

## *Construction de pseudo panel: le cas de la Mauritanie*

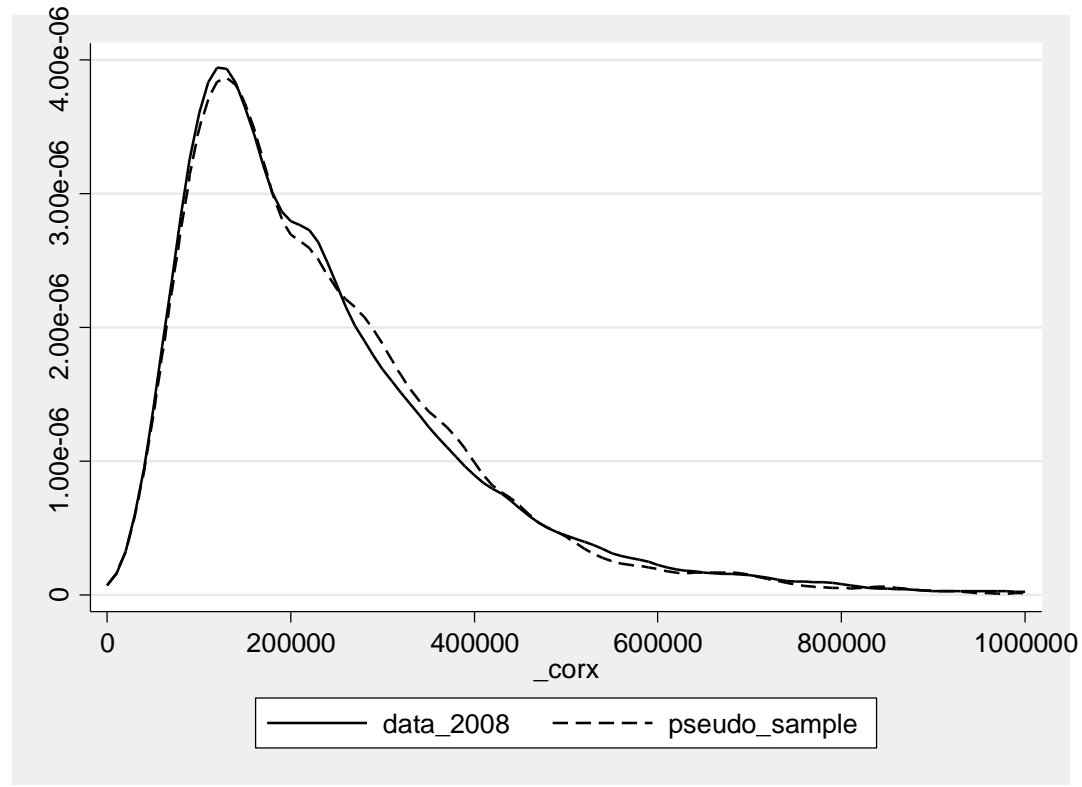
- Les données d'enquêtes auprès des ménages de type panel sont rarement disponibles dans les pays africains, comme c'est le cas également pour la Mauritanie.
- Pour certaines analyses distributives, les données de type panel sont requises. C'est le cas par exemple pour estimer la pauvreté transitoire et chronique.
- Au cours de la dernière décennie, la Mauritanie (ONS) a mené deux enquêtes de type transversal, et qui sont riches et représentatives au niveau national (Enquête nationale sur les conditions de vie des ménages en Mauritanie 2008 et en 2014).
- Le fichier des données de 2008 contient 13 738 observations, alors que celui de 2014 il contient 9557 observations.
- Dans la littérature, plusieurs approches ont proposé de construire ce que nous appelons le pseudo panel. La construction de ces panneaux repose principalement sur l'adaptation des unités synthétiques agrégées (grappes, cohortes, etc.) ou même sur les unités finales échantillonnées.
- De toute évidence, parmi les critiques formulées à l'égard de l'utilisation des unités agrégées, on peut citer la réduction de la variabilité des résultats d'intérêt (dépenses réelles par habitant dans notre cas).
- De plus, l'utilisation de modèles linéaires simples pour la prédiction des résultats peut soulever un biais significatif en omettant la variabilité attribuée à la composante résiduelle des modèles linéaires.

### ***La méthode d'appariement des scores (The Matching Pseudo-Panel)***

- Comme première approche pour construire le pseudo-panel mauritanien avec la population de référence de 2014, nous proposons d'utiliser une technique non paramétrique ou précisément la méthode d'appariement des scores (nn-1) -Voir l'annexe A pour l'estimation des scores-.
- Pour chaque ménage échantillonné en 2014, nous recherchons un ménage semblable en 2008. Comme le suggère Deaton (1985), il est important d'utiliser un ensemble de caractéristiques qui peuvent avoir une forte probabilité de rester constant entre les deux périodes (2008 et 2014).
- De plus, pour éviter l'endogénéité de l'utilisation des résultats comme variables explicatives dans les modèles d'appariement, on évite l'utilisation des variables d'analyses tels que le revenu, les classes sociales, les quintiles qui sont fondés sur les dépenses ou le revenu, ainsi que des variables explicatives fortement corrélées avec les variables d'analyses.
- Dans notre cas, l'unité échantillonnée finale est le ménage. Nous proposons un ensemble riche avec un grand nombre de covariantes qui sont supposés bien capter la composition du ménage, et ce, en fonction de l'âge et du sexe des membres du ménage (voir annexe B pour L'évolution de la composition démographique des heures supplémentaires). Le tableau suivant résume les variables sélectionnées:

<b><i>HH caractéristiques démographiques</i></b>	
i.comp_age	L'âge du chef de ménage en 2014: nous ajoutons six ans pour ceux de 2008
compa_age	Le nombre d'enfants qui ont entre 0 et 8 ans en 2008 et ceux qui ont entre 6 et 14 ans en 2014
av_hh_age_m	L'âge moyen du membre masculin du ménage en 2014. Pour l'année 2008, six ans où ajouté.
av_hh_age_f	L'âge moyen des femmes membres du ménage en 2014. Pour l'année 2008, six ans où ajouté.
i.hhsz	La taille du ménage
depend0014	Population de 0 à 14 ans
active1564	Population entre 15-64 ans
<b><i>HH zone de résidence</i></b>	
i.capital	Zone d'habitat (Urbain, semi-urbain, rural)
i.region	Nous supposons que le ménage a une forte probabilité de vivre dans sa région initiale de 2008.

- Supposons que le pseudo échantillon soit bien construit. Quelles sont les statistiques attendues de ce pseudo échantillon construit pour la vague de 2008?
- Étant donné que l'échantillon de 2014 est représentatif au niveau national, la distribution du bien-être du pseudo échantillon de 2008 doit converger vers l'échantillon collecté de 2008. Cela impliquera aussi que la distribution du pseudo panel sera semblable à celle de l'échantillon de 2008.
- Dans ce qui suit, nous présentons La densité des dépenses réelles par habitant (pcexpdr) du pseudo échantillon de 2008 et aussi celle de l'échantillon réel de 2008.



- Comme nous pouvons le constater, les deux densités sont très proches. En quelques mots, on peut dire que si les ménages de 2014 sont ceux qui étaient sélectionnés en 2008, ils donneront la même distribution que celle de l'échantillon de 2008.
- Dans ce qui suit, nous montrons la matrice de transition, Taux de pauvreté de 2008 et 2014, et ce, avec le pseudo-échantillon proposé en 2008.

poor_2008	poor_2014		Total
	0	1	
0	43.82	12.07	55.89
1	23.22	20.89	44.11
Total	67.03	32.97	100.00

- Bien sûr, la pauvreté estimée en 2014 est d'environ 33%, et c'est parce que notre approche non paramétrique assure ce résultat et où le groupe traité est l'échantillon collecté de 2014. Pour l'année 2008, l'effectif estimé basé sur le pseudo Échantillon est d'environ 44.11%, ce qui s'approche de celui basé sur les données recueillies de 2008 (environ 44.50%).
- De plus, on peut noter que la proportion de ceux qui sortent de la pauvreté est plus élevée que celle de l'entrée dans la pauvreté. En outre, environ 20,9% de la population reste pauvre dans les deux périodes. Cela suggère des politiques spécifiques pour éliminer cette forme chronique de pauvreté.

### *L'approche probabiliste (The Probabilistic Pseudo-Panel)*

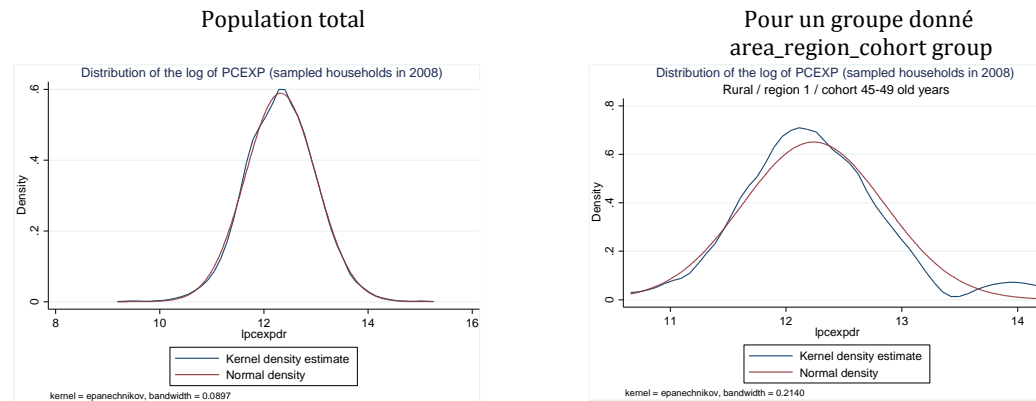
- Dans la première approche, nous avons suivi une technique d'appariement non paramétrique, qui consiste à sélectionner un ménage semblable à partir des données de 2008 pour chaque ménage échantillonné en 2014.
- La question naturelle qui peut être soulevé concerne l'impact de l'hypothèse de l'erreur nulle dans le choix du ménage correspondant.
- Nous indiquons ici que notre hypothèse était que les erreurs des moyens des variables d'intérêt du pseudo-panel convergeront à zéro, et ceci en considérant le grand nombre de ménages utilisés pour construire le pseudo panel en 2008.
- Même avec cette explication, la question raisonnable qui peut être soulevé est de savoir si une autre approche peut donner des résultats similaires à celle basée sur la technique d'appariement.
- Comme deuxième approche, nous nous baserons sur un modèle probabiliste simple. Principalement, nous proposons de construire des groupes de population cohérents basés sur les cohortes des chefs de ménage (10 groupes d'âge), la zone de vie (rurale / urbaine) et les régions (13 régions). Ainsi, à partir des 13738 observations, nous construirons environ 260 groupes de population pour l'année 2008.
- On peut noter que dans cette approche nous définissons d'avance le résultat d'intérêt, et qui est dans notre cas les dépenses réelles par habitant.
- Comme on le sait, le logarithme des dépenses par habitant suit de près une distribution normale. C'est toujours le cas pour les groupes construits de la population.
- Même si nous n'avons pas le niveau exact de la variables d'intérêt en 2008 pour chaque ménage échantillonné en 2014, nous avons en contrepartie générer une distribution probabiliste du revenu en 2008.

- Formellement, si  $g_{i,j,k}$  dénotes le groupe de population de 2008 vivant en zone  $i$  en région  $j$  et pour le cohorte  $k$ , la distribution de la variable d'intérêt d'un ménage sélectionné aléatoirement de ce groupe doit suivre une loi normal  $N(\mu_{i,j,k}, \sigma_{i,j,k})$ .
- En utilisant les données de 2008, on estime les deux paramètres de la distribution de chaque groupe  $(\mu_{i,j,k}, \sigma_{i,j,k})$  Ainsi, dans cette approche probabiliste, pour chaque ménage échantillonné en 2014, nous commençons par identifier son groupe en 2008, puis nous lui attribuons la répartition potentielle de son revenu en 2008.
- Étant donné que notre objectif est d'évaluer les statistiques de la pauvreté dynamique, soit la matrice de transition de la pauvreté, nous commençons par évaluer la probabilité d'être pauvre dans chaque période, et ce, pour chaque ménage.
- Pour la période de 2014, nous disposons de l'information réelle et la probabilité prend la valeur de un si le ménage est pauvre en 2014 et de zéro autrement.
- Pour la période de 2008, nous devons évaluer la probabilité d'être pauvre, et cela en fonction de la répartition spécifique du revenu du ménage en 2008. Basé sur cela, probabilité d'être pauvre est simplement la distribution normale cumulée du revenu per capita jusqu'à entre zéro et le du seuil de pauvreté.
- Soit  $P_{h,1}$  la probabilité d'être pauvre en 2008 et  $P_{h,2}$  est celle de 2014. Sur cette base, nous pouvons construire les quatre probabilités d'intérêt suivantes:

1.  $PP_h = P_{h,1} * P_{h,2};$
2.  $NP_h = (1 - P_{h,1}) * P_{h,2};$
3.  $PN_h = P_{h,1} * (1 - P_{h,2});$
4.  $PP_h = (1 - P_{h,1}) * (1 - P_{h,2});$

- Une fois que la probabilité de ces quatre statues est estimée au niveau du ménage, les valeurs attendues régénéreront la matrice de transition.

## Résultats



- Les graphiques ci-haut montrent la répartition du logarithme des dépenses par habitant pour l'ensemble de la population et pour un groupe donné parmi les 260 groupes construits. Comme nous pouvons le constater, l'hypothèse d'une distribution normale proche est vérifiée.

		<i>Approche probabiliste</i>			<i>Approche d'appariement</i>		
		Non pauvre	Pauvre	Totale - lignes	Non pauvre	Pauvre	Totale - lignes
2008		2014			2014		
	Non pauvre	41.7	15.59	57.29	43.82	12.07	55.89
	Pauvre	25.27	17.44	42.71	23.22	20.89	44.11
	Totale colonnes	66.97	33.03	100	67.03	32.97	100

- Comme on peut le constater, l'approche probabiliste donne des résultats pratiquement similaires à celle de l'appariement. Évidemment, cela conforte la robustesse des