ISSN 0825-0305

MODÉLISATION DES DONNÉES DE LA PAUVRETÉ PAR LA FAMILLE SINGH-MADDALA

MOHAMED CHEIKH HAIDARA

Abstract. Nous abordons la modélisation des données de la pauvreté, en particulier, celles des bases de données sénégalaises de 1996 à 2001, par la famille des fonctions de répartition de Singh-Maddala. Les résultats sont bien meilleurs que l'ajustage classique Lognormal. Ces résultats permettent le calcul fiable des indicateurs de pauvreté en termes d'intervalles de confiance en vue d'un suivi efficace de la pauvreté, un des objectifs du Millénaire pour le développement.

1. Introduction

L'analyse statistique de la pauvreté et de l'inégalité repose sur des modèles paramétriques et non paramétriques de la loi des revenus ou des dépenses des ménages. L'indicateur de pauvreté ou d'inégalité est une grandeur échantillonnée sur la population sous la forme empirique $I(F_n)$, dépendant de la fonction de répartition empirique F_n et de la taille n de l'échantillon tiré sans remise. On peut alors calculer la forme asymptotique exacte de cet indicateur $I(F)$, calculée en fonction de la distribution F supposée ou posée. Des résultats de ce type sont disponibles dans [6], [4], [9] pour les indicateurs de pauvreté et dans [3] pour les inégalités, par exemple. Mais l'utilisation de ces estimations en vue d'un suivi efficace du bien être des populations étudiées requiert aussi une grande précision basée sur des intervalles de confiances, eux même dérivés des résultats de normalité asymptotique sous la forme

$$\sqrt{n}(I(F_n) - I(F)) \simeq \mathcal{N}(0, \sigma^2(F)),$$

où $\mathcal{N}(0, \sigma^2(F))$ désigne la loi gaussienne centrée et de variance $\sigma^2(F)$. L'application de tels résultats repose explicitement sur le calcul de $\sigma^2(F)$. On peut toujours recourir à des méthodes non paramétriques. Cependant ce type de méthodes ne donne pas des résultats satisfaisants dans ce domaine. Les méthodes paramétriques semblent se révéler plus pertinentes ici. Ceci nous motive à trouver les meilleurs modèles de familles paramétriques ajustant au mieux les données de la pauvreté

2000 *Mathematics Subject Classification.* Primary 62-07, 62H12.
Key words and phrases. Mesure de pauvreté, distribution des revenus et des dépenses, familles paramétriques et non paramétriques, ajustage de lois.

Received the 24th, November 2009. Revised version received the 15th, January 2009.

ISSN 0825-0305

disponibles dans les bases de données actuelles en vue d'un suivi efficace de la pauvreté.

Suivant Aitchison et Brown(1957), Weiss(1972), le modèle lognormal est relativement correct pour ajuster les revenus de populations homogènes. Dans [10], la normalité asymptotique de l'indicateur général de pauvreté a été établie. Dans une approche applicative aux bases de données existantes, du Sénégal par exemple, ces résultats ont pu être largement utilisés. Le modèle qui y a été retenu et qui a donné satisfaction - et a été ensuite pris comme référence - est celui de la loi lognormale, suivant en cela aussi les conclusions de Brachman [2] pour des données similaires en Allemagne.

Le problème de trouver une famille paramétrique qui ajuste avec la plus grande précision les données des revenus ou dépenses trouve aussi sa justification dans le contexte de la réalisation des Objectifs du Millénaire pour le Développement (OMD). En effet, l'un des principaux OMD concerne la réduction de moitié de la pauvreté à l'horizon 2015. Ceci repose, bien sûr, sur la fiabilité et la rigeur de la mesure de pauvreté.

Revenant à la modélisation des revenus, Brachmann ([2]), signale l'excellence de l'ajustage de la loi de Singh-Maddala [14] pour les données Allemandes. A sa suite, nous nous proposons d'aborder la modélisation des données Sénégalaises de la pauvreté et de proposer une méthodologie pour l'exploitation des bases de données relatives aux enquêtes socio-économiques des pays à revenus faibles. En effet, un succès de l'ajustage par cette loi permet, sur ces bases, une étude rigoureuse de la répartition géographique de la pauvreté et de l'évolution de celle-ci.

Précisément, la loi étudiée ici a été indoduite par Singh et Maddala ([14]) sous la forme

$$F(x) = 1 - \frac{1}{(1 + ax^b)^c} \tag{1.1}$$

et dépend des trois paramètres positifs. Il s'agit d'une loi de Paréto généralisée puisqu'elle peut encore s'écrire sous la forme

$$1 - F(x) = x^{-bc} L(x),$$

où, pour tout $\lambda > 0$,

$$\lim_{x \to \infty} L(\lambda x)/L(x) = 1.$$

Nous montrons dans cette note que cette famille paramétrique est un très bon modèle, aussi bien pour le revenu et la dépense, pour les bases de données sénégalaises de 1996 à 2001, pour toutes les régions. Cette constante dans l'adéquation nous fait croire qu'elle s'étendrait naturellement aux bases de données des pays de même nature. Notons que de plus en plus, l'utilisation du revenu pour l'étude de la pauvreté est délaissée au profit de la dépense, plus accessible dans les enquêtes auprès des ménages. Il est donc rassurant que le modèle soit aussi valable pour elle. Dans la suite, nous ne parlerons plus que de la variable

ISSN 0825-0305

dépense Y de fonction de répartition G. Cependant, la méthodologie reste applicable au revenu.

Voici comment notre article est organisé. Dans la section 2, nous décrivons les méthodes d'ajustage et les outils d'appréciation. Dans la section 3, nous introduisons les échelles d'équivalence-adulte nécessaires pour définir les variables de travail. Enfin dans la section 4, les résultats de la modélisation sont exposés, illustrés et commentés. Une conclusion termine l'article.

ISSN 0825-0305

2. Ajustage des données.

Les indicateurs de pauvreté, en général dépendent des dépenses les plus faibles. Cela nous pousse à regarder attentivement la queue inférieure de la distribution. A cet effet, nous utilisons le plus souvent la transformation de la variable dépense Y de fonction de répartition G, en

$$X = 1/(Y - y_0),$$

où

$$y_0 = \inf\{x, G(x) > 0\},$$

est la borne inférieure de G. La fonction de répartition de X est alors, pour tout $x \geq 0$,

$$F(x) = 1 - G(1 - 1/x).$$

Dans [11], la modélisation de F par une famille Lognormal s'est révélée assez bonne au sens suivant. En comparant la fonction de répartition empirique $\widehat{F}$ issue des observations avec celle d'une loi Lognormal F_0, dont les paramètres sont estimés par les méthodes classiques (méthode du maximun de vraisemblance et méthode des moments), l'erreur uniforme

$$\left\|\widehat{F} - F_0\right\|_\infty = \sup_{-\infty<x<+\infty} \left|\widehat{F}(x) - F_0(x)\right|;$$

ne dépasse guère 6% pour toutes les régions étudiées, pour le revenu et la dépense de 1996 à 2001. Les p-values du test de Kolmogorov-Smirnov supportent aussi l'ajustage.

Nous essayons d'améliorer ces résultats en vue d'applications plus pertinentes. Pour cela, nous utilisons la densité de Singh-Maddala

$$f(x) = \frac{abcx^{b-1}}{(1 + ax^b)^{c+1}}. \tag{2.1}$$

Pour ajuster les données $x_1, ..., x_n$, selon ce modèle, nous devons commencer par estimer les paramètres par la méthode du maximum de vraisemblance au moyen des équations de vraisemblance données par :

$$\begin{cases} \frac{n}{a} - \sum_{i=1}^{n}(c+1)\frac{x_i^b}{1+ax_i^b} &= 0 \\ \frac{n}{b} - \sum_{i=1}^{n} log(x_i)\left[1 - \frac{(c+1)ax_i^b}{1+ax_i^b}\right] &= 0 \\ \frac{n}{c} - \sum_{i=1}^{n} log(1 + ax_i^b) &= 0 \end{cases}$$

Remarquons que dans ce cas d'espèce, il est impossible de trouver les formes explicites des estimateurs du maximum de vraisemblance $\widehat{a}$, $\widehat{b}$ et $\widehat{c}$. On a recours aux méthodes numériques, en particulier au package *rootSolve* du locigiel R et à un programme reposant sur l'algorithme de

ISSN 0825-0305

Newton-Raphson pour la résolution de ce système non linéaire. Pour les tailles de nos bases, (3700 pour la base de 1996, 6000 pour la base de 2001), les calculs sont complétés en très peu de temps.

Pour mesurer l'efficacité de l'ajustage, nous comparons d'abord la fonction de répartition empirique $\widehat{F}$ et la fonction de répartition de Singh-Maddala $F(x|\widehat{a},\widehat{b},\widehat{c})$ pour les paramètres estimés. La déviation maximale

$$\left\|\widehat{F}-F\right\|_{\infty}=\sup_{-\infty<x<+\infty}\left|\widehat{F}(x)-F(x|\widehat{a},\widehat{b},\widehat{c})\right|,$$

mesure la qualité de l'ajustage, tout comme les p-values de Kolmogorov-Smirnov et de Wilcox.

Ensuite, nous vérifions aussi la qualité de l'ajustage pour les densités. A cet effet, suivant [2], nous considérons l'estimateur à noyau de Parzen

$$\hat{f}_n(x)=\frac{1}{nh}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right),$$

où $K\geq 0$ est la densité gaussienne standard et $h>0$ est la largeur de la fenêtre à celle de Singh-Maddala $f(|\widehat{a},\widehat{b},\widehat{c})$.

Nous allons appliquer ces méthodes aux bases de données actuelles. Cependant, il est important de respecter certains préalables avant l'application. En effet, les bases de données socio-économiques des ménages ont été standardisées. Il est donc important de situer cette étude dans ce cadre cohérent, de façon à permettre à un lecteur intéressé de la reproduire. Pour cela, nous devons d'abord introduire les échelles d'équivalence adultes avant d'exposer les résultats sur les données puisque celles-ci sont issues de leur exploitation.

3. Equivalence Adulte d'un ménage

Sur le plan théorique, l'équivalent-adulte se calcule selon une fonction de bien être U dépendant du revenu du ménage et de sa taille

$$U(R,N).$$

L'échelle d'équivalence-adulte m vérifie

$$U(\frac{R}{m},1)=U(R,N).$$

Cela veut dire qu'en divisant le revenu par m, le ménage se comporte comme un seul adulte du point de vue du bien être tel que décrit par U. Le modèle est bien plus complexe car la fonction peut ne pas être la même pour toutes les zones. Elle dépendrait des facteurs socio-économiques et géographiques. Plusieurs auteurs se sont intéressés à la question. On peut citer par exemple Pollack et Wales [12], Bradbury [5], Duclos et Mercatier-Pats [7], Hourrietz et Olivier [8] et les références incluses. L'unanimité est loin d'être faite sur le choix de la fonction

ISSN 0825-0305

de bien être. Plusieurs modèles existent dont ceux linéaires. Selon les approches, nous pouvons distinguer les échelles suivantes.

3.1. **Echelles d'Oxford.** Le premier adulte est compté avec l'unité, le second adulte avec le coefficient 0.7 et les enfants de moins de 15 ans avec 0.5.

3.2. **Echelle FAO et OMS.** Ici la pondération applique 0,5 aux personnes âgées de moins de 15 ans, 0.8 aux femmes de plus de 15 ans et l'unité aux hommes de plus de 15 ans. Notons que cette échelle se fonde en partie sur les besoins nutritionnels, alimentaires et de santé.

3.3. **Echelle de Duclos, Mercadier-Pats.** Cette équivalence adulte fait partie de la famille de la forme

$$E = (N_a + c\ N_e)^s,$$

où N_a désigne le nombre d'adultes et N_e le nombre d'enfants, $0 \leq c \leq 1$, $0 \leq s \leq 1$. La valeur c=1 est choisie lorsqu'on considère que les enfants ont les mêmes besoins que les adultes. La valeur $s = 0$ est irréaliste. Cela revient à comparer des ménages sans tenir compte de la taille du ménage. Les valeurs $c = 1$ et $s = 1$ reviennent simplement à considérer que les enfants comptent autant que les adultes et que la taille du ménage suffit pour pondérer le revenu. Remarquons que cette échelle contient celle de FAO-OMS. Cependant, il faudra estimer c et s avant toute application. Cette échelle réaliste possède un intérêt théorique. On peut en effet jouer sur les valeurs prises par s et c pour voir l'impact sur la répartition spatiale de la pauvreté et de son évolution. La variabilité très grande des résultats inspire de prendre très au sérieux cette question de l'échelle d'équivalence adulte.

Il existe d'autres échelles intéressantes comme l'échelle RDA. Ces différentes échelles seront construites pour nos bases de travail puis appliquées à l'estimation de la pauvreté.

4. Application aux donnes Sénégalaises.

Les Enquêtes Sénégalaises Auprès des Ménages (ESAM) ont été menées en 1996 et 2001, en vue de sonder les conditions de vie des populations. Celle de 1996 sera appelée ici Esam I et celle de 2001, Esam II. Dans cette étude, nous exploitons les variables suivantes : la dépense totale du ménage YT, la structure de l'âge du ménage et bien sûr la région. Pour les deux enquêtes, les régions du pays étaient Dakar, Ziguinchor, Diourbel Saint-Louis, Tambacounda, Kaolack, Thiès, Fatick, Kolda, Louga.

Nous avons écarté la variable revenu non disponible pour Esam II. En effet, il a été constaté que le revenu était très difficilement quantifiable pour une société comme celle-ci en raison des fortes solidarités et de l'informel ambiant.

Pour chaque base, le fichier dénommé *personne* contient les données par individu. La variable *âge* du fichier *personne* permet le calcul des échelles. Nous avons établi les échelles Oxford, FAO-OMS, notées ici

ISSN 0825-0305

EQADUL. L'échelle de Duclos-Marcadiers-Pats, bien que programmée et calculée, n'a pas été utilisée. Pour toutes les régions, pour toute la contrée du Sénégal, la distribution de la variable individualisée

$$Y = YT/EQADUL$$

a été ajustée par la loi de Singh-Maddala. Les résultats sont remarquables et sont les suivantes :

4.1. **Estimatons des paramètres.** Il faut d'abord remarquer que l'application de la méthode d'ajustage décrite ci-haut est très peu concluante pour la variable dépense Y non transformée. L'algorithme ne converge pas ou donne des coefficients négatifs. Par contre la variable X transformée, utilisée ici, permet une estimation très précise des paramètres a, b et c. Pour l'ensemble des deux bases, pour les dix régions, les valeurs estimées de a, b et c sont respectivement de l'ordre de grandeur de 10^9, 2 et 1. On retient principalement que la valeur de a est exagérément grande face à celles de b et c qui sont de l'ordre des unités. Le tableau (1) donne les estimations des paramètres pour les deux échelles Oxford et FAO pour les deux périodes 1996 et 2001. Tandis que les tableaux (2) et (3) donnent les estimations des paramètres pour les dix régions correspondant à l'échelle FAO.

TABLE 1. Estimation des paramètres pour tout le SENEGAL pour les deux échelles

Les échelles	**a**	**b**	**c**
Echelle Oxford de 1996	7.702056e+09	1.917506e+00	1.373297e+00
Echelle Oxford de 2001	4.160192e+09	1.840678e+00	1.440699e+00
Echelle FAO de 1996	7.302428e+09	1.920449e+00	1.377922e+00
Echelle FAO de 2001	4.070514e+09	1.846035e+00	1.448704e+00

4.2. **Estimation de la fonction de répartition.** L'erreur commise

$$\left\|\widehat{F} - F\right\|_\infty = \sup_{-\infty<x<+\infty} \left|\widehat{F}(x) - F(x)\right|$$

par l'ajustage de la loi Singh-Maddala pour l'ensemble des dix régions et pour tout le sénégal ne dépasse guère 4,8%. Dans la majorité des cas, cette erreur est souvent bien plus petite, allant jusqu'à 5 pour mille pour le cas global du Sénégal pour 2001 avec les deux échelles. Les résultats sont résumés dans le tableau 4.

4.3. **Ajustage graphique.** La comparaison des graphes empiriques et théoriques, aussi bien pour les fonctions de répartition et les densités sont remarquables. Nous en donnons un échantillon : pour le cas général du Sénégal, pour la région la moins pauvre qui est Dakar et la plus pauvre qui se trouve être Kolda.

ISSN 0825-0305

TABLE 2. Estimation des paramètres par zone : échelles FAO-OMS de 1996

Les villes	**a**	**b**	**c**
Dakar	1.209516e+09	1.738288e+00	1.602333e+00
Ziguinchor	5.779392e+09	1.922398e+00	1.028830e+00
Diourbel	3.819616e+10	2.096158e+00	9.519214e-01
Saint-Louis	3.805696e+14	2.828004e+00	1.039388e+00
Tambacounda	3.229531e+19	3.698893e+00	3.523681e-01
Kaolack	3.194975e+11	2.258974e+00	6.880990e-01
Thiès	6.968884e+12	2.462923e+00	6.565597e-01
Louga	6.769311e+08	1.784471e+00	1.229630e+00
Fatick	3.393204e+10	2.070252e+00	8.694582e-01
Kolda	4.001668e+10	2.164509e+00	1.221513e+00

TABLE 3. Estimation des paramètres par zone : échelles FAO-OMS de 2001

Les villes	**a**	**b**	**c**
Dakar	2.151867e+10	1.894671e+00	1.500083e+00
Ziguinchor	1.916468e+12	2.372918e+00	7.695925e-01
Diourbel	8.675809e+12	2.462806e+00	8.492158e-01
Saint-Louis	5.818505e+11	2.251517e+00	9.603796e-01
Tambacounda	1.931342e+12	2.439416e+00	1.346718e+00
Kaolack	2.260478e+11	2.215662e+00	8.130504e-01
Thiès	3.846155e+12	2.407733e+00	1.203162e+00
Louga	7.130493e+12	2.446495e+00	1.005775e+00
Fatick	3.077499e+14	2.783397e+00	7.277847e-01
Kolda	4.010263e+14	2.792526e+00	5.387255e-01

A titre d'exemple, illustrons l'adéquation des fonctions de répartition pour Kolda (1996), voir figure (1).

4.4. **Test de Kolmogorov et Wilcox.** Ces deux tests sont très concluants. Pour le premier, les p-values sont de l'ordre de 75% pour tous les cas cités en haut. Pour le second, elles sont de l'ordre de 87%.

5. AJUSTAGE DES DENSITÉS.

Les densités s'ajustent de manière remarquablement bien. Illustrons cela pour la région de Kolda, la plus pauvre du Sénégal.

ISSN 0825-0305

FIGURE 1. Estimation de la distribution (Kolda)

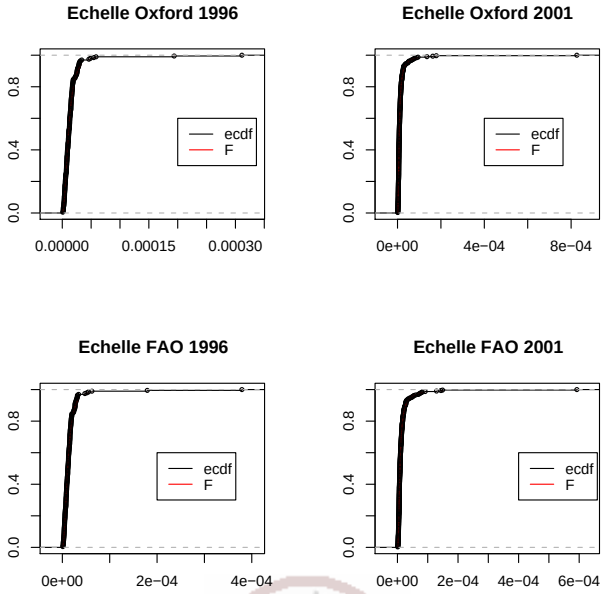
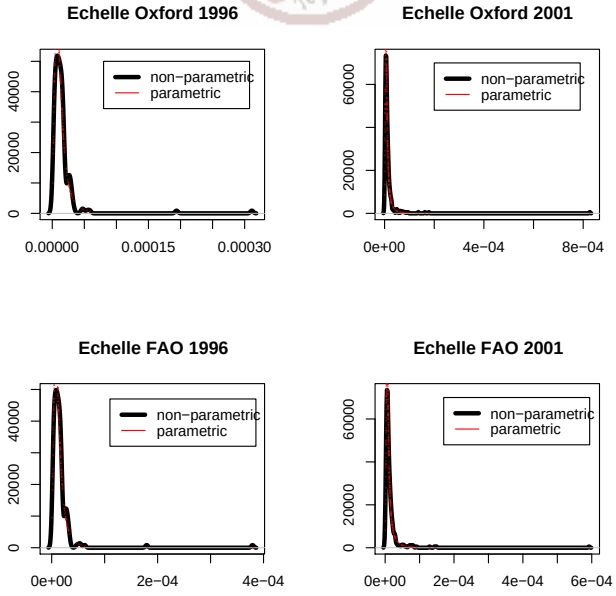


FIGURE 2. Estimation de la densité (Kolda)



ISSN 0825-0305

TABLE 4. Estimation de la déviation maximale

Les zones	deptot_ox1	deptot_ox2	deptot_fao1	deptot_fao2
Sénégal	0.010333028	0.005432949	0.010785011	0.005404051
Dakar	0.015609693	0.012592142	0.015722290	0.010670438
Ziguinchor	0.035582137	0.024623516	0.029287521	0.024265616
Diourbel	0.043002276	0.022354797	0.041931108	0.022851314
Saint-Louis	0.025212457	0.032243219	0.023569558	0.033055975
Tambacounda	0.042737043	0.023054213	0.042809574	0.024039482
Kaolack	0.026176166	0.018113849	0.026851291	0.020551484
Thiès	0.031753135	0.029308601	0.030989468	0.033671526
Louga	0.035057545	0.022522477	0.037840000	0.024527269
fatick	0.046435499	0.022838090	0.047942943	0.028714437
Kolda	0.046972388	0.027224172	0.043801580	0.027933053

6. CONCLUSION

Tous les outils mentionnés ci-haut (estimations des erreurs, graphes, tests) permettent de conclure sur la très bonne modélisation des dépenses de nos bases de données par le modèle de Singh-Maddala. Les résultats sont supérieurs à ceux de [11] avec le modèle Log-normal. Nous recommandons alors l'utilisation de la loi de Singh-Maddala pour ré-examiner les estimations des indicateurs de pauvreté et d'inégalité pour le Sénégal et les pays semblables. Dans un annexe, nous proposons quelques graphes supplémentaires.

REFERENCES

[1] Aitchison, J., Brown, J.A.C.,1957. *The Lognormal Distribution*. Cambridge University Press, London.

[2] Von Klaus Brachmann and Andreas Stich und Mark trede, köln : Evaluating Parametric Income Distribution Models. *Allgemeines Statistisches Archiv.*, 80, 317-285-298.

[3] Barrett G. and Donald, S. (2000). Statistical Inference with Generalized Gini Indices of Inequality and Poverty. Available at : (http://www.eco.utexas.edu/~donald/research/genginir.pdf)

[4] Bishop, J.A., Formby, J. P. and Zheng B.(1997). Statistical Inference and the Sen Index of Poverty. *International Economic Review*, Vol.38, n^o 2, pp.381-387.

[5] Bradbury Bruce (2003). The Welfare interpretation of consumer equivalence scales. *International Journal of Social Economics*, vol. 30, n^o 7, pp.770-787.

[6] Dia, G.(2005). Répartition Ponctuelle Aléatoire des Revenus et Estimation de l'Indice de Pauvreté. *Afrika Statistika*, Vol. 1 (1), p.47-66

ISSN 0825-0305

[7] Duclos J. Y. et Mercadier-Pats, M. (1996). *Households needs and poverty : with Application to Spain and UK*. Laval, MIMEO, Université de Laval, Canada.

[8] Hourriez J. M. and Olivier L.(1997). Niveau de vie et taille du ménage : estimation d'une échelle d'équivalence. *Economie et Statistique*, n^o 308-309-310

[9] Kakwani, N.(1993). Statistical inference in the measurement of poverty. *The Review of Economics and Statistics*, Vol. 75, n^o 4. (Nov., 1993), pp. 632-639.

[10] Lo, G. S., Sall, S.T. and Seck, C. T. (2007). The General Asymptotic Theory of Poverty Measures. *Working Paper, Social Sciences Research Network*. http://ssrn.com/abstract=1004391 Submitted.

[11] Lo, G. S. (2008). Estimation asymptotique des indices de pauvreté : modélisation continue et analyse spatio-temporelle de la pauvreté au Sénégal. *A paraître dans Journal Africain de Communication Scientifique et de Technologies*.
Disponible à : http://www.ufrsat.org/lerstad/pub/gslo_isr2008.pdf.

[12] Pollack A. R. and Wales Terence J.(1979). Welfare comparison and Equivalence scales. *American Economics Association*, vol.69.

[13] Ravallion M.(1992). *Poverty Compairisons. A Guide to Concepts and Methods*. Lsms, Working Paper, n^o 88, WorldBank.

[14] Singh.S.K and Maddala.G.S(1976) : A Function for Size Distribution of Incomes, *Econometrica*, Vol.44, 5, pp.963-970.

[15] Zheng, B.(1997). Aggregate Poverty Measures. *Journal of Economic Surveys*, 11 (2), 123-162.

[16] Weiss,Y.,1972. The risk element in occupational and educational choices. *Journal of Political Economy 80, 1203-1213*.

ISSN 0825-0305

7. APPENDICES

FIGURE 3. Estimation de la distribution (SENEGAL)

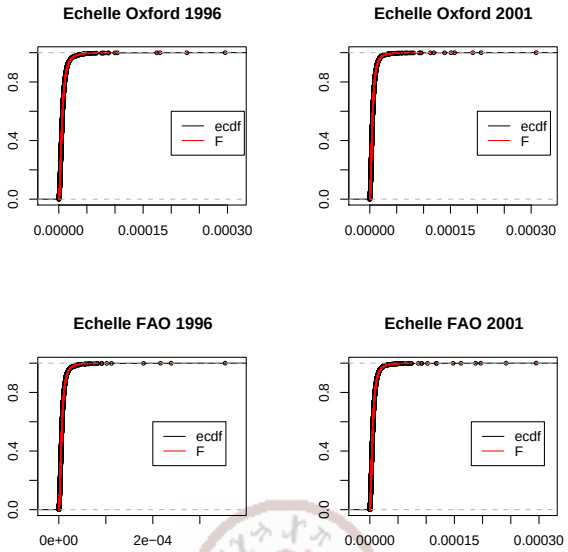
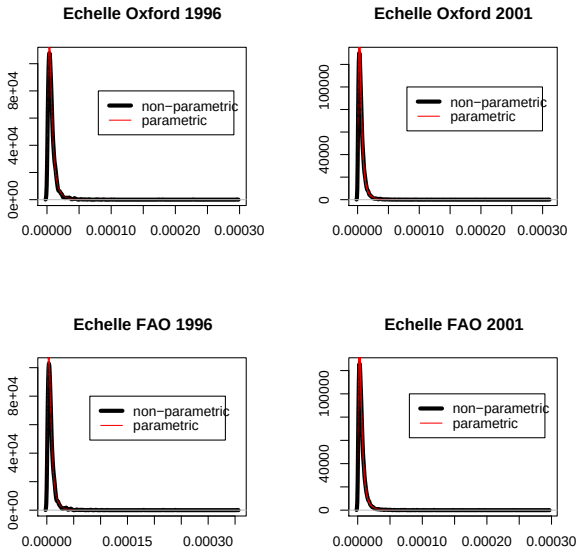


FIGURE 4. Estimation de la densité (SENEGAL)



ISSN 0825-0305

FIGURE 5. Estimation de la distribution (Dakar)

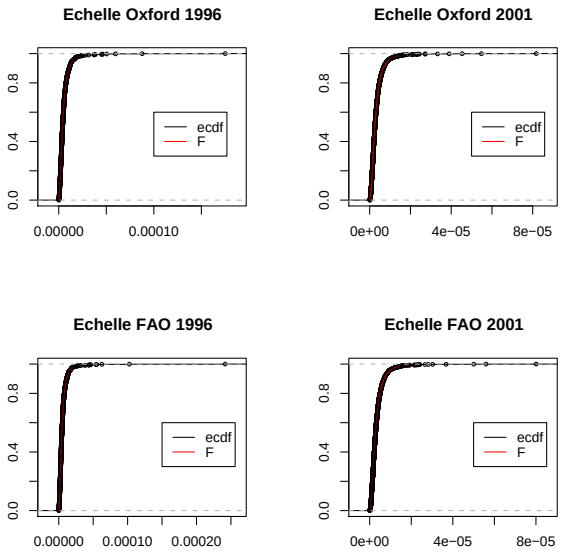
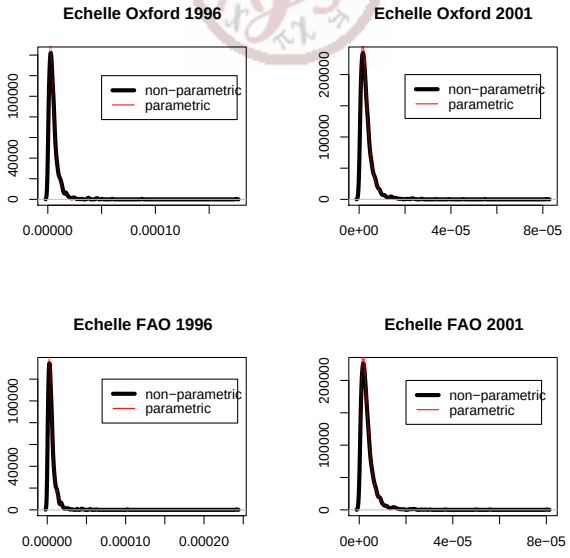


FIGURE 6. Estimation de la densité (Dakar)



LERSTAD, UNIVERSITÉ GASTON BERGER, SENEGAL
E-mail address: chaidara@ufrsat.org