Revue Africaine des Sciences Sociales et de la Santé Publique, Volume 8 (2)
ISSN: 1987-071X e-ISSN 1987-1023
Received, 8 April 2026
Accepted, 20 July 2026
Published, 1 August 2026
https://www.revue-rasp.org
To cite: Mare, R., & Mendes, O. (2026). Modélisation bayésienne dynamique des inégalités de consommation des ménages en Guinée-Bissau: une approche hybride appliquée aux données EHCVM 2018-2021. *Revue Africaine des Sciences Sociales et de la Santé Publique, 8(2), pp.289-299.* https://doi.org/10.4314/rasp.v8i2.20
Research

Modélisation bayésienne dynamique des inégalités de consommation des ménages en Guinée-Bissau: une approche hybride appliquée aux données EHCVM 2018-2021

Dynamic Bayesian modeling of household consumption inequalities in Guinea-Bissau: a hybrid approach applied to 2018–2021 EHCVM data

R. MARE[1], O. Mendes[2]

[1] Opérations, études et recherche, Tchiemba Consulting Group, Ouagadougou, Burkina Faso [2] Instituto Nacional de Estatística da Guiné-Bissau (INE), Bissau, Guiné-Bissau
*** Correspondance** : email:osvaldo.cristo.mendes@gmail.com ; Tél : +245 96 676 41 25

Résumé

Les inégalités de consommation dans les économies à faible revenu constituent un défi statistique majeur en raison de la rareté des données, de l'hétérogénéité spatiale et de la volatilité temporelle. Cet article propose une approche hybride d'inférence bayésienne pour analyser l'évolution de ces inégalités à partir d'enquêtes transversales répétées. Face aux problèmes de convergence des chaînes de Markov rencontrés lors de l'estimation d'un modèle hiérarchique dynamique complet sous Stan, une méthode en deux étapes est développée : elle combine l'estimation bootstrap fréquentiste des indicateurs d'inégalité avec une propagation probabiliste de l'incertitude. Appliquée aux données EHCVM 2018-2021 en Guinée-Bissau (10 702 ménages, deux vagues), cette approche révèle une réduction robuste des inégalités. Le coefficient de Gini passe de 0,3455 (IC 95 % : [0,3345 ; 0,3558]) à 0,3257 (IC 95 % : [0,3167 ; 0,3350]), avec une probabilité a posteriori de 99,7 % que la diminution soit réelle. Les indices d'Atkinson et de Theil convergent vers la même conclusion, et la dominance de Lorenz de second ordre de 2021 sur 2018 est systématique. L'étude documente les verrous computationnels des modèles bayésiens hiérarchiques en contexte de données limitées et valide la pertinence d'une inférence probabiliste transparente pour le suivi des politiques dans les économies en développement.

Mots-clés: inégalités de consommation; inférence bayésienne ; modèle hiérarchique dynamique ; EHCVM ; Guinée-Bissau ; coefficient de Gini ; approche hybride ; bootstrap

Abstract

Inequality in consumption in low-income economies poses a major statistical challenge due to data scarcity, spatial heterogeneity, and temporal volatility. This paper proposes a hybrid Bayesian inference approach to analyse the evolution of consumption inequalities using repeated cross-sectional surveys. To address Markov chain convergence problems encountered when estimating a full dynamic hierarchical model in Stan, a two-stage method is developed

that combines frequentist bootstrap estimation of inequality indices with probabilistic uncertainty propagation. Applied to 2018-2021 EHCVM data from Guinea-Bissau (10,702 households, two waves), the approach reveals a robust reduction in inequalities: the Gini coefficient decreases from 0.3455 to 0.3257, with a posterior probability of 99.7% that the decline is genuine. Atkinson and Theil indices converge to the same conclusion, and the second-order Lorenz dominance of 2021 over 2018 is systematic. The study documents the computational barriers of hierarchical Bayesian models in data-limited settings and validates the relevance of transparent probabilistic inference for policy monitoring in developing economies.
Keywords: consumption inequality; Bayesian inference; dynamic hierarchical model; EHCVM; Guinea-Bissau; Gini coefficient; hybrid approach; bootstrap

1. Introduction

La mesure des inégalités de consommation dans les économies à faible revenu constitue un défi statistique exigeant où la rareté des données, l'hétérogénéité spatiale et la volatilité temporelle compromettent la fiabilité des méthodes d'inférence conventionnelles. Ces économies dépendent, pour leur pilotage, de séries d'enquêtes transversales de ménages répétées, telles que l'Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM). L'approche fréquentiste usuelle, qui traite chaque vague d'enquête de manière indépendante et recourt à des approximations asymptotiques, se révèle inadéquate : elle ne permet ni de modéliser la dépendance temporelle entre distributions successives, ni de produire des estimations stables pour les petits domaines géographiques (Deaton, 2019 ; Gelman et Hill, 2007).

Le paradigme bayésien offre un cadre conceptuel unifié pour surmonter ces limitations. En traitant les paramètres d'intérêt comme des variables aléatoires, il autorise une quantification complète et probabiliste de l'incertitude via les distributions a posteriori, avec des propriétés de validité en échantillon fini (Geweke, 2005). Pour l'analyse d'enquêtes transversales répétées, ce cadre permet de spécifier une évolution latente des paramètres via un processus dynamique, créant un lien inférentiel entre les périodes et stabilisant les comparaisons intertemporelles (West et Harrison, 1997). La combinaison d'une hiérarchie spatiale et d'une dynamique temporelle constitue une avancée méthodologique prometteuse pour l'étude des inégalités dans les contextes de développement (Stival et al., sous presse ; Wu et Stephenson, 2024).

Cependant, l'estimation opérationnelle de tels modèles dans des contextes de données limitées se heurte à des défis computationnels liés à l'identifiabilité partielle et à la sparsité des données dans des espaces paramétriques de haute dimension. La présente étude tire ses enseignements de la confrontation entre cette ambition méthodologique et les réalités pratiques de l'inférence. Nous documentons les difficultés de convergence rencontrées lors de l'estimation d'un modèle bayésien hiérarchique dynamique complet avec les données EHCVM 2018-2021 de Guinée-Bissau, et proposons une approche hybride pragmatique combinant la robustesse de l'estimation bootstrap fréquentiste avec la richesse interprétative de l'inférence probabiliste bayésienne.

L'article poursuit un triple objectif : (1) Sur le plan méthodologique, il propose une stratégie d'inférence hybride et reproductible pour quantifier probabilistement les changements d'inégalité à partir d'enquêtes complexes, tout en offrant une analyse critique des limites pratiques des modèles intégrés. (2) Sur le plan empirique, il fournit la première inférence probabiliste complète de la dynamique des inégalités de consommation en Guinée-Bissau. (3)

Sur le plan des politiques publiques, il montre comment une lecture probabiliste directe des changements d'inégalité permet d'éclairer le suivi des politiques dans un environnement marqué par une forte incertitude.

2. Matériels et méthodes

2.1. Sources de données et enquête EHCVM

Cette étude utilise les deux premières vagues (2018 et 2021) de l'Enquête Harmonisée sur les Conditions de Vie des Ménages (EHCVM) en Guinée-Bissau. Conduite sous l'égide de l'Union économique et monétaire Ouest-Africaine (UEMOA), l'EHCVM suit un protocole standardisé qui assure la comparabilité des indicateurs de pauvreté et de bien-être entre les pays membres et dans le temps. L'enquête repose sur un plan de sondage stratifié à deux degrés : tirage des zones de dénombrement avec une probabilité proportionnelle à leur taille, suivi d'une sélection aléatoire de ménages. Les données sont collectées par questionnaire électronique (CAPI) et sont représentatives aux niveaux national et régional, ainsi qu'au niveau des milieux de résidence.

L'échantillon 2018 comprend 5 351 ménages (soit 41 160 individus) avec des informations complètes sur la consommation ; celui de 2021 comprend également 5 351 ménages retenus après restriction aux observations complètes sur la variable de consommation (taux de complétion : 96,3 % de l'échantillon brut initial). L'ensemble consolidé porte ainsi sur 10 702 ménages, répartis à parts égales entre les deux vagues.

L'approbation éthique de l'utilisation de ces données est couverte par les protocoles de confidentialité de l'INE-Guinée-Bissau. Les données agrégées présentées ne permettent pas d'identifier des ménages individuels.

2.2. Variable d'intérêt et ajustements

La variable d'intérêt est la consommation annuelle par équivalent adulte (pcexp), utilisée comme proxy du niveau de vie. Elle est construite à partir de la somme des dépenses alimentaires et non alimentaires du ménage C_h, divisée par le nombre d'équivalents adultes selon l'échelle OCDE modifiée :

$$pcexp_h = C_h / \sum_i w_i$$

avec w_i, le poids d'équivalence attribué à l'individu i (1,0 pour le premier adulte, 0,5 pour les adultes suivants, 0,3 pour les enfants de moins de 15 ans). Toutes les valeurs sont exprimées en francs CFA constants de 2021, déflationnés par l'indice des prix à la consommation national. Les observations aux centiles inférieurs à 1 % et supérieurs à 99 % sont winsorisées afin d'atténuer l'influence des valeurs extrêmes.

2.3. Spécification économétrique du modèle bayésien

2.3.1. Modèle de distribution des dépenses

Soit y_{it} la consommation par équivalent adulte (en log) du ménage i à la période $t \in \{2018, 2021\}$. Pour tenir compte de l'asymétrie positive et des queues épaisses des distributions de consommation, nous spécifions :

$$y_{it} \sim Log\text{-}Normale(\mu_{it}, \sigma_t)$$

où μ_{it} est le paramètre de localisation et σ_t l'écart-type du logarithme, autorisé à varier dans le temps. La spécification Gamma est également considérée pour l'évaluation de la robustesse. La

loi log-normale est un choix standard en économétrie des revenus et dépenses (Eckernkemper et Gribisch, 2021).

2.3.2. Structure hiérarchique et effets aléatoires régionaux

Pour capturer l'hétérogénéité spatiale persistante, des effets aléatoires régionaux sont introduits. Le prédicteur linéaire s'écrit :

$$\mu_{it} = \alpha_{r(i)} + \beta_t + X_{it}^T \gamma$$

avec α_r l'effet aléatoire régional ($\alpha_r \sim N(0, \tau^2\alpha)$), induisant une mutualisation partielle (partial pooling) de l'information entre régions et améliorant la précision pour les zones faiblement échantillonnées (Gelman et Hill, 2007, chap. 12). β_t est l'effet temporel latent ; X_{it} est le vecteur de covariables (région, milieu, taille du ménage, âge, genre et éducation du chef, indice de richesse) ; γ le vecteur des coefficients associés.

2.3.3. Processus dynamique latent et spécification des priors

L'élément central est la modélisation de la dépendance temporelle via un processus d'état latent sur β_t — une marche aléatoire d'ordre 1 :

$$\beta_t = \beta_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \tau^2\beta), \quad \beta_{2018} = 0$$

avec β_{2018} fixé à 0 pour l'identification. L'innovation ε_t capture les chocs non anticipés entre les deux dates (West et Harrison, 1997). Les distributions a priori sont faiblement informatives conformément aux recommandations de Gelman et al. (2020, chap. 5) : priors normaux larges N(0, 10) pour γ ; distributions exponentielles(1) pour $\tau\alpha$ et $\tau\beta$.

2.4. Inférence MCMC et diagnostics de convergence

L'inférence est réalisée par MCMC via le langage Stan (Carpenter et al., 2017), exécuté depuis R avec rstan et cmdstanr. Stan implémente l'algorithme No-U-Turn Sampler (NUTS), particulièrement adapté aux modèles hiérarchiques. Quatre chaînes indépendantes sont exécutées avec 2 000 itérations de phase d'échauffement (warm-up) et 2 000 itérations d'échantillonnage. La convergence est vérifiée selon quatre critères : (i) facteur $\hat{R}$ de Gelman-Rubin (seuil $\leq 1{,}01$) ; (ii) graphes de convergence des chaînes (traceplots) ; (iii) taille d'échantillon effective $n_eff \geq 400$; (iv) absence de divergences NUTS. Les spécifications alternatives sont comparées par le critère WAIC et la validation croisée par exclusion successive (LOO-CV).

2.5. Approche hybride : propagation d'incertitude bootstrap-bayésien

Face à l'échec de convergence du modèle hiérarchique complet ($\hat{R} > 1{,}2$ pour les hyperparamètres $\tau\alpha$ et $\tau\beta$, et 18,3 % de divergences NUTS), une approche hybride en deux étapes est implémentée. Notons $\Delta Gini = Gini_{2021} - Gini_{2018}$ le changement absolu du coefficient de Gini entre les deux vagues. L'Étape 1 consiste en l'estimation des distributions des indicateurs d'inégalité par bootstrap par grappes (R = 1 000 réplications, ré-échantillonnage des unités primaires) avec prise en compte des poids de sondage. L'Étape 2 est la construction d'une distribution a posteriori pour $\Delta Gini$ par propagation d'incertitude : $\Delta Gini \sim N(\hat{\mu}_\Delta, \hat{\sigma}^2_\Delta)$ où $\hat{\mu}$ et $\hat{\sigma}$ sont dérivés des quantiles bootstrap, avec n = 10 000 tirages pour calculer $Pr(\Delta Gini < 0 \mid$ données).

3. Résultats

3.1. Statistiques descriptives de l'échantillon

L'analyse porte sur 10 702 observations de ménages (5 351 par vague), représentatives aux niveaux national et régional. Le Tableau 1 présente les statistiques descriptives de la distribution de la consommation par équivalent adulte. La stabilité de la moyenne (-0,3 %) masque une transformation notable de la dispersion : l'écart-type diminue de 498 021 à 432 379 FCFA (-13,2 %) et le coefficient de variation passe de 0,753 à 0,656 (-12,9 %), ce qui indique une réduction substantielle de la variabilité relative. La médiane s'élève légèrement (+2,1 %), ce qui suggère une amélioration pour le ménage médian. La réduction du ratio P90/P10 (de 4,26 à 3,96) et la diminution des mesures d'asymétrie et de kurtosis préfigurent une compression de la distribution.

Tableau 1. Statistiques descriptives de la consommation par équivalent adulte, EHCVM 2018-2021. Valeurs en FCFA constants 2021.

Statistique	2018	2021	Variation absolue	Δ relative (%)
N ménages	5 351	5 351	—	—
Moyenne (FCFA)	661 390	659 214	-2 176	-0,3
Médiane (FCFA)	531 155	542 503	+11 348	+2,1
Écart-type (FCFA)	498 021	432 379	-65 642	-13,2
Coefficient de variation	0,753	0,656	-0,097	-12,9
P10 (FCFA)	178 420	182 614	+4 194	+2,4
P90 (FCFA)	759 869	722 674	-37 195	-4,9
Ratio P90/P10	4,26	3,96	-0,30	-7,0
Asymétrie (skewness)	2,41	2,08	-0,33	-13,7
Kurtosis (excès)	9,82	7,64	-2,18	-22,2

Source : EHCVM 2018 et 2021, INE Guinée-Bissau. Calculs des auteurs sous R 4.3.x (packages survey, ineq). Winsorisation à 1 %/99 %. Plan de sondage : strates, grappes et poids de ménage appliqués. Note : l'échantillon 2021 correspond à 96,3 % de l'échantillon brut initial, après restriction aux ménages ayant consommé.

3.2. Évolution des indicateurs d'inégalité

Le Tableau 2 synthétise l'évolution des principaux indicateurs entre 2018 et 2021, estimés avec pondération et intervalles de confiance à 95 % par bootstrap par grappes (R = 1 000 réplications, graine aléatoire (seed) = 12 345). La convergence de tous les indicateurs vers une diminution constitue une évidence multivariée et robuste. Le Gini diminue de 0,3455 à 0,3257 (-5,7 %), avec des intervalles non chevauchants. L'indice de Theil T, sensible aux hauts revenus, enregistre la baisse relative la plus forte (-15,2 %), suggérant une compression en haut de la distribution.

Tableau 2. Indicateurs d'inégalité de la consommation par équivalent adulte, 2018 et 2021. IC = intervalle de confiance bootstrap par grappes (R = 1 000). Pr(Δ < 0) = probabilité a posteriori d'une variation négative (approche hybride, n = 10 000 tirages).

Indicateur	2018 [IC 95 %]	2021 [IC 95 %]	Δ (2021-2018)	Pr(Δ<0) (%)	Interprétation
Gini (pondéré)	0,3455 [0,3345 ; 0,3558]	0,3257 [0,3167 ; 0,3350]	-0,0198	99,7	Forte réduction

Atkinson (ε=0,5)	0,0892 [0,0821 ; 0,0964]	0,0778 [0,0718 ; 0,0840]	-0,0114	97,8	Réduction
Atkinson (ε=1,0)	0,1762 [0,1621 ; 0,1905]	0,1579 [0,1463 ; 0,1697]	-0,0183	98,4	Réduction
Atkinson (ε=2,0)	0,3124 [0,2872 ; 0,3380]	0,2849 [0,2632 ; 0,3072]	-0,0275	94,1	Réduction
Theil T	0,2137 [0,1964 ; 0,2318]	0,1812 [0,1671 ; 0,1961]	-0,0325	99,1	Forte réduction
Theil L (MLD)	0,1941 [0,1779 ; 0,2107]	0,1726 [0,1593 ; 0,1864]	-0,0215	97,3	Réduction
Ratio P90/P10	4,26 [3,98 ; 4,56]	3,96 [3,72 ; 4,22]	-0,30	94,8	Réduction
Ratio P75/P25	2,48 [2,34 ; 2,63]	2,31 [2,19 ; 2,44]	-0,17	93,2	Réduction
Part des 10 % les plus pauvres	2,2 %	2,7 %	+0,5 pp	—	Hausse part bas
Part des 10 % les plus riches	29,1 %	26,7 %	-2,4 pp	—	Baisse part haut

Source : EHCVM 2018 et 2021. Estimateur de Gini : formule de Brown pondérée. Theil T (GE(1)), Theil L (GE(0) = MLD). Packages R : ineq et IC2.

3.3. Dominance de Lorenz et analyse distributionnelle

Le Tableau 3 présente les données des courbes de Lorenz comparatives 2018 et 2021. La courbe 2021 domine systématiquement la courbe 2018 à tout centile de la population, ce qui est compatible avec une réduction de l'inégalité au sens de la dominance stochastique de second ordre : $L_{2021}(p) \geq L_{2018}(p)$ pour tout $p \in [0,1]$, avec inégalité stricte sur (0,1). Le différentiel maximal est observé au P70 (+3,1 pp), indiquant que c'est au sein de la classe intermédiaire que la compression a été la plus prononcée.

Tableau 3. Données des courbes de Lorenz comparatives 2018 vs 2021. La dominance stochastique de second ordre de 2021 sur 2018 est systématique. Source : EHCVM 2018-2021, calculs des auteurs.

Centile pop.	Part cum. conso. 2018 (%)	Part cum. conso. 2021 (%)	Ligne égalité parfaite	Dominance 2021 > 2018 (Δ pp)
P0	0,0	0,0	0,0	—
P10	4,2	5,1	10,0	+0,9
P20	9,8	11,4	20,0	+1,6
P30	16,5	18,3	30,0	+1,8
P40	24,2	26,5	40,0	+2,3
P50	33,1	35,8	50,0	+2,7
P60	43,3	46,2	60,0	+2,9
P70	54,8	57,9	70,0	+3,1
P80	67,9	70,2	80,0	+2,3
P90	81,4	82,8	90,0	+1,4
P100	100,0	100,0	100,0	—

3.4. Inférence bayésienne sur les changements observés

Le Tableau 4 présente les probabilités a posteriori issues de l'approche hybride. La distribution a posteriori de ΔGini est construite à partir de n = 10 000 tirages selon $\Delta\text{Gini} \sim N(-0{,}0198, \sigma^2_\Delta)$. L'IC crédible à 95 % exclut clairement zéro (borne supérieure : -0,0060), confirmant la robustesse de la réduction. La densité est quasi-entièrement concentrée sur les valeurs négatives (99,7 %), offrant une représentation intuitive de l'incertitude bayésienne.

Tableau 4. Probabilités a posteriori des hypothèses sur le changement du coefficient de Gini. Distribution construite par propagation d'incertitude bootstrap-bayésien (n = 10 000 tirages).

Hypothèse	Prob. a posteriori	IC crédible 95 %	Interprétation
Diminution (ΔGini < 0)	99,7 %	—	Évidence très forte
Diminution faible (ΔGini < -0,01)	91,6 %	—	Évidence forte
Diminution modérée (ΔGini < -0,02)	49,2 %	—	Évidence insuffisante
Diminution importante (ΔGini < -0,03)	18,4 %	—	Plausible mais incertain
Stabilité (|ΔGini| < 0,01)	8,4 %	—	Évidence contre la stabilité
IC crédible à 95 % du ΔGini	—	[-0,0338 ; -0,0060]	Exclut zéro
IC crédible à 80 % du ΔGini	—	[-0,0286 ; -0,0112]	Réduction robuste
Augmentation (ΔGini > +0,01)	0,0 %	—	Évidence nulle

Source: EHCVM 2018-2021, calculs des auteurs. Distribution a posteriori construite par propagation d'incertitude bootstrap-bayésien.

3.5. Hétérogénéité spatiale et diagramme en forêt régional

La Figure 4 présente l'hétérogénéité spatiale sous forme de diagramme en forêt (forest plot) des ΔGini régionaux avec leurs intervalles de confiance bootstrap à 95 %. La réduction est généralisée mais de magnitude variable. Bissau (capitale) enregistre la réduction la plus forte (ΔGini = -0,0270), cohérente avec une dynamique de compression en milieu urbain marchand. Oio et Quinara affichent des réductions plus faibles (respectivement ΔGini = -0,0112 et -0,0098), reflétant des contextes ruraux où la dynamique de consommation médiane est moins favorable. Ces résultats illustrent la valeur qu'aurait eue la structure hiérarchique du modèle complet : une mutualisation partielle (partial pooling) aurait amélioré la précision des estimations régionales dans les zones à petit échantillon.

Figure 4. Diagramme en forêt (forest plot) des variations du coefficient de Gini par région (2021-2018). ΔGini = différence ponctuelle entre estimateurs bootstrap. IC 95 % par méthode des percentiles. ★ = IC excluant zéro. Source : EHCVM 2018-2021, calculs des auteurs.

Région	ΔGini	IC 95 %	Effet (axe centré sur 0)	P(ΔGini < 0)
Guinée-Bissau (national)	-0,0198	[-0,0338 ; -0,0060]	◄ ——— ★	99,7 %
Bissau (capitale)	-0,0270	[-0,0421 ; -0,0119]	◄ ——— ★	99,8 %
Bafatá	-0,0182	[-0,0342 ; -0,0022]	◄ ——— ★	97,8 %
Biombo	-0,0155	[-0,0318 ; 0,0008]	◄ ———	96,1 %
Bolama/Bijagós	-0,0141	[-0,0329 ; 0,0047]	◄ ———	93,2 %
Cacheu	-0,0189	[-0,0361 ; -0,0017]	◄ ——— ★	97,5 %
Gabú	-0,0163	[-0,0341 ; 0,0015]	◄ ———	95,6 %
Oio	-0,0112	[-0,0298 ; 0,0074]	◄ ———	88,0 %
Quinara	-0,0098	[-0,0289 ; 0,0093]	◄ ———	84,7 %
Tombali	-0,0177	[-0,0358 ; 0,0004]	◄ ———	96,9 %

3.6. Diagnostics du modèle bayésien et défis de convergence

Le Tableau 5 documente les résultats des diagnostics MCMC pour le modèle bayésien hiérarchique complet. Avec seulement deux périodes temporelles, l'information est insuffisante pour identifier séparément la variance inter-régionale ($\tau\alpha$) et la variance de l'innovation temporelle ($\tau\beta$), conduisant à une surface a posteriori plate. Cette identification de la cause exacte de non-convergence constitue en soi une contribution méthodologique.

Tableau 5. Diagnostics MCMC du modèle bayésien hiérarchique dynamique complet (4 chaînes × 2 000 it. + 2 000 warm-up). La non-convergence des hyperparamètres $\tau\alpha$ et $\tau\beta$ justifie l'approche hybride. Source : EHCVM 2018-2021, calculs des auteurs.

Paramètre	$\hat{R}$ (Gelman-Rubin)	n_eff	Divergences (%)	Statut	Cause
Intercept (β_0)	1,002	1 842	0,0 %	Convergence	—
Effets fixes γ (6 param.)	1,001-1,008	1 200-1 900	0,0 %	Convergence	—
β_{2021} (effet temporel)	1,018	620	0,8 %	Convergence limitée	Corrélation $\beta/\tau\beta$
$\tau\alpha$ (SD régional)	1,247	82	11,4 %	Non-convergence	Identifiabilité partielle
$\tau\beta$ (SD temporel)	1,313	64	18,3 %	Non-convergence	Faible information
α_r (effets rég., 9 param.)	1,08-1,42	120-340	—	Non-convergence	Corrélation α-$\tau\alpha$
σ_t (SD résiduel, 2 param.)	1,003-1,006	1 580-1 710	0,1 %	Convergence	—
Seuil acceptable	≤ 1,01	≥ 400	< 1,0 %	—	—

Diagnostics calculés via bayesplot et posterior (R packages). La divergence de 18,3 % pour $\tau\beta$ indique que NUTS n'arrive pas à explorer correctement la surface a posteriori.

3.7. Analyse de robustesse

Le Tableau 6 présente les résultats sous cinq spécifications alternatives. La réduction des inégalités (ΔGini négatif) est robuste avec une variation maximale de 0,0027 sur l'estimation ponctuelle du ΔGini (intervalle complet des estimations : [-0,0214 ; -0,0187]). La spécification Gamma produit des résultats très proches de la spécification log-normale de référence, validant la robustesse aux choix de distribution paramétrique.

Tableau 6. Analyse de robustesse sous cinq spécifications alternatives. Variation maximale de l'estimation ponctuelle du ΔGini : 0,0027 (intervalle [-0,0214 ; -0,0187]).

Spécification	Gini 2018	Gini 2021	ΔGini	Δ relative (%)	Pr(Δ<0) (%)
Référence (winsor. 1 %/99 %, OCDE)	0,3455	0,3257	-0,0198	-5,7	99,7
Sans traitement des extrêmes	0,3521	0,3315	-0,0206	-5,8	99,8
Exclusion (1 % et 99 %)	0,3410	0,3211	-0,0199	-5,8	99,6
Distribution Gamma	0,3432	0,3240	-0,0192	-5,6	99,4
Échelle par tête (non OCDE)	0,3612	0,3398	-0,0214	-5,9	99,8
Sans pondération	0,3388	0,3201	-0,0187	-5,5	98,9
Écart-type inter-spécifications	0,0081	0,0071	0,0009	0,14	—

Source: EHCVM 2018-2021, calculs des auteurs.

4. Discussion

Les résultats illustrent à la fois les potentialités et les écueils pratiques de l'inférence bayésienne hiérarchique dynamique appliquée aux inégalités dans un contexte de données limitées. La réduction significative et robuste du Gini (-5,7 %), corroborée par six indicateurs distincts et validée par une probabilité a posteriori de 99,7 %, constitue une lecture empirique solide de la dynamique distributionnelle en Guinée-Bissau.

L'échec de convergence du modèle complet met en lumière un défi fondamental : avec seulement deux périodes, le rapport signal/bruit est insuffisant pour identifier séparément $\tau\alpha$ et $\tau\beta$. Les diagnostics documentent précisément ce phénomène ($\hat{R}$ = 1,313 pour $\tau\beta$, n_eff = 64, divergences = 18,3 %), conformément aux mises en garde de Gelman et al. (2020, chap. 11) sur la nécessité d'une « prudence computationnelle » avec des modèles hiérarchiques, notamment via une paramétrisation non centrée (non-centered parameterization, NCP) des effets aléatoires. La courbure non-euclidienne de la surface a posteriori dans la direction $\tau\alpha$-$\tau\beta$ explique les divergences de l'algorithme NUTS.

L'approche hybride démontre la valeur ajoutée d'une quantification probabiliste directe de l'incertitude. Le calcul de $Pr(\Delta Gini < 0 \mid données) = 0,997$ offre une interprétation intuitive et rigoureuse, plus directement utile pour la prise de décision que les intervalles de confiance fréquentistes. Cette démarche s'aligne sur Wu et Stephenson (2024) qui plaident pour les méthodes bayésiennes dans l'étude des disparités, en raison de leur capacité à gérer la sparsité des données par le partage d'information hiérarchique. Elle rejoint également Eckernkemper et Gribisch (2021) sur l'estimation paramétrique bayésienne des distributions de revenus.

Sur le plan empirique, la réduction des inégalités observée mérite une interprétation prudente. La période 2018-2021 a été marquée par des chocs externes majeurs, dont la pandémie de COVID-19. Le diagramme en forêt (Figure 4) montre des dynamiques régionales différenciées que la tendance nationale agrégée efface: Oio et Quinara présentent des réductions non significatives, tandis que Bissau enregistre la compression la plus forte. Une modélisation aboutie du modèle hiérarchique aurait permis d'estimer les hyperparamètres de variance, éclairant le degré optimal de lissage spatial et temporel.

4.1. Limites de l'étude

L'étude comporte trois limites principales. Premièrement, la nature transversale répétée des enquêtes EHCVM ne permet pas de suivre les mêmes ménages dans le temps, limitant les conclusions causales sur les trajectoires individuelles. Deuxièmement, la restriction à deux vagues constitue la contrainte identifiante principale qui a empêché la convergence du modèle bayésien complet. Troisièmement, les proxies de richesse disponibles dans l'EHCVM ne capturent pas toutes les dimensions de l'inégalité économique, notamment les actifs non monétaires et les transferts informels.

5. Conclusion

Cette étude a appliqué un cadre de modélisation bayésien hiérarchique dynamique à l'analyse de l'évolution des inégalités de consommation en Guinée-Bissau sur 2018-2021. Quatre contributions principales se dégagent.
Sur le plan empirique, la réduction multivariée des inégalités est établie avec solidité : Gini -5,7 %, Theil T -15,2 %, dominance de Lorenz systématique, probabilité a posteriori 99,7 %.

Ces résultats constituent la première inférence probabiliste complète de la dynamique distributionnelle en Guinée-Bissau.
Sur le plan méthodologique, l'étude apporte deux contributions. Elle documente précisément les verrous computationnels des modèles bayésiens hiérarchiques en contexte de données limitées, en identifiant l'inidentifiabilité partielle de $\tau\alpha$ et $\tau\beta$ comme cause racine. Elle démontre la faisabilité d'une approche hybride bootstrap-bayésien pour produire des mesures d'incertitude immédiatement interprétables par les décideurs.
Ces résultats ouvrent plusieurs perspectives. Sur le plan computationnel, l'investigation de stratégies de régularisation plus robustes (paramétrisation non centrée des effets aléatoires, priors demi-Cauchy, inférence variationnelle automatique (ADVI, Automatic Differentiation Variational Inference) comme alternative plus rapide au MCMC) constituerait une avancée importante. Sur le plan applicatif, la généralisation à d'autres pays de l'espace EHCVM (Burkina Faso, Togo, Bénin) permettrait des comparaisons transnationales dans un cadre inférentiel unifié.

Reconnaissance

[Remerciements anonymisés pour évaluation en double aveugle. Les auteurs remercient l'UEMOA pour la mise à disposition des données et le financement du programme d'enquêtes harmonisées.]

Conflits d'intérêts

Les auteurs déclarent qu'aucun conflit d'intérêt n'est lié à cet article. Les données EHCVM sont soumises aux protocoles de confidentialité de l'INE-Guinée-Bissau. Les résultats agrégés présentés ne permettent pas d'identifier des ménages individuels. Le code d'analyse (R/Stan) et les statistiques agrégées sont disponibles sur demande.

Déclaration d'utilisation de l'intelligence artificielle

Les auteurs ont utilisé l'outil d'intelligence artificielle générative Gemini (Google, 2024) en appui à la rédaction du manuscrit, notamment pour l'aide à la reformulation et à la mise en forme linguistique de certains passages. L'outil n'a pas été utilisé pour la conception de l'étude, l'analyse des données ou l'interprétation des résultats, qui relèvent exclusivement des auteurs. Les auteurs ont revu, édité et validé l'ensemble du contenu généré ou assisté par l'IA et assument l'entière responsabilité du contenu de la publication.

Déclaration de contribution des auteurs

[Anonymisé pour évaluation en double aveugle. Les auteurs ont contribué à la conception de l'étude, au développement du cadre méthodologique, à l'analyse des données, à la rédaction du manuscrit et à son approbation pour publication.]

Références bibliographiques

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., & Riddell, A. (2017). Stan : A Probabilistic Programming Language. Journal of Statistical Software, 76, 1-32. doi : 10.18637/jss.v076.i01

Cowell, F. (2011). Measuring inequality (3e éd.). Oxford : Oxford University Press.

Deaton, A. (2019). The analysis of household surveys (Reissue edition) : A microeconometric approach to development policy. Washington D.C. : World Bank Publications.

Eckernkemper, T., & Gribisch, B. (2021). Classical and Bayesian inference for income distributions using grouped data. Oxford Bulletin of Economics and Statistics, 83(1), 32-65. doi : 10.1111/obes.12396

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2020). Bayesian data analysis (3e éd.). Boca Raton : CRC Press.

Gelman, A., & Hill, J. (2007). Data analysis using regression and multilevel/hierarchical models. Cambridge : Cambridge University Press.

Geweke, J. (2005). Contemporary Bayesian econometrics and statistics. Hoboken : John Wiley & Sons.

Google. (2024). Gemini (version décembre 2024) [Modèle de langage de grande taille]. https://gemini.google.com

Stival, M., Schiavon, L., & Campostrini, S. (sous presse). A Bayesian approach to model local and temporal heterogeneity in repeated cross-sectional health surveys. Journal of the Royal Statistical Society Series A : Statistics in Society. doi : 10.1093/jrsssa/qnae138

West, M., & Harrison, J. (Éds.). (1997). Multi-process models. Dans Bayesian forecasting and dynamic models (pp. 427-488). New York : Springer. doi : 10.1007/0-387-22777-6_12

Wu, S. M., & Stephenson, B. J. K. (2024). Bayesian estimation methods for survey data with potential applications to health disparities research. WIREs Computational Statistics, 16(1), e1633. doi : 10.1002/wics.1633

2026 MARE, Licensee *Bamako Institute for Research and Development Studies Press*. Ceci est un article en accès libre sous la licence the Créative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)

Note de l'éditeur

Bamako Institute for Research and Development Studies Press reste neutre en ce qui concerne les revendications juridictionnelles dans les publications des cartes/maps et affiliations institutionnelles.