IARIW 2025

IARIW 2025

Thursday, October 2 & Friday, October 3

Reconstructing Two Decades of Inequality in the Sahel Region

Gianni Betti

(University of Siena)

Federico Crescenzi

(University of Tuscia)

Hai-Anh Dang

(World Bank)

Vasco Molini

(World Bank)

Lorenzo Mori

(University of Bologna)

Paper prepared for the IARIW–World Bank–UEB/VNU Conference on "Improving Well-being Measurement in Data-challenged Environments in Developing Countries for Better Evidence-based Policies" October 2-3, 2025

Session 3A Poverty and Inequality

Time Slot: Friday, October 3, 3:30- 5:30 PM

Reconstructing Two Decades of Inequality in the Sahel Region

Betti,[1] Gianni, Crescenzi[2], Federico, Dang[3], Hai-Anh, Molini[3], Vasco, Mori[4], Lorenzo

Abstract: This paper reconstructs the evolution of inequality in the Sahel region over the past two decades through an innovative framework that integrates Survey-to-Survey Imputation Techniques (SSITs) with Generalized Additive Models for Location, Scale, and Shape (GAMLSS). The measurement of inequality in West Africa is often constrained by the limited availability and irregular collection of household consumption data. To address this challenge, we impute consumption onto labor force surveys conducted in eight countries between 2003 and 2021, allowing for the production of consistent and comparable estimates of inequality. The findings highlight pronounced regional disparities, persistent levels of inequality, and a clear association between inequality patterns and episodes of conflict or political instability. This study offers a dual contribution: on the methodological side, it introduces a flexible SSIT-GAMLSS model that incorporates two levels of random effects; on the substantive side, it addresses a significant gap in the literature by providing new evidence on inequality trends in francophone West Africa, a region that remains underrepresented in empirical economic research.

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 rather 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, posing serious challenges for farmers compared to those in the more coastal areas of Western Africa. Farmers in the Sudan-Sahel region face significant rainfall constraints, which hinder their agricultural productivity. 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. This disparity has contributed to the widening regional divide within Western Africa.

Additionally, episodes of insurgency affecting countries such as Mali, Burkina Faso, Niger, and Chad have further deteriorated the socio-economic conditions of these regions. These conflicts have likely 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 the disruption of infrastructure and the negative impact on economies in these regions are quite significant. Based on this preliminary evidence, it is crucial to monitor the well-being of these countries in a timely manner to address and mitigate the adverse effects of these challenges that are happening more and more frequently.

Collecting poverty and inequality that can document the socio-economic impact of these events in a reliable and timely way is unfortunately too costly and complicated for many Western African countries. Not many of these manage collecting data quite regularly, and even in these cases, the interval between surveys is often very long: it is very difficult to have estimates of poverty and inequality every 5/6 years and almost impossible to have them on an annual basis. To overcome the problem, scholars have

[1] Department of of Economics and Statistics, University of Siena, Italy
[2] Department of Economics, Statistics and Business, Universitas Mercatorum, Italy
[3] World Bank
[4] Department of Statistical Sciences, University of Bologna, Italy

focused on the development of accurate methods to compare over time welfare indicators from surveys characterized by limited 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 and Lanjouw 2023, Dang et al. 2019, Dang et. 2025a, 2025b). These techniques, which involve imputing income onto censuses, have been widely applied in developing countries to provide geographically disaggregated estimates of poverty. Recently, there has been a trend towards survey-to-survey imputations, mapping 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., 2014).

While this approach is well-established for poverty estimates, there is limited attention in the literature to potential problems in obtaining accurate inequality measures using this method. Standard SSITs, based on regression and the assumption of normally distributed residuals, are more accurate in predicting central moments of a distribution rather than the shape of the tails, which is crucial for inequality prediction (Schluter, 2012).

Demombynes et al. (2007) found that correlations between estimated and true welfare at the 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 Labor 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 adding to Stata's s2s (povimp) command an option to draw residuals from the empirical distribution of the error terms rather than from a normal, mitigating potential bias.

Betti et al. (2024) address the limitations of SSITs in measuring inequality by using 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 the inclusion of covariates and random effects in each parameter of the distribution, even when the distribution does not belong to the exponential family. In their original study, the authors applied this methodology to estimate inequality in Morocco using a GB2 model with covariates in all four parameters and random effects on the location and scale parameters. In this paper, we build on their approach by incorporating two distinct random effects: one that captures differences among areas within the same country and another that accounts for differences across countries. The main goal is the estimation of 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 an innovative methodology to address the problem of inequality estimation, this paper is innovative also regarding the regional focus. 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 economic literature, leading to a smaller evidence base for local policymakers in these regions. The francophone West African countries, in particular, suffer from a scarcity of economic literature. This 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 between surveys that include the consumption variable and those where it is missing (Appendix A provides a complete summary of the datasets). Section 3 briefly reviews the SSIT-GAMLSS methodology and presents the extension with a double random-effect structure. In Section 4 we present simulation analysis to assess the model performances. Section 5 presents the results on inequality across the Sahel region over a 20-year period. Section 6 concludes the paper.

2. Data

The data used as a reference for the consumption model are 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 the member countries of the WAEMU 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; however, other socioeconomic and demographic variables remain available for these countries.

Harmonized data collection is essential for the WAEMU Commission to monitor national economies and support the 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 labor 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 households covered ranging from 6,000 to 8,000. The surveys are conducted in two waves to account for the seasonality of consumption: the first wave is conducted between October and December, while the second wave is conducted between April and July. Each wave covers half of the sample, ensuring that the data collected reflects seasonal variations in living conditions and consumption patterns.

The labor force surveys span a longer time than the WAEMU data and are collected more frequently than the normal household budget surveys. However, since their scope is not to capture well-being, they lack information on consumption. In this paper, we impute consumption on these surveys to provide a more regular picture of poverty and inequality over the last two decades. However, not all available labor force surveys from these countries 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 select 17 labor force surveys on which we can impute consumption plus 3 WAEMU surveys for which consumption is not available. The surveys by country and year are reported in Table 1 and indicated by the number 1. The survey with the x are those form WAEMU 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 within each country used for statistical purposes. Two asterisks (**) indicate that NUTS 2 information is available only for some statistical units within the dataset. If no asterisk is present, 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 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 labor 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 provide a consistent and reliable source for cross-country and temporal comparisons. The harmonized dataset is structured into four modules: (1) Module P (poverty-related variables), (2) Module H (household-level variables excluding poverty indicators), (3) Module I (individual-level variables excluding labor force data), and (4) Module L (labor force variables).

For all countries listed in Table 1, we report a summary of selected key variables that will be used in subsequent analyses. In the paper, we present only the mean, standard deviation, and median for each variable, while the Appendix A include tables form A2 to A10 containing additional statistics such as quartiles, minimum, and maximum values. Notably, for Chad in 2021, although WAEMU expenditure data are available, other variables of interest are missing. This is not a limitation for the present analysis, as in that year only expenditure data are essential.

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

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			Waterpipe			Sex			Age Class			Ever Attendance			Literacy		
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
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

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			Waterpipe			Sex			Age Class			Ever Attendance			Literacy		
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 aim of SSIT is to estimate H(Y) even for years in which Y is not observed, by imputing it using a model-based procedure (see Dang et al., 2024). In the following, we present a strategy to impute Y based on GAMLSS.

GAMLSS were proposed by Rigby and Stasinopoulos (2005) and incorporate a location parameter, a scale parameter, and up to two shape parameters, with the possibility of using covariates and random effects in each parameter of the chosen distribution. The distribution does not need to belong to the exponential family, encompassing a wide range of commonly encountered distribution types.

GAMLSS assume 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, $k = 1, \dots, p$ $(\theta\_i\hat{}T = (\theta\_i1, \dots, \theta\_ik, \dots, \theta\_ip))$. More formally, let $y^T = (y_1, \dots, y_n)$ be the n length vector of the response variable. Let $g_k(\cdot)$ be a known monotonic link functions relating the p distribution parameters to explanatory variables by:

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

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 (1). Changing the definition of the matrix Z_m^k is possible to include P-spline, cubic splines, random-effects, non-parametric random effects, and many others.

Starting from GAMLSS Betti et al. propose the SSIT-GAMLSS which are typically employed to generate reliable estimates either at the national level or at a more detailed disaggregated level, aligning with the surveys' intended purpose. SSITs predominantly incorporate fixed effects and random effects. Moving from (1) we propose a SSIT-GAMLSS by considering regional specific random-effect, i.e. at the geographic regional level, and limiting our attention to four or less parameter distributions ($k = 1, \dots, 4$). In particular, the variable of interest $\boldsymbol{y^T} = (y_{11}, \dots, y_{ij}, \dots, y_{nJ})$ is indexed by i the i-th unit in the sample, i = 1,…, n, that lives in the area j with j = 1,…, J:

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

where μ, σ, ν *and* τ represents, respectively, the location, scale and, if any, two addition shape parameters of the considered distribution. In (2) random effects are $\gamma_j^\kappa \sim N(0, \boldsymbol{\Psi_k})$ for $k = 1, \dots, 4$. The variance-covariance matrix $\boldsymbol{\Psi}$ of the multivariate Normal involves the variance of the random effects σ_k^2. The estimated parameters of (2) are then used to impute the variable of interest in the new survey which must share the same covariates with the survey in which the model (2) is estimated. Furthermore, Betti et al. (2024) relax the assumption of constant regression parameters over time by introducing a weighting scheme. The core idea is that the model in equation (2) can be estimated for two different years, and when 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 for a smoother and more realistic transition of parameters over time. To conclude, in the paper authors propose also the use of Monte Carlo (MC) approximation to estimate the inequality indices and a non-parametric bootstrap for the estimation of the Mean Square Error (MSE).

3.2 Adding a double geographical partition

As previously said here we are interested in estimating a unique model for all the considered countries that has to take into account both the variation between areas in the same country, i.e. regional variation, and variation between countries. To do this we add a second random effect to model (2). Let us 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, now, i is the i-th unit in the sample, i = 1,…, n, that lives in the area j with $j_h = 1, \dots, J_h$, in the country h, h = 1,…, H. This means that the we can now merge together all the single datasets and estimate a unique regression. The model (2) become:

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

where the vectors of the random effects are $\boldsymbol{\gamma_{j_h}} = \left(\gamma^{\mu}_{j_h}, \gamma^{\sigma}_{j_h}, \gamma^{v}_{j_h}, \gamma^{\tau}_{j_h}\right)^T$ *and* $\boldsymbol{\lambda_h} = \left(\lambda^{\mu}_{h}, \lambda^{\sigma}_{h}, \lambda^{v}_{h}, \lambda^{\tau}_{h}\right)^T$ 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. components are independent, with a matrix such that $\boldsymbol{\Psi_\gamma} = diag(\sigma^2_{\gamma^\mu}, \sigma^2_{\gamma^\sigma}, \sigma^2_{\gamma^v}, \sigma^2_{\gamma^\tau})$ and $\boldsymbol{\Psi_\lambda} = diag(\sigma^2_{\lambda^\mu}, \sigma^2_{\lambda^\sigma}, \sigma^2_{\lambda^v}, \sigma^2_{\lambda^\tau})$. 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 (3) can be estimate by using a slightly modified version of the Fisher scoring algorithm the so called Cole and Green algorithm (Rigby and Stasinopoulos, 2005).

Subsequent to the estimation of model (3), 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 unobserved.

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

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

where $\pi_{(i)}$ and $y_{(i)}$ are, respectively, 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 $T_0 = 0$. It's important to note two key 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 performances and robustness check

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, here as in the application, is the per-capita expenditure expressed in Purchasing Power Parity (PPP) and that was divided by a region specific spatial price index. Note that here we have 7 countries the 8 that has no 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 for the dependent variable (household expenditure). Both 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 both the 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 to the same indices calculated directly from the 2021 survey expenditure data. The chosen indicators, as said, 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 2 compares the cross-country mean of each inequality index computed from observed and imputed expenditure. Point estimates are very close across the two sources and standard errors are essentially identical, indicating stable precision. For instance, the Gini is 0.324 with survey data and 0.325 with imputed expenditure; the P90/P10 ratio differs by only 0.141. Entropy-based measures are slightly higher under imputation (e.g., GE(1): 0.174 vs 0.183), whereas GE(2) is marginally lower (0.235 vs 0.232).

To assess whether these small level differences are statistically meaningful, we test equality of the two estimates using paired (resampling-based) t-tests and report the test statistic and two-sided p-value in the last columns. At conventional levels, none of the indices shows a statistically significant difference at the 5% level, with the sole borderline case being P80/P20 ($p = 0.050$). Differences for Gini ($p = 0.197$), GE(0) ($p = 0.328$), GE(1) ($p = 0.085$), GE(2) ($p = 0.384$), P90/P10 ($p = 0.060$) and S80/S20 ($p = 0.096$) are not statistically significant.

Table 3 summarizes the mean relative bias and mean absolute bias at the NUTS 2 level, again computed across all countries. The relative bias is generally below 0.12 for all indices, and the mean absolute bias remains modest even for the more sensitive percentile and share ratios.

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 both central tendency and dispersion. These visual impressions are corroborated by formal Kolmogorov Smirnov tests run country by country, with permutation based p-values equal to 1.00, implying failure to reject equality and, for practical purposes, indistinguishable distributions.

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

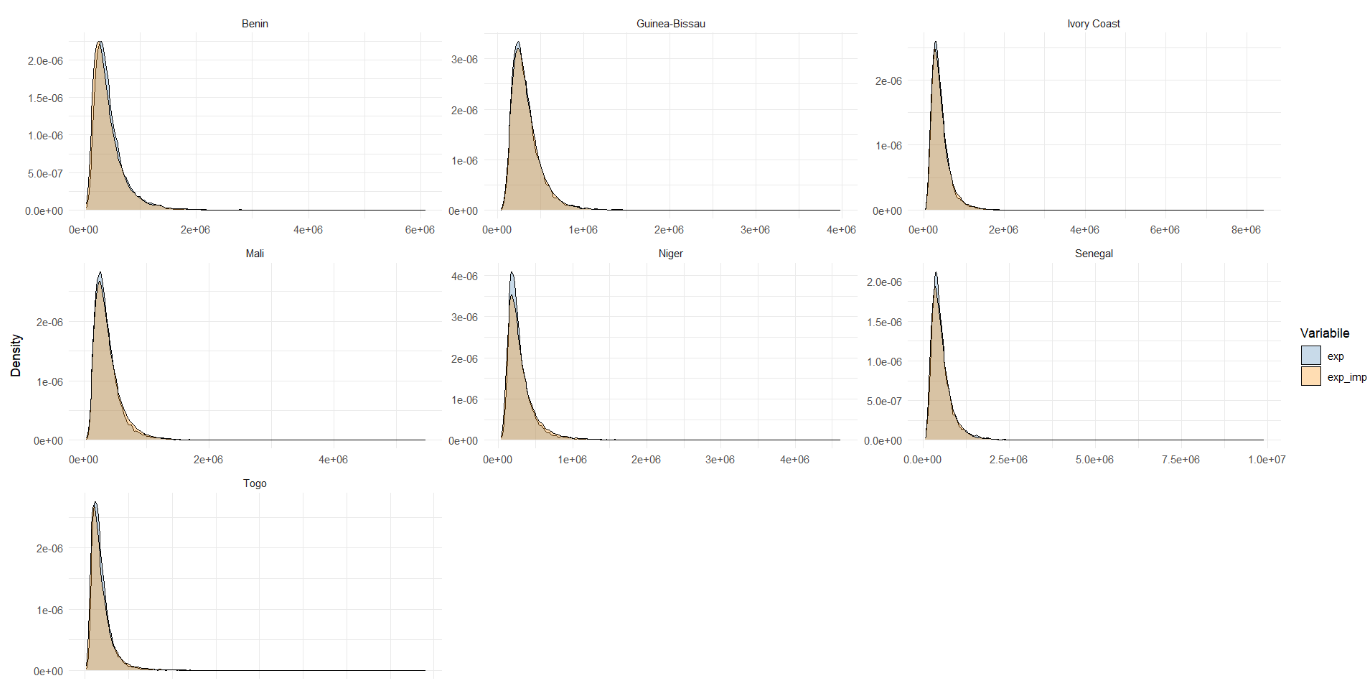
Table 2. Estimated indices on the basis of real data and imputed expenditure.

	Expenditure	SE Expenditure	Imputed Expenditure	SE Imputed Expenditure	Test statistics	P-value
Gini	0.324	0.001	0.325	0.001	-14.993	0.197
GE(0)	0.161	0.001	0.174	0.001	-15.61	0.328
GE(1)	0.174	0.001	0.183	0.001	-7.33	0.085
GE(2)	0.235	0.003	0.232	0.002	0.87	0.384
P80/P20	2.464	0.005	2.494	0.005	-4.59	0.050
P90/P10	3.967	0.011	4.108	0.012	-9.38	0.060
S80/S20	10.969	0.022	10.670	0.019	10.68	0.096

Table 3. Mean Relative Bias and Mean Absolute Bias computed over all countries at NUTS 2 level.

	Relative Bias	Absolute Bias
Gini	0.068	0.024
GE(0)	0.115	0.023
GE(1)	0.100	0.025
GE(2)	0.120	0.045
P80/P20.80%	0.065	0.181
P90/P10.90%	0.104	0.408
S80/S20	0.075	0.798

Figure 1. Expenditure Vs Imputed Expenditure distribution



4.1 Model robustness

In the measurement of economic wellbeing, inequality estimates are particularly sensitive to the presence of extreme values, much more so 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.

From this reference population, we drew, mimicking the original sample scheme, 200 bootstrap samples, each representing a possible realization of the sampling process, with a sample size equal to 1% of the population (around 4,000 units). In each sample, we introduced artificial outliers to simulate three levels of contamination: (i) low contamination: 5 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 1,000.

For each contaminated sample, the model was re-estimated using the same 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 on the original, unperturbed data.

Table 4 (renumbered from original robustness table) shows that the presence of extreme values consistently leads to a mild underestimation of inequality across all metrics, but the magnitude of the bias remains limited, even under severe contamination. For instance, in the low-contamination scenario, the Gini coefficient has an absolute bias of −0.224 points (−1.012% relative bias), while GE(1) and GE(2) show relative biases of −1.9% and −1.8%, respectively. Even the more sensitive S80/S20 index changes by only −1.24%.

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

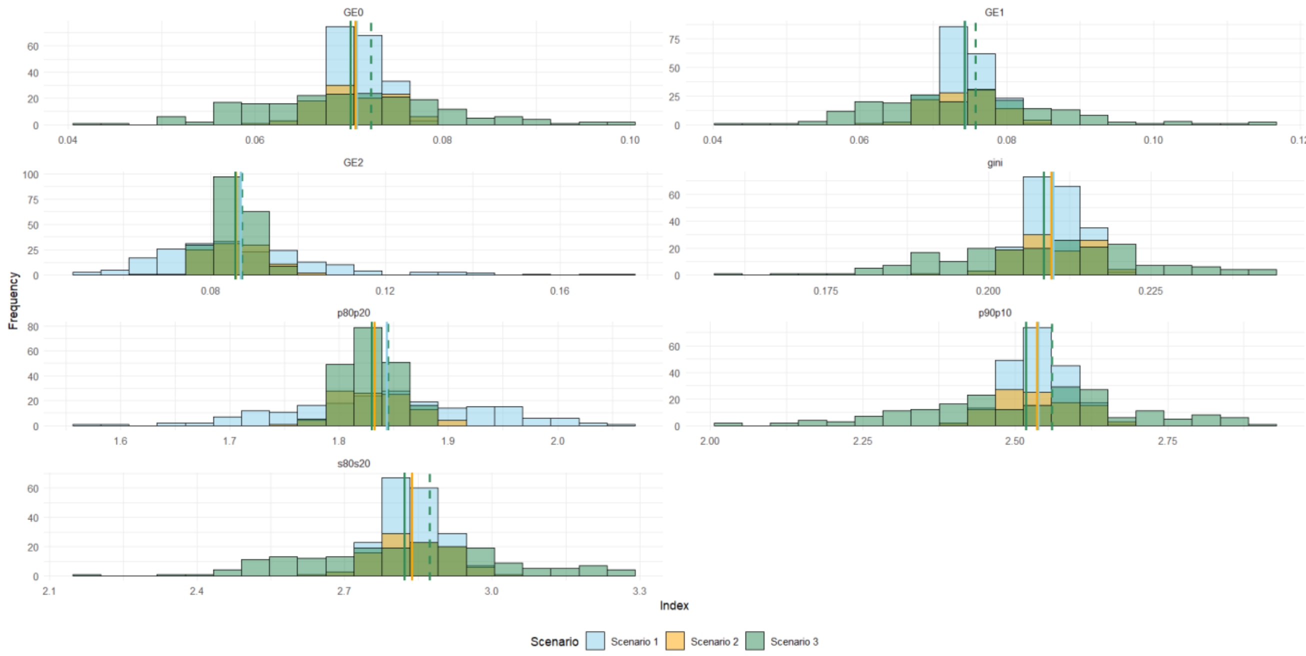
Figure 2 displays the distributions of the bootstrap-derived inequality indices across the three contamination scenarios. The vertical dashed lines mark the corresponding "true" coefficients from the uncontaminated model. In all scenarios, the distributions remain narrow and centered close to the true values, indicating high stability of the estimates even when the data are heavily perturbed. Although greater contamination produces a slight widening of the distributions, especially for P90/P10, P80/P20, and S80/S20, the displacement from the true values is minimal. These patterns confirm the numerical findings in Table 4, showing that the GAMLSS model 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 model, with both location and scale components and random effects at regional and national levels, is highly robust to outliers. Although extreme values introduce a systematic downward bias, the effect is never substantial, even in unrealistic contamination scenarios. This reinforces the model's reliability for empirical applications in settings where anomalous observations may occur.

Table 4. 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 lines is the true coefficients



.

5. The inequality in the Sahel Area

To implement our methodology and analyze inequality trends in the Sahel, we apply the SSIT-GAMLSS framework using harmonized household and labor force survey data for eight WAEMU countries. This application aims to generate annual estimates of household consumption, which are 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 incorporates both national and sub-national variation through a dual random-effects structure. This allows 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. In what follows, we present the variables included in the model and discuss how they contribute to predicting household consumption across time and space. The model we estimate is based on the 2018 WAEMU wave, in which data for Ivory Coast and Guinea-Bissau were incorporated by taking their 2021 figures and adjusting them to 2018 prices. Note that the regions are always known for these countries in this year (see Table 1), and it is not a problem if in some years they are unknown when predicting expenditure. The random-effect value for those units without a region year will be equal to 0. Before to move on, it is important to emphasize that, for each year and country, the index was computed based on 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 measured based on consumption, the Gini index tends to be lower by about 20 percent than the one estimated in income (Clementi et al., 2021, 2023).

The covariates available for this analysis are: (i) sex, coded as 1 for male and 0 for female; (ii) age class, a variable that categorizes respondents into six age groups; (iii) Hhsize, household size, divided into seven categories; (iv) Rural/Urban, a variable coded as 1 if the household resides in an urban area; (v) waterpipe, indicating whether the household has running water; (vi) bathroom, indicating whether the household has a bathroom; (v) waterpipe, indicating whether the household has a waterpipe; (vi) everattd, indicating whether the respondent has never attended school and, (vi) literacy, indicating whether the respondent know write and read. Aditionally, we need to incorporate two random effects: one based on the country and the other based on the region.

Based on these covariates and leveraging the ability of the GAMLSS framework to select the distribution that best fits the data from a wide range of candidates, we follow the procedure outlined by Mori and Ferrante (2025) to determine the most appropriate distribution for the data.

4.1 The consumption distribution in Sahel

Figure 3 suggests that the distribution to be fitted must be skewed, with a right log tail. Following Mori and Ferrante (2025), we begin by analyzing the Akaike Information Criterion (AIC) for a range of potential candidate distributions. It is important to note that we examine the distribution of per-capita expenditure without adjusting for PPP, as we prefer to model the data in its original monetary form and subsequently convert each predicted value into PPP based on the year and country. Table5 reports the AIC for a number of different distributions (for the parameterization of the distributions see Stasinopoulus et al. 2019). Results clearly show how the Log-Normal distribution and the Generalized Beta of the Second Kind (GB2) are the two with the lowest AIC. Between them the Log-Normal distribution is the one that best performs. This result is in line with Barigozzi and Speciale (2011).

Table 5. Akaike Information Criterion for selected distributions

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

Once we select the distribution that best fits data we estimate our model based on the Log-Normal distribution and using both covariates and random-effects on both the parameters. Table 6 reports the summary of 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 are statistically different from 0 for both parameters and the pseudo-R^2 is 0.51. The residual analysis confirms that the model is accurate.

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

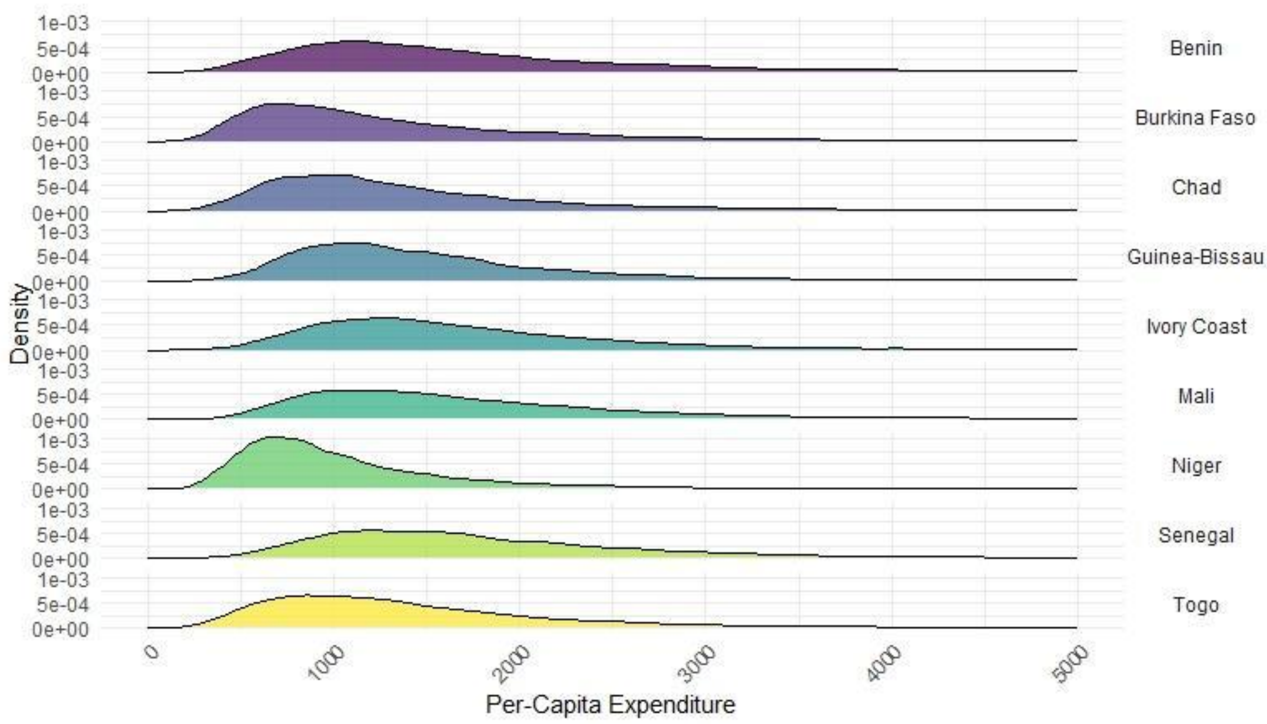


Table 6: Regression summary

4.2 Inequality: National level

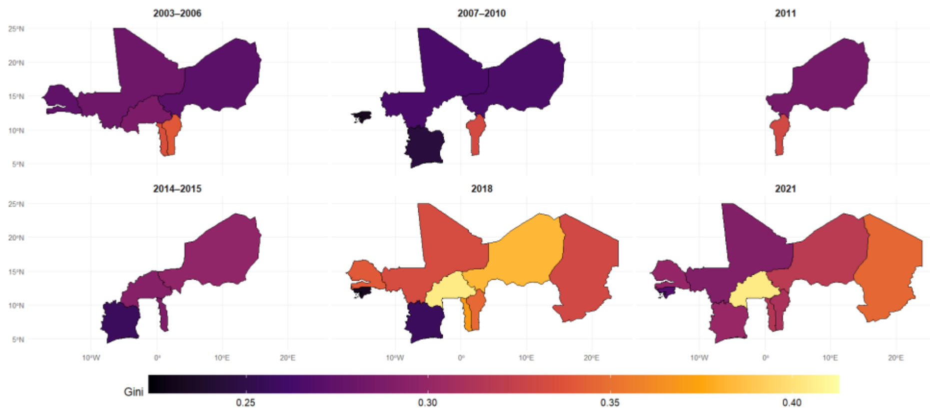
***μ* Coefficients**						***σ* Coefficients**					
	Estimate	**Std. Error**	**t value**	**Pr(>|t|)**			**Estimate**	**Std. Error**	**t value**	**Pr(>|t|)**	
(Intercept)	13.46923	0.0162	831.434	0.00000	***	(Intercept)	-0.71647	0.0349	-20.5292	0.00000	***
Hhsize2	-0.33615	0.0179	-18.7793	0.00000	***	Hhsize2	-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	0.132039	0.0051	25.89	0.00000	***	Urban	0.084082	0.006	14.01367	0.00000	***
Waterpipe	0.137144	0.002	68.572	0.00000	***	Waterpipe	0.035765	0.0077	4.644805	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	0.044834	0.0048	9.340417	0.00000	***	Age_class2	-0.00742	0.0054	-1.37407	0.17249	
Age_class3	0.057677	0.0086	6.706628	0.00000	***	Age_class3	-0.00593	0.0034	-1.74412	0.08421	*
Age_class4	0.080717	0.0078	10.34833	0.00000	***	Age_class4	0.007978	0.0113	0.706018	0.48182	
Age_class5	0.079774	0.0077	10.36026	0.00000	***	Age_class5	0.005825	0.0101	0.576733	0.56542	
Age_class6	0.068437	0.024	2.851542	0.00529	***	Age_class6	0.03113	0.0619	0.502908	0.61613	
Everattd	0.05969	0.0015	39.79333	0.00000	***	Everattd	-0.01506	0.0143	-1.05315	0.29481	
Literacy	0.123617	0.0073	16.93384	0.00000	***	Literacy	0.0361	0.0117	3.08547	0.00263	***
Random-effect variances											
$\sigma^2_{\gamma^\mu}$	0.140179	0.016268	7.84022	0.00000	***	$\sigma^2_{\gamma^\sigma}$	0.083582	0.00698	11.96430	0.00000	***
$\sigma^2_{\lambda^\mu}$	0.127547	0.019650	7.13371	0.00000	***	$\sigma^2_{\lambda^\sigma}$	0.079902	0.01075	7.432959	0.00000	***
R2: 0.5121											
Residuals analysis: mean = 0.0002845879 variance = 1.000002 coef. of skewness = 0.1378414 coef. of kurtosis = 3.298156											

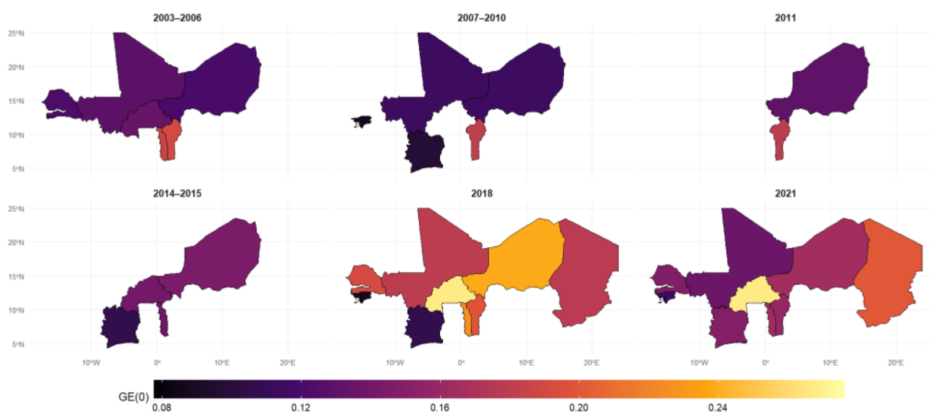
At the national level (see Figure 4 and 5 for the estimates and Figure B1 and B2 for the coefficients of variation), the inequality measures estimated for the Sahel countries between 2003 and 2021 reveal both structural patterns and marked shifts linked to political and economic turbulence. In the following, we highlight that for each country we focus mainly on one index, but similar conclusions can also be drawn from the other indices, as the interested reader can see in Figures 4 and 5. In addition, Figure 6 reports, as an illustrative example, the national Gini index based on the same underlying data used in the maps. We included this figure because in this sub-section, and in the following one, we frequently refer to events such as coups d'état, conflicts, or economic crises, and it provides a visual reference to identify the years affected by these episodes and their potential repercussions on inequality dynamics.

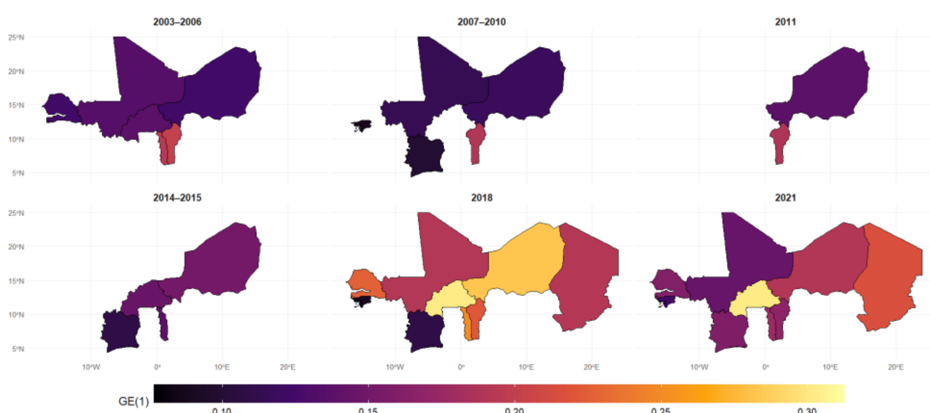
The Gini index shows values ranging from 0.20 to 0.45, with coefficients of variation consistently below the 16.6% threshold recommended by Statistics Canada, which is widely used as a benchmark for acceptable reliability. Other measures, such as GE(0), which is more sensitive to changes at the lower 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 from 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 between 2005 and 2014 around 2.25, then rose steadily to 2.69 in 2018. Burkina Faso, represented in 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 appears in 2006, 2015, 2018 and 2021, has maintained comparatively stable national inequality levels, with the Gini index fluctuating only modestly 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, with the GE(2) index, which heavily weights the highest incomes, averaging 0.263 over the period with a peak in 0.33 in 2018. This reflects structural disparities between the economically dominant coastal areas and the less developed northern regions. Ivory Coast, included for 2018, 2015, 2018 and 2021 following political stabilisation under President Alassane Ouattara, displays relatively modest changes in national inequality, with the P90/P10 ratio fluctuating within a narrow range around 3.35 with the lowest peak in 2008 of 3.07. 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 the context of political fragility. Senegal, which is represented in 2005, 2018 and 2021, is widely considered one of the most politically stable countries in the region and has experienced steady GDP growth 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 have been grouped and plotted together.


2003–2006
2007–2010
2011
2014–2015
2018
2021
Gini
0.25
0.30
0.35
0.40


2003–2006
2007–2010
2011
2014–2015
2018
2021
GE(0)
0.08
0.12
0.16
0.20
0.24


2003–2006
2007–2010
2011
2014–2015
2018
2021
GE(1)
0.10
0.15
0.20
0.25
0.30

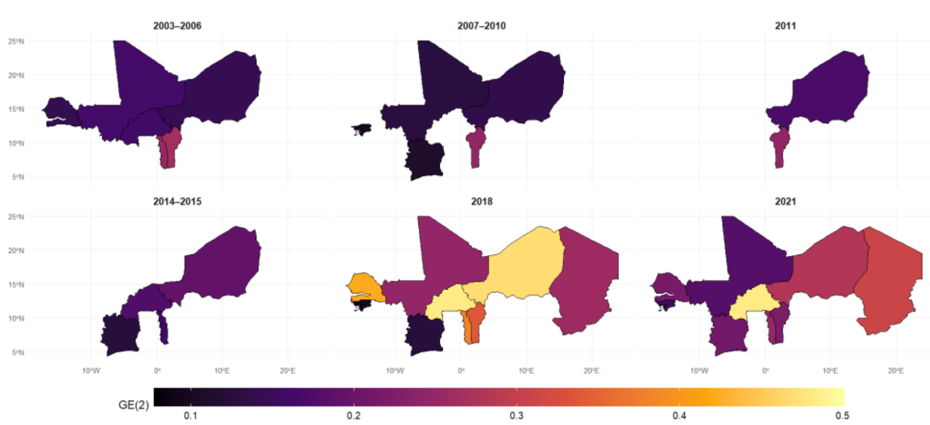

2003–2006
2007–2010
2011
2014–2015
2018
2021
GE(2)
0.1
0.2
0.3
0.4
0.5

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 have been grouped and plotted together.

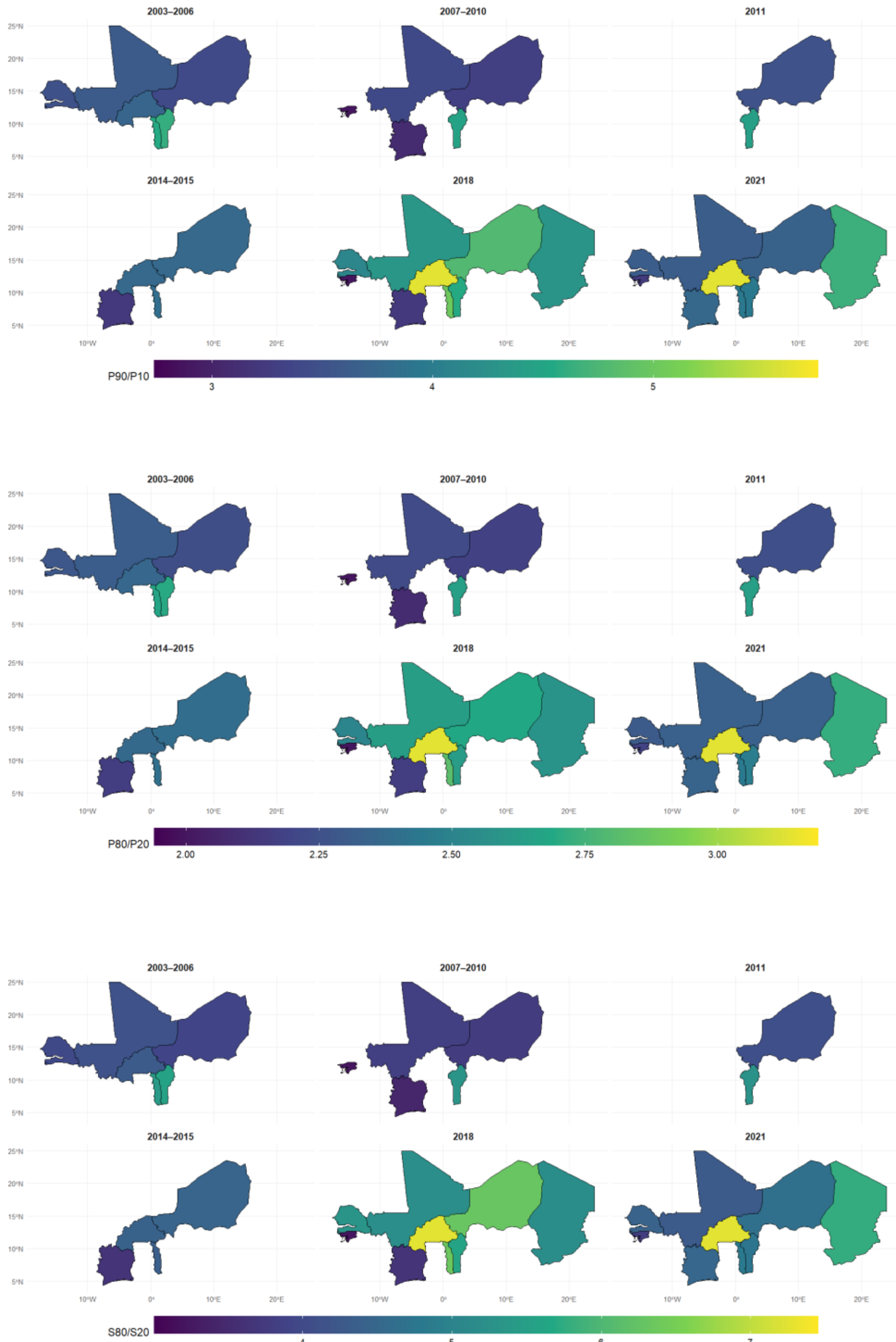
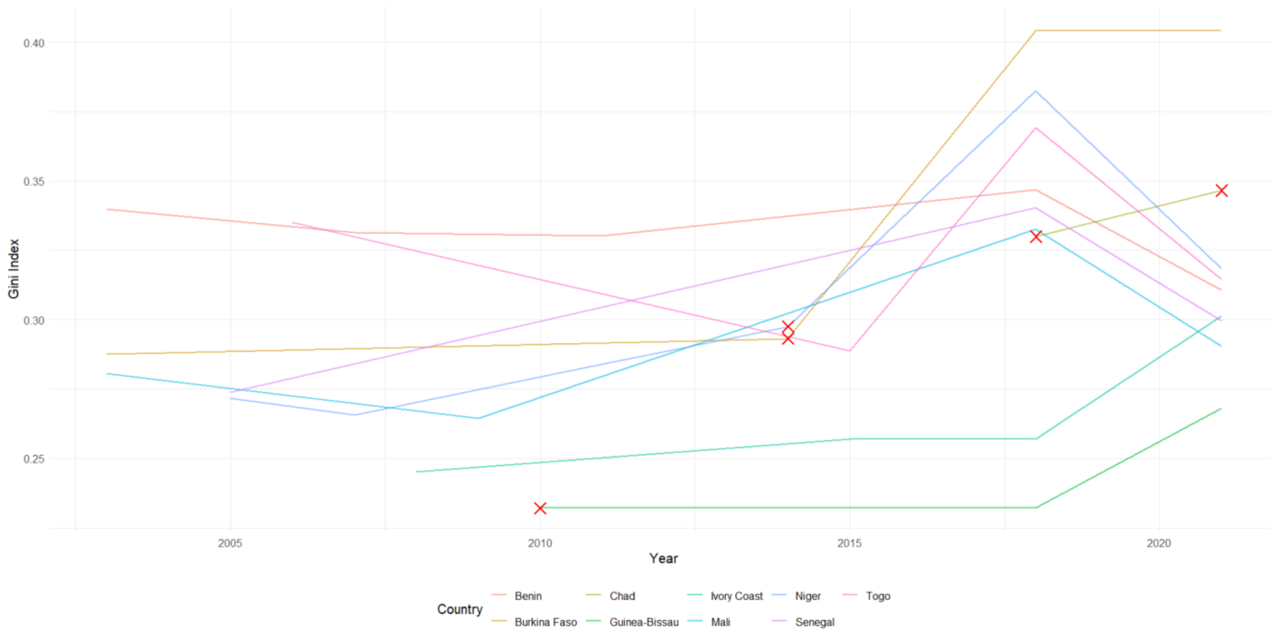


Figure 6. National Gini index trends for Sahel countries (2003–2021). Each colored line represents one country. Red crosses mark years of political or economic instability (e.g., coups, conflicts, financial crises).



Before to move on, it is important to emphasize that our inequality and poverty estimates are derived from a measure of household expenditure that explicitly accounts for regional variation in prices within 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 better 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 remains both necessary and informative. In particular, the World Bank's Poverty and Inequality Platform (PIP) and the UNU-WIDER World Income Inequality Database (WIID) 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. It is also important to note, however, that PIP and WIID figures often diverge substantially from one another, reflecting differences in survey harmonization, welfare aggregates such as treatment of price adjustments. 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 decrease form 0.34 in 2003 to 0.31 in 2021 showing a little improvement in the 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, declines from 0.38 to 0.34 and in WIID from 0. 53 to 0.50. Similarly, in Mali our estimates indicate a quite 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 the converging trends, adds credibility to our approach and underlines the value of producing country-specific and spatially sensitive measures of welfare distribution

4.3 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 the national level in Figure 6, which highlights the years marked by coups, conflicts, or economic crises, such events are also relevant in interpreting regional inequality dynamics. Additional results are provided in Appendix B: Figure B3 shows choropleth maps of inequality indices at the regional level (GE(0), GE(1), GE(2)); Figure B4 presents choropleth maps of ratio indices at the 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 the regional level are naturally higher than at the national level, approaching 10% percent in sparsely populated areas, but they remain within the limits considered acceptable for analytical purposes.

Let us first consider the countries that experienced riots, wars, or coups between 2003 and 2021. In Mali, for example, a similar trend to the one observed at National level, is observed also at the regional level, where the Bamako region shows the highest level of inequality, while Taoudénit experiences the least. 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, both Mali and Niger 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 increased 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-Ouest region in 2014 to 0.37 in the Centre region in 2018. Similar to 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 exhibited higher levels of inequality in 2018 compared to 2003.

For Chad, data is available only for the years 2018 and 2021, both of which correspond to the period when President Idriss Déby Itno was still in power. During this period, inequality increased slightly. The lowest Gini index value of 0.26 was recorded in 2021 in the Sila region, while the highest, 0.36, was observed in the Logone Orientale region in 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. Despite having undergone a coup d'état, Togo’s inequality remained relatively stable throughout the years considered. The only region to experience a notable change in inequality is the Grand Lom area, where the Gini index increased from 0.30 in 2006 to 0.34 in 2021, following an initial decrease in 2015.

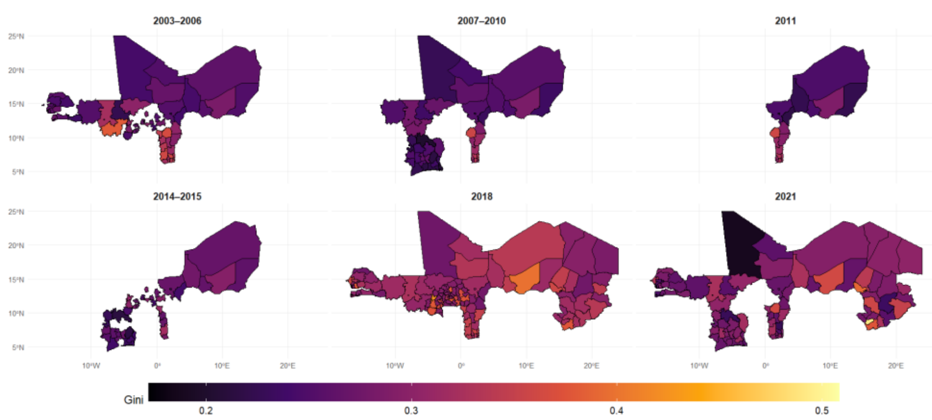
Among the more stable countries, Benin exhibits a higher level of inequality compared to Mali, with the Atacora region reaching a Gini index of 0.37 in 2018. Generally, inequality remains high across all regions, with the Collines region recording the lowest value of 0.25. Inequality has remained relatively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

In Haut-Sassandra, the most populous region of Ivory Coast, the Gini is quite stable between 2008 and 2015. By 2021, the HCR rose slightly to 0.27, signaling 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.

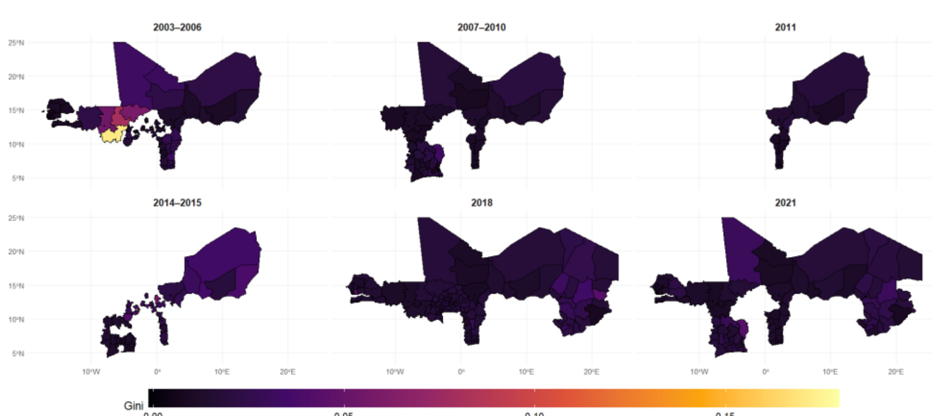
Guinea-Bissau is one of the countries in the region with the lowest levels of inequality, with the only region, Donga, approaching a Gini index of 0.28 in 2018. This is indicative of the country's increased stability after 2010.

Senegal is considered one of the most stable countries in the region, with 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 inequality of Gini index (on the left) and its coefficient of variation (on the right) at regional level from 2003 to 2021 in the Sahel region. Years without overlapping countries have been grouped and plotted together.



Gini index



Coefficient of Variation

6. Conclusion

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

The use of GAMLSS, with its flexibility to incorporate distributional assumptions beyond the exponential family and the integration of both regional and national random effects, has proven particularly effective in capturing the heterogeneity of consumption distributions across space and time. Furthermore, the introduction of time-varying parameters via a weighted approach allowed us to model smoother transitions in welfare indicators between survey years, improving the accuracy of intertemporal estimates.

Empirically, our findings highlight 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 some countries such as Senegal and Benin exhibit relatively stable inequality levels over time, others—particularly Mali, Niger, and Burkina Faso—show a marked increase in inequality, often linked to episodes of violence, displacement, or economic disruption. The granular, region-level estimates further reveal 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 notable 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 twenty years, this paper contributes to building a more robust evidence base for monitoring inclusive growth in the Sahel.

Future research should explore the extension of this methodology to other welfare dimensions, such as multidimensional poverty or vulnerability to shocks. Moreover, improved data availability, especially through more frequent and harmonized household surveys, would significantly enhance the reliability and scope of this type of analysis. In the meantime, the SSIT-GAMLSS framework remains a valuable tool for producing timely and policy-relevant estimates in data-scarce environments.

References

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. H., 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. H., & 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., Kilic, T., Abanokova, K., & Carletto, C. (2024). Imputing Poverty Indicators without Consumption Data: An Exploratory Analysis. IZA-Institute of Labor Economics.

Dang, H. A., Kilic, T., Hlasny, V., Abanokova, K., & Carletto, C. (2025). *Using Survey-to-Survey Imputation to Fill Poverty Data Gaps at a Low Cost: Evidence from a Randomized Survey Experiment* (No. 1392). GLO Discussion Paper.

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

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 labor force surveys: a primer. *The 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. *The World Bank Economic Review*, ***18***, 401-421.

Fosu, A. K. (2009). Inequality and the impact of growth on poverty: comparative evidence for sub-Saharan Africa. *The 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.

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. *The 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. *The review of economics and statistics*, ***91***, 773-792.

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'Ivore	2002	Enquête Niveau de Vie des Ménages 2002	N	10799
Cote D'Ivore	2008	Enquête Niveau de Vie des Ménages 2008	N	12600
Cote D'Ivore	2015	Enquête Niveau de Vie des Ménages 2015	N	12899
Cote D'Ivore	2018	Enquête Harmonisee sur les Conditions de Vie des Menages	N	12992
Cote D'Ivore	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

Country	Year	Survey	Consumption (Y = yes, N = no)	Sample size HH
Niger	1995	National Survey on Household Living Conditions and Agriculture 1995	N	4383
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
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
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
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
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
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 region. Years without overlapping countries have been grouped and plotted together.

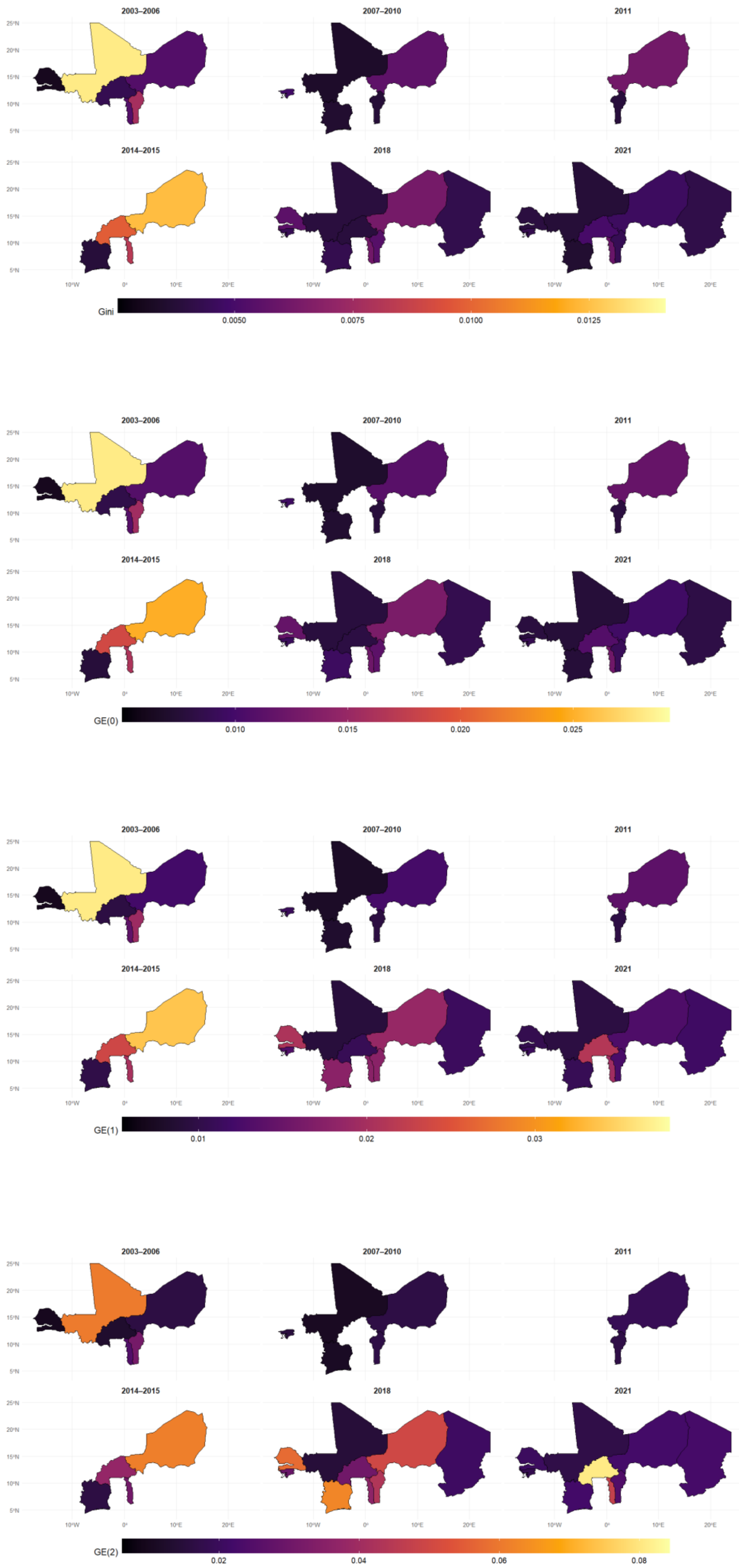


Figure B2: Choropleth map of coefficient of variation of ratio indices (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel region. Years without overlapping countries have been grouped and plotted together.

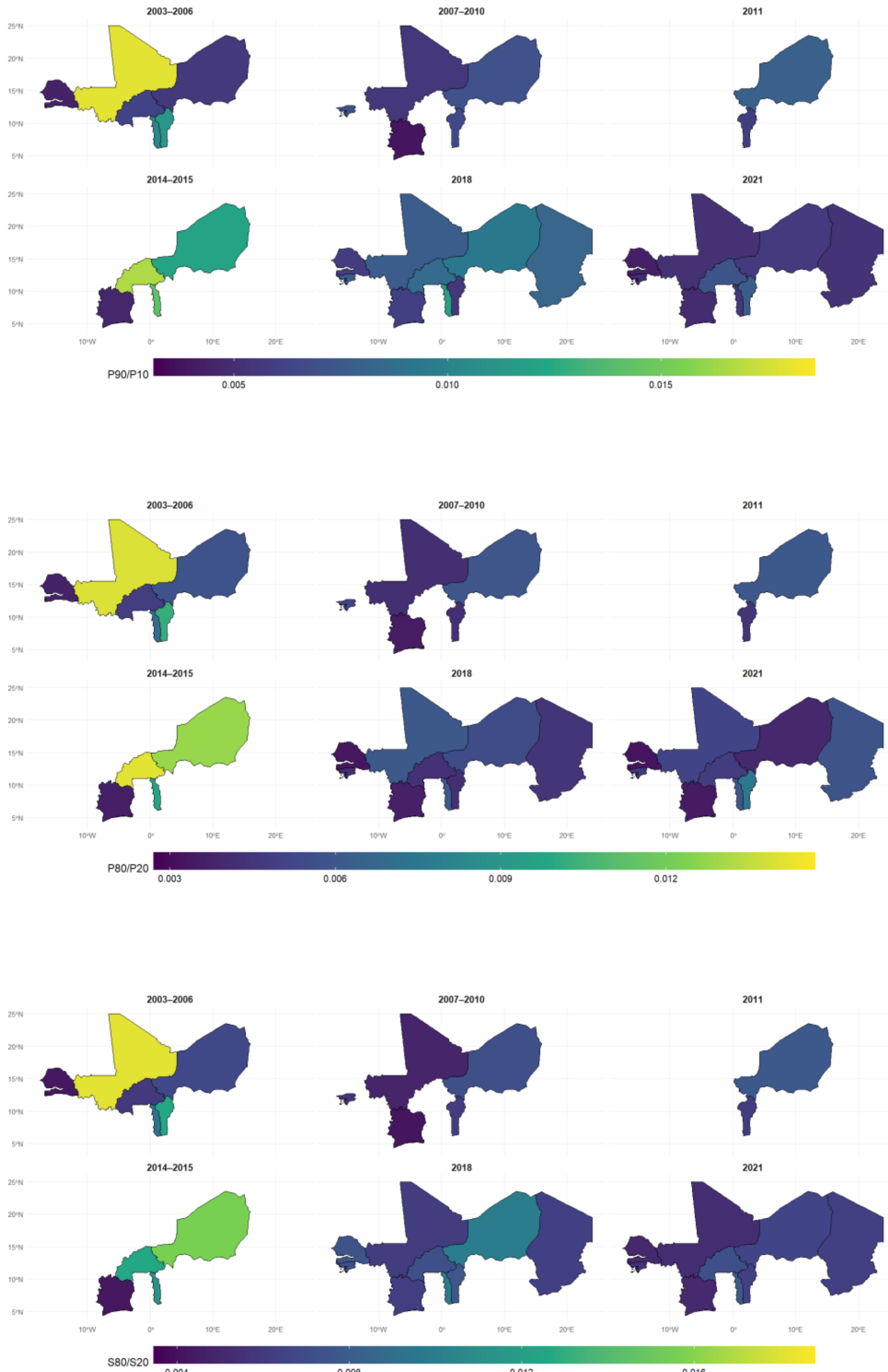


Figure B3: Choropleth map of inequality indices at regional level (GE(0), GE(1), GE(2)) from 2003 to 2021 in the Sahel region. Years without overlapping countries have been grouped and plotted together.

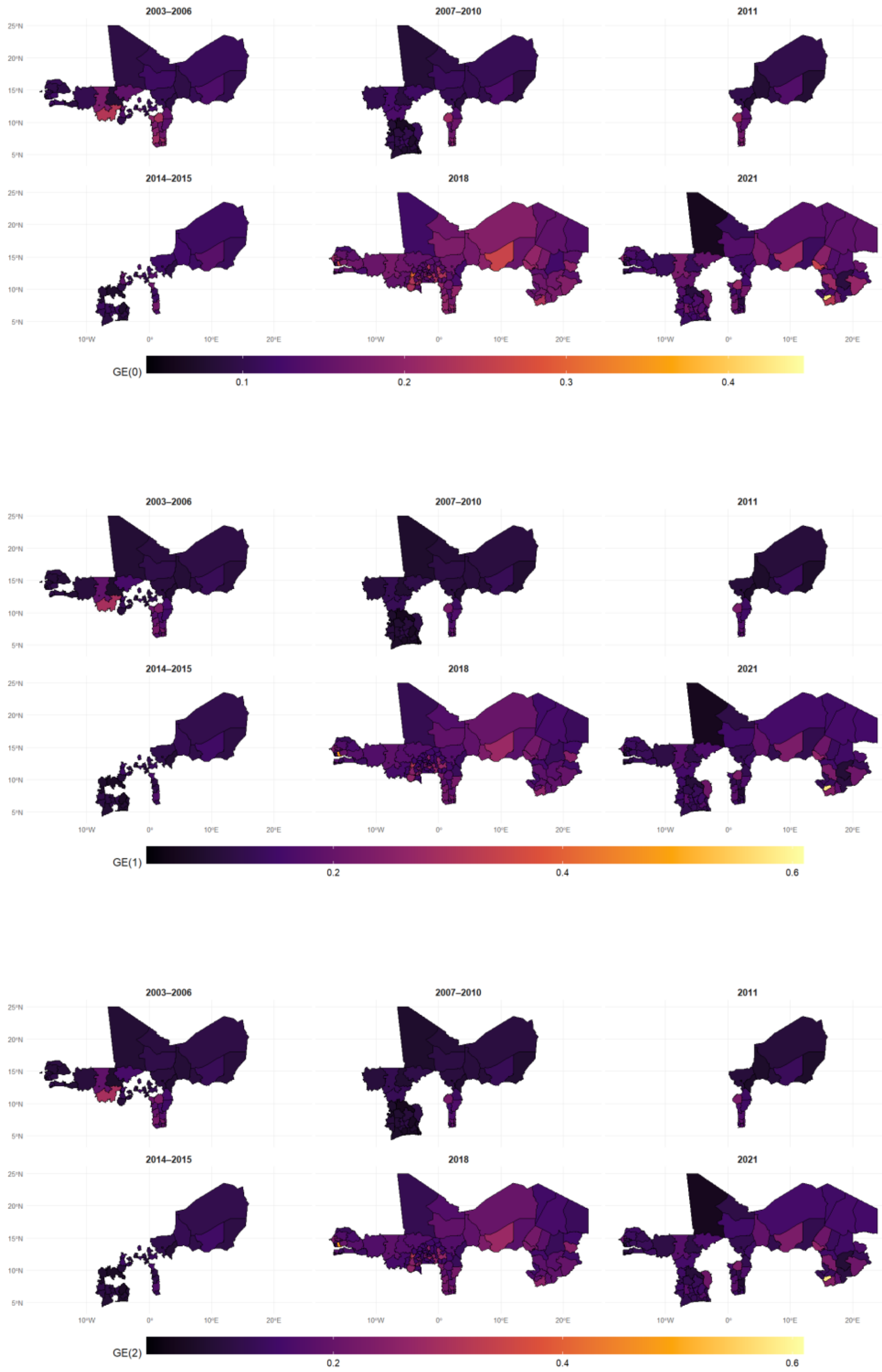


Figure B4: Choropleth map of ratio indices at regional level (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel region. Years without overlapping countries have been grouped and plotted together.

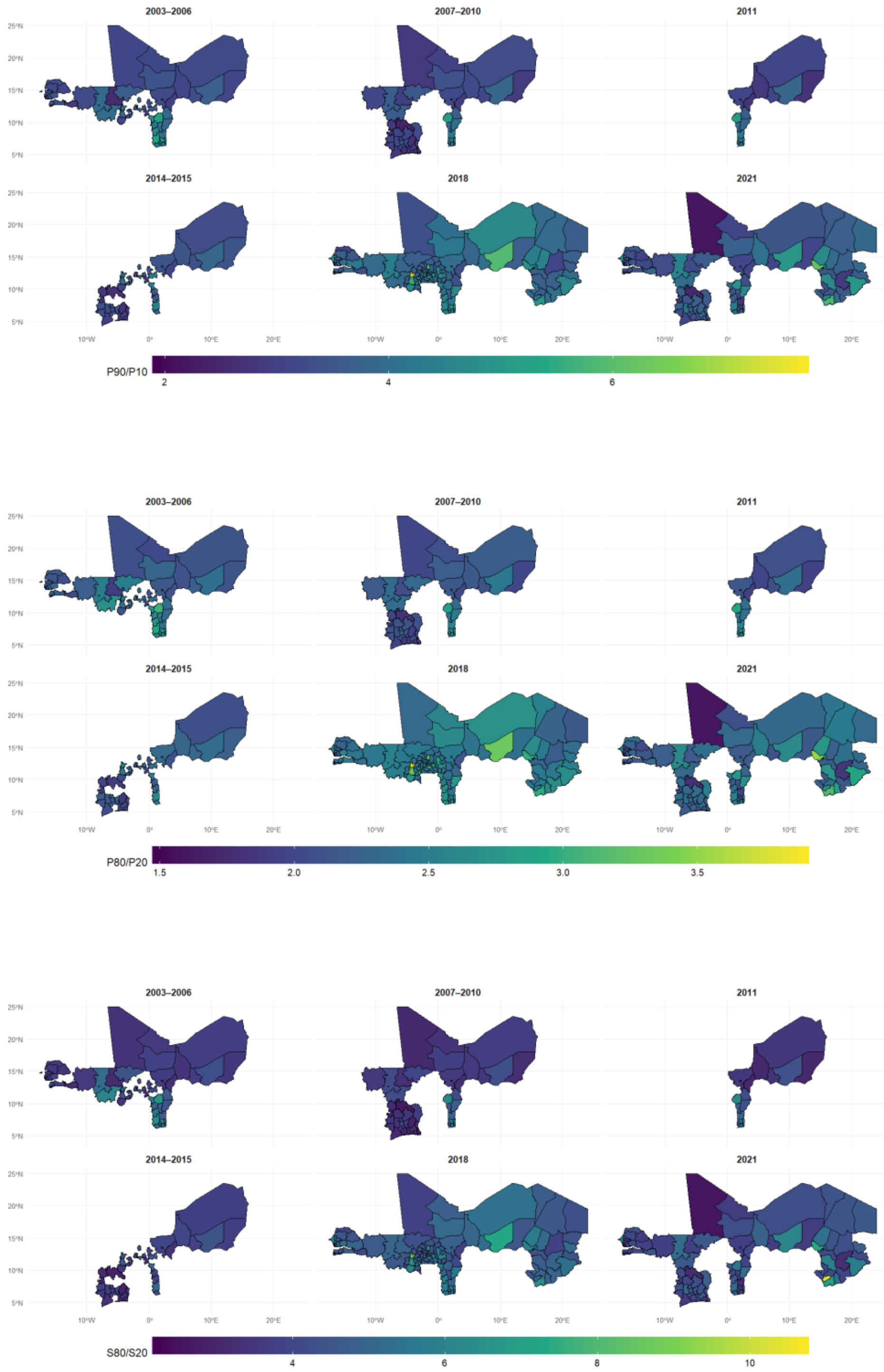


Figure B5: Choropleth map of coefficient of variation of inequality indices at regional level (GE(0), GE(1), GE(2)) from 2003 to 2021 in the Sahel region. Years without overlapping countries have been grouped and plotted together.

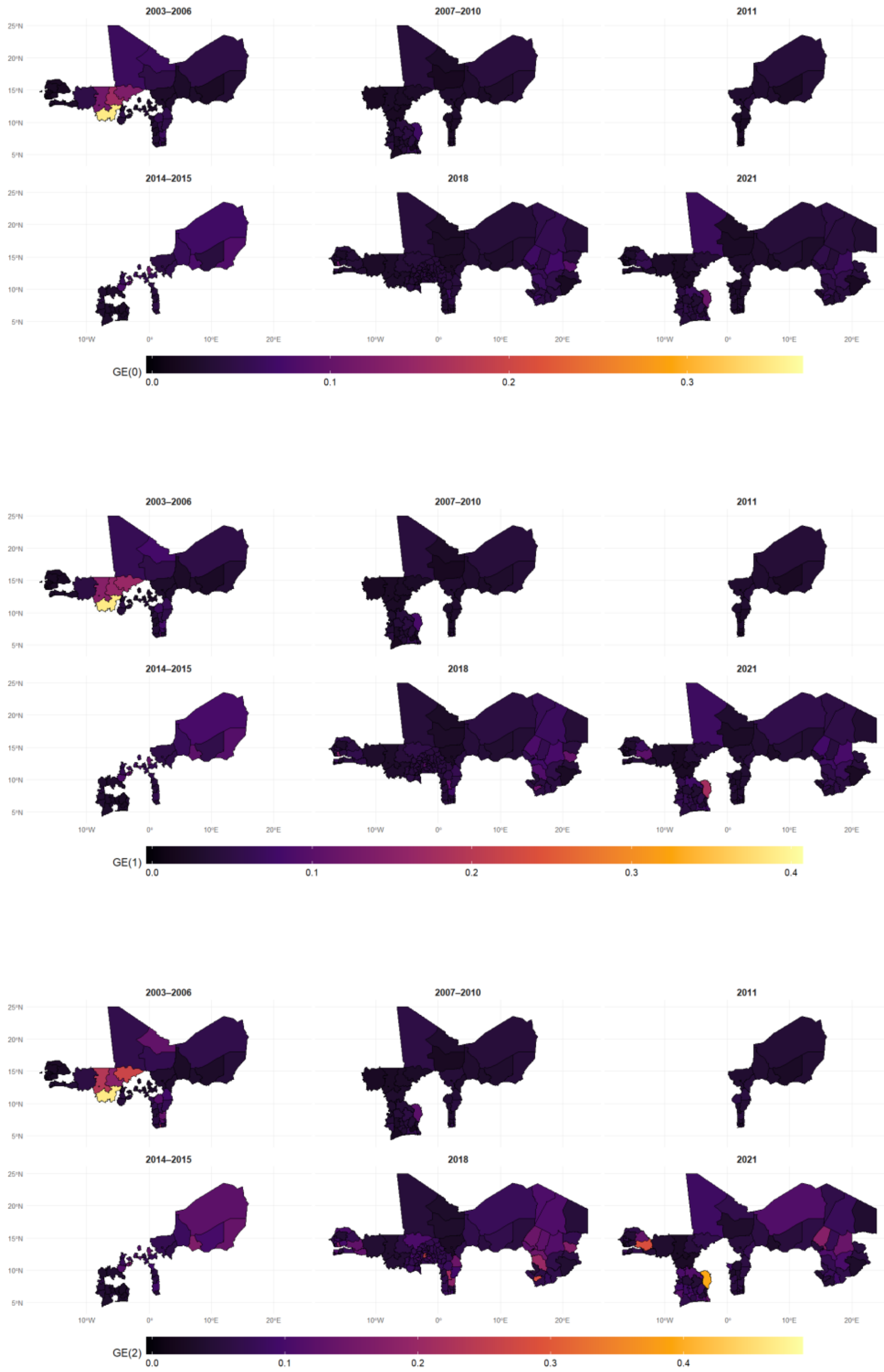


Figure B6: Choropleth map of coefficient of variation of ratio indices at regional level (P90/P10, P80/P20, S80/S20) from 2003 to 2021 in the Sahel region. Years without overlapping countries have been grouped and plotted together.