LAMETA
Laboratoire Montpelliérain d'Economie Théorique et Appliquée
UMR
Unité Mixte de Recherche

DOCUMENT de RECHERCHE

«Test d'inférence statistique de l'indice multidimensionnel flou de pauvreté appliqué à l'Argentine»

María Noel PI ALPERIN,
Michel TERRAZA

DR n°2007-01

Faculté de Sciences Economiques - Espace Richter
Avenue de la Mer - Site de Richter C.S. 79606
34960 MONTPELLIER CEDEX 2
Tél: 33(0)467158495 Fax: 33(0)467158467
E-mail: lameta@lameta.univ-montp1.fr

Université Montpellier 1
agro Montpellier Ecole Nationale Supérieure Agronomique
CNRS Centre National de la Recherche Scientifique
INRA

Test d’inférence statistique de l’indice multidimensionnel flou de pauvreté appliqué à l’Argentine

María Noel PI ALPERIN♣ **Michel TERRAZA**

LAMETA, Université Montpellier I
Av. de la Mer - Site de Richter - C.S. 79606
34960 Montpellier Cedex 2 - France
Tel: 33 (0)4 67 15 84 93
♣E-mail: pialperin@lameta.univ-montp1.fr

Décembre 2006

Résumé: La loi de probabilité des indices décomposés de pauvreté, établis sur la théorie des ensembles flous, est inconnue. Comme le calcul des contributions est normalisée et bornée dans [0,1], les changements dans les indices concernent des valeurs très petites et il est difficile de conclure sur la signification des coefficients calculés. Nous proposons dans cet article une construction des intervalles de confiance utilisant la technique du bootstrap pour vérifier la signification statistique des indices de pauvreté calculés pour l’Argentine en mai 2003.

Mots clés: Bootstrap, Pauvreté multidimensionnelle, Ensembles flous, Argentine.

Classification JEL: C14, D31, D63, I32.

1. Introduction

En 1990 Cerioli et Zani développent une première méthode multidimensionnelle pour mesurer la pauvreté basée sur la théorie des ensembles flous. Dagum et Costa (2004) ont ensuite introduit les indices unidimensionnels pour mesurer l'état de privation de chaque attribut pour l'ensemble de la population, permettant de mesurer la contribution de chaque dimension à la pauvreté globale. Mussard et Pi Alperin (2006), ont pu alors proposer une décomposition synthétique qui combine à la fois le rôle des groupes d'une population et les dimensions de la pauvreté dans l'explication de la pauvreté totale. Comme les calculs des contributions sont normalisées et bornées dans [0,1], les changements dans les indices concernent des valeurs très petites. Il est donc indispensable d'en vérifier la signification statistique sachant que la loi de probabilité associée à l'échantillon des ménages sélectionnés est inconnue. Il est nécessaire de ce fait de recourir à une méthode d'inférence statistique. La technique du bootstrap retenue dans cet article permet, entre autres, de construire des intervalles de confiance de paramètres estimés et répond de ce fait au problème posé.

Nous l'employons pour analyser la signification statistique des indices de pauvreté calculés pour l'Argentine en mai 2003. A notre connaissance aucun test utilisant cette méthodologie n'a été appliqué pour les indices multidimensionnels basés sur la théorie des ensembles flous. Cet essai est novateur même si une application à un indicateur unidimensionnel a été réalisée par Biewen en 2002.

Dans ce travail nous présentons brièvement les principales notions de l'approche multidimensionnelle de la pauvreté basée sur la théorie des ensembles flous et la décomposition synthétique de l'indice de pauvreté multidimensionnel (Section 2). Nous exposons la méthode du bootstrap retenue (Section 3) et nous l'appliquons à l'indice de pauvreté argentin (Section 4).

2. L'indice multidimensionnel flou de pauvreté

Résumons brièvement les principales notions concernant l'approche multidimensionnelle de la pauvreté basée sur la théorie des ensembles flous [*Cf.* Dagum et Costa (2004)].

Soit $A = \{a_1, ..., a_i, …, a_n\}$ l'ensemble des ménages situés dans un espace économique et soit $X = \{X_1, …, X_j, …, X_m\}$ un vecteur d'ordre m des attributs socio-économiques sélectionnés pour étudier l'état de pauvreté de A. Appelons B un sous-ensemble flou de A tel que chaque $a_i \in B$ présente un degré de privation dans au moins un des m attributs inclus en X.

La fonction d'appartenance au sous-ensemble flou B du i-ème ménage ($i = 1,…, n$) par rapport au j-ème attribut ($j = 1,…, m$) est définie de la manière suivante :

$$x_{ij} : \mu_B\left(X_j\left(a_i\right)\right),\ 0 \le x_{ij} \le 1. \tag{1}$$

Avec :

- $x_{ij} = 1$, si le i-ème ménage n'a pas le j-ème attribut ;
- $x_{ij} = 0$, si le i-ème ménage possède le j-ème attribut ;
- $0 < x_{ij} < 1$, si le i-ème ménage a le j-ème attribut avec une intensité comprise entre (0,1).

La fonction d'appartenance du i-ème ménage au sous-ensemble flou B, mesure le ratio de pauvreté multidimensionnel du ménage a_i où w_j est le poids attaché au j-ème attribut. Cette fonction peut être définie comme le poids moyen de x_{ij} :

$$\mu_B\left(a_i\right) = \sum_{j=1}^{m} x_{ij} w_j \Big/ \sum_{j=1}^{m} w_j, \qquad 0 \le \mu_B\left(a_i\right) \le 1. \tag{2}$$

Plus précisément:

- $\mu_B(a_i) = 0$, si a_i possède les m attributs ;
- $\mu_B(a_i) = 1$, si a_i est totalement dépourvu des m attributs ;
- $0 < \mu_B(a_i) < 1$, si a_i est partiellement ou totalement privé de quelques attributs mais pas totalement démuni de tous les attributs.

Le poids w_j retenu est celui proposé par Betti et Verma (1999). Il est fondé sur deux principes : (*i*) il représente l'intensité de privation liée à l'attribut X_j ; sa valeur est une fonction inverse du degré de privation de cet attribut pour la population des ménages, (*ii*) il est construit de manière à limiter l'influence des attributs qui sont hautement corrélés :

$$w_j = w_j^a * w_j^b, \tag{3}$$

où w_j^a dépend de la distribution du j-ème attribut dans la population et w_j^b dépend de la corrélation entre X_j et les autres attributs.

Plus précisément, w_j^a est déterminé par la dispersion du *j*-ème attribut dans la population, il est représenté par le coefficient de variation :

$$w_j^a = \left[\sum_{i=1}^{n} \left(x_{ij} - \bar{x}_j \right)^2 \Big/ n \right]^{1/2} \Bigg/ \left(\sum_{i=1}^{n} x_{ij} \Big/ n \right). \tag{3'}$$

La composante w_j^b est établie à partir de la formule :

$$w_j^b = \left[\frac{1}{1 + \sum_{j'=1}^{m} \rho_{j,j'} \Big/ \rho_{j,j'} < \rho_H} \right] * \left[\frac{1}{\sum_{j'=1}^{m} \rho_{j,j'} \Big/ \rho_{j,j'} \geq \rho_H} \right], \tag{3''}$$

où $\rho_{j,j'}$ représente le niveau de corrélation calculé à partir du coefficient de corrélation linéaire simple entre deux attributs. Dans l'équation, la somme est effectuée sur tous les attributs possédant un niveau de corrélation avec le *j*-ième attribut inférieur ou supérieur à une valeur ρ_H qui correspond à l'écart le plus important entre deux coefficients de corrélation ordonnés.

La théorie des ensembles flous permet aussi de déterminer un indice unidimensionnel représentant le degré de privation du *j*-ème attribut pour la population des *n* ménages :

$$\mu_B(X_j) = \sum_{i=1}^{n} x_{ij}\, g(a_i) \Big/ \sum_{i=1}^{n} g(a_i). \tag{4}$$

où $g(a_i) \Big/ \sum_{i=1}^{n} g(a_i)$ est la fréquence relative associée à l'observation de l'échantillon a_i de la population.

L'indice de pauvreté multidimensionnel de la population *A* peut être défini comme une moyenne pondérée de $\mu_B(a_i)$ donnée par (2), et aussi comme une moyenne pondérée des indices unidimensionnels pour chaque attribut [$\mu_B(X_j)$]:

$$\mu_B = \sum_{i=1}^{n} \mu_B(a_i) g(a_i) \Big/ \sum_{i=1}^{n} g(a_i) = \sum_{j=1}^{m} \mu_B(X_j) w_j \Big/ \sum_{j=1}^{m} w_j . \tag{5}$$

L'analyse des résultats obtenus en (4), pour $j = 1, \ldots, m$, donne la possibilité aux décideurs d'identifier les caractéristiques de la pauvreté et d'intervenir structurellement pour la réduire.

2.1. Décomposition de l'indice multidimensionnel de pauvreté

Nous présentons trois types de décomposition : (*i*) la décomposition en groupes de population ; (*ii*) la décomposition en attributs ; et, finalement (*iii*) la décomposition multidimensionnelle. L'indice de pauvreté multidimensionnel basé sur la théorie des ensembles flous satisfait ces trois types de décomposition[1].

Décomposition en groupes de population

Une autre manière d'évaluer la structure de la pauvreté est de proposer une décomposition en groupes de population. Divisons l'espace économique en k groupes, S_k ,de taille n_k (k = 1,…, s). L'intensité de la pauvreté du i-ème ménage de S_k est donnée par :

$$\mu_B(a_i^k) = \sum_{j=1}^{m} x_{ij}^k \, w_j \Big/ \sum_{j=1}^{m} w_j \ , \tag{6}$$

où x_{ij}^k est la fonction d'appartenance au sous-ensemble flou B du i-ème ménage (i = 1,…, n_k) de S_k par rapport au j-ème attribut (j = 1,…, m)[2]. L'indice de pauvreté multidimensionnel associé au groupe S_k est alors défini de la manière suivante :

$$\mu_B^k = \sum_{i=1}^{n_k} \mu_B\left(a_i^k\right) g\left(a_i^k\right) \Big/ \sum_{i=1}^{n_k} g\left(a_i^k\right). \tag{7}$$

avec $g\left(a_i^k\right) / \sum_{i=1}^{n_k} g\left(a_i^k\right)$ la fréquence relative représentée par l'observation de l'échantillon a_i^k de S_k.

D'après (7), l'indice de pauvreté global peut être calculé comme une moyenne pondérée du niveau de pauvreté à l'intérieur de chaque groupe :

$$\mu_B = \sum_{k=1}^{s} \sum_{i=1}^{n_k} \mu_B\left(a_i^k\right) g\left(a_i^k\right) \Big/ \sum_{i=1}^{n} g\left(a_i\right). \tag{8}$$

Il est aussi possible de diviser chaque groupe S_k, (k = 1,…, s), en b sous-groupes S_{bk} (b = 1,…, p) de taille n_{bk}. L'intensité de la pauvreté du i-ème ménage de S_{bk} est la suivante:

[1] Confer Mussard et Pi Alperin (2006).

[2] Pour satisfaire la propriété de décomposabilité en groupes de population, une condition nécessaire est d'affecter un même poids pour chaque dimension et pour chaque groupe après la décomposition. Un poids différent peut être calculé mais il ne permet pas la comparaison entre les groupes.

$$\mu_B\left(a_i^{bk}\right)=\sum_{j=1}^{m} x_{ij}^{bk}\ w_j \Big/ \sum_{j=1}^{m} w_j\ , \tag{9}$$

où x_{ij}^{kb} est la fonction d'appartenance au sous-ensemble flou B du i-ème ménage (i = 1,…, n_{bk}) de S_{bk} par rapport au j-ème attribut (j = 1,…, m). L'indice de pauvreté multidimensionnel de chaque sous-groupe est:

$$\mu_B^{bk} = \sum_{i=1}^{n_{bk}} \mu_B\left(a_i^{bk}\right) g\left(a_i^{bk}\right) \Big/ \sum_{i=1}^{n_{bk}} g\left(a_i^{bk}\right). \tag{10}$$

avec $g\left(a_i^{bk}\right)/\sum_{i=1}^{n_k} g\left(a_i^{bk}\right)$ la fréquence relative représentée par l'observation de l'échantillon a_i^{bk} de S_{bk}..

L'indice de pauvreté flou global peut être calculé comme une moyenne pondérée de l'intensité de la pauvreté existant dans chaque groupe de la deuxième partition:

$$\mu_B = \sum_{b=1}^{p} \sum_{k=1}^{s} \sum_{i=1}^{n_{bk}} \mu_B\left(a_i^{bk}\right) g\left(a_i^{bk}\right) \Big/ \sum_{i=1}^{n} g\left(a_i\right). \tag{11}$$

Cette décomposition en multi-niveau permet aux décideurs de réduire la pauvreté en identifiant les groupes les plus affectés (régions, sexe, etc.).

Décomposition par attributs : Dagum et Costa (2004)

Dagum et Costa (2004) ont introduit la décomposition par attribut en démontrant qu'il est possible de calculer un indice unidimensionnel pour chaque attribut pour la population de ménages [*Cf.* équation (4)].

De la même manière il est possible de calculer l'indice unidimensionnel de pauvreté du j-ème attribut pour le k-ème groupe :

$$\mu_B\left(X_j^k\right)=\sum_{i=1}^{n_k} x_{ij}^k g\left(a_i^k\right) \Big/ \sum_{i=1}^{n_k} g\left(a_i^k\right) \tag{12}$$

Si deux partitions sont prises en compte nous pouvons aussi calculer l'indice de pauvreté unidimensionnel du j-ème attribut de S_{bk} :

$$\mu_B\left(X_j^{bk}\right)=\sum_{i=1}^{n_{kb}} x_{ij}^{bk} g\left(a_i^{bk}\right) \Big/ \sum_{i=1}^{n_{kb}} g\left(a_i^{bk}\right). \tag{13}$$

Ce deuxième type de décomposition enrichit l'information sur les différentes caractéristiques de la pauvreté, nécessaires aux décideurs pour améliorer les politiques socioéconomiques de réduction de l'état d'exclusion sociale.

Décomposition multidimensionnelle

En 1998, Chakravarty, Mukherjee et Ranade ont introduit une classe d'indices de pauvreté simultanément décomposables par attribut et par groupe. Comme il a été démontré [*Cf.* Mussard et Pi Alperin (2006)], l'indice de pauvreté flou satisfait cette propriété.

Il est possible de définir l'indice de pauvreté comme une fonction pondérée des indices unidimensionnels du *j*-ème attribut dans le *k*-ème groupe :

$$\mu_B = \sum_{k=1}^{s} \sum_{j=1}^{m} \mu_B\left(X_j^k\right) w_j \Big/ \sum_{j=1}^{m} w_j \ . \tag{14}$$

Considérons deux partitions de la population et les indices de pauvreté unidimensionnels du *j*-ème attribut de S_{bk}, l'indice de pauvreté multidimensionnel peut s'écrire:

$$\mu_B = \sum_{k=1}^{s} \sum_{b=1}^{p} \sum_{j=1}^{m} \mu_B\left(X_j^{bk}\right) w_j \Big/ \sum_{j=1}^{m} w_j \ . \tag{15}$$

Cette décomposition simultanée donne toutes les combinaisons « attribut/groupe » et « attribut/sous-groupe » qui contribuent à l'état de pauvreté globale. En définitive, on trouve toute l'information nécessaire pour réduire l'intensité de la pauvreté.

3. Le principe du bootstrap

L'analyse et la mesure de la pauvreté sont réalisées en se basant sur des échantillons observés qui, par construction, sont soumis à des erreurs d'échantillonnage. L'inférence statistique traite ce problème. Elle permet de déterminer si les mesures estimées de la pauvreté représentent les vrais paramètres de la population.

Considérons un échantillon aléatoire $X = (x_1,...,x_n)$ à partir d'une distribution de probabilité inconnue *F*. Nous voulons estimer le paramètre $\xi = t(F)$ à partir de X ; et soit $\hat{\xi} = s(X)$ un estimateur. L'analyse de la précision de l'estimateur peut être réalisée en utilisant la méthode du bootstrap introduite par Efron en 1979. C'est une technique statistique basée sur le rééchantillonnage avec remise. Chaque échantillon bootstrap, X^*, est un échantillon de taille *n*, aléatoire et indépendant des autres. Il est construit à partir d'une distribution empirique $\hat{F}$ où chaque observation x_i, $i = 1,...,n$, possède la probabilité $1/n$ d'appartenance. Les éléments de l'échantillon bootstrap sont les mêmes que ceux de l'échantillon original[3]. A chaque échantillon bootstrap correspond une réplication de $\hat{\xi} : \hat{\xi}^* = s(X^*)$. A partir de ces échantillons, et de ses réplications, il est possible de construire différents types d'intervalles de confiance pour tester la signification statistique des paramètres estimés.

La méthode bootstrap et l'indice multidimensionnel de pauvreté

Pour appliquer la méthode du bootstrap on considère une densité de probabilité *F* dont la loi est inconnue et un échantillon de *n* ménages choisis aléatoirement. L'échantillon $\hat{F}$ est construit en affectant la probabilité $1/n$ à chaque observation, a_i, qui est ainsi liée à un vecteur de *m* variables $\{a_i = (x_{1i}, x_{2i},..., x_{mi})\}, \forall i = 1,...,n$.

A partir de $\hat{F}$ nous prélevons une série de *R* échantillons aléatoires avec remise des *n* observations (ménages) dans l'échantillon initial. Ces échantillons successifs sont notés :

$$a_i^{*r} = (x_{1i},..., x_{ij},...x_{mi}), \ \forall i = 1,...,n \text{ et } \forall r = 1,...R .$$

où chaque ménage conserve son propre vecteur des variables qui lui est associé dans chaque échantillonnage.

[3] Dans chaque échantillon bootstrap il y aura des observations qui apparaissent une fois, d'autres deux fois, d'autres trois fois, etc., et d'autres zéro fois.

Les intervalles de confiance[4]

Il est possible de mesurer *R* multi-décompositions de l'indice de pauvreté multidimensionnel en utilisant les a_i^{*r} ménages, appelés réplications bootstraps, $\hat{\mu}^* = f\left(a_i^*\left(x_{1i},...,x_{mi}\right)\right)$. Soit $\hat{\psi}$ la distribution cumulative de $\hat{\mu}^*$. Nous avons retenu la méthode des percentiles simples car contrairement à la méthode de l'erreur-standard (*Cf.* Annexe A.1), il n'est pas nécessaire que la distribution d'échantillonnage du paramètre étudié obéisse à une loi normale. Le nombre de rééchantillonnages *R* doit être plus élevé que dans le cas de la méthode de l'erreur standard car il faut un plus grand nombre d'observations pour estimer, avec une précision suffisante, un percentile que pour estimer un écart-type. L'intervalle basé sur la méthode de percentiles simples est défini par les $100.\alpha/2$-ième et $100.(1-\alpha/2)$-ième percentiles de $\hat{\psi}$:

$$\left(\hat{\mu}_{inf}\,;\hat{\mu}_{sup}\right) = \left(\hat{\psi}^{-1}(\alpha/2);\hat{\psi}^{-1}(1-\alpha/2)\right)$$

Par définition $\hat{\psi}^{-1}(\alpha/2) = \hat{\mu}^{*(\alpha/2)}$, c'est-à-dire le 100.α/2-ième percentile de la distribution bootstrap. Alors:

$$\left(\hat{\mu}_{inf}\,;\hat{\mu}_{sup}\right) = \left(\hat{\mu}^{*(\alpha/2)};\hat{\mu}^{*(1-\alpha/2)}\right)$$

représente l'intervalle de confiance construit pour déterminer la signification statistique des paramètres obtenus à partir des données de l'échantillon initial[5].

5. Application à l'Argentine

L'indice multidimensionnel flou de pauvreté est calculé pour l'Argentine à partir de la base de données « Encuesta Permanent de Hogares » (EPH) de l'INDEC (Institut National de Statistique et de Recensement de l'Argentine). Elle contient 16 924 ménages et concerne le mois de mai 2003.

L'information disponible dans l'EPH permet de sélectioner les attributs suivants: revenu équivalent disponible du ménage[6] (X_1) ; taille du foyer (X_2) ; installation d'eau dans le foyer

[4] Voir l'annexe pour une revue des différents intervalles de confiance.

[5] Où $\hat{\mu}^*$ peut représenter les indices multidimensionnels pour les différents groupes de population aussi bien que les indices unidimensionnels pour l'ensemble de la population ou pour les différents groupes après la décomposition.

(X_3) ; caractéristiques des sanitaires (X_4) ; matériaux de construction du logement (X_5) ; statut actuel d'occupation du logement (X_6) ; indice du nombre de personnes ayant un salaire/taille du foyer (X_7) ; sexe, âge et catégorie socioprofessionnelle du chef du ménage (X_8) ; pensions et autres transferts perçus par les personnes possèdent un emploi (X_9) ; niveau d'éducation maximal atteint par le chef du ménage (X_{10}).

La décomposition standard

L'indice multidimensionnel de pauvreté (IMP) calculé indique que 10,19% des ménages sont structurellement pauvres. Les indices unidimensionnels de pauvreté, (IUP), résultat de la décomposition par attribut de Dagum et Costa (2004), permettent d'identifier les principales caractéristiques de ces ménages pauvres. Ils montrent que les pensions et autres transferts (X_9), le niveau d'éducation du chef du ménage (X_{10}), et le niveau de revenu disponible (X_1) sont les attributs les plus significatifs (*Cf.* Tableau 1).

Tableau 1 : IUP pour l'ensemble de la population

Attributs	$\mu_B(X_j)$	**Intervalles de confiance bootstrap**
Revenu équivalent disponible (X_1)	**0,3993**	[0,3885 - 0,4119]
Taille du foyer (X_2)	0,091	[0,0846 - 0,0971]
Installation d'eau (X_3)	0,0137	[0,0107 - 0,0172]
Caractéristiques des sanitaires (X_4)	0,0211	[0,0181 - 0,0245]
Matériaux de construction (X_5)	0,0165	[0,0144 - 0,0185]
Statut actuel d'occupation (X_6)	0,1704	[0,1630 - 0,1787]
Indice: n° salariés / taille du foyer (X_7)	0,2164	[0,2059 - 0,2267]
Catégorie socioprofessionnelle (X_8)	0,1888	[0,1802 - 0,1979]
Pensions et autres transferts (X_9)	**0,7692**	[0,7578 - 0,7808]
Niveau d'éducation (X_{10})	**0,5887**	[0,5770 - 0,5999]
Total	**0,1019**	

Source : Elaboration personnelle en utilisant l' « EPH », mai 2003

La méthode des percentiles simples, retenue pour calculer les intervalles de confiance de nos coefficients, indique que tous les coefficients estimés sont significativement différents de zéro (*Cf.* Tableau 1) puisqu'ils se trouvent à l'intérieur de l'intervalle de confiance calculé.

[6] Ajusté pour sa valeur correspondante d'échelle d'équivalence. Voir Dagum et Costa (2004) pour plus de détailles sur cette méthode.

La décomposition par groupe de population

La décomposition par groupes de population concerne ici le sexe du chef du ménage. Le Tableau 2 fournit : (*i*) les indices multidimensionnels de pauvreté pour chaque groupe après la décomposition ; et (*ii*) les intervalles de confiance bootstrap construits pour étudier leur signification statistique.

Tableau 2 : IMP pour les groupes après décomposition

Décomposition		μ_B^k	**Intervalles de confiance bootstrap**
Sexe	**Homme**	0,0999	[0,0975-0,1022]
	Femme	0,1067	[0,1012-0,1121]

Source : Elaboration personnelle en utilisant l' « EPH », mai 2003

La décomposition par sexe montre que le groupe des femmes est plus pauvre que le groupe des hommes, respectivement 10,67% et 9,99% des ménages sont structurellement pauvres. Les intervalles de confiance construits montrent la signification statistique des indices multidimensionnels de pauvreté mesurés pour chaque groupe.

Tableau 3 : IUP par attribut et par groupe

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
H	0,3727 [0,3596 0,3869]	0,1051 [0,0970 0,1133]	0,0137 [0,0101 0,0176]	0,0220 [0,0183 0,0262]	0,0176 [0,0151 0,0202]	0,1662 [0,1576 0,1753]	0,2214 [0,2091 0,2347]	0,1610 [0,1508 0,1717]	0,7364 [0,7230 0,7511]	0,5726 [0,5586 0,5861]
F	0,4614 [0,4392 0,4826]	0,0579 [0,0495 0,0672]	0,0140 [0,0084 0,0207]	0,0191 [0,0142 0,0252]	0,0139 [0,0107 0,0175]	0,1801 [0,1660 0,1955]	0,2045 [0,1842 0,2245]	0,2537 [0,2379 0,2706]	0,8459 [0,8262 0,8634]	0,6263 [0,6070 0,6450]

Source : Elaboration personnelle en utilisant l' « EPH », mai 2003

Le Tableau 3 rassemble deux types d'information : les indices unidimensionnels de pauvreté pour chaque groupe, et les limites supérieures et inférieures des intervalles de confiance créées pour chacun de ces indices unidimensionnels. Nous observons différentes intensités pour les caractéristiques de pauvreté pour chaque groupe. Les pensions et autres transferts (X_9), le niveau d'éducation du chef du ménage (X_{10}) et le niveau de revenu disponible (X_1) sont les trois principales caractéristiques aux intensités différentes de la pauvreté dans chaque groupe. D’autres dimensions ressortent comme le numéro des salariés / taille foyer et la catégorie socioprofessionnelle respectivement pour le groupe des hommes et le groupe des femmes.

En observant les intervalles de confiance nous constatons que tous les couples « attribut *j* / groupe *k* » (*Cf.* Tableau 3) sont significatifs.

La décomposition en multi-niveau

Nous avons réalisé une partition en multi-niveau de la population par sexe et par âge du chef du ménage (*Cf.* Tableau 4). Cette deuxième partition montre que les sous-groupes des hommes de tous les âges sont plus pauvres que les sous-groupes de femmes.

Tableau 4 : IMP par sous-groupe de population

Décomposition		μ_B^{bk}	**Intervalles de confiance bootstrap**
Homme	**< 25**	0,1097	[0,1097-0,1618]
	25 - 45	**0,1022**	**[0,1177-0,1295]**
	46 - 65	**0,0997**	**[0,1149-0,1283]**
	> 65	**0,1016**	**[0,1172-0,1453]**
Femme	**< 25**	**0,0928**	**[0,0554-0,0917]**
	25 - 45	**0,0971**	**[0,0627-0,0759]**
	46 - 65	**0,0978**	**[0,0610-0,0744]**
	> 65	**0,0968**	**[0,0583-0,0726]**

Source : Elaboration personnelle en utilisant l’ « EPH », mai 2003

Cependant, si on observe les intervalles de confiance calculés, nous remarquons qu’à l’exception du sous-groupe des hommes de moins de 25 ans, les valeurs calculées pour les indices multidimensionnels par sous-groupe ne sont pas significatives. Ce résultat peut être expliqué par le fait qu’après cette deuxième partition de la population, la taille de l’échantillon de chaque sous-groupe est trop petite et n’est pas nécessairement représentative de la vraie population.

Tableau 5 : IUP par attribut et par sous-groupe de population

		X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
		0,4424	0,0656	0,0338	0,0207	0,0121	0,1849	0,2268	0,2142	0,7724	0,6176
	< 25	[0,3342	[0,0565	[0,0019	[0,0078	[0,0070	[0,1252	[0,1557	[0,1400	[0,6981	[0,5271
		0,4717]	0,1312]	0,0374]	0,0443]	0,0309]	0,2173]	0,2820]	0,2428]	0,8387]	0,6608]
		0,3950	0,0853	0,0149	**0,0273**	0,0178	0,1634	0,2097	0,1940	0,7727	0,5899
	25 - 45	[0,3782	[0,0801	[0,0089	**[0,0160**	[0,0131	[0,1568	[0,1982	[0,1733	[0,7484	[0,5698
H		0,4197]	0,1025]	0,0204]	**0,0271]**	0,0205]	0,1842]	0,2350]	0,2049]	0,7892]	0,6079]
		0,3837	0,0913	0,0130	0,0203	0,0164	0,1709	0,2093	0,1805	0,7605	0,5798
	46 - 65	[0,3778	[0,0785	[0,0080	[0,0154	[0,0128	[0,1557	[0,1954	[0,1708	[0,7482	[0,5667
		0,4243]	0,1043]	0,0209]	0,0280]	0,0210]	0,1857]	0,2406]	0,2064]	0,7919]	0,6090]
		0,3995	0,1122	0,0140	0,0221	0,0167	0,1614	**0,1734**	0,2079	0,8030	0,5883
	> 65	[0,3632	[0,0712	[0,0047	[0,0114	[0,0105	[0,1445	**[0,1803**	[0,1618	[0,7315	[0,5523
		0,4360]	0,1129]	0,0244]	0,0318]	0,0236]	0,1934]	**0,2511]**	0,2175]	0,8059]	0,6247]
		0,3392	0,1039	0,0022	0,0118	0,0324	0,1325	0,1687	0,1944	0,7471	0,5722
	< 25	[0,3221	[0,0496	[0,0011	[0,0056	[0,0055	[0,1179	[0,1441	[0,1246	[0,6811	[0,5106
		0,4858]	0,1438]	0,0406]	0,0511]	0,0350]	0,2346]	0,2958]	0,2526]	0,8479]	0,6651]
		0,3661	0,0758	0,0180	0,0142	0,0179	0,1623	0,2050	0,1772	0,7511	0,5836
	25 - 45	[0,3629	[0,0715	[0,0046	[0,0122	[0,0107	[0,1453	[0,1816	[0,1579	[0,7293	[0,5565
F		0,4350]	0,1122]	0,0260]	0,0327]	0,0240]	0,1966]	0,2506]	0,2181]	0,8089]	0,6271]
		0,3986	0,0803	0,0112	0,0149	0,0171	0,1524	0,2186	0,1843	0,7587	0,5884
	46 - 65	[0,3668	[0,0722	[0,0055	[0,0127	[0,0105	[0,1465	[0,1839	[0,1632	[0,7315	[0,5516
		0,4362]	0,1122]	0,0260]	0,0325]	0,0235]	0,1961]	0,2518]	0,2184]	0,8031]	0,6224]
		0,4238	0,0996	0,0230	0,0249	0,0193	0,1518	0,2261	0,1883	0,7796	0,6020
	> 65	[0,3594	[0,0681	[0,0045	[0,0113	[0,0100	[0,1428	[0,1785	[0,1572	[0,7274	[0,5448
		0,4392]	0,1148]	0,0254]	0,0322]	0,0246]	0,1975]	0,2552]	0,2220]	0,8097]	0,6270]

Source : Elaboration personnelle en utilisant l' « EPH », mai 2003

La décomposition multidimensionnelle en multi-niveau fournit les valeurs des indices unidimensionnels de pauvreté pour chaque sous-groupe (*Cf.* Tableau 5). Nous pouvons constater que les pensions et autres transferts, le niveau d'éducation du chef du ménage et le revenu équivalent disponible, sont les trois principales dimensions aux intensités différentes d'exclusion sociale dans tous les sous-groupes. Parmi ces résultats deux couples « attributs/sous-groupe » ne sont pas significatifs (« X_4/Hommes âgés entre 25 et 45 ans » et « X_7/Hommes de plus de 65 ans ») ce qui empêche toute conclusion définitive.

Ces derniers résultats montrent bien l'intérêt de vérifier la signification des coefficients de la décomposition à l'aide des méthodes d'inférence statistique comme celle proposée dans ce travail. L'absence de ce type d'étude peut amener à des conclusions erronées surtout lorsque les tailles des échantillons sont insuffisantes. Cette première recherche reste cependant partielle sur le plan méthodologique. Des améliorations utilisant des méthodes Bootstrap plus sophistiquées sont à envisager.

Bibliographie

Athanasopoulos G. et Vahid F. (2003), « Statistical Inference on Changes in Income Inequality in Australia », *The Economic Record,* Vol. 79 (247), p. 412-424.

Biewen M. (2002), « Bootstrap inference for inequality, mobility and poverty measurement », *Journal of Econometrics*, 108 (2002) 317-342.

Efron B. and Tibshirani R. (1993), *Introduction to the Bootstrap*, New York, Springer-Verlag.

Efron, B. (1979), « Bootstrap Methods: Another Look at the Jackknife», *Annals of Statistics*, **7**, 1-26

Efron B. (1995), « *Le Bootstrap et ses Applications. Discrimination et Régression* ». C.I.S.I.A.

Hall P. (1994), « Methodology and Theory for the Bootstrap », Engle and McFaden (eds.), *Handbook of Applied Econometrics*, North Holland.

Mussard S. et Pi Alperin M.N. (2006), « Multidimensional Decomposition of Poverty: A Fuzzy Set Approach ». Working Paper : 05-06, GREDI, Université de Sherbrooke, Québec, Canada.

Palm R. (2002), « Utilisation du Bootstrap pour les problèmes statistiques liés à l'estimation des paramètres », *Biotechnol. Agron. Soc. Environ.* 2002 **6** (3), 143-153.

Annexe : Méthodes Bootstrap

Dans cette annexe nous présentons brièvement quelques méthodes de construction des intervalles de confiance pour tester la signification de paramètres estimés à l'aide de la méthode bootstrap.

A.1. La méthode de l'erreur-standard

Efron et Tibshiriani (1993) proposent pour l'estimation de l'erreur standard d'un paramètre l'algorithme bootstrap suivant,

$$s\hat{e}_B = \left\{ \sum_{r=1}^{B} \left[\hat{\xi}^*(b) - \hat{\xi}^*(.) \right]^2 \Big/ (B-1) \right\}^{1/2}$$

où $\hat{\xi}^{*}(.) = \sum_{r=1}^{B} \hat{\xi}^{*}(b) \Big/ B$ et $b = 1,\ldots,B$

Nous pouvons noter que la déviation standard empirique s'approche de celle de la population quand le nombre de réplications bootstraps *B* tend vers l'infini.

La méthode de l'erreur-standard est construite de la manière suivante :

$$\hat{\xi} \pm \mu_{1-\alpha/2} s\hat{e}_{\hat{\xi}^{*}},$$

où $\mu_{1-\alpha/2}$ est le $100.1-\alpha/2$-ième percentile de la distribution normale réduite ; $1-\alpha$ étant le degré de confiance retenu ; et $s\hat{e}_{\hat{\xi}^{*}}$ étant l'erreur-standard des *B* réplications qui a été définie précédemment.

Pour que cette approche soit satisfaisante, il est nécessaire que la distribution d'échantillonnage du paramètre étudié soit approximativement normale, que l'estimateur soit non biaisé, et que $s\hat{e}_{\hat{\xi}^{*}}$ soit une bonne estimation de l'erreur-standard de la distribution du paramètre.

A.2. La méthode des percentiles corrigés pour le biais

L'intervalle de confiance défini par la méthode de percentiles simples corrigés pour le biais avec un degré de signification de α est défini de la manière suivante:

$$CI_{\alpha} : \left(\hat{\xi}_{\inf}, \hat{\xi}_{\sup}\right) = \left(\hat{\xi}^{*}(\alpha_1); \hat{\xi}^{*}(\alpha_2)\right),$$

avec $\alpha_1 = \Phi\left(2\hat{z}_0 + z^{\left(\alpha/2\right)}\right)$

$$\alpha_2 = \Phi\left(2\hat{z}_0 + z^{\left(1-\alpha/2\right)}\right),$$

où Φ est la fonction de répartition de la distribution normale, $z^{(\alpha/2)}$ et $z^{(1-\alpha/2)}$ sont les $100.\alpha/2$-ième et $100.1-\alpha/2$-ième percentiles respectivement de la distribution normale.

La valeur de $\hat{z}_0$, représentant la correction du biais, peut être calculée en déterminant le nombre de réplications bootstraps ($\hat{\xi}^{*r}$) possédant une valeur inférieur à la valeur estimée initialement ($\hat{\xi}$):

$$\hat{z}_0 = \Phi^{-1}\left\{\frac{\#\left\{\xi^{*}(r) < \hat{\xi}\right\}}{B}\right\}$$

où Φ^{-1} représente la fonction inverse de la densité cumulée de la distribution normale.

A.3. La méthode des percentiles avec correction pour le biais et l'accélération

La méthode précédente peut être étendue de manière à tenir compte d'un éventuel changement de l'erreur-standard de $\hat{\xi}$ lorsque ξ varie. Cet intervalle de confiance est nommé BC_α. Les limites de confiance sont données par les percentiles de la distribution bootstrap. Dans cette méthode, les percentiles dépendent de deux paramètres $\hat{z}_0$ et $\hat{a}$ qui sont respectivement appelés correction de biais et accélération.

L'intervalle BC_α avec un degré de signification de $(1-\alpha)$ est donné par l'équation suivante :

$$BC_\alpha := \left(\xi^*_{(\alpha_1)}, \xi^*_{(\alpha_{21})}\right),$$

$$\text{où } \alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha/2)}}{1-\hat{a}\left(\hat{z}_0 + z^{(\alpha/2)}\right)}\right)$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha/2)}}{1-\hat{a}\left(\hat{z}_0 + z^{(1-\alpha/2)}\right)}\right)$$

et où Φ est la fonction de répartition de la distribution normale, $z^{(\alpha/2)}$ et $z^{(1-\alpha/2)}$ sont les $100.\alpha/2$-ième et $100.1-\alpha/2$-ième respectivement de la distribution normale.

La valeur $\hat{z}_0$ est calculée de la même manière que précédemment (A.I.3). En revanche, la valeur $\hat{a}$ est définie comme une valeur *jackknife*[7] de la statistique ξ^*. Soit $x_{(i)}$ l'échantillon brut moins la *i*-ème observation, x_i, et posons $\hat{\xi}_{(.)} = \sum_{i=1}^{n} \hat{\xi}_{(i)} \Big/ n$, alors :

$$\hat{a} = \sum_{i=1}^{n}\left(\xi_{(.)} - \hat{\xi}_{(i)}\right)^3 \Big/ 6\left\{\sum_{i=1}^{n}\left(\hat{\xi}_{(.)} - \hat{\xi}_{(i)}\right)^2\right\}^{3/2}$$

nous retrouvons une expression qui permet de corriger l'accélération.

A.4. La méthode de percentiles proposé par Hall (1994)

[7] La méthode *jackknife* est une autre forme de rééchantillonnage permettant d'estimer l'erreur standard et le biais d'un paramètre. Avec cette technique on calcule *n* fois la valeur du paramètre à partir d'un échantillon de *n-1* observations, chacune des observations étant éliminée à son tour de l'échantillon.

Il s'agit d'un autre intervalle de confiance corrigé par le biais qui a été proposé par Hall (1994) et qui est écrit de la manière suivante,

$$\left(2\hat{\xi} - \xi^*_{\inf}, 2\hat{\xi} - \xi^*_{\sup}\right)$$

où $\xi^*_{\inf}$ et $\xi^*_{\sup}$ représentent respectivement les $100.\alpha/2$-ième et $100.1-\alpha/2$-ième percentiles supérieur et inférieur de la distribution bootstrap de ξ^*.

Cette méthode est plus appropriée que la méthode des percentiles simples. En revanche, les deux méthodes vont produire le même intervalle si $\hat{\xi}$ représente la valeur centrale entre ξ^*_{lo} et ξ^*_{hi}.

A.5. La méthode des percentiles pour la différence entre deux distributions [Halls (1994)]

Supposons deux distributions de taille n_1 et n_2 tirées de deux distributions de probabilité inconnues $\hat{F}_1$ et $\hat{F}_2$. Supposons aussi que $\Delta\hat{\xi}$ soit la différence entre les estimateurs des paramètres de deux distributions $\Delta\hat{\xi} = \hat{\xi}_1 - \hat{\xi}_2$. La méthode bootstrap simule la distribution de la différence entre les estimateurs basés sur des échantillons de taille n_1 et n_2 tirés des distributions empiriques $\hat{F}_1$ et $\hat{F}_2$. Nous prélevons B échantillons bootstraps qui produisent:

$$\Delta\xi^{*r} = \xi_1^{*r} - \xi_2^{*r} \text{ pour } r = 1,\ldots, B$$

Selon Hall (1994), l'intervalle percentile de confiance pour les différences entre deux distributions réalisé par Bootstrap est :

$$\Pr\left(2\Delta\hat{\xi} - \Delta\xi^*_{\sup} \leq \Delta\xi \leq 2\Delta\hat{\xi} - \Delta\xi^*_{\inf}\right) = (100-\alpha)/100,$$

où $\Delta\xi^*_{\sup}$ et $\Delta\xi^*_{\inf}$ sont respectivement le 100.α/2-ième et le 100.1-α/2-ième percentiles de la distribution bootstrap définissant la différence des deux distributions.

Contacts :

Thierry BLAYAC : blayac@lameta.univ-montp1.fr

Valérie CLEMENT : clement@lameta.univ-montp1.fr