	13.66%
Mother education	6.76%	0.00%	4.91%	5.53%	6.45%
Religion	10.91%	22.42%	10.05%	8.58%	10.94%

In the cohort analysis, a striking detail emerges from the oldest generation, where the impact of a mother's education on literacy inequality registers at zero. This phenomenon is not an artifact of measurement or an anomaly in data but rather a reflection of the socio-educational context of the time. For this generation[12], none of the mothers had received formal education. This uniform absence of educational background among mothers in the 1967 cohort eliminates mothers' education as a variable contributing to literacy inequality

[12] 2610 individuals out of 2694 stated that the mother had no level of education

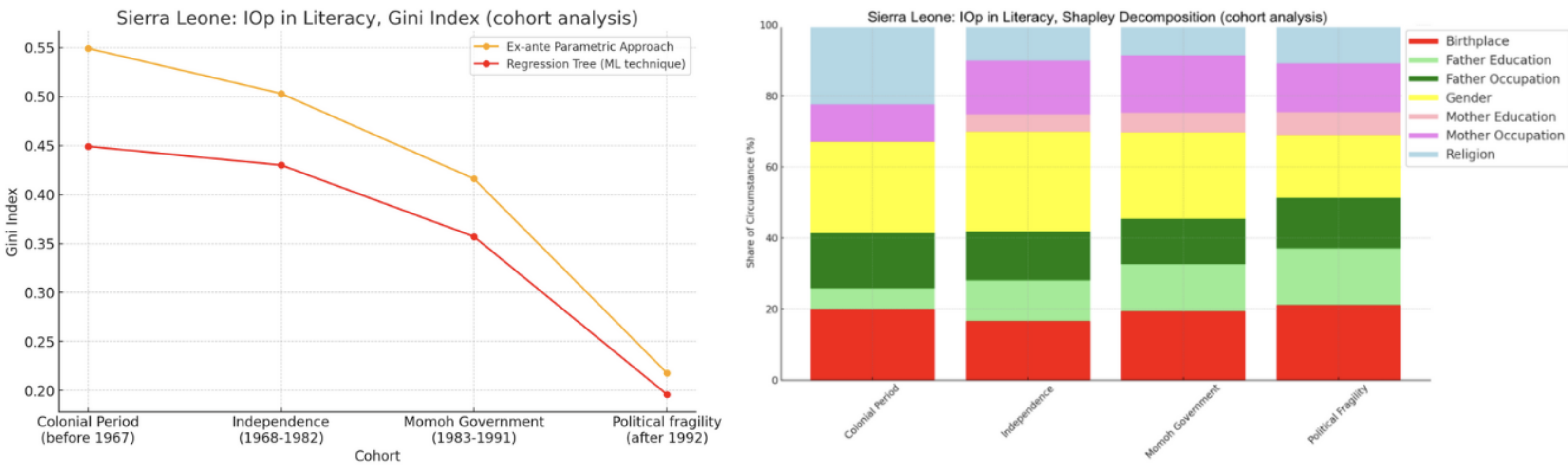


Figure 5: Sierra Leone: on the left the trend of the IOp Gini Indeces across cohorts; on the right the Shapley decomposition per each cohort

within this group. The implications of this homogeneity are profound, providing a stark contrast to subsequent generations where the educational attainment of mothers plays an increasingly significant role in shaping literacy outcomes. It underscores the transformative social changes that have taken place over the decades, leading to a diversification in the educational profiles of mothers and, as a result, a more variegated landscape of literacy opportunities for their children.

5 Conclusion

The Gambia and Sierra Leone, both located in sub-Saharan Africa, provide salient examples of poverty and inequality in the region, particularly with regard to inequality of opportunity. In The Gambia, a high Gini index in 2015 highlighted the broad challenges to be addressed, with more than 35 percent of this inequality attributable to circumstances beyond individual control, such as region of birth and ethnicity. Normative and machine learning approaches were used to examine these inequalities, revealing that region of birth and ethnicity significantly influence income distribution and opportunity. One encouraging finding is that cohort analysis of literacy levels indicates a promising decline in illiteracy rates among younger generations, driven primarily by place of birth and father's social status. In contrast, older generations are influenced by gender and region of birth. The increase in the percentage of literate women and the reduction in non-literate individuals underscore the critical role of targeted educational interventions to improve literacy and achieve gender equality, which are essential to promote socioeconomic development.

In Sierra Leone, economic and political challenges similarly affected inequality, with the inequality index quite high (33 percent overall inequality of which 55 percent was due to opportunity) in 2018. Again, using the same methodologies showed how the results follow the same trend. Shapley's decomposition showed that region of birth and factors such as religion and mother's occupation have a significant impact on inequality of opportunity. Despite low literacy rates in older cohorts, there has been a marked improvement in younger generations, reflecting the effectiveness of educational policies such as the Bunumbu Project. However, while the impact of region of birth on literacy inequality has reduced in younger generations, gender remains a significant factor, highlighting the persistent challenges women face in a patriarchal society.

Both countries stress the urgent need for targeted policy interventions to mitigate the impact of region of birth and ethnicity on opportunity, emphasizing equitable access to quality education and health care in all sectors. Policies should aim to bridge the gap between different ethnic groups, foster national unity, and promote shared progress. The Gambia's focus on addressing educational disparities and Sierra Leone's holistic approach to addressing entrenched inequalities offer insights into promoting a more equitable future. By tackling these issues head-on, both nations can set a precedent for other countries in sub-Saharan Africa, reversing the trend of inequality and paving the way for sustainable development and social cohesion.

References

Aguilar, R Andres Castaneda, Diaz-Bonilla, Carolina, Fujs, Tony HMJ, Jolliffe, Dean, Lakner, Christoph, Mahler, Daniel G, Nguyen, Minh C, Schoch, Marta, Tetteh-Baah, Samuel K, Mendoza, Martha C Viveros, *et al.* 2022. Poverty and Inequality Platform (PIP).

Anderson, John O, Chiu, Mei-Hung, & Yore, Larry D. 2010. First cycle of PISA (2000–2006)—International perspectives on successes and challenges: Research and policy directions. *International Journal of Science and Mathematics Education*, **8**, 373–388.

Atamanov, Aziz, Cuevas, Pablo Facundo, Lebow, Jeremy Aaron, & Mahler, Daniel Gerszon. 2024. *New Evidence on Inequality of Opportunity in Sub-Saharan Africa: More Unequal Than We Thought*. Tech. rept. The World Bank.

Bank, World. 2022. *Poverty and shared prosperity 2022: correcting course*. The World Bank.

Becker, Gary S, *et al.* 1964. *Human capital*.

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.

Bradbury, Katharine, & Triest, Robert K. 2016. Inequality of opportunity and aggregate economic performance. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, **2**(2), 178–201.

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, Hufe, Paul, & Mahler, Daniel Gerszon. 2018. The roots of inequality: estimating inequality of opportunity from regression trees. *World Bank Policy Research Working Paper*.

Brunori, Paolo, Palmisano, Flaviana, & Peragine, Vitorocco. 2019a. Inequality of opportunity in sub-Saharan Africa. *Applied Economics*, **51**(60), 6428–6458.

Brunori, Paolo, Peragine, Vito, & Serlenga, Laura. 2019b. 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.

Buhmann, Brigitte, Rainwater, Lee, Schmaus, Guenther, & Smeeding, Timothy M. 1988. Equivalence scales, well-being, inequality, and poverty: sensitivity estimates across ten countries using the Luxembourg Income Study (LIS) database. *Review of income and wealth*, **34**(2), 115–142.

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

Chantreuil, Frédéric, & Trannoy, Alain. 1997. Inequality decomposition values. *Unpublished mimeograph, Universite de Cergy-Pontoise, Cergy-Pontoise.*

Chantreuil, Frédéric, & Trannoy, Alain. 1999. *Inequality decomposition values: the trade-off between marginality and consistency.* Ph.D. thesis, auto-saisine.

Chantreuil, Frédéric, & Trannoy, Alain. 2011. Inequality decomposition values. *Annals of Economics and Statistics/Annales d'Économie et de Statistique*, 13–36.

Checchi, Daniele. 2004. Does educational achievement help to explain income inequality? *Inequality, growth and poverty in an era of liberalization and globalization.*

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.

Condron, Dennis J. 2011. Egalitarianism and educational excellence: Compatible goals for affluent societies? *Educational Researcher*, **40**(2), 47–55.

Cowell, Frank A, & Van Kerm, Philippe. 2015. *Wealth inequality: A survey.* London.

Crowder, Michael. 2013. *Colonial West Africa: Collected Essays.* Routledge.

Dao, Minh Quang, *et al.* 2013. Population and Economic Growth in Africa-Popolazione e crescita economica in Africa. *Economia Internazionale/international Economics*, **66**(2), 215–230.

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

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, & Gignoux, Jérémie. 2014. The measurement of educational inequality: Achievement and opportunity. *The World Bank Economic Review*, **28**(2), 210–246.

Ferreira, Francisco HG, & Peragine, Vito. 2013. Equality of opportunity: Theory and evidence.

Ferreira, Francisco HG, & Walton, Michael. 2005. *World development report 2006: equity and development*. Vol. 28. World Bank Publications.

Ferreira, Francisco HG, Lakner, Christoph, Lugo, Maria Ana, & Özler, Berk. 2018. Inequality of opportunity and economic growth: how much can cross-country regressions really tell us? *Review of Income and Wealth*, **64**(4), 800–827.

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

Földvári, Péter, & van Leeuwen, Bas. 2011. Should less inequality in education lead to a more equal income distribution? *Education Economics*, **19**(5), 537–554.

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.

Keller, Katarina RI. 2010. How can education policy improve income distribution? An empirical analysis of education stages and measures on income inequality. *The Journal of Developing Areas*, 51–77.

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).

Lebby, Sam J, & Lutz, Jack. 1982. Education and Productive Work: The Bunumbu Approach. *Prospects: Quarterly Review of Education*, **12**(4), 485–93.

LEONE, SIERRA, ALLEN, SR MATO, WILLIAMS, A, FERRON, OM, GREEN, TL, BIRCH, TW, NEW, DOES NEW CONTENT NEED, KARR, SHARON, & DYASI, HM. 1975. SIERRA LEONE JOURNAL OF EDUCATION. *Sierra Leone Journal of Education*, **5**.

Marrero, Gustavo A, & Rodríguez, Juan G. 2013. Inequality of opportunity and growth. *Journal of Development Economics*, **104**, 107–122.

McFerson, Hazel M. 2012. Women and post-conflict society in Sierra Leone. *Journal of International Women's Studies*, **13**(1), 46–67.

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*.

Pai, Grace. 2013. Particularizing universal education in postcolonial Sierra Leone. *Current Issues in Comparative Education*, **16**(1).

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.

Petcu, Claudia. 2014. Does Educational Inequality Explain Income Inequality Across Countries?

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.)*.

Reardon, Sean F. 2011. The widening academic achievement gap between the rich and the poor: New evidence and possible explanations. *Whither opportunity*, **1**(1), 91–116.

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.

Schlosser, Lisa, Hothorn, Torsten, Stauffer, Reto, & Zeileis, Achim. 2019. Distributional regression forests for probabilistic precipitation forecasting in complex terrain.

Schultz, Theodore W. 1961. Education and economic growth. *Teachers College Record*, **62**(10), 46–88.

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

Shorrocks, Anthony F. 1982. Inequality decomposition by factor components. *Econometrica: Journal of the Econometric Society*, 193–211.

Sleight, GF. 1964. The development programme in education for Sierra Leone, 1964-1970. *(No Title)*.

Smith, Timothy L. 1978. Religion and ethnicity in America. *The American Historical Review*, **83**(5), 1155–1185.

Tetteh-Baah, Samuel, Harttgen, Kenneth, Lahoti, Rahul, & Günther, Isabel. 2024. Horizontal Inequalities in Africa. *Review of Income and Wealth*.

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

Weiss, Andrew. 1995. Human capital vs. signalling explanations of wages. *Journal of Economic perspectives*, **9**(4), 133–154.

Young, H Peyton. 1985. *Cost allocation: methods, principles, applications*. North Holland Publishing Co.

Appendix

Table 8: 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 9: 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 11: Socio-demographics variables - Sierra Leone 2018

	Analysis sample (N=10947)
Gender	
Female	6275 (57.3%)
Male	4672 (42.7%)
Region of birth	
Bo	881 (8.0%)
Bombali	1778 (16.2%)
Bonthe	611 (5.6%)
Kailahun	669 (6.1%)
Kambia	668 (6.1%)
Kenema	882 (8.1%)
Koinadugu	765 (7.0%)
Kono	653 (6.0%)
Moyamba	561 (5.1%)
Port Loko	1039 (9.5%)
Pujehun	663 (6.1%)
Tonkolili	761 (7.0%)
Western Rural Area	34 (0.3%)
Western Urban Area	982 (9.0%)
Religion	
Ahmadis Muslim	1068 (9.8%)
Anglican	89 (0.8%)
Bahai	30 (0.3%)
Catholic	739 (6.8%)
Methodist	264 (2.4%)
No religion	8 (0.1%)
Other Christian	773 (7.1%)
Other Muslim	7323 (66.9%)
Pentacostal	582 (5.3%)
SDA	51 (0.5%)
Traditional	20 (0.2%)

Table 10: 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 12: Parental Education - Sierra Leone 2018

	Analysis sample (N=10947)
Mother's education (levels)	
No school	9898 (90.4%)
Primary incomplete	203 (1.9%)
Primary complete	205 (1.9%)
Secondary incomplete	305 (2.8%)
Secondary complete	141 (1.3%)
Post-secondary professional	121 (1.1%)
Vocational	23 (0.2%)
First degree or more	51 (0.5%)
Father's education (levels)	
No school	9050 (82.7%)
Primary incomplete	219 (2.0%)
Primary complete	208 (1.9%)
Secondary incomplete	567 (5.2%)
Secondary complete	359 (3.3%)
Post-secondary professional	230 (2.1%)
Vocational	41 (0.4%)
First degree or more	273 (2.5%)

Table 14: 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

Table 13: Parental Occupation - Sierra Leone 2018

Analysis sample (N=10947)	
Mother's occupation (categories)	
Accommodation and food service activities	46 (0.4%)
Administrative and support service activities	24 (0.2%)
Agriculture, foresting, fishing	8503 (77.7%)
Construction	6 (0.1%)
Education	179 (1.6%)
Electricity and water supply	2 (0.0%)
Financial activities	8 (0.1%)
Manufacturing	128 (1.2%)
Mining, quarrying	19 (0.2%)
Professional	21 (0.2%)
Public administration and defense	40 (0.4%)
Social activities	14 (0.1%)
Transportation and storage	4 (0.0%)
Wholesale and retail trader, repair of motor vehicles and motorcycles	1953 (17.8%)
Father's occupation (categories)	
Accommodation and food service activities	16 (0.1%)
Administrative and support service activities	113 (1.0%)
Agriculture, foresting, fishing	8597 (78.5%)
Construction	281 (2.6%)
Education	337 (3.1%)
Electricity and water supply	39 (0.4%)
Financial activities	49 (0.4%)
Manufacturing	163 (1.5%)
Mining, quarrying	226 (2.1%)
Professional	46 (0.4%)
Public administration and defense	237 (2.2%)
Social activities	29 (0.3%)
Transportation and storage	210 (1.9%)
Wholesale and retail trader, repair of motor vehicles and motorcycles	604 (5.5%)

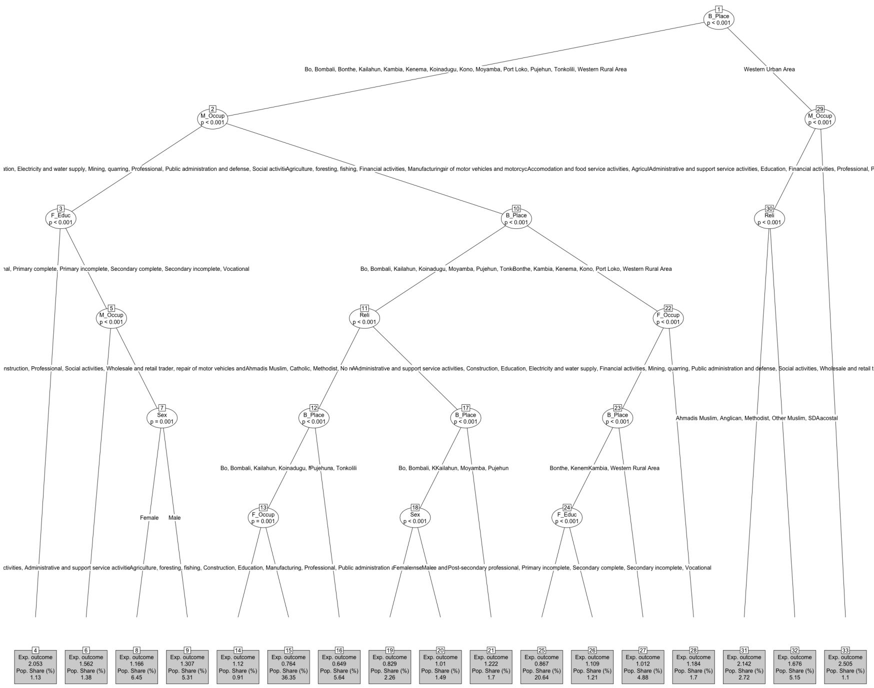


Figure 6: Sierra Leone: Conditional Regression Tree - IOp in consumption

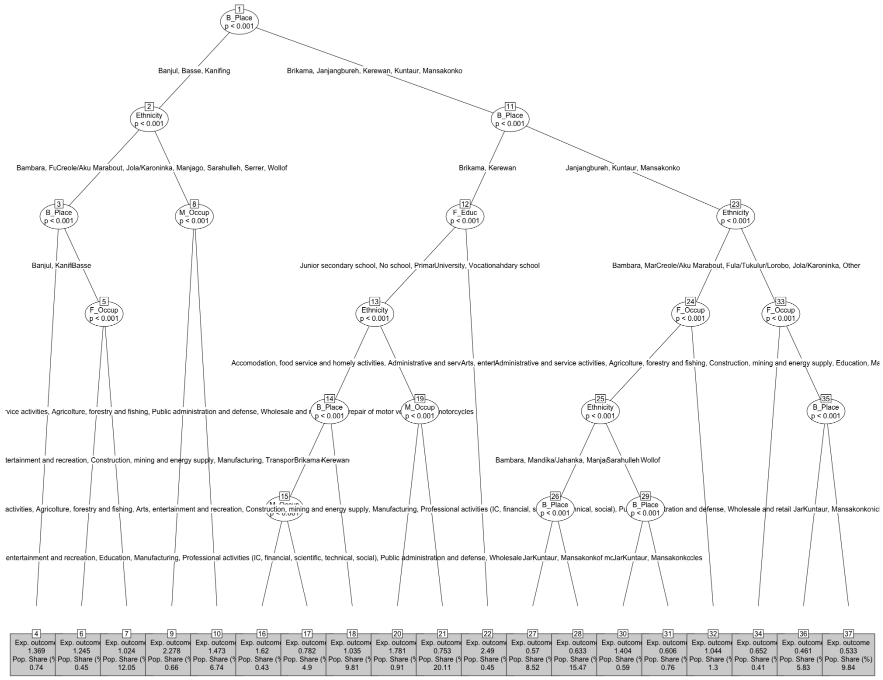


Figure 7: The Gambia: Conditional Regression Tree - IOp in income

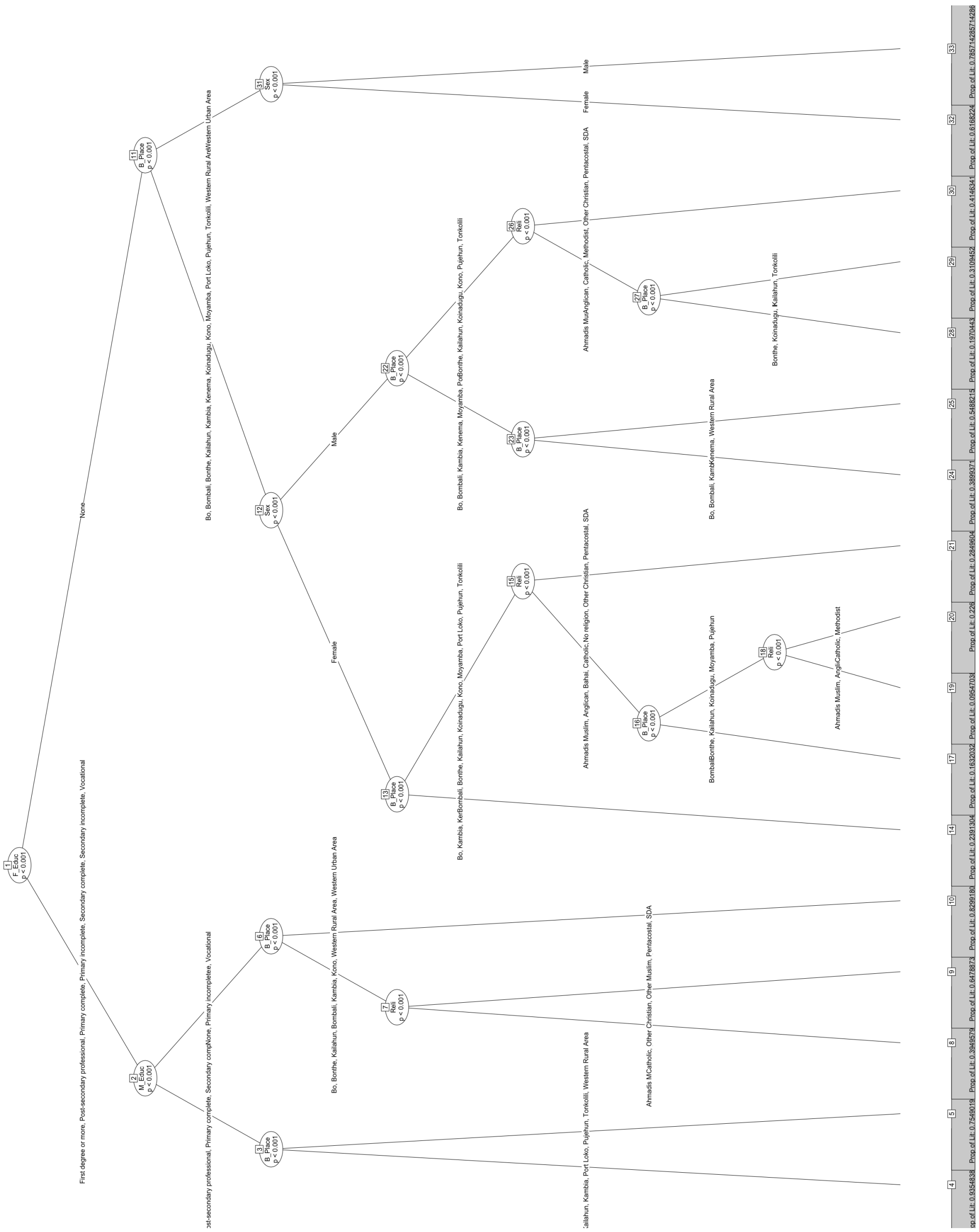


Figure 8: Sierra Leone: Conditional Tree - Literacy - All Sample

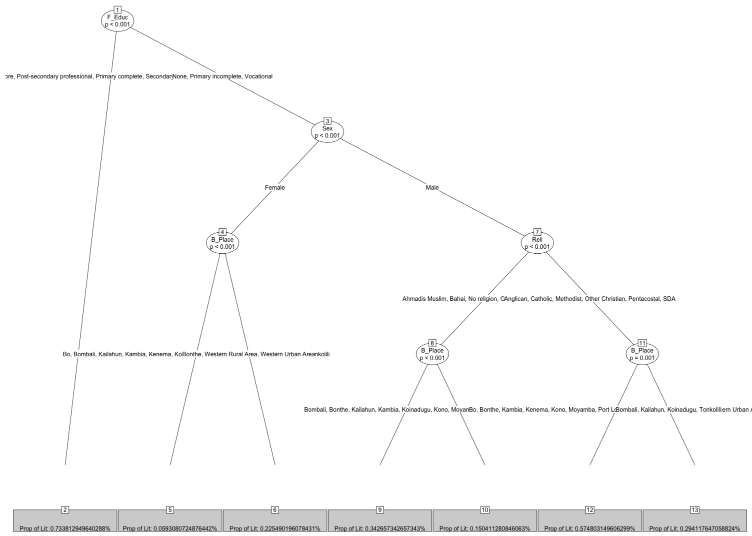


Figure 9: Sierra Leone: Conditional Tree - Literacy - Cohort before 1967

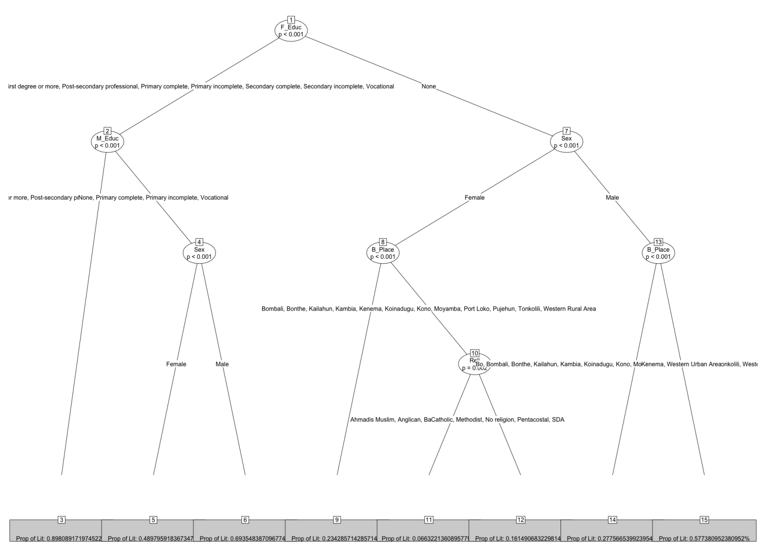


Figure 10: Sierra Leone: Conditional Tree - Literacy - Cohort 1968-1982

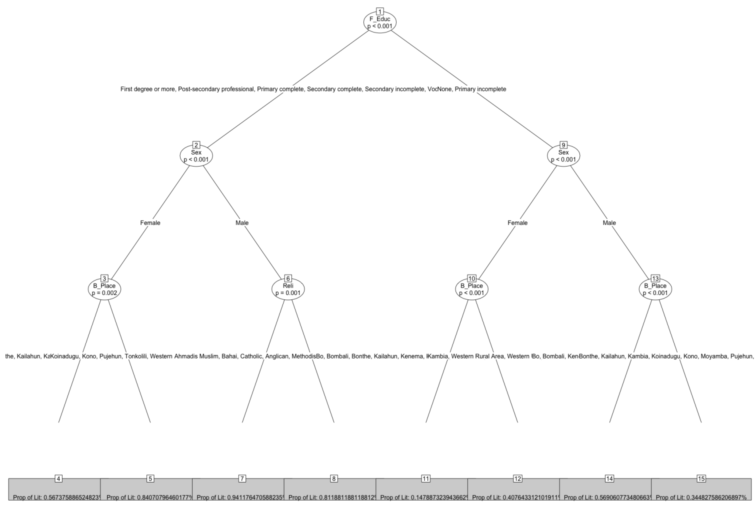


Figure 11: Sierra Leone: Conditional Tree - Literacy - Cohort 1983-1991

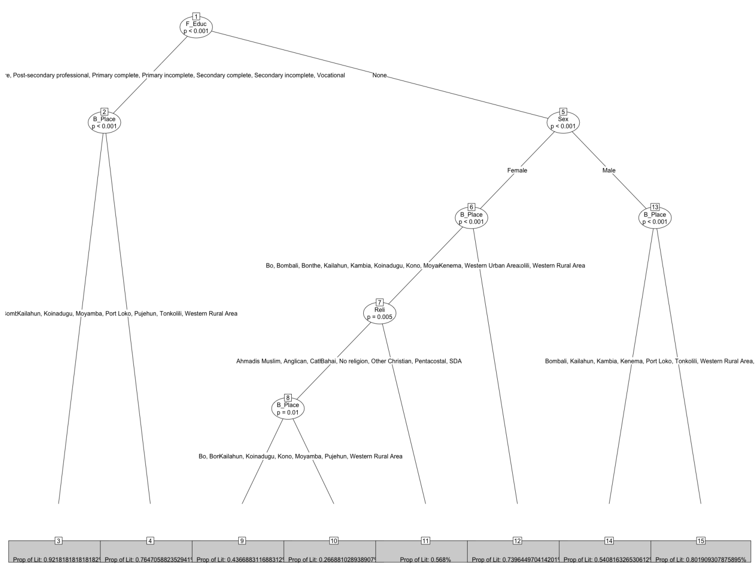


Figure 12: Sierra Leone: Conditional Tree - Literacy - Cohort after 1992

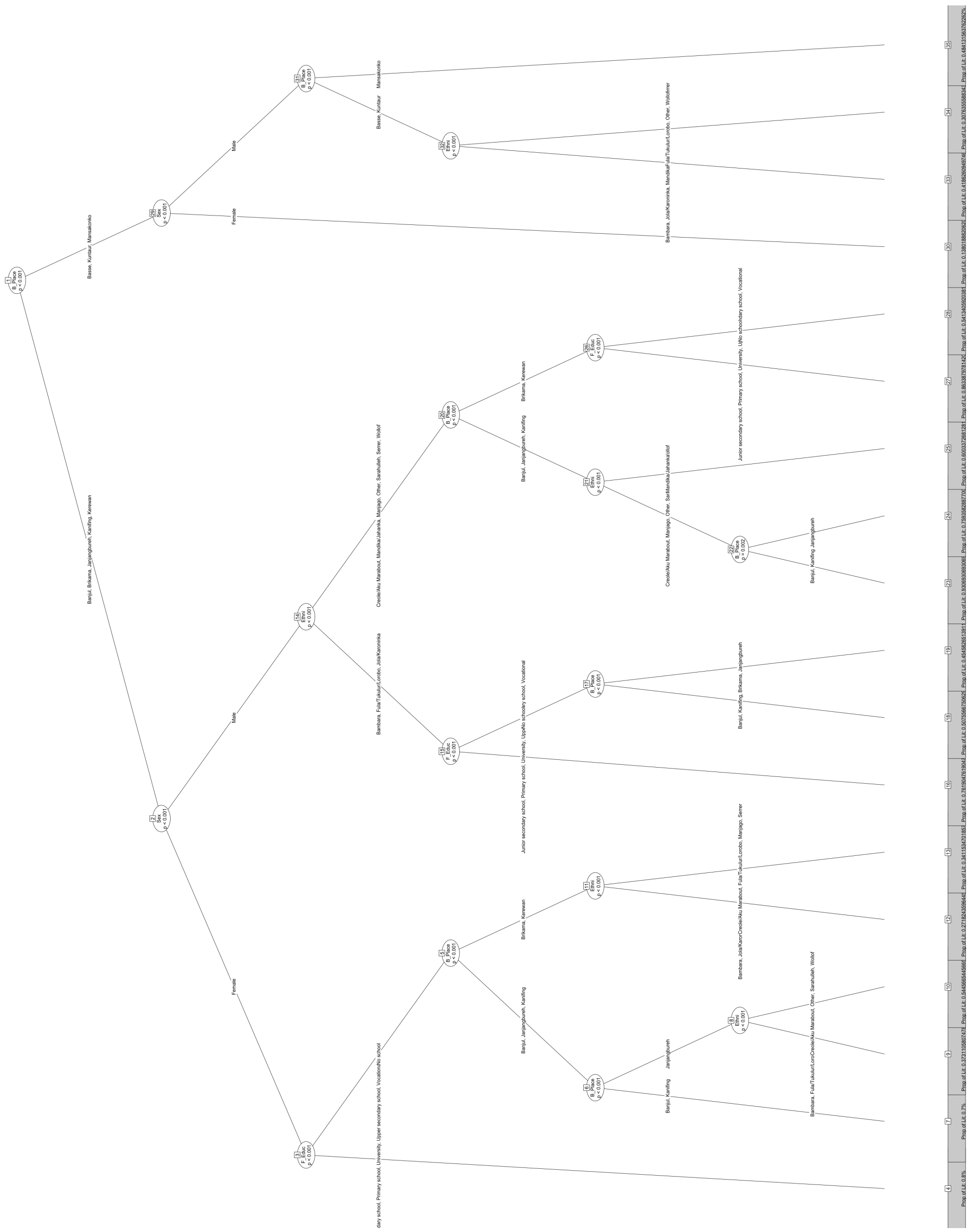


Figure 13: The Gambia: Conditional Tree - Literacy - All Sample

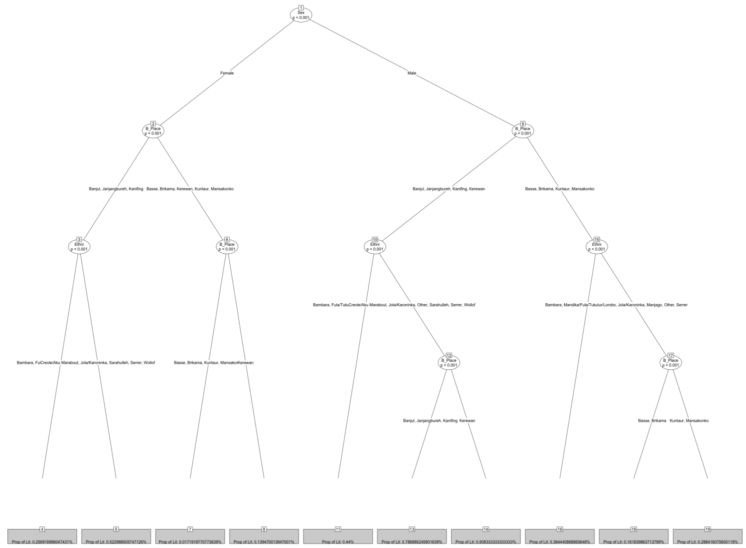


Figure 14: The Gambia: Conditional Tree - Literacy - Cohort before 1965

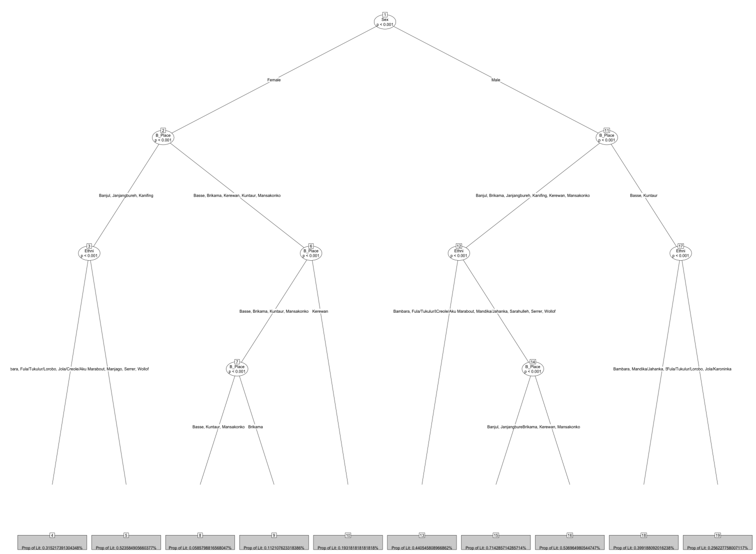


Figure 15: The Gambia: Conditional Tree - Literacy - Cohort 1966-1981

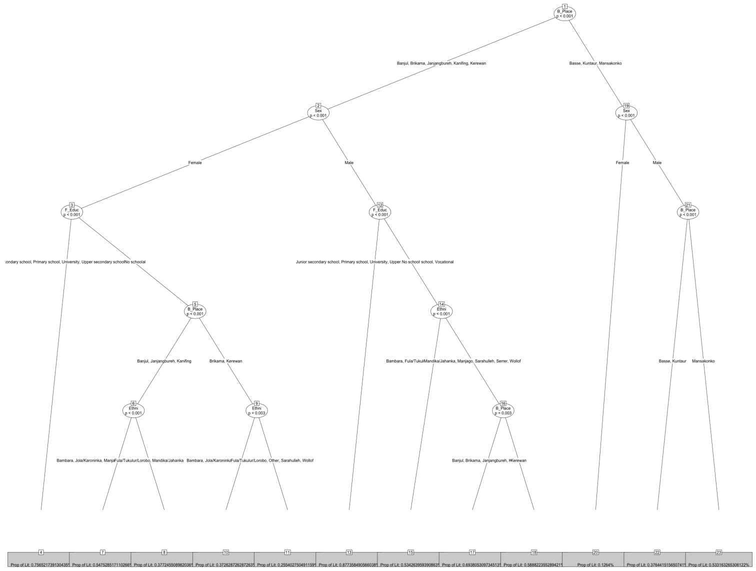


Figure 16: The Gambia: Conditional Tree - Literacy - Cohort 1982-1990

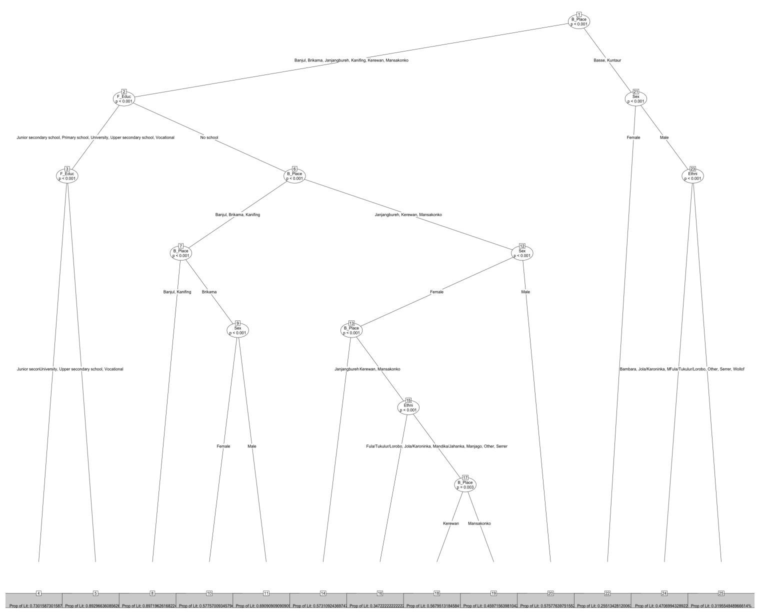


Figure 17: The Gambia: Conditional Tree - Literacy - Cohort after 1991