DISCUSSION PAPER SERIES

IZA DP No. 18286

Reconstructing Two Decades of Inequality in the Sahel Region

Gianni Betti
Federico Crescenzi
Hai-Anh Dang
Vasco Molini
Lorenzo Mori

NOVEMBER 2025

I Z A Institute of Labor Economics

Initiated by Deutsche Post Foundation

DISCUSSION PAPER SERIES

IZA DP No. 18286

Reconstructing Two Decades of Inequality in the Sahel Region

Gianni Betti
University of Siena

Federico Crescenzi
Universitas Mercatorum

Hai-Anh Dang
World Bank, IZA and Indiana University

Vasco Molini
World Bank, IZA and Indiana University

Lorenzo Mori
University of Bologna

NOVEMBER 2025

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Reconstructing Two Decades of Inequality in the Sahel Region*

Measuring inequality in West Africa is a challenging task that is constrained by the limited availability and irregular collection of household consumption data. To address this challenge, we reconstructed the evolution of inequality in the Sahel region using an innovative framework that combines Survey-to-Survey Imputation Techniques (SSITs) with Generalized Additive Models for Location, Scale and Shape (GAMLSS), based on labour force surveys conducted in eight countries between 2003 and 2021. The findings highlight pronounced regional disparities, persistent levels of inequality, and a clear association between inequality patterns and episodes of conflict or political instability. Our contribution is twofold: methodologically, we introduce a flexible SSIT-GAMLSS model that incorporates two levels of random effects; substantively, we provide new evidence of inequality trends in francophone West Africa, a region largely underrepresented in empirical research.

JEL Classification: C15, I32, O15

Keywords: inequality, imputation, labour force surveys, West Africa

Corresponding author:
Hai-Anh H. Dang
World Bank Development Data Group
1818 H Street
Washington, DC
USA
E-mail: hdang@worldbank.org

* We thank Daniel Mahler and participants at the IARIW-WB-UEB/VNU conference (Hanoi) for helpful feedback on previous versions. Dang would like to thank the UK Foreign Commonwealth and Development Office (FCDO)'s Data and Evidence for Tackling Extreme Poverty (DEEP) program for funding support.

1. Introduction

Over the last decade, the growth pattern of many Western African countries has been marked by limited inclusiveness and an increasing regional divide. This significant divide is not a recent phenomenon but has been observed to be the result of various agro-ecological and socioeconomic factors (*inter alia* Fosu, 2009; Fosu, 2015; Cornia, 2017; Fosu, 2017a, 2017b; Odusola et al., 2017; Fosu, 2018).

The semiarid Sudan-Sahel agro-ecological zone, for instance, is characterized by limited and erratic rainfall patterns, which pose serious challenges for farmers compared to those in the more coastal areas of West Africa. Farmers in the Sudan-Sahel region face significant rainfall constraints that hinder their agricultural productivity (Alfani et al. 2019). In contrast, the coastal and southern parts of West Africa, which started with better initial conditions, have further improved their socioeconomic situation over the years (Alfani et al. 2019). This disparity has contributed to the widening regional divide in West Africa.

Episodes of insurgency affecting countries such as Mali, Burkina Faso, Niger and Chad have further depressed the socio-economic conditions of these regions. Such conflicts have presumably exacerbated spatial inequalities, making it even more challenging for these areas to achieve economic stability and growth. Several assessments conducted by the World Bank between 2010 and 2019 indicate that disruption of infrastructure and the resulting negative impact on economies in these regions are quite significant (Marc et al. 2015; David et al. 2025). This preliminary evidence suggests that it is crucial to monitor the well-being of these countries in a timely manner to address and mitigate the adverse effects of these increasingly frequent challenges.

It is unfortunately too costly and complicated for many West African countries to collect household surveys for poverty and inequality estimates to document the socio-economic impact of these events in a reliable and timely way. Few of these countries manage to collect survey data regularly, and even for those that do, the interval between surveys is often very long. Indeed, it is difficult to obtain estimates of poverty and inequality every five or so years and it is almost impossible to have these estimates on an annual basis. To overcome this challenge, a new economic literature on imputation methods has recently been developed that compares welfare indicators from surveys characterized by limited availability and comparability.

Survey-to-survey imputation techniques (SSITs) in economics are largely built on the poverty map literature (Elbers et al., 2003; Tarozzi and Deaton, 2009; Dang et al. 2019; Dang and Lanjouw 2023). These techniques, which involve imputing income from censuses, have been widely applied in developing countries to obtain geographically disaggregated estimates of poverty. Recently, there have been survey-to-survey imputations that map data from surveys with consumption information to those with other outcomes of interest but without standard welfare aggregates (Elbers et al., 2004; Dabalen et al., 2014; Dang et al., 2017; Dang et al. 2025b and forthcoming).

Yet, despite the increasingly common use of imputation methods for poverty estimates, the literature has paid limited attention to the challenge of obtaining accurate inequality measures based on imputation. Standard SSITs—typically based on regression analysis and the assumption of normally distributed residuals—are more accurate at predicting central parts of a distribution rather than the shape of the tails, which is crucial for inequality prediction (Schluter, 2012).

Just a handful of existing studies briefly discuss these challenges. Demombynes et al. (2007) found that correlations between estimated and true welfare at local level are highest for mean expenditure and poverty measures but lower for inequality measures. Douidich et al. (2016) obtained accurate quarterly poverty rate estimates using a classical survey-to-survey imputation method but warned that incorrect normal-error assumptions or ignoring heteroskedasticity could bias poverty or inequality estimates. Krafft et al. (2019) imputed consumption from Household Budget Surveys onto Labour Force Surveys and found similar measures of consumption, poverty and inequality across survey pairs, particularly in Jordan and Egypt. Dang et al. (2017) addressed this normality issue by providing a practical option to draw residuals from the empirical distribution of the error terms rather than from a normal one that mitigates potential bias (offered by the Stata "*s2s*" and "*povimp*" commands (Dang et al., 2025a)).

Betti et al. (2024) explicitly addressed the limitations of SSITs in measuring inequality by means of a Generalized Additive Model for Location, Scale and Shape (GAMLSS). In essence, this parametric model extends classical SSITs by relaxing the assumption of normality and allowing inclusion of covariates and random effects in the parameters of the distribution, even for distributions not belonging to the exponential family. In their original study, the authors used this method to estimate inequality in Morocco with a GB2 model having covariates in all four parameters and random effects on the location and scale parameters.

In this paper, we build on Betti et al.'s (2024) approach by incorporating two distinct random effects: one that captures differences between areas in the same country and another that accounts for differences across countries. Our main goal was to estimate the following inequality measures: Gini coefficient, Generalized Entropy indices GE(0), GE(1) and GE(2), and three percentile- or share-based ratios: P90/P10, P80/P20 and S80/S20. Besides proposing this methodology innovation, we also provide new empirical evidence about inequality. Economic research in Africa is generally limited and tends to focus on a few countries. According to a recent study (Porteous, 2022), 45% of all economics journal articles and 65% of articles in the top five economics journals are about just five countries, which account for only 16% of the continent's population. This uneven distribution of research means that many African countries are underrepresented in the economic literature, leading to a smaller evidence base for local policymakers in these regions. The francophone West African countries, in particular, have received even less attention in the economic literature, despite the region's significant economic activity, its population size and its relative proximity to developed countries. This paper is one of the first (see also Nikiema, 2025) to make use of the Harmonised Living Standard Household Survey of the West African Economic and Monetary Union (WAEMU) that covers eight countries in the region.

The remainder of the paper is organized as follows. Section 2 introduces the data, distinguishing surveys that include the consumption variable and those that do not (Appendix A provides a complete summary of the datasets). Section 3 briefly reviews the SSIT-GAMLSS methodology and presents an extension with a double random-effect structure. Section 4 offers simulation analysis to assess model performance. Section 5 presents the inequality results across the Sahel region over a 20-year period. Finally, Section 6 concludes the paper.

2. Data

The data used as a reference for the consumption model is derived from the West African Economic and Monetary Union (WAEMU) household surveys, also known as the Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM). These surveys are a series of harmonized studies conducted across WAEMU member countries in 2018 and 2021. We specifically focus on the 2018 wave, as it is the closest to the year for which we intend to impute data. It is worth noting that while all countries participated in both survey rounds, Côte d'Ivoire and Guinea-Bissau did not collect consumption data in 2018, and Burkina Faso did not collect consumption data in 2021; other socioeconomic and demographic variables were, however, available for these countries.

Harmonized data collection is essential if the WAEMU Commission is to monitor national economies and support alignment of national statistical systems with international standards. This harmonized approach ensures that the data collected is comparable across member states, facilitating regional analysis and policymaking. Such comparability is crucial for addressing cross-border issues such as income convergence, external tariffs, regional investments, financial inclusion, resilience to shocks, and labour mobility.

The WAEMU household surveys involve the following eight member countries: Benin, Burkina Faso, Côte d'Ivoire, Guinea-Bissau, Mali, Niger, Senegal and Togo. The surveys are designed to be nationally representative, covering both urban and rural areas, with the number of sampled households ranging from 6000 to 8000. They are conducted in two waves to capture consumption seasonality: the first wave between October and December, the second wave between April and July. Each wave covers half the sample, ensuring that the data collected reflects seasonal variations in living conditions and consumption patterns.

The labour force surveys span a longer interval than the WAEMU data and are collected more frequently than the regular household budget surveys. However, since their objective is not to capture well-being, they lack information on consumption. Consequently, we impute consumption in these surveys to provide a more regular picture of poverty and inequality over the last two decades. Not all the available labour force surveys from these countries, however, are suitable for imputation. Using the following three criteria: i) surveys collected after 2000; ii) surveys with a minimal number of missing data points; and iii) surveys that share a sufficient number of covariates with the WAEMU surveys, we selected 17 labour force surveys and 3 WAEMU surveys. Since these surveys have no consumption data, we impute consumption for these surveys. Surveys by country and year are reported in Table 1 and indicated with the number 1. Surveys that are marked “x” are those from the WAEMU surveys, for which the consumption variable is available. An asterisk (*) next to the number indicates that the dataset also contains the corresponding NUTS 2 (Nomenclature of Units for Territorial Statistics, level 2) geographic information, which refers to a standardized classification of regions in each country used for statistical purposes. Two asterisks (**) indicate that NUTS 2 information is available only for some statistical units in the dataset. No asterisk means that NUTS 2 geographic information is not included in the dataset.

Table 1. Dataset to impute. Surveys by country and year. The number “1” indicates survey availability but without consumption data; “x” marks WAEMU surveys with consumption data. An asterisk (*) denotes availability of NUTS-2 regional information for all units, while two asterisks (**) indicate partial availability. Absence of an asterisk means that no NUTS-2 information is included.

Country	2003	2005	2006	2007	2008	2009	2010	2011	2014	2015	2018	2021
BEN	1*	0	0	1*	0	0	0	1*	0	0	x*	x*
BFA	1**	0	0	0	0	0	0	0	1**	0	x*	1
CIV	0	0	0	0	1*	0	0	0	0	1**	1	x*
GNB	0	0	0	0	0	0	1	0	0	0	1*	x*
MLI	1*	0	0	0	0	1*	0	0	0	0	x*	x*
NER	0	1*	0	1*	0	0	0	1*	1*	0	x*	x*
SEN	0	1**	0	0	0	0	0	0	0	0	x*	x*
TCD	0	0	0	0	0	0	0	0	0	0	x*	x**
TGO	0	0	1*	0	0	0	0	0	0	1*	x*	x*

Both sets of data are harmonized and stored in a repository created by the Sub-Saharan Team for Statistical Development (SSATSD) of the World Bank. Approximately 200 variables from existing household surveys are extracted and harmonized from household budget surveys and labour force surveys. These variables cover key aspects such as household consumption, infrastructure access, employment status, education and health. Since survey questions vary across datasets, standardizing variable definitions is a major challenge.

The harmonized household survey data provides a consistent and reliable source for cross-country and temporal comparisons. The harmonized dataset is structured in four modules: 1) Module P (poverty-related variables), 2) Module H (household-level variables excluding poverty indicators), 3) Module I (individual-level variables excluding labour force data), and 4) Module L (labour force variables).

For all countries listed in Table 1, we report a summary of selected key variables that will be used in subsequent analyses (Table 2). In the paper, we present only the mean, standard deviation, and median of each variable, while Appendix A includes Tables A2 to A10 that contain additional statistics such as quartiles, minimums, and maximums. Note that although WAEMU expenditure data is available for Chad in 2021, other variables of interest are missing. This is not a limitation for the present analysis, because only expenditure data for that year is essential.

Table 2: Mean, Standard Deviation (SD) and Median of selected variables.

Country	Year	Household size			Rural/Urban			Open Defecation			Access to piped water			Sex			Age Class			School attendance			Literacy		
		Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
Benin	2003	5.35	1.84	6.00	0.62	0.49	1.00	0.33	0.47	0.00	0.10	0.30	0.00	0.51	0.50	1.00	2.92	1.05	3.00	0.40	0.49	0.00	0.07	0.25	0.00
Benin	2007	5.57	1.67	6.00	0.65	0.48	1.00	0.99	0.10	1.00	0.08	0.27	0.00	0.51	0.50	1.00	2.14	1.26	2.00	0.48	0.50	0.00	0.45	0.50	0.00
Benin	2011	5.68	1.67	7.00	0.60	0.49	1.00	0.96	0.19	1.00	0.05	0.23	0.00	0.51	0.50	1.00	2.26	1.27	2.00	0.57	0.50	1.00	0.50	0.50	0.00
Benin	2018	5.67	1.59	6.00	0.47	0.50	0.00	0.55	0.50	1.00	0.29	0.45	0.00	0.49	0.50	0.00	2.08	1.19	2.00	0.58	0.49	1.00	0.48	0.50	0.00
Benin	2021	5.61	1.55	6.00	0.57	0.50	1.00	0.42	0.49	0.00	0.34	0.47	0.00	0.55	0.50	1.00	1.95	1.05	2.00	0.94	0.24	1.00	1.00	0.00	1.00
Burkina Faso	2003	6.99	0.12	7.00	0.70	0.46	1.00	0.98	0.13	1.00	0.08	0.27	0.00	0.52	0.50	1.00	2.50	1.19	2.00	0.29	0.45	0.00	0.28	0.45	0.00
Burkina Faso	2014	6.97	0.26	7.00	0.66	0.47	1.00	1.00	0.00	1.00	0.10	0.31	0.00	0.53	0.50	1.00	2.09	1.22	2.00	0.47	0.50	0.00	0.35	0.48	0.00
Burkina Faso	2018	6.11	1.38	7.00	0.39	0.49	0.00	0.34	0.47	0.00	0.41	0.49	0.00	0.48	0.50	0.00	2.16	1.26	2.00	0.48	0.50	0.00	0.43	0.50	0.00
Chad	2018	5.80	1.55	7.00	0.50	0.50	1.00	0.49	0.50	0.00	0.20	0.40	0.00	0.48	0.50	0.00	2.00	1.19	2.00	0.64	0.48	1.00	0.29	0.45	0.00
Guinea-Bissau	2010	7.00	0.00	7.00	0.55	0.50	1.00	0.85	0.35	1.00	0.06	0.25	0.00	0.53	0.50	1.00	2.81	1.05	2.00	0.53	0.50	1.00	0.53	0.50	1.00
Guinea-Bissau	2021	6.40	1.16	7.00	0.36	0.48	0.00	0.11	0.32	0.00	0.47	0.50	0.00	0.48	0.50	0.00	2.23	1.23	2.00	0.69	0.46	1.00	0.53	0.50	1.00
Ivory Coast	2008	7.00	0.02	7.00	0.49	0.50	0.00	0.87	0.34	1.00	0.23	0.42	0.00	0.49	0.50	0.00	2.27	1.16	2.00	0.52	0.50	1.00	0.52	0.50	1.00
Ivory Coast	2015	6.93	0.43	7.00	0.55	0.50	1.00	0.64	0.48	1.00	0.30	0.46	0.00	0.50	0.50	0.00	2.26	1.22	2.00	0.43	0.49	0.00	0.37	0.48	0.00
Ivory Coast	2021	5.49	1.68	6.00	0.26	0.44	0.00	0.31	0.46	0.00	0.29	0.46	0.00	0.50	0.50	0.00	2.18	1.27	2.00	0.53	0.50	1.00	0.42	0.49	0.00
Mali	2003	6.73	0.86	7.00	0.64	0.48	1.00	0.63	0.48	1.00	0.20	0.40	0.00	0.09	0.29	0.00	4.12	0.96	4.00	0.25	0.43	0.00	0.29	0.45	0.00
Mali	2009	7.00	0.10	7.00	0.62	0.48	1.00	0.95	0.23	1.00	0.35	0.48	0.00	0.53	0.50	1.00	2.94	1.08	3.00	0.30	0.46	0.00	0.31	0.46	0.00
Mali	2018	6.21	1.27	7.00	0.42	0.49	0.00	0.16	0.37	0.00	0.45	0.50	0.00	0.48	0.50	0.00	2.35	1.32	2.00	0.54	0.50	1.00	0.42	0.49	0.00

Country	Year	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median	Mean	SD	Median
		Household size			Rural/Urban			Open Defecation			Access to piped water			Sex			Age Class			School attendance			Literacy		
Mali	2021	6.26	1.20	7.00	0.43	0.50	0.00	0.09	0.29	0.00	0.53	0.50	1.00	0.49	0.50	0.00	2.20	1.31	2.00	0.56	0.50	1.00	0.43	0.49	0.00
Niger	2005	6.98	0.21	7.00	0.67	0.47	1.00	0.95	0.22	1.00	0.14	0.35	0.00	0.51	0.50	1.00	2.74	1.10	2.00	0.42	0.49	0.00	0.38	0.49	0.00
Niger	2007	7.00	0.00	7.00	0.53	0.50	1.00	0.96	0.19	1.00	0.19	0.39	0.00	0.52	0.50	1.00	2.15	1.21	2.00	0.57	0.50	1.00	0.35	0.48	0.00
Niger	2011	7.00	0.00	7.00	0.00	0.00	0.00	0.96	0.21	1.00	0.19	0.39	0.00	0.51	0.50	1.00	2.12	1.23	2.00	0.56	0.50	1.00	0.30	0.46	0.00
Niger	2014	6.78	0.75	7.00	0.36	0.48	0.00	0.91	0.29	1.00	0.42	0.49	0.00	0.42	0.49	0.00	2.08	1.11	2.00	0.99	0.09	1.00	1.00	0.00	1.00
Niger	2018	5.92	1.46	7.00	0.26	0.44	0.00	0.65	0.48	1.00	0.37	0.48	0.00	0.48	0.50	0.00	2.04	1.22	2.00	0.54	0.50	1.00	0.31	0.46	0.00
Niger	2021	5.87	1.44	7.00	0.39	0.49	0.00	0.55	0.50	1.00	0.52	0.50	1.00	0.48	0.50	0.00	2.07	1.25	2.00	0.60	0.49	1.00	0.38	0.48	0.00
Senegal	2005	6.95	0.39	7.00	0.37	0.48	0.00	0.63	0.48	1.00	0.50	0.50	0.00	0.54	0.50	1.00	2.86	1.06	2.00	0.41	0.49	0.00	0.43	0.50	0.00
Senegal	2018	6.63	0.99	7.00	0.53	0.50	1.00	0.08	0.27	0.00	0.64	0.48	1.00	0.46	0.50	0.00	2.23	1.29	2.00	0.68	0.47	1.00	0.48	0.50	0.00
Senegal	2021	6.53	1.08	7.00	0.61	0.49	1.00	0.04	0.20	0.00	0.70	0.46	1.00	0.50	0.50	1.00	2.16	1.15	2.00	0.95	0.23	1.00	1.00	0.00	1.00
Togo	2006	5.24	1.76	6.00	0.65	0.48	1.00	0.91	0.28	1.00	0.29	0.45	0.00	0.51	0.50	1.00	2.91	1.05	3.00	0.61	0.49	1.00	0.54	0.50	1.00
Togo	2015	7.00	0.00	7.00	0.39	0.49	0.00	0.78	0.41	1.00	0.36	0.48	0.00	0.51	0.50	1.00	2.20	1.26	2.00	0.75	0.43	1.00	0.61	0.49	1.00
Togo	2018	5.23	1.76	6.00	0.32	0.46	0.00	0.52	0.50	1.00	0.18	0.39	0.00	0.48	0.50	0.00	2.21	1.28	2.00	0.69	0.46	1.00	0.59	0.49	1.00
Togo	2021	5.21	1.71	5.00	0.34	0.47	0.00	0.51	0.50	1.00	0.29	0.45	0.00	0.47	0.50	0.00	2.27	1.34	2.00	0.73	0.44	1.00	0.63	0.48	1.00

3. Methods

3.1 An overview of GAMLSS and SSIT-GAMLSS

The goal of SSIT is to predict a general parameter $H(Y)$, which is a function of a random variable Y of interest—usually consumption—and typically represents poverty indices, such as the headcount ratio, or inequality indices, as in this case. Specifically, the objective of SSIT is to estimate *H(Y)* for years in which *Y* is not observed, by imputing it based on a model-based procedure (see Dang and Lanjouw, 2023). We now present a strategy to impute *Y* by GAMLSS.

Proposed by Rigby and Stasinopoulos (2005), GAMLSS incorporates a location parameter, a scale parameter and up to two shape parameters. Covariates and random effects can be used in each parameter of the chosen distribution. The distribution does not need to belong to the exponential family but can encompass a wide range of commonly encountered distribution types.

GAMLSS assumes independent observations $y_i,\ i = 1, \dots, n$ from a random variable Y, with probability density function (PDF) $f(Y \mid \theta_i)$, conditional on a vector of p distribution parameters, $k = 1, \dots, p$ $(\theta\_i\text{^}T = (\theta\_i1, \dots, \theta\_ik, \dots, \theta\_ip))$. More formally, let $y^T = (y_1, \dots, y_n)$ be the length (n) vector of the response variable. Let $g_k(\cdot)$ be a known monotonic link function relating the p distribution parameters to the explanatory variables as follows:

$$g_k(\theta_k) = X^k\beta_k + \sum_{m=1}^{M_k} Z_m^k \gamma_m^k, \text{ with } k = 1, \dots, p, \ (1)$$

where $\theta_k^T = (\theta_{1k}, \dots, \theta_{nk})$ is a vector of length n, $\beta_k^T = (\beta_{1k}, \dots, \beta_{kM_k})$ is a parameter vector of length M_k, X^k is a matrix of known covariates of order $n \times M_k$, Z_m^k is a fixed known $n \times q_{mk}$ design matrix and γ_m^k is a q_{mk}-dimensional random variable. A number of different additive smoothing terms are allowed in Equation (1). By changing the definition of the matrix Z_m^k, it is possible to include P-spline, cubic splines, random-effects, non-parametric random effects and many others.

Betti et al. (2024) proposed the SSIT-GAMLSS, which is typically employed to generate reliable estimates at national or a more detailed disaggregated level, in line with the intended purpose of the surveys. SSITs predominantly incorporate fixed effects and random effects. Building on Equation (1), we propose a SSIT-GAMLSS that considers a regional specific random-effect, i.e. at geographic regional level, and that employs no more than four parameter distributions ($k = 1, \dots, 4$). In particular, the variable of interest $\boldsymbol{y}^{\boldsymbol{T}} = (y_{11}, \dots, y_{ij}, \dots,\ y_{nJ})$ is indexed by i, the i-th unit of the sample, i = 1,…, n, that lives in area j for j = 1,…, J:

$$\begin{cases} g_\mu(\mu_{ij}) = X_{ij}^\mu \beta_\mu + \gamma_j^\mu \\ g_\sigma(\sigma_{ij}) = X_{ij}^\sigma \beta_\sigma + \gamma_j^\sigma \\ g_v(v_{ij}) = X_{ij}^v \beta_v + \gamma_j^v \\ g_\tau(\tau_{ij}) = X_{ij}^\tau \beta_\tau + \gamma_j^\tau \end{cases}, \quad (2)$$

In Equation (2), μ, σ, v *and* τ are the location, scale and two, if any, additional shape parameters of the distribution. The random effects are $\gamma_j^K \sim N(0, \boldsymbol{\Psi_k})$ for $k = 1, \dots, 4$. The variance-covariance matrix $\boldsymbol{\Psi}$ involves the variance of the random effects σ_k^2. The estimated parameters of Equation (2) are then used to impute the variable of interest in the new survey, which must share the same covariates as the survey in which Equation (2) is estimated. Furthermore, Betti et al. (2024) relaxed the assumption of constant regression parameters over time by introducing a weighting scheme. The core idea is that Equation (2) can be estimated for two different years, and in imputing values for an intermediate year, the estimated parameters should reflect a weighted average of those from the two reference years. The closer the imputation year is to one of the reference years, the greater the weight assigned to that year's parameters, and the smaller the weight given to the more distant year. This approach allows a smoother and more realistic transition of parameters over time. Betti et al. (2024) also proposed using a Monte Carlo (MC) approximation to estimate the inequality indices and a non-parametric bootstrap for the Mean Square Error (MSE).

3.2 Adding a double geographical partition

As discussed above, we set out to develop a single model for the Sahel countries, which incorporates variations within (regional) and between countries. To achieve this, we add a second random effect to model Equation (2). We rewrite the vector of the sampled units y^T as $\boldsymbol{y^T} = (y_{11_1 1}, \dots, y_{ij_h h}, \dots, y_{nJ_h H})$ where i is now the i-th unit in the sample, i = 1,…, n, that lives in area j for $j_h = 1, \dots, J_h$, in country h, h = 1,…, H. This means that we can merge the single datasets and estimate a single regression. Equation (2) becomes:

$$\begin{cases} g_\mu(\mu_{ij_h h}) = X_{ij_h h}^\mu \beta_\mu + \gamma_{j_h}^\mu + \lambda_h^\mu \\ g_\sigma(\sigma_{ij_h h}) = X_{ij_h h}^\sigma \beta_\sigma + \gamma_{j_h}^\sigma + \lambda_h^\sigma \\ g_v(v_{ij_h h}) = X_{ij_h h}^v \beta_v + \gamma_{j_h}^v + \lambda_h^v \\ g_\tau(\tau_{ij_h h}) = X_{ij_h h}^\tau \beta_\tau + \gamma_{j_h}^\tau + \lambda_h^\tau \end{cases}, \quad (3)$$

where the vectors of the random effects are $\boldsymbol{\gamma_{j_h}} = \left(\gamma_{j_h}^\mu, \gamma_{j_h}^\sigma, \gamma_{j_h}^v, \gamma_{j_h}^\tau\right)^T$ *and* $\boldsymbol{\lambda_h} = (\lambda_h^\mu, \lambda_h^\sigma, \lambda_h^v, \lambda_h^\tau)^T$ with $\boldsymbol{\gamma_{j_h}} \sim N(0, \boldsymbol{\Psi_\gamma})$ and $\boldsymbol{\lambda_{j_h}} \sim N(0, \boldsymbol{\Psi_\lambda})$ where $\boldsymbol{\Psi_\gamma}$ and $\boldsymbol{\Psi_\lambda}$ are diagonal matrices, i.e. the components are independent, with a matrix such that $\boldsymbol{\Psi_\gamma} = diag(\sigma_{\gamma^\mu}^2, \sigma_{\gamma^\sigma}^2, \sigma_{\gamma^v}^2, \sigma_{\gamma^\tau}^2)$ and $\boldsymbol{\Psi_\lambda} = diag(\sigma_{\lambda^\mu}^2, \sigma_{\lambda^\sigma}^2, \sigma_{\lambda^v}^2, \sigma_{\lambda^\tau}^2)$. The variance-covariance $\boldsymbol{\Gamma}$ is:

$$\boldsymbol{\Gamma} = \begin{bmatrix} \boldsymbol{I_J} \otimes \boldsymbol{\Psi_\gamma} & 0 \\ 0 & \boldsymbol{I_H} \otimes \boldsymbol{\Psi_\lambda} \end{bmatrix},$$

where $\otimes$ is the product of Kronecker and $\boldsymbol{I_J}$ and $\boldsymbol{I_H}$ are identity matrices of dimension $J \times J$ and $H \times H$, respectively. The model in Equation (3) can be estimated using a slightly modified version of the Fisher scoring algorithm, the so-called Cole and Green algorithm (Rigby and Stasinopoulos, 2005).

Once model (3) is estimated, the imputed values $\hat{y}_i$, specifically the imputed value for unit i of the variable of interest, can be calculated. This calculation is performed using the estimated coefficients, random effects and covariates obtained from a second survey in which y is not observed.

We are interested in estimating the parameter H(Y) by imputing Y on the basis of the estimated model parameter. Let us now focus on the Gini index (G) though the same strategy can be used to estimate other inequality indices. Given the first order inclusion probability π_i for each unit i and the corresponding $\hat{y}_i$, we can estimate the Gini index from each dataset in which the imputation based on Equation (3) is possible:

$$\hat{G} = 1 - \frac{1}{\bar{y}N} \sum_{i=1}^{n} \frac{1}{\pi_{(i)}} \hat{y}_{(i)} \frac{(T_{i-1} + T_i)}{N} ,$$

where $\pi_{(i)}$ and $y_{(i)}$ are the first order inclusion probability and the predicted value of the i-th sorted unit, $T_i = \sum_{j=1}^{i} \frac{1}{\pi_{(j)}}$ with $T_0 = 0$, respectively. It is important to note two things: i) if N is unknown, this quantity can be replaced by an estimate derived from the probability of inclusion; ii) to mitigate the impact of prediction errors as effectively as possible, a MC procedure can be utilized. This method involves predicting $\hat{y}_i$ and estimating $\hat{G}$ repeatedly, then averaging the results across all MC runs. The variance of the estimates can be computed by adopting the non-parametric bootstrap suggested by Betti et al. (2024).

4. Model performance and robustness check

4.1 Model performance

We first assess the accuracy of our imputation approach by estimating the model on the 2018 WAEMU survey data and applying it to predict household expenditure in 2021. Our dependent variable is the per-capita expenditure expressed in Purchasing Power Parity (PPP) adjusted for region-specific spatial price index. Note that we have seven countries (i.e. 8 countries without 0 value in Table 1 for 2021 minus Chad for which we do not have the full set of covariates). The imputation model was estimated within the GAMLSS framework, assuming a log-normal distribution of the dependent variable (household expenditure). Both the location and scale components were specified with the same set of covariates to explicitly model heteroskedasticity. Covariates included household size, rural/urban residence, access to water, open defecation, sex and age of the household head, school attendance, literacy status, and random effects at region and country levels to account for contextual heterogeneity.

We then computed a set of standard inequality measures from the imputed expenditure values and compared them with the same indices calculated directly from the 2021 survey expenditure data. The chosen indicators, as discussed above, are the Gini coefficient, the indices GE(0), GE(1) and GE(2), and three percentile- or share-based ratios: P90/P10, P80/P20 and S80/S20.

Table 3 compares the country-specific indicators computed from the observed and imputed expenditures. The point estimates are very close across the two sources and standard errors are essentially identical, indicating stable precision. For instance, the Gini index in Benin is 0.317 with survey data and 0.325 with imputed expenditure; the P90/P10 ratio differs by only -0.257. Entropy-based measures are slightly higher under imputation (e.g. GE(1): 0.176 vs 0.182), whereas GE(2) is almost identical (0.234 vs 0.235). These results are similar for all the other countries.

To assess whether these (small) differences are statistically meaningful, we test equality of the two estimates using paired (resampling-based) t-tests and report the test statistics and two-sided p-value in the last columns. At conventional levels, almost none of the indices shows a statistically significant difference at 5% level.

Table 4 reports the mean relative bias for all the inequality indices, computed at national and NUTS 2 (regional) level. Two alternative model specifications are compared: i) the Sahel-level model, the pooled model estimated jointly on all countries using the specification described in Equation (3), which leverages a larger combined dataset and shared parameters across the region; and ii) the country-level model, which assumes a log-normal distribution and the same set of covariates but is estimated separately for each country, including only a regional random effect. For each country, the table presents the relative bias at national level and the mean regional relative bias at NUTS 2 level. Overall, the results show that the Sahel-level model consistently outperforms the country-level specification, yielding smaller relative biases across most inequality indicators and aggregation levels. As expected, bias values tend to increase from national to regional estimates, reflecting the smaller sample size and greater variability of subnational data. These findings confirm that the pooled Sahel model improves precision, effectively borrowing strength across countries in the small-area estimation framework.

Figure 1 illustrates the distributions of real and imputed expenditure values. The histograms overlap closely, indicating that the model reproduces the shape of the true expenditure distribution very well. The similarity is consistent across the range of values, suggesting that the imputation preserves central tendency and dispersion. These visual impressions are corroborated by formal Kolmogorov Smirnov tests run country by country, with permutation-based p-values of almost 1.00, implying failure to reject equality and, for practical purposes, indistinguishable distributions.

Overall, these results indicate that when trained on the 2018 data, the GAMLSS-based imputation model is capable of generating 2021 expenditure distributions and inequality measures that are encouragingly close to those derived from the actual survey data. This provides strong evidence of the model's validity in cross-year imputation tasks.

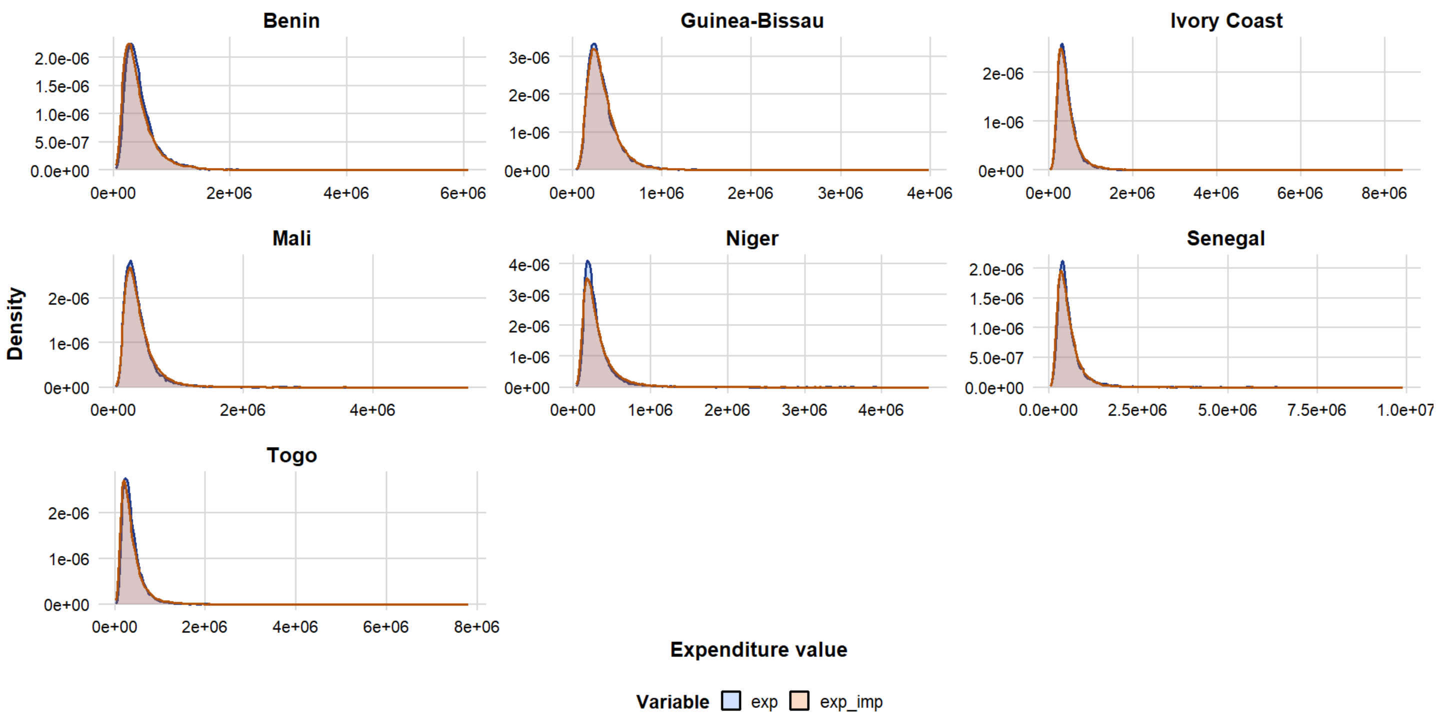
Table 3. Indices estimated on the basis of real data and imputed expenditure.

Country	Index	Expenditure	SE Expenditure	Imputed Expenditure	SE Imputed Expenditure	diff	SE diff	t-stat	p-value
BEN	Gini	0.317	0.015	0.325	0.015	-0.008	0.021	-0.377	0.706
	GE(0)	0.165	0.015	0.174	0.015	-0.009	0.021	-0.424	0.671
	GE(1)	0.176	0.020	0.182	0.020	-0.006	0.028	-0.212	0.832
	GE(2)	0.234	0.045	0.235	0.040	-0.001	0.060	-0.017	0.987
	P80/P20	2.472	0.100	2.606	0.125	-0.134	0.160	-0.837	0.403
	P90/P10	4.041	0.225	4.298	0.270	-0.257	0.351	-0.731	0.465
	S80/S20	4.874	0.305	5.151	0.325	-0.277	0.446	-0.621	0.534
CIV	Gini	0.279	0.010	0.279	0.010	0.000	0.014	0.000	1.000
	GE(0)	0.127	0.010	0.126	0.005	0.001	0.011	0.089	0.929
	GE(1)	0.133	0.010	0.132	0.010	0.001	0.014	0.071	0.944
	GE(2)	0.167	0.035	0.161	0.015	0.006	0.038	0.158	0.875
	P80/P20	2.241	0.065	2.234	0.060	0.007	0.088	0.079	0.937
	P90/P10	3.411	0.115	3.473	0.140	-0.062	0.181	-0.342	0.732
	S80/S20	4.037	0.150	4.028	0.150	0.009	0.212	0.042	0.966
GNB	Gini	0.268	0.005	0.267	0.010	0.001	0.011	0.089	0.929
	GE(0)	0.116	0.005	0.115	0.005	0.001	0.007	0.141	0.888
	GE(1)	0.123	0.010	0.121	0.010	0.002	0.014	0.141	0.888
	GE(2)	0.153	0.020	0.144	0.015	0.009	0.025	0.360	0.719
	P80/P20	2.177	0.065	2.172	0.060	0.005	0.088	0.057	0.955
	P90/P10	3.226	0.110	3.286	0.115	-0.060	0.159	-0.377	0.706
	S80/S20	3.782	0.130	3.785	0.130	-0.003	0.184	-0.016	0.987
MLI	Gini	0.291	0.010	0.303	0.005	-0.012	0.011	-1.073	0.283
	GE(0)	0.138	0.005	0.150	0.005	-0.012	0.007	-1.697	0.090
	GE(1)	0.145	0.010	0.154	0.010	-0.009	0.014	-0.636	0.525
	GE(2)	0.180	0.020	0.188	0.015	-0.008	0.025	-0.320	0.749
	P80/P20	2.336	0.060	2.452	0.065	-0.116	0.088	-1.311	0.190
	P90/P10	3.627	0.130	3.965	0.160	-0.338	0.206	-1.640	0.101
	S80/S20	4.285	0.150	4.599	0.165	-0.314	0.223	-1.408	0.159
NER	Gini	0.320	0.010	0.340	0.010	-0.020	0.014	-1.414	0.157
	GE(0)	0.166	0.015	0.188	0.015	-0.022	0.021	-1.037	0.300
	GE(1)	0.191	0.020	0.211	0.025	-0.020	0.032	-0.625	0.532
	GE(2)	0.281	0.050	0.311	0.075	-0.030	0.090	-0.333	0.739
	P80/P20	2.344	0.075	2.586	0.100	-0.242	0.125	-1.936	0.053
	P90/P10	3.704	0.160	4.320	0.215	-0.616	0.268	-2.298	0.022
	S80/S20	4.667	0.235	5.315	0.280	-0.648	0.366	-1.773	0.076
SEN	Gini	0.308	0.010	0.328	0.010	-0.020	0.014	-1.414	0.157
	GE(0)	0.155	0.010	0.176	0.015	-0.021	0.018	-1.165	0.244
	GE(1)	0.170	0.020	0.192	0.020	-0.022	0.028	-0.778	0.437
	GE(2)	0.233	0.055	0.267	0.050	-0.034	0.074	-0.457	0.647
	P80/P20	2.332	0.085	2.528	0.090	-0.196	0.124	-1.583	0.113
	P90/P10	3.790	0.170	4.176	0.205	-0.386	0.266	-1.449	0.147
	S80/S20	4.596	0.235	5.095	0.260	-0.499	0.350	-1.424	0.154
TGO	Gini	0.316	0.015	0.335	0.010	-0.019	0.018	-1.054	0.292
	GE(0)	0.165	0.015	0.186	0.010	-0.021	0.018	-1.165	0.244
	GE(1)	0.181	0.025	0.192	0.015	-0.011	0.029	-0.377	0.706
	GE(2)	0.263	0.095	0.245	0.035	0.018	0.101	0.178	0.859
	P80/P20	2.462	0.105	2.705	0.125	-0.243	0.163	-1.489	0.137
	P90/P10	3.980	0.180	4.630	0.270	-0.650	0.324	-2.003	0.045
	S80/S20	4.846	0.270	5.477	0.290	-0.631	0.396	-1.593	0.111

Table 4. Mean Relative Bias computed at national and NUTS 2 level using a country-level model and a Sahel-level model.

Country	Index	Country		Regions	
		Country level	Sahel level	Country level	Sahel level
BEN	Gini	14.126	2.524	26.058	13.753
	GE(0)	11.384	5.455	26.878	13.660
	GE(1)	9.197	3.409	33.987	15.824
	GE(2)	6.566	0.427	11.568	6.393
	P80/P20	10.601	5.421	9.179	5.883
	P90/P10	14.745	6.360	16.890	8.915
	S80/S20	13.183	5.683	12.532	6.797
CIV	Gini	14.054	0.001	0.553	-1.957
	GE(0)	14.940	-0.787	-1.984	-3.086
	GE(1)	15.766	-0.752	-5.208	-4.717
	GE(2)	7.120	-3.593	-0.103	-1.168
	P80/P20	5.535	-0.312	0.371	-2.041
	P90/P10	10.727	1.818	2.548	0.551
	S80/S20	10.076	-0.223	2.743	0.652
GNB	Gini	-1.254	-0.373	16.651	1.677
	GE(0)	-4.816	-0.862	18.932	2.787
	GE(1)	-10.212	-1.626	25.466	4.677
	GE(2)	-0.867	-5.882	8.432	1.392
	P80/P20	0.905	-0.230	5.853	0.510
	P90/P10	3.375	1.860	9.874	0.609
	S80/S20	0.395	0.079	6.145	-0.245
MLI	Gini	8.232	4.124	17.864	16.290
	GE(0)	6.388	8.696	15.999	14.599
	GE(1)	4.674	6.207	15.940	14.562
	GE(2)	6.817	4.444	8.024	7.333
	P80/P20	4.379	4.966	8.806	8.028
	P90/P10	9.136	9.319	12.758	10.734
	S80/S20	7.632	7.328	8.054	7.507
NER	Gini	12.953	6.250	13.312	10.898
	GE(0)	13.093	13.253	7.962	7.263
	GE(1)	-8.074	10.471	1.840	3.819
	GE(2)	6.240	10.676	6.009	4.874
	P80/P20	11.906	10.324	9.523	6.445
	P90/P10	18.042	16.631	15.844	13.013
	S80/S20	14.564	13.885	11.693	9.537
SEN	Gini	8.897	6.494	1.018	11.577
	GE(0)	-13.120	13.548	-2.935	10.530
	GE(1)	-12.579	12.941	-8.688	10.280
	GE(2)	0.583	14.592	0.270	5.407
	P80/P20	8.531	8.405	3.546	5.115
	P90/P10	15.866	10.185	2.434	6.260
	S80/S20	10.005	10.857	3.219	6.197
TGO	Gini	29.746	6.013	28.928	12.539
	GE(0)	27.061	12.727	29.515	9.823
	GE(1)	23.583	6.077	35.943	7.935
	GE(2)	13.635	-6.844	13.069	5.565
	P80/P20	15.189	9.870	13.298	7.414
	P90/P10	24.705	16.332	20.381	12.169
	S80/S20	25.837	13.021	15.622	9.263

Figure 1. Expenditure vs Imputed Expenditure distribution



4.2 Model robustness

In the measurement of economic wellbeing, inequality estimates are particularly sensitive to the presence of extreme values, much more than poverty or other welfare indicators, because they assign greater weight to the tails of the distribution. To evaluate the robustness of the fitted model to the presence of extreme values, we conducted a model-based simulation study treating the observed sample as the true underlying population. This 'pseudo-population' approach assumes that the original dataset, which includes complex household-level and contextual variables, adequately captures the distributional features of the target population.

Mimicking the original sample scheme, we drew 200 bootstrap samples, each representing a possible realization of the sampling process, from this reference population. Sample size was equal to 1% of the population (around 4000 units). We introduced artificial outliers in each sample to simulate three levels of contamination: i) low contamination: five extreme observations with expenditure multiplied by 10, ii) moderate contamination: 50 extreme observations with expenditure multiplied by 100, and iii) high contamination: 500 extreme observations with expenditure multiplied by 1000.

For each contaminated sample, the model was re-estimated using the log-normal GAMLSS specification described earlier. Predictions were generated over the full population and inequality measures were computed from these predicted values. Each index was then compared to its corresponding value from the model fitted to the original unperturbed data.

Table 5 shows that the presence of extreme values consistently led to slight underestimation of inequality across all metrics, but the magnitude of the bias remained limited, even under severe contamination. For instance, in the low-contamination scenario, the Gini coefficient had an absolute bias of −0.224 points (−1.012% relative bias), while GE(1) and GE(2) showed relative biases of −1.9% and −1.8%, respectively. Even the more sensitive S80/S20 index changed by only −1.24%.

Under moderate contamination, biases increased slightly (e.g. GE(0) relative bias −2.33%), but still remained small. In the high-contamination setting, no measure showed signs of breakdown, and relative biases remained under −3% for all indices.

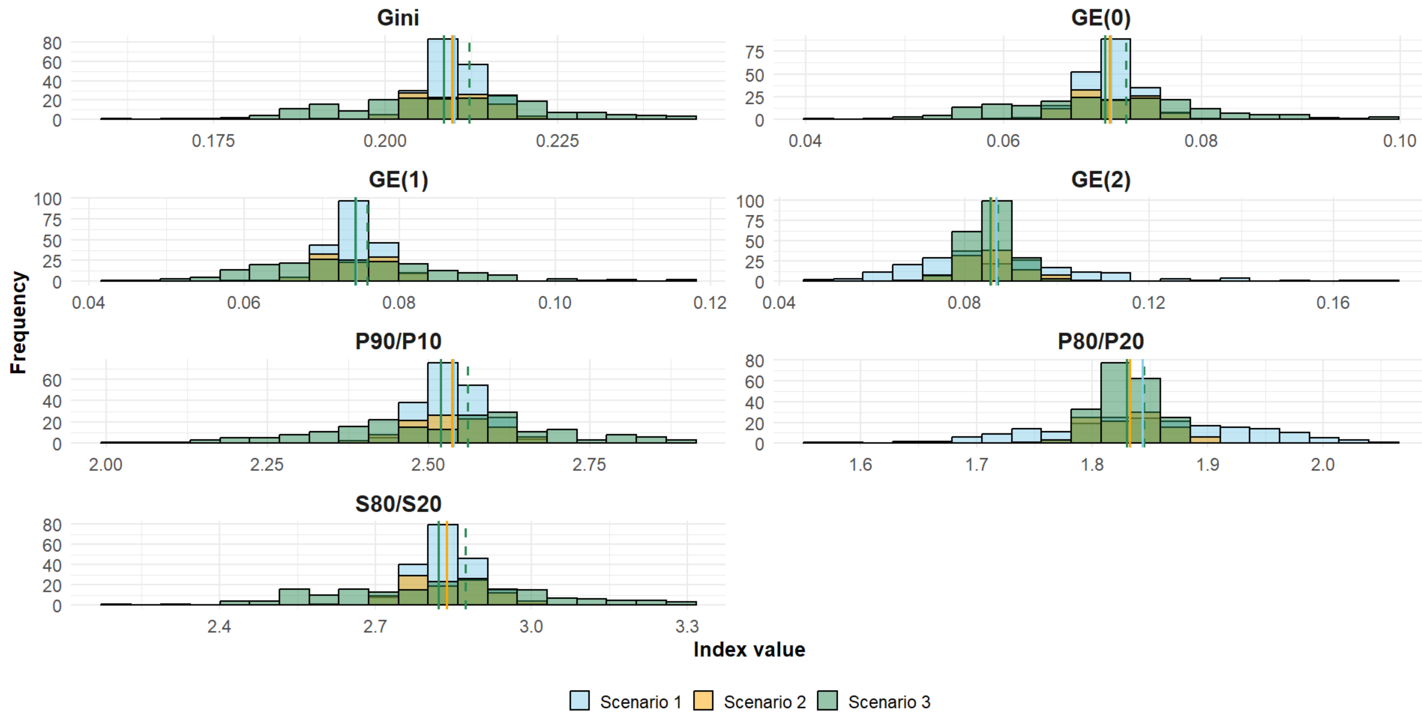
Figure 2 displays the distributions of the bootstrap-derived inequality indices across the three contamination scenarios. The vertical dashed lines mark the corresponding "true" coefficients from the uncontaminated model. In all scenarios, the distributions remain narrow and centred close to the true values, indicating high stability of the estimates even when the data is heavily perturbed. Although greater contamination produces a slight widening of the distributions, especially for P90/P10, P80/P20 and S80/S20, the differences from the true values appears negligible. These patterns confirm the numerical findings in Table 5, showing that GAMLSS maintains strong resistance to the influence of extreme values across a variety of inequality measures.

Taken together, these findings demonstrate that the log-normal GAMLSS, including location and scale components and random effects at the regional and national levels, is highly robust to outliers. Although extreme values introduced a systematic downward bias, the effect was never substantial, even in unrealistic contamination scenarios. This reinforces the model's reliability for empirical applications in settings where anomalous observations may occur.

Table 5. Bias and relative bias (%) of the Gini coefficient computed on model predictions under three contamination scenarios.

Scenario	1		2		3	
	Bias	Relative Bias	Bias	Relative Bias	Bias	Relative Bias
Gini	-0.224	-1.012	-0.254	-1.197	-0.376	-1.777
GE(0)	-0.154	-2.076	-0.168	-2.326	-0.215	-2.970
GE(1)	-0.147	-1.899	-0.154	-2.036	-0.151	-2.044
GE(2)	-0.161	-1.846	-0.147	-1.685	-0.017	-0.343
P90/P10	-2.254	-0.866	-2.425	-0.947	-4.576	-1.639
P80/P20	-1.183	-0.653	-1.204	-0.804	-1.411	-1.442
S80/S20	-3.572	-1.243	-3.649	-1.270	-5.218	-1.816

Figure 2. Histograms of bootstrap indices computed on predicted values from 200 bootstrap replications under three contamination scenarios. The vertical dashed line indicates the true coefficient



5. Inequality in Sahel

To implement our method and analyse inequality trends in the Sahel, we applied the SSIT-GAMLSS framework using harmonized household and labour force survey data for eight WAEMU countries. This application aimed to generate annual estimates of household consumption, which were then used to compute inequality indicators: Gini index, GE(0), GE(1) and GE(2) and ratios (P90/P10, P80/P20 and S80/S20). Given the heterogeneity across countries and regions, the model incorporated both national and sub-national variation through a dual random-effects structure. This allowed us to account for differences not only between countries but also within them, capturing region-specific effects that are often critical in explaining inequality dynamics. We now present the variables included in the model and discuss how they contribute to predicting household consumption across time and space. The model we estimated is based on the 2018 WAEMU wave, where data for Ivory Coast and Guinea-Bissau in 2021 were adjusted to 2018 prices. Note that the regional data was always known for these countries in this year (see Table 1), and it is not a problem if in some years it was unknown when predicting expenditure. The random-effect value of units without region year data equals 0. Before we proceed, it is important to emphasize that the index was computed for each year and country based on the imputed expenditure expressed in PPP dollars using a spatial price index. The use of expenditures instead of income has implications for the magnitude of inequality indexes; when estimated using consumption, the Gini index tends to be lower by about 20 percent than when estimated using income (Clementi et al., 2021, 2023).

The covariates available for this analysis are: i) sex, coded 1 for male and 0 for female; ii) age class, a variable that categorizes respondents into six age groups; iii) Hhsize, household size in seven categories; iv) Rural/Urban, coded 1 for urban households; v) water access, indicating whether a household has piped water on or off premises; (vi) bathroom?, indicating whether a household has an indoor toilet; vii) school attendance, indicating whether the respondent has ever attended school, and vii) literacy, indicating whether the respondent knows how to read and write. Additionally, we need to incorporate two random effects: one based on country and the other on region.

Using these covariates and leveraging the ability of the GAMLSS framework to select the distribution that best fits the data from a wide range of options, we followed the procedure outlined by Mori and Ferrante (2025) to determine the most appropriate type of data distribution.

5.1 Consumption distribution in Sahel

Figure 3 suggests that the distribution must be skewed, with a long right tail. Following Mori and Ferrante (2025), we began by analysing the Akaike Information Criterion (AIC) for a range of potential distributions. We examined the distribution of per-capita expenditure without adjusting for PPP, as we preferred to model the data using the original local currency before subsequently converting each predicted value into PPP values based on the year and country. Table 6 reports the AIC for a number of different distributions (for the parameterization of the distributions see Stasinopoulus et al. 2019). The results clearly show that Log-Normal and Generalized Beta of the Second Kind (GB2) were the two with the lowest AIC, the former performing better than the latter. This result is in line with Barigozzi and Speciale (2011) in the context of documented and undocumented immigrants in Italy.

Table 6. Akaike Information Criterion for selected distributions

Dist	AIC	Dist	AIC
Log-Normal	16261607	Dagum	16335614
GB2	16263452	Gamma	16350006
Skew-*t*	16272788	Weibull	16445245
Inverse Gamma	16277186	Pareto	16623002
Fisk	16294668	Normal	16824507

After selecting the distribution of best fit to the data we estimated our model based on the Log-Normal distribution, using covariates and random-effects on both parameters. Table 7 summarises the model. The standard errors reported in the table are robust, computed via nonparametric bootstrap with 100 iterations to account for potential heteroskedasticity and model uncertainty (Gonçalves and White, 2005). Covariates and the variances of the random effects were statistically different from 0 for both parameters and pseudo-R^2 was 0.51. Residual analysis confirmed that the model is accurate.

Figure 3: Total annual per capita expenditure in $PPP.

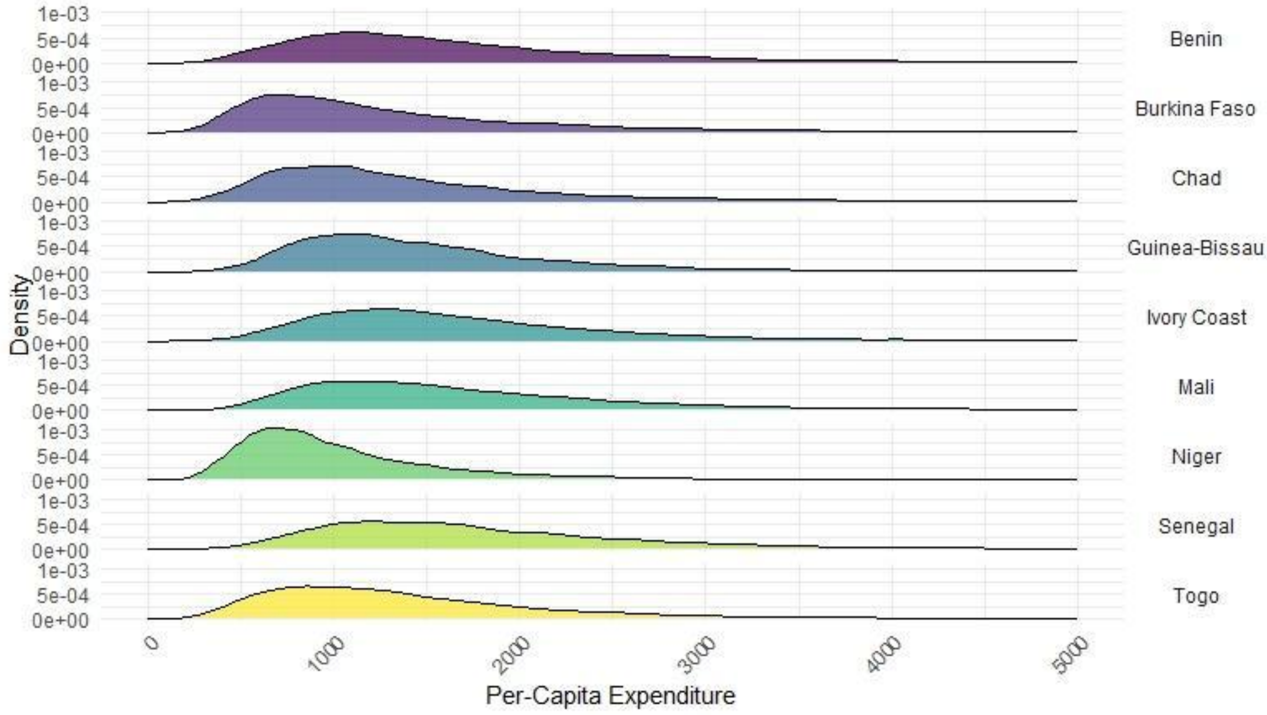


Table 7: Regression summary

***μ* Coefficients**						***σ* Coefficients**					
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**			**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	
(Intercept)	13.4692	0.0162	831.434	0.00000	***	(Intercept)	-0.71647	0.0349	-20.5292	0.00000	***
Hhsize2	-0.33615	0.0179	-18.7793	0.00000	***	Hhsize2	-0.02456	0.0498	-0.49317	0.62297	
Hhsize3	-0.56306	0.0106	-53.1189	0.00000	***	Hhsize3	-0.0868	0.0302	-2.87417	0.00495	***
Hhsize4	-0.70939	0.0259	-27.3896	0.00000	***	Hhsize4	-0.11803	0.0373	-3.16434	0.00206	***
Hhsize5	-0.81619	0.0185	-44.1184	0.00000	***	Hhsize5	-0.11286	0.0302	-3.73709	0.00031	***
Hhsize6	-0.90444	0.0147	-61.5265	0.00000	***	Hhsize6	-0.13202	0.0398	-3.31709	0.00127	***
Hhsize7	-1.10796	0.018	-61.5533	0.00000	***	Hhsize7	-0.13792	0.0316	-4.36456	0.00003	***
Urban	0.13203	0.0051	25.89	0.00000	***	Urban	0.084082	0.006	14.0136	0.00000	***
Waterpipe	0.13714	0.002	68.572	0.00000	***	Waterpipe	0.035765	0.0077	4.64480	0.00001	***
Bathroom	-0.18311	0.0025	-73.244	0.00000	***	Bathroom	-0.03817	0.0083	-4.5988	0.00001	***
sex_1	-0.01074	0.0031	-3.46452	0.00078	***	sex_1	-0.00548	0.0037	-1.48108	0.14173	
Age_class 2	0.04483	0.0048	9.34041	0.00000	***	Age_class2	-0.00742	0.0054	-1.37407	0.17249	
Age_class3	0.05767	0.0086	6.70662	0.00000	***	Age_class3	-0.00593	0.0034	-1.74412	0.08421	*
Age_class4	0.08071	0.0078	10.3483	0.00000	***	Age_class4	0.007978	0.0113	0.70601	0.48182	
Age_class5	0.07977	0.0077	10.3602	0.00000	***	Age_class5	0.005825	0.0101	0.57673	0.56542	
Age_class6	0.06843	0.024	2.85154	0.00529	***	Age_class6	0.03113	0.0619	0.50290	0.61613	
Schooling	0.05969	0.0015	39.79333	0.00000	***	Schooling	-0.01506	0.0143	-1.05315	0.29481	
Literacy	0.12361	0.0073	16.9338	0.00000	***	Literacy	0.0361	0.0117	3.08547	0.00263	***
Random-effect variances											
$\sigma^2_{\gamma^\mu}$	0.14017	0.01626	7.84022	0.00000	***	$\sigma^2_{\gamma^\sigma}$	0.083582	0.00698	11.9643	0.00000	***
$\sigma^2_{\lambda^\mu}$	0.12754	0.01965	7.13371	0.00000	***	$\sigma^2_{\lambda^\sigma}$	0.079902	0.01075	7.43295	0.00000	***
R2: 0.5121											
Residuals analysis: mean = 0.0002845879 variance = 1.000002 coef. of skewness = 0.1378414 coef. of kurtosis = 3.298156											

5.2 Inequality: National level

At the national level (see Figure 4 and 5 for the estimates and Figure B1 and B2 for the coefficients of variation), the inequality measures estimated for the Sahel countries between 2003 and 2021 reveal structural patterns and marked shifts linked to political and economic turbulence. We mostly focus on one index, but similar conclusions can also be drawn from the others, as the reader can see in Figures 4 and 5. In addition, Figure 6 reports the national Gini index based on the data used to plot the maps by way of example. This figure helps illustrate the years affected by key events (such as coups d'état, conflicts and economic crises) and their possible repercussions on inequality dynamics. The Gini index shows values ranging from 0.20 to 0.45, with coefficients of variation consistently below the 16.6% threshold recommended by Statistics Canada, which is widely used as a benchmark for acceptable reliability. Other measures, such as GE(0), which is more sensitive to changes at the lower end of the distribution, and P90/P10, which reflects the gap between the richest and poorest deciles, confirm the same general picture. National averages for GE(0) are around 4.5, while P90/P10 values typically oscillate between 2.5 and 5.5, indicating that income concentration has remained a persistent feature of these economies.

Mali, which is represented in the dataset for 2003, 2009, 2018 and 2021, provides a clear example of how political instability can shape inequality dynamics. The national Gini index declined from 0.28 in 2003 to 0.26 in 2009, a period of relative political stability, before rising to 0.33 in 2018 and 0.29 in 2021. The S80/S20 ratio in Mali followed a similar trajectory, reflecting a widening gap between the richest and poorest income quintiles with a minimum of 3.81 in 2009 and a maximum of 5.20 in 2018.

Niger, which is covered for 2005, 2007, 2010, 2014, 2018 and 2021, exhibits a slower but still significant increase in inequality from 0.26 to 0.31 with a peak of 0.38 in 2018. The P80/P20 ratio remained relatively stable around 2.25 between 2005 and 2014, then rose steadily to 2.69 in 2018. Burkina Faso, with data from 2003, 2014, 2018 and 2021, has one of the highest inequality profiles in the WAEMU region. The GE(1) index, which assigns more weight to disparities at the top of the distribution, increased from 0.14 in 2014 to 0.30 in 2021.

Togo, which offers data in 2006, 2015, 2018 and 2021, has maintained comparatively stable national inequality levels, the Gini index fluctuating slightly around 0.30. The most notable shifts occurred after 2015, reflecting urban-centred economic growth, particularly in the capital area, where infrastructure investment and service sector expansion have not been evenly distributed. Benin, covered in 2003, 2007, 2011, 2018 and 2021, maintains persistently high inequality, the GE(2) index, which weights the highest incomes heavily, averaging 0.263 over the period and showing a peak of 0.33 in 2018. This reflects structural disparities between the economically dominant coastal areas and the less developed northern regions.

Ivory Coast, included in 2018, 2015, 2018 and 2021 after political stabilisation under President Alassane Ouattara, displays relatively small changes in national inequality, the P90/P10 ratio fluctuating in a narrow range around 3.35 and showing a low peak of 3.07 in 2008. Guinea-Bissau, with data for 2010, 2018 and 2021, remains one of the most equal countries in the sample, with national Gini index values close to 0.24, indicating a relatively compressed income distribution even in a context of political fragility.

Senegal, which has data for 2005, 2018 and 2021, is widely considered one of the most politically stable countries in the region and has experienced steady growth in GDP over the period. Nevertheless, its national GE(0) index rose from 0.12 in 2005 to 0.15 in 2021, showing that economic expansion did not automatically translate into a more equal distribution of income.

Figure 4: Choropleth map of inequality indices (Gini, GE(0), GE(1), GE(2)) from 2003 to 2021 in the Sahel region. Years without overlapping countries are grouped and plotted together.

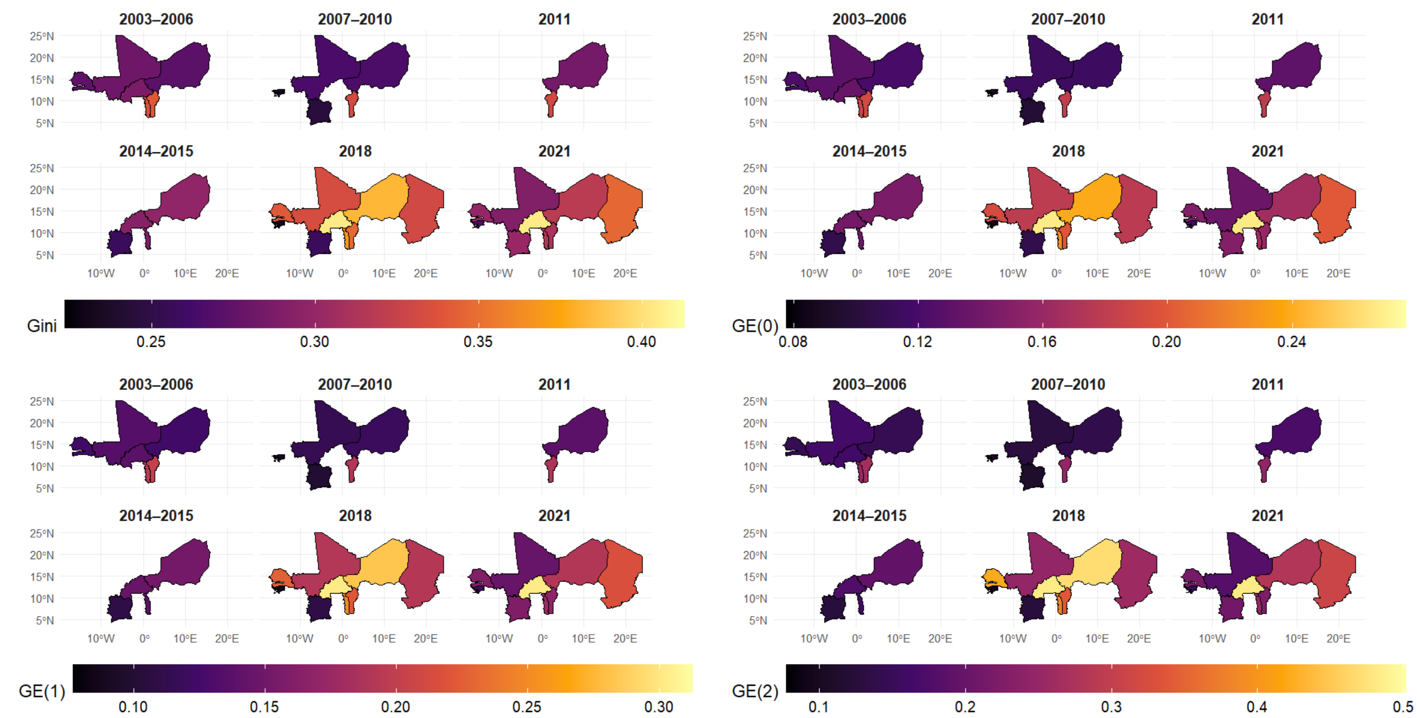


Figure 5: Choropleth map of ratio indices (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel region. Years without overlapping countries are grouped and plotted together.

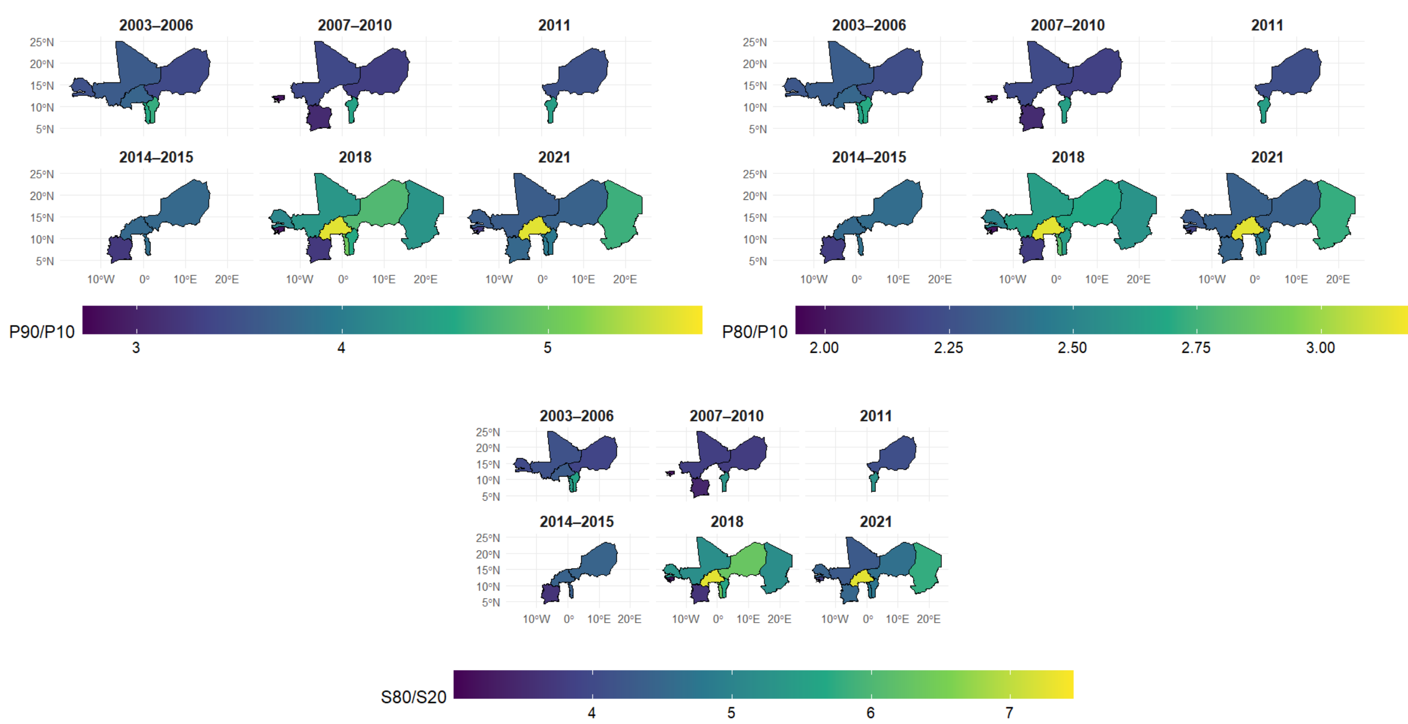
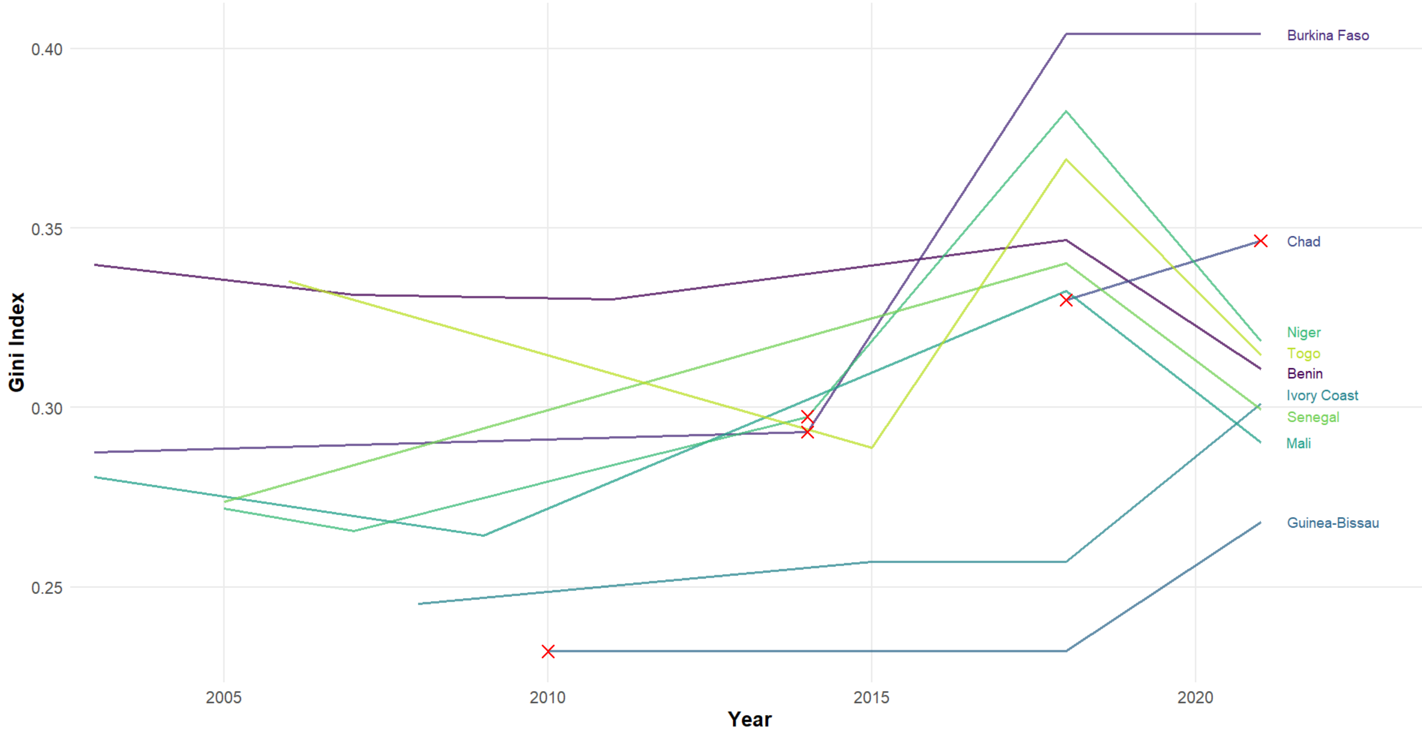


Figure 6. National Gini index trends for Sahel countries (2003–2021). Each coloured line represents a country. Red crosses mark years of political or economic instability (e.g. coups, conflicts, financial crises).



Notably, our inequality and poverty estimates are derived from a measure of household expenditure that accounts explicitly for regional variation in prices in each country. As discussed earlier, our approach divides household expenditure by a spatial price index that captures subnational differences in the cost of living. This feature ensures that our estimates reflect real differences in welfare across heterogeneous regions, but it also implies that they are not directly comparable with standard estimates from international databases that rely on nationally uniform deflators.

Nevertheless, benchmarking our results against external sources is both necessary and informative. In particular, the World Bank's Poverty and Inequality Platform (PIP, World Bank, 2025) and the UNU-WIDER World Income Inequality Database (WIID, UNU-WIDER, 2025) are widely respected references in the field. Comparing our estimates with those provided in these two datasets allows us to contextualize our methodology within the broader empirical literature. However, PIP and WIID figures often diverge substantially from one another, reflecting differences in survey harmonization and welfare aggregates such as treatment of price adjustments. (Interestingly, comparison of several inequality databases shows substantial disagreements (Ferreira et al., 2015)). This divergence reinforces the notion that inequality estimates are highly sensitive to methodological choices.

What emerges as most striking is not the absolute level of inequality reported in the different sources, but the similarity of the trends over time. Despite differences in levels, our estimates generally track the same direction of change as those in PIP, and to a somewhat lesser extent, those in WIID. This suggests that while methodological refinements, such as our adjustment for regional prices, may shift the magnitude of inequality, they do not alter the fundamental story told by the data about its evolution.

For instance, in Benin our estimate of the Gini index decreases from 0.34 in 2003 to 0.31 in 2021 showing a small improvement in inequality. While the levels differ, all three sources document a steady decline over the period, reinforcing confidence in the robustness of the observed trend. In PIP, in-fact, the Gini index declines from 0.38 to 0.34 and in WIID from 0. 53 to 0.50. Similarly, in Mali our estimates indicate quite a stable Gini index until 2018, closely matching the PIP and WIID trend over the same period. Once again, the common direction of change highlights the consistency of the dynamics across data sources, even if differences in methodology translate into distinct levels.

Taken together, these comparisons show that our estimates, while methodologically distinct due to the incorporation of regional price indices, remain aligned with the best available international evidence in terms of trend detection. This dual perspective, recognizing both the unavoidable level differences and converging trends, adds credibility to our approach and underlines the value of producing country-specific and spatially sensitive measures of welfare distribution

5.3 Inequality: Regional level

The regional breakdown of inequality indicators shows patterns that are often masked by national averages. For brevity and clarity, the regional analysis focuses solely on the Gini index, reported in Figure 7 together with its coefficient of variation. As already illustrated at national level in Figure 6, which highlights the years marked by coups, conflicts and economic crises, such events are also relevant for interpreting regional inequality dynamics. Additional results are provided in Appendix B: Figure B3 shows choropleth maps of inequality indices at regional level (GE(0), GE(1), GE(2)); Figure B4 shows choropleth maps of ratio indices at regional level (P90/P10, P80/P20, S80/S20); Figure B5 depicts the coefficient of variation of the inequality indices (GE(0), GE(1), GE(2)); and Figure B6 displays the coefficient of variation of the ratio indices (P90/P10, P80/P20, S80/S20). These subnational results highlight the importance of accounting for spatial heterogeneity when assessing inequality, as local dynamics frequently diverge from national trends. Coefficients of variation at regional level are naturally higher than at national level, approaching 10% percent in sparsely populated areas, but they remain within the limits considered acceptable for analytical purposes.

We first consider the countries that experienced riots, wars or coups between 2003 and 2021. In Mali, for example, the trend observed at regional level, where the Bamako region shows the highest level of inequality and Taoudénit the lowest, is similar to the one observed at national level. The year-to-year fluctuations in the index can be explained by historical events. In 2012 and 2013, Mali underwent a coup d'état that destabilized the economy and reversed the declining trend in inequality. Additionally, Mali and Niger both faced challenges related to Islamic fundamentalist insurgencies. In the capital region of Niger, Niamey, the estimated Gini coefficient shows stability from 0.28 in 2005 to 0.29 in 2011 and 2014. This downward trajectory suggests a gradual reduction in income inequality during this period. However, between 2014 and 2018, the Gini index increases sharply to 0.40, indicating a significant widening of income disparities. The subsequent drop to 0.36 in 2021 points to a partial reversal of this inequality surge, though the value remains higher than in the mid-2000s. Overall, the trend reflects a period of relative equality gains in the 2000s followed by a notable setback in the late 2010s.

Burkina Faso demonstrates a relatively high level of inequality across all regions and years, with the Gini index ranging from 0.24 in the Centre-West region in 2014 to 0.37 in the Centre region in 2018. Like Mali, Burkina Faso experienced political unrest, including riots and coups, starting in 2014, which contributed to the increase in inequality. As a result, all regions show higher levels of inequality in 2018 than in 2003.

For Chad, data is only available for the years 2018 and 2021, when President Idriss Déby Itno was still in power. Inequality increased slightly in this period. The lowest Gini index, 0.26, was recorded in 2021 in the Sila region, while the highest, 0.36, was observed in the Logone Orientale region the same year. These values highlight the considerable regional variation in inequality.

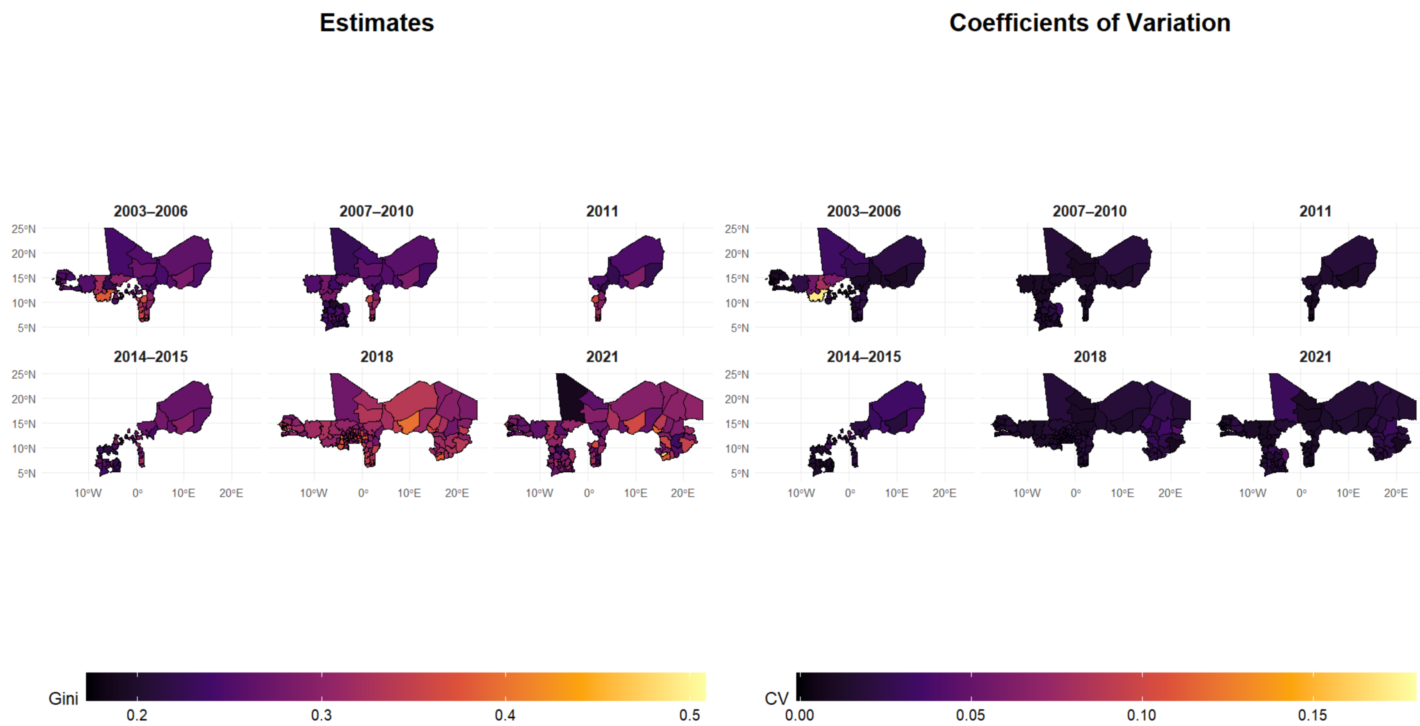
Togo serves as a link between the countries experiencing instability and those that are relatively stable. Although it underwent a coup d'état, Togo's inequality remained relatively stable throughout the years considered. The only region to experience a notable change in inequality was Grand Lom, where the Gini index increased from 0.30 in 2006 to 0.34 in 2021, after an initial decrease in 2015.

Among the more stable countries, Benin shows a higher level of inequality than Mali, the Atacora region scoring a Gini index of 0.37 in 2018. In general, inequality remains high across all regions, with the Collines recording the lowest value of 0.25. Inequality has remained relatively stable over the past two decades, reflecting the political and socio-economic context of the country.

In Haut-Sassandra, the most populous region of the Ivory Coast, the Gini index is quite stable between 2008 and 2015. By 2021, the HCR has risen slightly to 0.27, signalling a partial reversal of previous gains. Given the region's demographic weight, changes in poverty levels here have a substantial influence on national poverty figures, making trends in Haut-Sassandra particularly critical for policy planning and resource allocation.

Senegal is considered one of the most stable countries in the region, and shows a constant increase in gross domestic product. However, inequality remains relatively high, ranging from a minimum of 0.22 in 2005 in the Thies region to a maximum of 0.32 in 2021 in the Kedougou region. A slight increase in inequality is observed across all regions of Senegal.

Figure 7: Choropleth map of Gini index inequality (left) and its coefficient of variation (right) at regional level from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.



6. Conclusion

This paper offers a comprehensive reconstruction of inequality dynamics in the Sahel over the last two decades, addressing both methodological and empirical gaps in the literature. By leveraging SSITs enhanced by a GAMLSS framework, we overcame the challenges posed by limited and irregular consumption data, a major obstacle to robust welfare analysis in many West African countries.

The flexibility of GAMLSS in incorporating distributional assumptions beyond the exponential family and integrating regional and national random effects proved particularly effective for capturing the heterogeneity of consumption distributions across space and time. The introduction of time-varying parameters by a weighted approach also allowed us to model smoother transitions in welfare indicators between survey years, improving the accuracy of intertemporal estimates.

Empirically, our findings highlighted substantial spatial and temporal variations in inequality, with clear evidence of divergence between relatively stable countries and those affected by conflict or political instability. While countries like Senegal and Benin exhibit relatively stable inequality levels over time, others—particularly Mali, Niger and Burkina Faso—show a sharp increase in inequality, often linked to episodes of violence, displacement or economic disruption. Granular region-level estimates further revealed the importance of sub-national dynamics, underscoring the need for disaggregated data in policy design.

In addition to its methodological contributions, this study helps fill a void in empirical research on francophone West Africa. The underrepresentation of this region in economic literature has long limited the availability of actionable evidence for policymakers. By producing a consistent and comparable set of inequality estimates across eight countries and nearly 20 years, this paper contributes to building a more robust evidence base for monitoring inclusive growth in the Sahel.

Future research could extend this methodology to other welfare dimensions, such as multidimensional poverty or vulnerability to shocks. Improved data availability, especially from more frequent and harmonized household surveys, would significantly enhance the reliability and scope of this type of analysis.

References

Alfani F., Dabalen A., Fisker P., Molini V. (2019). Vulnerability to stunting in the West African Sahel. *Food Policy*, ***83***, pp. 39 – 47

Barigozzi, M., & Speciale, B. (2011). Immigrants' legal status, permanence in the destination country and the distribution of consumption expenditure. *Applied Economics Letters*, ***18***, 1341-1347.

Betti, G., Molini, V., & Mori, L. (2024). An attempt to correct the underestimation of inequality measures in cross-survey imputation through generalized additive models for location, scale and shape. *Socio-Economic Planning Sciences*, ***91***, 101784.

Clementi, F., Fabiani, M., Molini, V., & Schettino, F. (2023). Is inequality systematically underestimated in Sub-Saharan Africa? A proposal to overcome the problem. *A Proposal to Overcome the Problem. World Bank Policy Research Working Paper*, *10058*.

Clementi, F., Fabiani, M., Molini, V., & Schettino, F. (2021). Response to 'Polarization and Poverty Reduction in Africa: The Devil Is in the Choice of Equivalence Relation'. *Journal of African Economies*, ***31***, 176-179.

Cornia, G. A. (2017). Inequality Levels, Trends and Determinants in Sub-Saharan Africa: an overview of main changes since the early 1990s. *Income Inequality Trends in Sub-Saharan Africa: Divergence, Determinants and Consequences*, 23-51.

Dabalen, A., Narayan, A., Saavedra-Chanduvi, J., & Suarez, A. H. (2014). *Do African children have an equal chance?: A human opportunity report for sub-Saharan Africa*. World Bank Publications.

Dang, H. A.., Lanjouw, P. F., & Serajuddin, U. (2017). Updating poverty estimates in the absence of regular and comparable consumption data. *Oxford Economic Papers*, ***69***, 939-962.

Dang, H. A., Jolliffe, D., & Carletto, C. (2019). Data gaps, data incomparability and data imputation: A review of poverty measurement methods for data-scarce environments. *Journal of Economic Surveys*, ***33***, 757-797.

Dang, H. A.., & Lanjouw, P. F. (2023). Regression-based imputation for poverty measurement in data-scarce settings. In *Research handbook on measuring poverty and deprivation* (pp. 141-150). Edward Elgar Publishing.

Dang, H. A.., Abanokova, K., Kilic, T., & Carletto, C. (2025a). *Guidebook for Using Survey-to-Survey Imputation for Poverty Measurement (English).* LSMS Guidebook Washington, D.C. : World Bank Group. http://documents.worldbank.org/curated/en/099711008012519694

Dang, H. A.., Kilic, T., Abanokova, K., & Carletto, C. (2025b). Poverty imputation in contexts without consumption data: a revisit with further refinements. *Review of Income and Wealth*, ***71***, e12714.

Dang, H. A., Kilic, T., Hlasny, V., Abanokova, K., & Carletto, C. (forthcoming). Using Survey-to-Survey Imputation to Fill Poverty Data Gaps at a Low Cost: Evidence from a Randomized Survey Experiment. *World Bank Economic Review.*

David, A., Leibbrandt, M., Ranchhod, V., & Yasser, R. (2025). *Inequalities in Sub-Saharan Africa: Multidimensional Perspectives and Future Challenges*. World Bank Publications.

Demombynes, G., & Hoogeveen, J. G. (2007). Growth, inequality and simulated poverty paths for Tanzania, 1992–2002. *Journal of African Economies*, ***16***, 596-628.

Douidich, M., Ezzrari, A., Van der Weide, R., & Verme, P. (2016). Estimating quarterly poverty rates using labour force surveys: a primer. *World Bank Economic Review*, ***30***, 475-500.

Elbers, C., Lanjouw, J. O., & Lanjouw, P. (2003). Micro-level estimation of poverty and inequality. *Econometrica*, ***71***, 355-364.

Elbers, C., Lanjouw, P. F., Mistiaen, J. A., Özler, B., & Simler, K. (2004). On the unequal inequality of poor communities. *World Bank Economic Review*, ***18***, 401-421.

Ferreira, F. H., Lustig, N., & Teles, D. (2015). Appraising cross-national income inequality databases: An introduction. *Journal of Economic Inequality*, *13*(4), 497-526.

Fosu, A. K. (2009). Inequality and the impact of growth on poverty: comparative evidence for sub-Saharan Africa. *Journal of Development Studies*, ***45***, 726-745.

Fosu, A. K. (2017). Growth, inequality and poverty reduction in developing countries: Recent global evidence. *Research in Economics*, ***71***, 306-336.

Fosu, A. K. (2017). Instabilities and development in Africa. In *Towards Africa's Renewal* (pp. 209-224). Routledge.

Fosu, A. K. (2018). The recent growth resurgence in Africa and poverty reduction: the context and evidence. *Journal of African Economies*, ***27***, 92-107.

Gonçalves, S., & White, H. (2005). Bootstrap standard error estimates for linear regression. *Journal of the American Statistical Association*, **100**, 970-979.

Krafft, C., Assaad, R., Nazier, H., Ramadan, R., Vahidmanesh, A., & Zouari, S. (2019). Estimating poverty and inequality in the absence of consumption data: an application to the Middle East and North Africa. *Middle East Development Journal*, 11(1), 1-29.

Marc, A., Verjee, N., & Mogaka, S. (2015). *The challenge of stability and security in West Africa*. World Bank Publications.

Mori, L., & Ferrante, M. R. (2025). Small Area Estimation of Household Economic Indicators under Unit-Level Generalized Additive Models for Location, Scale and Shape. *Journal of Survey Statistics and Methodology*, ***13***, 160-196.

Nikiema, R. A. (2025). Intrahousehold Inequality, Economies of Scale and Poverty: Evidence from Eight Harmonised Living Standard Measurement Surveys Data in West Africa. *Journal of Development Studies*, 1-26.

Odusola, A. (2017). Fiscal space, poverty and inequality in Africa. *African Development Review*, ***29***, 1-14.

Porteous, O. (2022). Research deserts and oases: evidence from 27 thousand economics Journal Articles on Africa. *Oxford Bulletin of Economics and Statistics*, ***84***, 1235-1258.

Rigby, R. A., & Stasinopoulos, D. M. (2005). Generalized additive models for location, scale and shape. *Journal of the Royal Statistical Society Series C: Applied Statistics*, ***54***, 507-554.

Schluter, C. (2018). Top incomes, heavy tails and rank-size regressions. *Econometrics*, ***6***, 10.

Tarozzi, A., & Deaton, A. (2009). Using census and survey data to estimate poverty and inequality for small areas. *Review of Economics and Statistics*, ***91***, 773-792.

UNU-WIDER. United Nations University – World Institute for Development Economics Research.(2025). WIID – World Income Inequality Database (version May 2029) [Data set] UNU-WIDER https://www.wider.unu.edu/project/wiid-world-income-inequality-database Accessed November 17, 2025.

World Bank.(2025).Poverty and Inequality Platform (version 20250930_2021_01_02_PROD) [Data set] World Bank Group www.pip.worldbank.org Accessed November 17, 2025.

7. Appendix A

Table A1: Dataset available

Country	Year	Survey	Consumption (Y = yes, N = no)	Sample size HH
Benin	2003	Enquête Modulaire Intégrée sur les Conditions de Vie des Ménages 2003	N	5350
Benin	2007	Enquête Modulaire Intégrée sur les Conditions de Vie des Ménages 2007	N	17823
Benin	2011	Enquête Modulaire Intégrée sur les Conditions de Vie des Ménages 2011	N	17975
Benin	2015	Enquête Modulaire Intégrée sur les Conditions de Vie des Ménages 2015	N	21434
Benin	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	8012
Benin	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	8032
Burkina Faso	1994	Enquête Multisectorielle Continue 1994	N	8642
Burkina Faso	1998	Enquête Multisectorielle Continue 1998	N	8478
Burkina Faso	2003	Enquête Multisectorielle Continue 2003	N	8500
Burkina Faso	2005	Enquête Multisectorielle Continue 2005	N	8439
Burkina Faso	2009	Enquête Multisectorielle Continue 2009	N	8404
Burkina Faso	2014	Enquête Multisectorielle Continue 2014	N	10800
Burkina Faso	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	7010
Burkina Faso	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	N	7176
Chad	2003	Enquête sur la Consommation des Ménages et le Secteur Informel au Tchad 2003	N	6730
Chad	2011	Enquête sur la Consommation des Ménages et le Secteur Informel au Tchad 2011	N	9259
Chad	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	7493
Chad	2022	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	7532
Cote D'Ivoire	2002	Enquête Niveaui de Vie des Ménages 2002	N	10799
Cote D'Ivoire	2008	Enquête Niveau de Vie des Ménages 2008	N	12600
Cote D'Ivoire	2015	Enquête Niveau de Vie des Ménages 2015	N	12899
Cote D'Ivoire	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	N	12992
Cote D'Ivoire	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	13693
Guinea-Bissau	1993	Inquerito Ligeiro para a Avaliação da Pobreza 1993	N	3308
Guinea-Bissau	2010	Inquerito Ligeiro para a Avaliação da Pobreza 2010	N	3178
Guinea-Bissau	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	N	5351
Guinea-Bissau	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	5351
Mali	1994	Enquête Légère Intégrée auprès des Ménages 1994	N	9484
Mali	2003	Enquête Légère Intégrée auprès des Ménages 2003	N	4122
Mali	2009	Enquête Légère Intégrée auprès des Ménages 2009	N	9235
Mali	2010	Enquête Légère Intégrée auprès des Ménages 2010	N	2976
Mali	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	6602
Mali	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	6143
Niger	1995	National Survey on Household Living Conditions and Agriculture 1995	N	4383

Country	Year	Survey	Consumption (Y = yes, N = no)	Sample size HH
Niger	2002	National Survey on Household Living Conditions and Agriculture 2002	N	2500
Niger	2005	National Survey on Household Living Conditions and Agriculture 2005	N	6690
Niger	2007	National Survey on Household Living Conditions and Agriculture 2007	N	4000
Niger	2011	National Survey on Household Living Conditions and Agriculture 2011	N	3968
Niger	2014	National Survey on Household Living Conditions and Agriculture 2014	N	3699
Niger	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	6024
Niger	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	6632
Senegal	1995	Enquête de Suivi de la Pauvreté au Sénégal 1995	N	979
Senegal	2001	Enquête de Suivi de la Pauvreté au Sénégal 2001	N	6589
Senegal	2005	Enquête de Suivi de la Pauvreté au Sénégal 2005	N	13542
Senegal	2011	Enquête de Suivi de la Pauvreté au Sénégal 2011	N	17878
Senegal	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	7156
Senegal	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	7120
Togo	2001	Questionnaire des Indicateurs de Base du Bien-être 2001	N	2500
Togo	2006	Questionnaire des Indicateurs de Base du Bien-être 2006	N	7500
Togo	2011	Questionnaire des Indicateurs de Base du Bien-être 2011	N	5532
Togo	2015	Questionnaire des Indicateurs de Base du Bien-être 2015	N	2367
Togo	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	6172
Togo	2021	Enquête Harmonisee sur les Conditions de Vie des Menages	Y	6462

Table A2: Summary of variables: Benin

Year	**Variable**	**Mean**	**SD**	**Q25**	**Median**	**Q75**	**Min**	**Max**
2003	age_class	2.92	1.046	2	3	4	2	6
2007	age_class	2.144	1.257	1	2	3	1	6
2011	age_class	2.26	1.27	1	2	3	1	6
2018	age_class	2.082	1.194	1	2	3	1	6
2021	age_class	1.954	1.054	1	2	2	1	6
2003	everattd	0.398	0.489	0	0	1	0	1
2007	everattd	0.478	0.499	0	0	1	0	1
2011	everattd	0.566	0.496	0	1	1	0	1
2018	everattd	0.576	0.494	0	1	1	0	1
2021	everattd	0.937	0.243	1	1	1	0	1
2003	hhsize	5.354	1.836	4	6	7	1	7
2007	hhsize	5.574	1.674	4	6	7	1	7
2011	hhsize	5.681	1.674	5	7	7	1	7
2018	hhsize	5.668	1.593	5	6	7	1	7
2021	hhsize	5.61	1.546	5	6	7	1	7
2003	literacy	0.067	0.249	0	0	0	0	1
2007	literacy	0.45	0.498	0	0	1	0	1
2011	literacy	0.496	0.5	0	0	1	0	1
2018	literacy	0.479	0.5	0	0	1	0	1

Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
2021	literacy	1	0	1	1	1	1	1
2003	open_def	0.334	0.471	0	0	1	0	1
2007	open_def	0.989	0.103	1	1	1	0	1
2011	open_def	0.964	0.187	1	1	1	0	1
2018	open_def	0.551	0.497	0	1	1	0	1
2021	open_def	0.421	0.494	0	0	1	0	1
2003	rururb	0.617	0.486	0	1	1	0	1
2007	rururb	0.646	0.478	0	1	1	0	1
2011	rururb	0.6	0.49	0	1	1	0	1
2018	rururb	0.474	0.499	0	0	1	0	1
2021	rururb	0.569	0.495	0	1	1	0	1
2003	sex	0.507	0.5	0	1	1	0	1
2007	sex	0.507	0.5	0	1	1	0	1
2011	sex	0.512	0.5	0	1	1	0	1
2018	sex	0.489	0.5	0	0	1	0	1
2021	sex	0.552	0.497	0	1	1	0	1
2003	waterpipe	0.099	0.299	0	0	0	0	1
2007	waterpipe	0.077	0.267	0	0	0	0	1
2011	waterpipe	0.054	0.226	0	0	0	0	1
2018	waterpipe	0.289	0.453	0	0	1	0	1
2021	waterpipe	0.342	0.475	0	0	1	0	1

Table A3: Summary of variables: Burkina Faso

Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
2003	age_class	2.5	1.192	2	2	3	1	6
2014	age_class	2.089	1.218	1	2	3	1	6
2018	age_class	2.155	1.262	1	2	3	1	6
2003	everattd	0.289	0.453	0	0	1	0	1
2014	everattd	0.467	0.499	0	0	1	0	1
2018	everattd	0.482	0.5	0	0	1	0	1
2003	hhsize	6.995	0.115	7	7	7	2	7
2014	hhsize	6.974	0.257	7	7	7	1	7
2018	hhsize	6.108	1.379	5	7	7	1	7
2003	literacy	0.284	0.451	0	0	1	0	1
2014	literacy	0.349	0.477	0	0	1	0	1
2018	literacy	0.431	0.495	0	0	1	0	1
2003	open_def	0.981	0.135	1	1	1	0	1
2014	open_def	1	0	1	1	1	1	1
2018	open_def	0.344	0.475	0	0	1	0	1
2003	rururb	0.699	0.459	0	1	1	0	1
2014	rururb	0.658	0.474	0	1	1	0	1
2018	rururb	0.393	0.488	0	0	1	0	1
2003	sex	0.518	0.5	0	1	1	0	1
2014	sex	0.533	0.499	0	1	1	0	1
2018	sex	0.476	0.499	0	0	1	0	1
2003	waterpipe	0.078	0.268	0	0	0	0	1
2014	waterpipe	0.105	0.306	0	0	0	0	1
2018	waterpipe	0.408	0.492	0	0	1	0	1

Table A4: Summary of variables: Chad

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Chad	2018	age_class	2	1.188	1	2	3	1	6
Chad	2018	everattd	0.636	0.481	0	1	1	0	1
Chad	2018	hhsize	5.803	1.551	5	7	7	1	7
Chad	2018	literacy	0.291	0.454	0	0	1	0	1
Chad	2018	open_def	0.493	0.5	0	0	1	0	1

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Chad	2018	rururb	0.501	0.5	0	1	1	0	1
Chad	2018	sex	0.484	0.5	0	0	1	0	1
Chad	2018	waterpipe	0.199	0.399	0	0	0	0	1

Table A5: Summary of variables: Guinea-Bissau

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Guinea-Bissau	2010	age_class	2.811	1.052	2	2	3	1	6
Guinea-Bissau	2021	age_class	2.232	1.234	1	2	3	1	6
Guinea-Bissau	2010	everattd	0.528	0.499	0	1	1	0	1
Guinea-Bissau	2021	everattd	0.685	0.464	0	1	1	0	1
Guinea-Bissau	2010	hhsize	7	0	7	7	7	7	7
Guinea-Bissau	2021	hhsize	6.395	1.164	6	7	7	1	7
Guinea-Bissau	2010	literacy	0.527	0.499	0	1	1	0	1
Guinea-Bissau	2021	literacy	0.531	0.499	0	1	1	0	1
Guinea-Bissau	2010	open_def	0.855	0.353	1	1	1	0	1
Guinea-Bissau	2021	open_def	0.113	0.316	0	0	0	0	1
Guinea-Bissau	2010	rururb	0.548	0.498	0	1	1	0	1
Guinea-Bissau	2021	rururb	0.355	0.479	0	0	1	0	1
Guinea-Bissau	2010	sex	0.528	0.499	0	1	1	0	1
Guinea-Bissau	2021	sex	0.483	0.5	0	0	1	0	1
Guinea-Bissau	2010	waterpipe	0.065	0.246	0	0	0	0	1
Guinea-Bissau	2021	waterpipe	0.475	0.499	0	0	1	0	1

Table A6: Summary of variables: Ivory Coast

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Ivory Coast	2008	age_class	2.274	1.155	1	2	3	1	6
Ivory Coast	2015	age_class	2.26	1.219	1	2	3	1	6
Ivory Coast	2021	age_class	2.176	1.273	1	2	3	1	6
Ivory Coast	2008	everattd	0.525	0.499	0	1	1	0	1
Ivory Coast	2015	everattd	0.429	0.495	0	0	1	0	1
Ivory Coast	2021	everattd	0.53	0.499	0	1	1	0	1
Ivory Coast	2008	hhsize	7	0.02	7	7	7	4	7
Ivory Coast	2015	hhsize	6.934	0.43	7	7	7	1	7
Ivory Coast	2021	hhsize	5.495	1.679	4	6	7	1	7
Ivory Coast	2008	literacy	0.518	0.5	0	1	1	0	1
Ivory Coast	2015	literacy	0.371	0.483	0	0	1	0	1
Ivory Coast	2021	literacy	0.419	0.493	0	0	1	0	1
Ivory Coast	2008	open_def	0.866	0.341	1	1	1	0	1
Ivory Coast	2015	open_def	0.639	0.48	0	1	1	0	1
Ivory Coast	2021	open_def	0.311	0.463	0	0	1	0	1
Ivory Coast	2008	rururb	0.49	0.5	0	0	1	0	1
Ivory Coast	2015	rururb	0.549	0.498	0	1	1	0	1
Ivory Coast	2021	rururb	0.257	0.437	0	0	1	0	1
Ivory Coast	2008	sex	0.492	0.5	0	0	1	0	1
Ivory Coast	2015	sex	0.496	0.5	0	0	1	0	1
Ivory Coast	2021	sex	0.497	0.5	0	0	1	0	1
Ivory Coast	2008	waterpipe	0.228	0.419	0	0	0	0	1
Ivory Coast	2015	waterpipe	0.3	0.458	0	0	1	0	1
Ivory Coast	2021	waterpipe	0.294	0.455	0	0	1	0	1

Table A7: Summary of variables: Mali

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Mali	2003	age_class	4.116	0.963	3	4	5	2	6
Mali	2009	age_class	2.938	1.077	2	3	4	2	6

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Mali	2018	age_class	2.354	1.323	1	2	3	1	6
Mali	2021	age_class	2.203	1.308	1	2	3	1	6
Mali	2003	everattd	0.25	0.433	0	0	0	0	1
Mali	2009	everattd	0.305	0.46	0	0	1	0	1
Mali	2018	everattd	0.541	0.498	0	1	1	0	1
Mali	2021	everattd	0.559	0.497	0	1	1	0	1
Mali	2003	hhsize	6.731	0.857	7	7	7	1	7
Mali	2009	hhsize	6.997	0.104	7	7	7	1	7
Mali	2018	hhsize	6.211	1.266	6	7	7	1	7
Mali	2021	hhsize	6.263	1.204	6	7	7	1	7
Mali	2003	literacy	0.288	0.453	0	0	1	0	1
Mali	2009	literacy	0.313	0.464	0	0	1	0	1
Mali	2018	literacy	0.422	0.494	0	0	1	0	1
Mali	2021	literacy	0.427	0.495	0	0	1	0	1
Mali	2003	open_def	0.629	0.483	0	1	1	0	1
Mali	2009	open_def	0.945	0.228	1	1	1	0	1
Mali	2018	open_def	0.163	0.369	0	0	0	0	1
Mali	2021	open_def	0.091	0.288	0	0	0	0	1
Mali	2003	rururb	0.636	0.481	0	1	1	0	1
Mali	2009	rururb	0.624	0.484	0	1	1	0	1
Mali	2018	rururb	0.419	0.493	0	0	1	0	1
Mali	2021	rururb	0.43	0.495	0	0	1	0	1
Mali	2003	sex	0.093	0.291	0	0	0	0	1
Mali	2009	sex	0.534	0.499	0	1	1	0	1
Mali	2018	sex	0.484	0.5	0	0	1	0	1
Mali	2021	sex	0.494	0.5	0	0	1	0	1
Mali	2003	waterpipe	0.198	0.398	0	0	0	0	1
Mali	2009	waterpipe	0.349	0.477	0	0	1	0	1
Mali	2018	waterpipe	0.451	0.498	0	0	1	0	1
Mali	2021	waterpipe	0.526	0.499	0	1	1	0	1

Table A8: Summary of variables: Niger

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Niger	2005	age_class	2.742	1.103	2	2	3	1	6
Niger	2007	age_class	2.152	1.215	1	2	3	1	6
Niger	2011	age_class	2.121	1.234	1	2	3	1	6
Niger	2014	age_class	2.082	1.113	1	2	3	1	6
Niger	2018	age_class	2.036	1.224	1	2	3	1	6
Niger	2021	age_class	2.069	1.254	1	2	3	1	6
Niger	2005	everattd	0.425	0.494	0	0	1	0	1
Niger	2007	everattd	0.57	0.495	0	1	1	0	1
Niger	2011	everattd	0.557	0.497	0	1	1	0	1
Niger	2014	everattd	0.993	0.085	1	1	1	0	1
Niger	2018	everattd	0.544	0.498	0	1	1	0	1
Niger	2021	everattd	0.603	0.489	0	1	1	0	1
Niger	2005	hhsize	6.985	0.208	7	7	7	1	7
Niger	2007	hhsize	7	0	7	7	7	7	7
Niger	2011	hhsize	7	0	7	7	7	7	7
Niger	2014	hhsize	6.775	0.75	7	7	7	1	7
Niger	2018	hhsize	5.924	1.459	5	7	7	1	7
Niger	2021	hhsize	5.869	1.442	5	7	7	1	7
Niger	2005	literacy	0.38	0.485	0	0	1	0	1
Niger	2007	literacy	0.349	0.477	0	0	1	0	1
Niger	2011	literacy	0.3	0.458	0	0	1	0	1
Niger	2014	literacy	1	0	1	1	1	1	1
Niger	2018	literacy	0.31	0.463	0	0	1	0	1
Niger	2021	literacy	0.375	0.484	0	0	1	0	1
Niger	2005	open_def	0.951	0.216	1	1	1	0	1

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Niger	2007	open_def	0.963	0.188	1	1	1	0	1
Niger	2011	open_def	0.956	0.206	1	1	1	0	1
Niger	2014	open_def	0.909	0.288	1	1	1	0	1
Niger	2018	open_def	0.655	0.475	0	1	1	0	1
Niger	2021	open_def	0.549	0.498	0	1	1	0	1
Niger	2005	rururb	0.669	0.47	0	1	1	0	1
Niger	2007	rururb	0.526	0.499	0	1	1	0	1
Niger	2011	rururb	0	0	0	0	0	0	0
Niger	2014	rururb	0.363	0.481	0	0	1	0	1
Niger	2018	rururb	0.262	0.44	0	0	1	0	1
Niger	2021	rururb	0.387	0.487	0	0	1	0	1
Niger	2005	sex	0.507	0.5	0	1	1	0	1
Niger	2007	sex	0.517	0.5	0	1	1	0	1
Niger	2011	sex	0.508	0.5	0	1	1	0	1
Niger	2014	sex	0.424	0.494	0	0	1	0	1
Niger	2018	sex	0.482	0.5	0	0	1	0	1
Niger	2021	sex	0.48	0.5	0	0	1	0	1
Niger	2005	waterpipe	0.139	0.346	0	0	0	0	1
Niger	2007	waterpipe	0.187	0.39	0	0	0	0	1
Niger	2011	waterpipe	0.192	0.394	0	0	0	0	1
Niger	2014	waterpipe	0.425	0.494	0	0	1	0	1
Niger	2018	waterpipe	0.374	0.484	0	0	1	0	1
Niger	2021	waterpipe	0.524	0.499	0	1	1	0	1

Table A9: Summary of variables: Senegal

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Senegal	2005	age_class	2.861	1.063	2	2	3	1	6
Senegal	2018	age_class	2.227	1.288	1	2	3	1	6
Senegal	2021	age_class	2.16	1.152	1	2	3	1	6
Senegal	2005	everattd	0.405	0.491	0	0	1	0	1
Senegal	2018	everattd	0.675	0.468	0	1	1	0	1
Senegal	2021	everattd	0.945	0.227	1	1	1	0	1
Senegal	2005	hhsize	6.952	0.392	7	7	7	1	7
Senegal	2018	hhsize	6.629	0.995	7	7	7	1	7
Senegal	2021	hhsize	6.526	1.083	7	7	7	1	7
Senegal	2005	literacy	0.431	0.495	0	0	1	0	1
Senegal	2018	literacy	0.48	0.5	0	0	1	0	1
Senegal	2021	literacy	1	0	1	1	1	1	1
Senegal	2005	open_def	0.633	0.482	0	1	1	0	1
Senegal	2018	open_def	0.08	0.271	0	0	0	0	1
Senegal	2021	open_def	0.041	0.199	0	0	0	0	1
Senegal	2005	rururb	0.366	0.482	0	0	1	0	1
Senegal	2018	rururb	0.529	0.499	0	1	1	0	1
Senegal	2021	rururb	0.612	0.487	0	1	1	0	1
Senegal	2005	sex	0.536	0.499	0	1	1	0	1
Senegal	2018	sex	0.464	0.499	0	0	1	0	1
Senegal	2021	sex	0.503	0.5	0	1	1	0	1
Senegal	2005	waterpipe	0.496	0.5	0	0	1	0	1
Senegal	2018	waterpipe	0.639	0.48	0	1	1	0	1
Senegal	2021	waterpipe	0.704	0.457	0	1	1	0	1

Table A10: Summary of variables: Togo

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Togo	2006	age_class	2.908	1.054	2	3	4	2	6
Togo	2015	age_class	2.201	1.263	1	2	3	1	6
Togo	2018	age_class	2.209	1.279	1	2	3	1	6

Country	Year	Variable	Mean	SD	Q25	Median	Q75	Min	Max
Togo	2021	age_class	2.274	1.335	1	2	3	1	6
Togo	2006	everattd	0.605	0.489	0	1	1	0	1
Togo	2015	everattd	0.749	0.434	0	1	1	0	1
Togo	2018	everattd	0.691	0.462	0	1	1	0	1
Togo	2021	everattd	0.729	0.445	0	1	1	0	1
Togo	2006	hhsize	5.241	1.758	4	6	7	1	7
Togo	2015	hhsize	7	0	7	7	7	7	7
Togo	2018	hhsize	5.229	1.762	4	6	7	1	7
Togo	2021	hhsize	5.206	1.711	4	5	7	1	7
Togo	2006	literacy	0.539	0.498	0	1	1	0	1
Togo	2015	literacy	0.615	0.487	0	1	1	0	1
Togo	2018	literacy	0.594	0.491	0	1	1	0	1
Togo	2021	literacy	0.63	0.483	0	1	1	0	1
Togo	2006	open_def	0.912	0.283	1	1	1	0	1
Togo	2015	open_def	0.782	0.413	1	1	1	0	1
Togo	2018	open_def	0.522	0.5	0	1	1	0	1
Togo	2021	open_def	0.507	0.5	0	1	1	0	1
Togo	2006	rururb	0.65	0.477	0	1	1	0	1
Togo	2015	rururb	0.394	0.489	0	0	1	0	1
Togo	2018	rururb	0.315	0.465	0	0	1	0	1
Togo	2021	rururb	0.34	0.474	0	0	1	0	1
Togo	2006	sex	0.514	0.5	0	1	1	0	1
Togo	2015	sex	0.511	0.5	0	1	1	0	1
Togo	2018	sex	0.478	0.5	0	0	1	0	1
Togo	2021	sex	0.473	0.499	0	0	1	0	1
Togo	2006	waterpipe	0.288	0.453	0	0	1	0	1
Togo	2015	waterpipe	0.356	0.479	0	0	1	0	1
Togo	2018	waterpipe	0.183	0.387	0	0	0	0	1
Togo	2021	waterpipe	0.287	0.453	0	0	1	0	1

8. Appendix B

Figure B1: Choropleth map of coefficient of variation of inequality indices (Gini, GE(0), GE(1), GE(2)) from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.

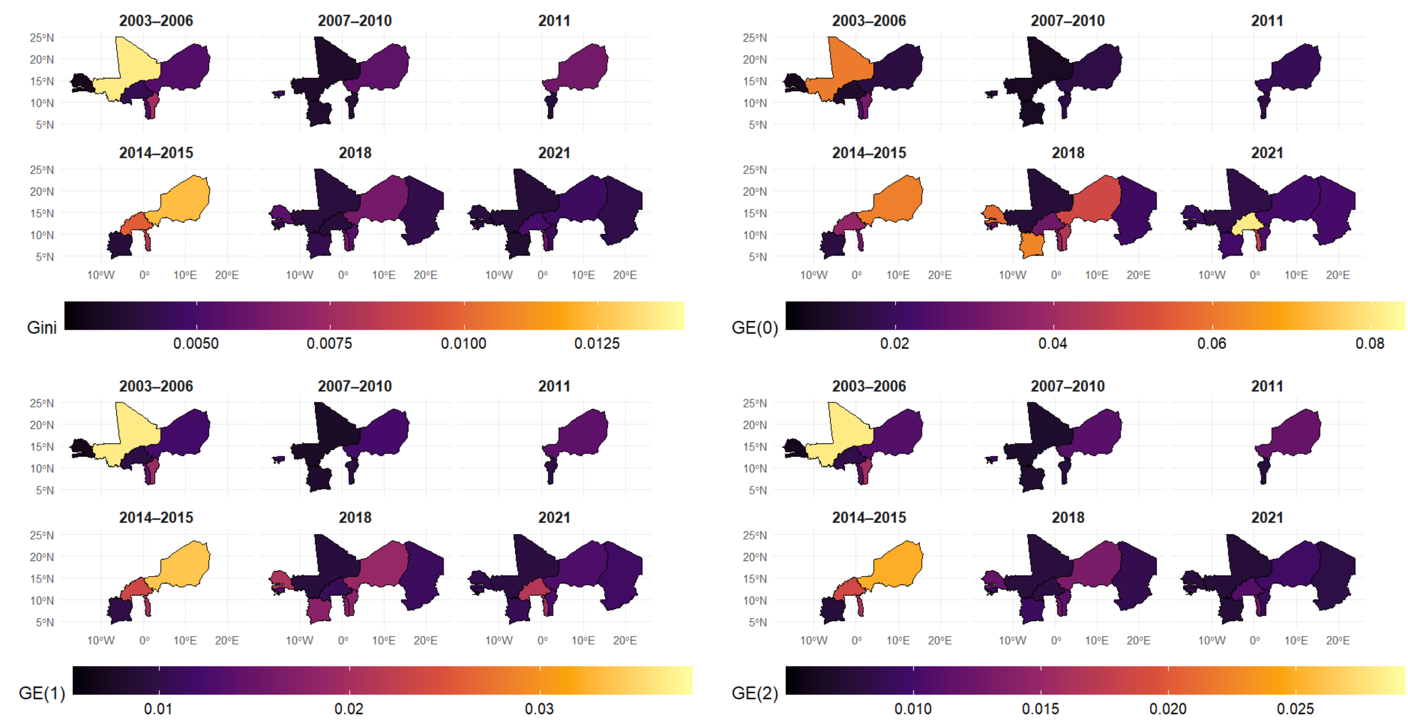


Figure B2: Choropleth map of coefficient of variation of ratio indices (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.

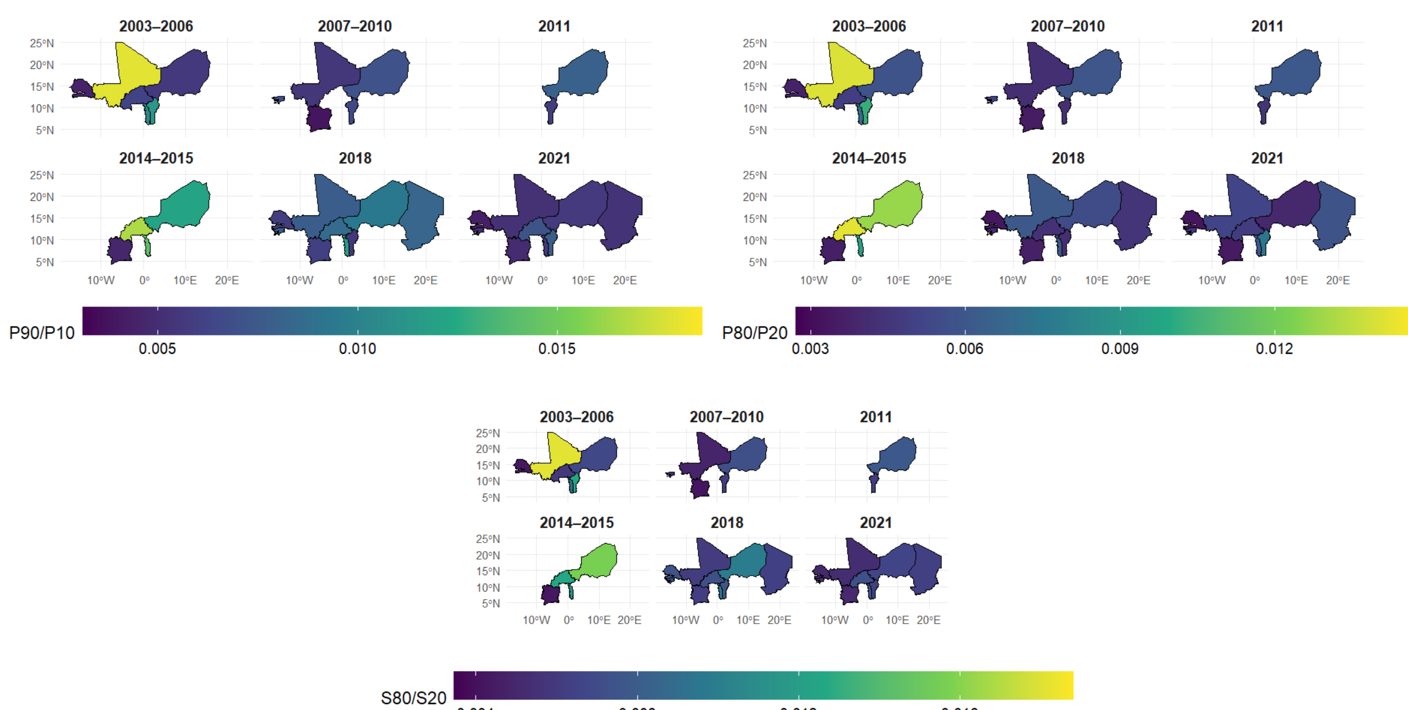


Figure B3: Choropleth map of inequality indices at regional level (Gini, GE(0), GE(1), GE(2)) from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.

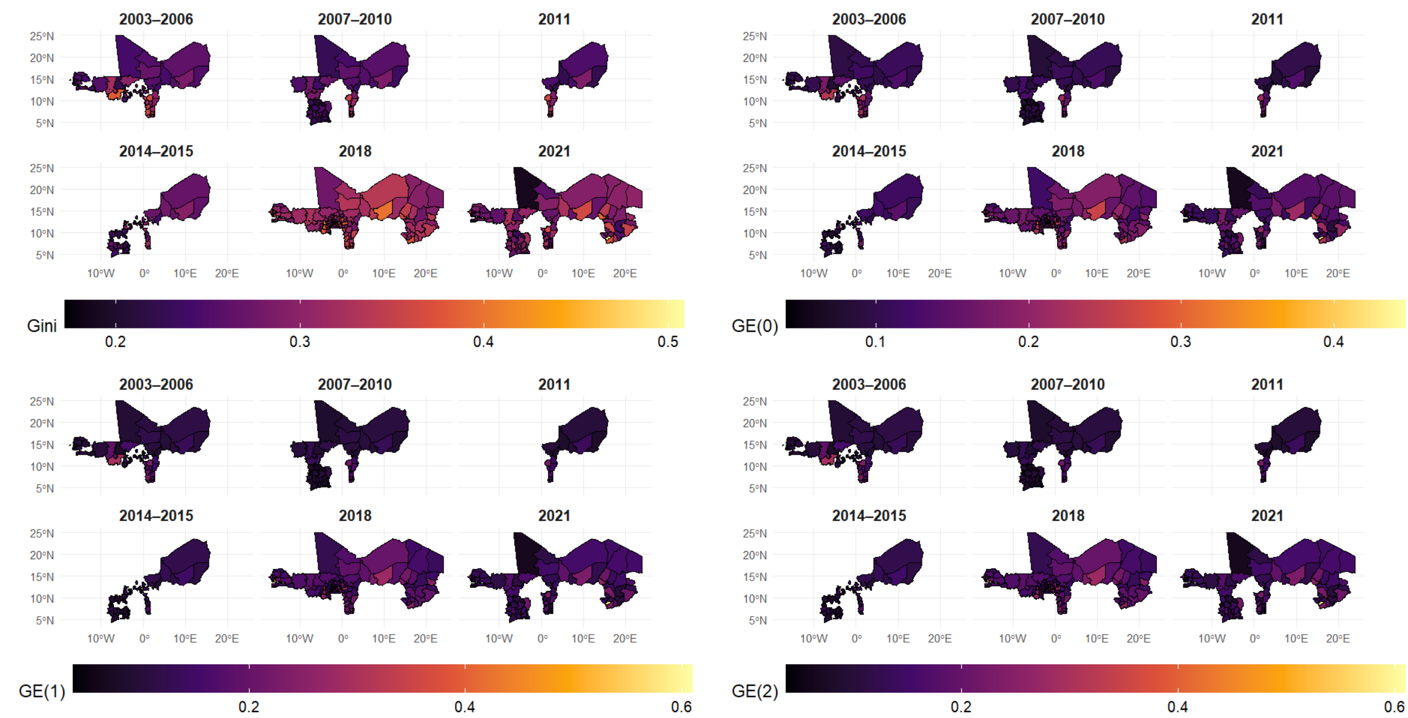


Figure B4: Choropleth map of ratio indices at regional level (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.

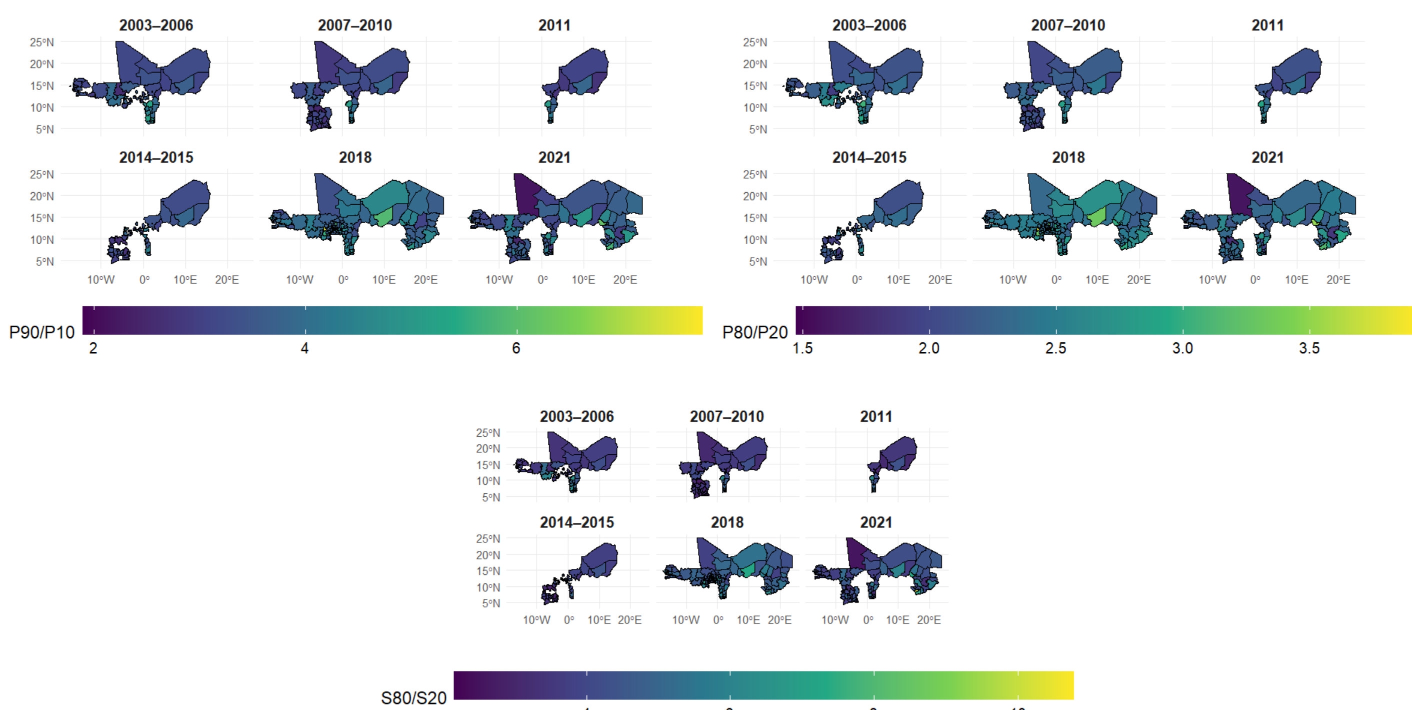


Figure B5: Choropleth map of coefficient of variation of inequality indices at regional level (Gini, GE(0), GE(1), GE(2)) from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.

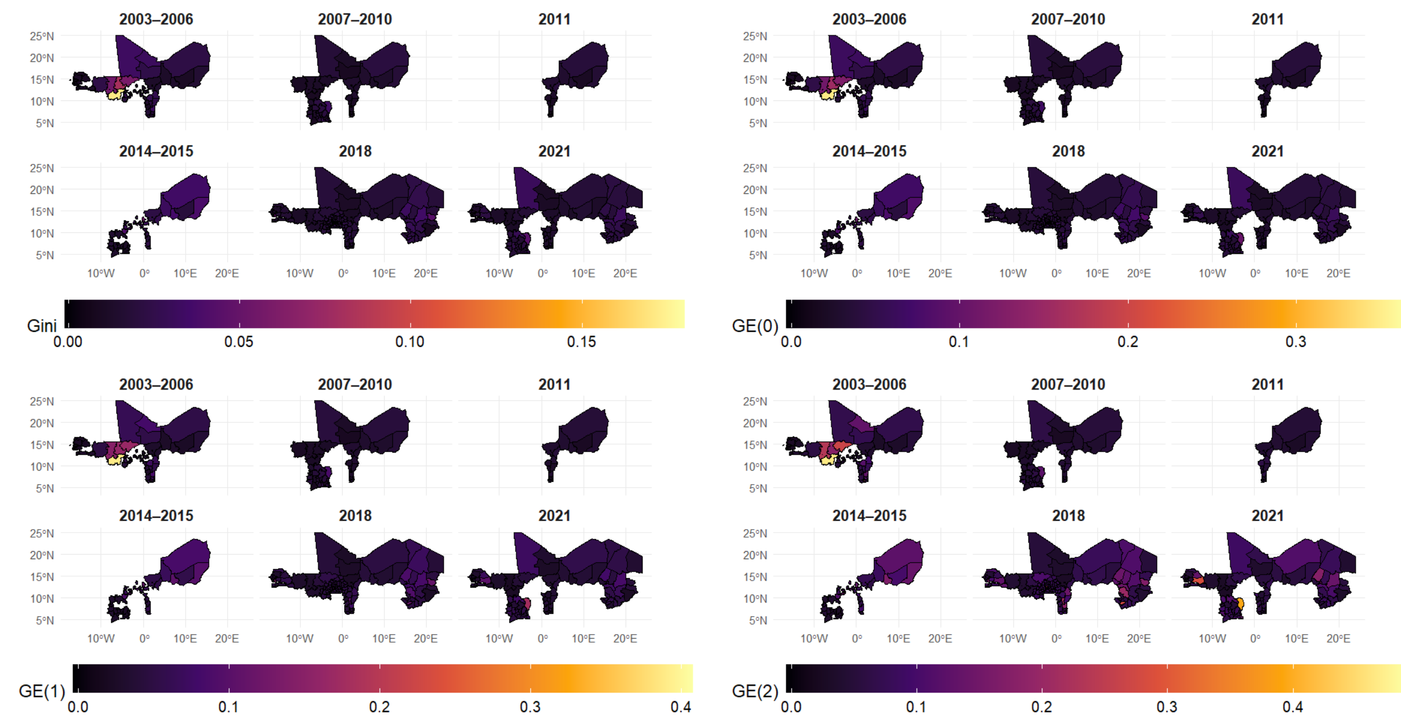


Figure B6: Choropleth map of coefficient of variation of ratio indices at regional level (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel. Years without overlapping countries are grouped and plotted together.