halshs-00277281, version 1 - 6 May 2008

GREQAM

Groupement de Recherche en Economie
Quantitative d'Aix-Marseille - UMR-CNRS 6579
Ecole des Hautes Etudes en Sciences Sociales
Universités d'Aix-Marseille II et III

Document de Travail
n°2008-09

INTRODUCTION A L'ECONOMETRIE DES MESURES DE PAUVRETE

Michel LUBRANO

Mars 2008

DT-GREQAM

halshs-00277281, version 1 - 6 May 2008

Introduction à l'économétrie des mesures de pauvreté

Michel Lubrano[1]

Mars 2008

Abstract

[**Résumé**] Le but de ce papier est de présenter une vue générale sur les problèmes économétriques liés à la mesure de la pauvreté. A partir d'une fonction de bien-être, on définit tout d'abord des mesures d'inégalité, puis de pauvreté, basées soit sur des indices, soit sur la notion de dominance stochastique. Ensuite, le choix des données (dépense ou revenu, échelles d'équivalence, etc...) peut inverser certaines conclusions empiriques quand il s'agit de comparer les pays entre eux, ou l'évolution de la pauvreté dans le temps. Le traitement statistique des données d'enquête met en jeu des outils qui sont simples au début (statistiques d'ordre, estimation de densités), mais qui deviennent vite complexes quand il s'agit d'estimer des écart-types ou plus encore de tester la dominance stochastique. Un certain nombre de modèles explicatifs vont permettre de passer d'une analyse essentiellement descriptive à une explicitation des ressorts de la pauvreté au moyen par exemple des régressions quantiles et des modèles dynamiques à facteurs. Enfin, l'analyse des phénomènes de pauvreté nécessite l'accès à des données individuelles d'enquête. Selon les pays, l'accès à ces données est très inégal pour les chercheurs, en particulier pour la France.

Abstract

The aim of this paper is to provide an overview of the econometric problems that are linked to the measurement of poverty. Starting from a welfare function, we first define inequality measures and then poverty measures which are based either on indices or on the notion of stochastic dominance. We then show how the choice data definitions (income or consumption spending, equivalence scales, etc...) can modify or invert empirical conclusions when comparing countries or the time evolution of poverty. The statistical treatment of surveys data implies the use of instruments that are simple at the beginning (order statistics, density estimation), but that can become rapidly complex when one has to estimate standard deviations or even more when one wants to test stochastic dominance. Most of statistical

[1]GREQAM-CNRS, Centre de la Vieille Charité, 2 rue de la Charité, 13002 Marseille. Email:lubrano@univmed.fr.
Une première version de ce texte a été présentée lors de l'Ecole Thématique CNRS *La Pauvreté dans les Pays Riches* qui s'est tenue à la Baume les Aix en Septembre 2006. La présente version a bénéficié des précieuses indications de Laurence Bouvard pour la dernière section sur l'accès aux données. Une relecture de ce texte par Russell Davidson a permis de corriger plusieurs imprécisions. Les erreurs restantes sont de ma seule responsabilité.

analysis concerning inequality and poverty is descriptive. Explicative models are needed for analysing the dynamics of poverty such as dynamic factor models or quantile regressions. Finally serious empirical analysis of poverty requires the use of survey data such as the UK Family expenditure survey or the French enquête sur le budget des familles. The access to these data set for scientific researchers is very unequal, in particular in France.

Mots clés : Econométrie, mesures d'inégalité, seuil de pauvreté, tests statistiques, banques de données.

Keywords: Econometrics, inequality measurement, poverty level, statistical tests, survey data sets.

JEL Classification: C10, C42, D31, D63.

halshs-00277281, version 1 - 6 May 2008

1 Introduction

Comment juger de la pauvreté d'un pays, d'une région, comment comparer la distribution des revenus de deux populations? Il s'agit tout d'abord de transcrire cette notion au moyen d'un outil mathématique qui la rende mesurable. La base théorique a été jetée dans Kolm (1969) et dans Atkinson (1970) et Atkinson (1987). Une axiomatique permet de construire des fonctions de bien-être d'où l'on va ensuite pouvoir tirer des indices d'inégalité, puis de pauvreté quand on aura défini une ligne ou seuil de pauvreté. Les indices de pauvreté sont en général des transformations intégrales de la distributions des revenus. La relation entre inégalité et pauvreté est controversée et mérite discussion dans un cadre normatif.

Pour autant tous les problèmes ne sont pas résolus, surtout quand on veut parler de la pauvreté en Europe. Les données statistiques concernent les ménages. Ceux-ci ont des compositions différentes et il faut alors passer par les échelles d'équivalence. Considérer les revenus ou les dépenses et ces dernières avant ou après les coûts de logement peut avoir une incidence marquée sur les résultats. Bref, il faut savoir ce que l'on mesure. Ce sont les premiers problèmes statistiques à traiter.

Sur la plan de la théorie statistique, la gamme des problèmes que l'on doit aborder est immense. S'il est relativement facile d'estimer une distribution et les différents indices d'inégalité et de pauvreté, l'estimation des densités est déjà un peu plus complexe. Ensuite, toutes ces quantités résultent de la réalisation de variables aléatoires qui possèdent bien sûr une moyenne, mais aussi un écart-type. Le calcul des écart-types et des tests de comparaison s'est longtemps heurté à des problèmes conceptuels et bien souvent ces calculs ne peuvent être résolus que par une approche par simulation. Enfin, le test de dominance stochastique est de loin le plus complexe à mettre en place. Davidson and Duclos (2000) présentent l'état actuel de la théorie statistique.

L'analyse de la distribution des revenus, de son évolution dans le temps et dans sa structure pose le problème crucial de la disposition de données d'enquêtes adéquates. La disposition de ces données se heurte à deux questions, à savoir la réalisation matérielle de ces enquêtes et ensuite l'accès que peuvent y avoir les chercheurs. Les situations sont très inégales entre les pays sur ce plan.

2 Les fonctions de bien-être et les indices d'inégalité et de pauvreté

Les données statistiques auxquelles on s'intéresse sont des données d'enquête sur les ménages. Ces enquêtes portent sur les caractéristiques micro-économiques des ménages. Il s'agit essentiellement de la composition des ménages (nombre d'enfants, statut marital du chef de ménage, présence d'autres adultes), des revenus du ménage (salaires, revenus financiers et immobiliers, transferts et impôts) et de la consommation (consommation courante, emprunts, loyers, investissements). On peut tirer de ces enquêtes des indications sur le niveau de vie et son évolution dans le temps. On va s'intéresser à la distribution de ces indicateurs parmi les ménages en ciblant les plus pauvres.

Une bonne partie de l'économétrie de la pauvreté sera consacrée à l'estimation de densités, tant de manière paramétrique que non-paramétrique. On devra ensuite comparer ces densités entre elles soit au moyen d'indices de pauvreté ou d'inégalité soit au moyen de la dominance stochastique. Pourquoi s'intéresser à la queue gauche de la distribution des revenus et donc porter une attention aux plus pauvres? Il faut expliciter l'aversion qu'une société peut avoir pour l'inégalité et la pauvreté. Atkinson (1970) a formalisé ce problème au moyen des fonctions de bien-être. C'est aussi l'approche adoptée par Deaton (1997) dans son chapitre 3.

2.1 Fonction de bien-être

On considère que la société est formée par une collection de n individus et l'on va s'intéresser à une mesure de bien-être pour l'ensemble de ces n éléments pris comme une entité. La mesure se fera à partir d'une quantité uni-dimensionnelle que l'on prendra égale soit au revenu, soit à la dépense de consommation que l'on va noter x_i pour l'individu i. On a donc la première donnée de

$$X = (x_1, x_2, \cdots, x_n) \tag{1}$$

qui représente la distribution des revenus (ou de toute autre caractéristique) au niveau de la population. On définit ensuite la fonction de bien-être comme une fonction à n arguments:

$$W(x) = V(x_1, \cdots, x_n). \tag{2}$$

Cette fonction a un aspect très normatif et sa construction répond à une série d'axiomes qui précisent les comparaisons que l'on s'autorise à faire entre les individus.

halshs-00277281, version 1 - 6 May 2008

halshs-00277281, version 1 - 6 May 2008

1. **Axiome de Pareto:** la fonction est croissante en chacun de ses termes. On peut affaiblir cet axiome en demandant à la fonction d'être simplement non-décroissante en ses termes. Alors, on peut construire une fonction de bien-être qui restera constante si le revenu des plus riches augmente et ne croîtra que si le revenu des plus pauvres augmente.

2. **Axiome de symétrie ou anonymat:** On doit pouvoir intervertir les individus sans que la valeur de la fonction change. Mais, il existe des problèmes soulevés par la composition des ménages. Les données concernent les ménages, alors que la notion de bien-être s'intéresse aux individus. Il y aura donc une incidence non triviale de la composition des ménages que l'on essayera de gommer au moyen des échelles d'équivalence.

3. **Principe du transfert:** La quasi concavité de la fonction de bien-être implique que si l'on transfère d'un riche vers un pauvre, le bien-être augmente, à condition que le transfert ne change pas l'ordre du classement des individus. Il s'agit du principe dit de Pigou-Dalton.

4. **Autres axiomes:** la littérature économique sur la construction des fonctions de bien-être et des indices d'inégalité est importante. Certains axiomes se recoupent. On peut chercher le nombre minimal d'axiomes qui conduise à la construction de la fonction de bien-être. On consultera à ce propos l'ouvrage de Sen (1997).

La conséquence de ces axiomes, c'est qu'une fonction de bien-être exprime l'aversion d'une société pour l'inégalité et que cette fonction sera maximale quand tous les ménages auront le même revenu. Tout un courant de la littérature s'intéresse à la mesure de l'aversion pour l'inégalité, voir en particulier Carlsson, Daruvala, and Johansson-Stenman (2005) et les références citées dans ce papier.

2.2 Inégalité et bien-être social

On veut passer de la fonction de bien-être à des mesures d'inégalité. On va supposer que la fonction est homogène de degré 1 ce qui permet de mettre en facteur le revenu moyen

$$W(x) = \mu V(x_1/\mu, \cdots, x_n/\mu). \tag{3}$$

On va ensuite normaliser $V(.)$ de telle sorte que $V(1, \cdots, 1) = 1$, ce qui fait que le bien-être ne peut être supérieur à μ. On peut alors factoriser la fonction de bien-être en

$$W(x) = \mu(1 - I) \tag{4}$$

où μI représente le coût de l'inégalité. Alors I devient une mesure d'inégalité. Remarquons que le bien-être augmente quand μ croit, ce qui fait que l'on peut avoir dans le même temps une croissance du bien-être et une croissance des inégalités. On rediscutera de la forme de cette fonction plus tard. Notons au passage le débat initié par Okun (1975) sur équité et efficience.

halshs-00277281, version 1 - 6 May 2008

2.3 Mesures d'inégalité

La prise en compte d'une fonction de bien-être pour définir un indice d'inégalité s'illustre au premier chef par l'indice d'Atkinson. On part de la fonction de bien-être suivante

$$W = \frac{1}{n}\sum_i \frac{x_i^{1-\epsilon}}{1-\epsilon} \tag{5}$$

où ϵ est le paramètre d'aversion à l'inégalité. On utilise en général pour ϵ des valeurs comprises entre 0 et 2. Pour $\epsilon = 1$, la forme est indéterminée et on lève l'indétermination en prenant

$$W = \frac{1}{n}\sum_i \log x_i. \tag{6}$$

Pour $\epsilon \to \infty$, on tombe sur une mesure Rawlsienne, Rawls (1972), où seul le sort du plus pauvre importe à la société. La mesure d'inégalité que l'on déduit de cette fonction de bien-être particulière s'écrit:

$$I_A = 1 - \left(\frac{1}{n}\sum_i (x_i/\mu)^{1-\epsilon}\right)^{1/(1-\epsilon)}. \tag{7}$$

La littérature a proposé toute une série d'indices d'inégalité dont on trouvera des exemples dans le chapitre écrit par Patrick Moyes dans ce livre. Dans la mesure où ces indices respectent le principe des transferts, on peut les poser comme un point de départ et effectuer la démarche inverse qui consiste à remonter à une fonction de bien-être au moyen de la formule $W = \mu(1-I)$. L'indice d'inégalité le plus courant est l'indice de Gini. Il est basé sur la moyenne de toutes les paires distinctes de différences de revenu prises en valeur absolue. Il y a $n(n-1)/2$ paires distinctes. On normalise par rapport à la moyenne, ce qui donne:

$$I_G = \frac{1}{\mu n(n-1)}\sum_{j=1}^{n-1}\sum_{i=j+1}^{n} |x_i - x_j|. \tag{8}$$

Cet indice est à valeur dans [0,1]. Il vaut 1 quand un individu a tout et les autres rien. La fonction de bien-être qui est associée à l'indice de Gini est celle qui pondère les observations par leur rang. C'est le plus pauvre qui recevra le poids le plus important. On peut généraliser cette fonction en

$$W = \mu(1 - I_G)^\sigma \tag{9}$$

pour σ entre 0 et 1.

Supposons maintenant que l'on connaisse la distribution cumulée des revenus que l'on va appeler F. On peut définir l'indice de Gini à partir de la donnée de F au moyen du calcul intégral, ce qui donne

$$I_G = \frac{1}{2\mu}\int\int |x - x'| dF(x)dF(x'). \tag{10}$$

Ce type de calcul intégral a été popularisé par Atkinson (1970). Pour certaines distributions simples, cet indice peut se calculer de manière analytique. Par exemple, la distribution Gamma

halshs-00277281, version 1 - 6 May 2008

est caractérisée par deux paramètres, un paramètre d'échelle β et un paramètre de forme α. Le coefficient de Gini associé à cette distribution est

$$I_G = \frac{\Gamma(\alpha + 1/2)}{\Gamma(\alpha + 1)\sqrt{(\pi)}} \tag{11}$$

où $\Gamma(.)$ est la fonction mathématique Gamma. Pour la Weibull, on obtiendrait

$$I_G = 1 - (1/2)^{1/\alpha}. \tag{12}$$

Quand on aura estimé les paramètres de F à partir des n observations, alors on pourra en déduire directement la valeur de l'indice de Gini.

On peut construire une classe d'indices d'inégalité à partir de la fonction de bien-être d'une autre façon qui est plus habituelle. Partons de la fonction W et de la distribution de revenu $X = x_1, \cdots, x_n$. $W(X)$ prend une certaine valeur. On va chercher le revenu équivalent ξ qui, distribué de manière uniforme vérifie $W(\xi) = W(X)$. Si le principe des transferts s'applique, alors on aura toujours $\xi \leq \mu$. On peut donc choisir comme indice d'inégalité la mesure

$$I = 1 - \frac{\xi}{\mu}. \tag{13}$$

On peut à ce moment là requérir deux axiomes alternatifs d'indépendance par rapport à l'échelle pour construire la fonction $\xi(x)$ qui donne la valeur de ξ en fonction de la distribution de X. Soit on choisit que $W(x_1, \cdots, x_n) = W(\lambda x_1, \cdots, \lambda x_n)$, alors la fonction $\xi(x)$ prendra la forme conduisant à l'indice d'Atkinson avec

$$\xi(x) = \left[\frac{1}{n}\sum_i x_i^{1-\epsilon}\right]^{1/(1-\epsilon)}. \tag{14}$$

Soit on choisit une autre forme d'axiome d'indépendance par rapport à l'échelle qui dit que $W(x_1, \cdots, x_n) = W(\delta + x_1, \cdots, \delta + x_n)$ et alors on obtient l'indice de Kolm (1976)

$$I_K = \frac{1}{\alpha}\log\left(\frac{1}{n}\sum_i \exp(\alpha[x_i - \mu(x)])\right) \tag{15}$$

avec $\alpha > 0$.

2.4 De l'inégalité à la pauvreté

L'examen de la forme de la fonction de bien-être (4) permet de voir que la croissance économique, c'est à dire l'augmentation conjointe de μ et de W peut s'accompagner d'une augmentation des inégalités: certains vont s'enrichir plus vite que les autres. C'est ce que l'on a constaté par exemple au Royaume Uni pendant la période Thatcher. Atkinson (2003) montre comment dans les années 1980, le revenu réel des plus pauvres est resté constant alors que la croissance des revenus

halshs-00277281, version 1 - 6 May 2008

a concerné les groupes moyens et surtout les groupes les plus riches. Malgré cela la mesure du bien-être a augmenté.

La pauvreté est ressentie comme un échec et cela justifie que l'on s'y intéresse plus particulièrement. La fonction de bien-être transforme une distribution complète en un nombre permettant d'analyser les effets de mesures de politique économique sur l'ensemble de la distribution des revenus. Si l'on veut concentrer son attention sur les plus pauvres, on va s'intéresser plus particulièrement à une partie de la distribution des revenus, celle qui concerne les plus pauvres, ne serait-ce que pour les compter. On va donc passer de l'analyse des inégalités à l'analyse de la pauvreté en concentrant son attention sur la queue gauche de la distribution des revenus.

Pour concenter son attention sur les plus pauvres, il faut définir ce que l'on appelle une ligne de pauvreté, c'est à dire un seuil en deçà duquel une personne (ou un ménage) sera considéré comme pauvre et au delà duquel il basculera dans la catégorie des non-pauvres. On mesure combien ce qu'un tel seuil a d'arbitraire. On peut le définir de deux façons

1. un seuil de pauvreté absolu se définit par rapport à un niveau minimum de subsistance. Le gouvernement indien par exemple a défini un nombre minimum de calories nécessaires en ville et qui est différent de celui nécessaire à la campagne. En passant par un indice de prix, il arrive à un seuil monétaire de pauvreté en ville et à la campagne. Sur la même base alimentaire, le gouvernement américain a défini un seuil absolu de pauvreté, mais en divisant celui-ci par la part de la nourriture dans le budget d'un ménage moyen. Le RMI (revenu minimum d'insertion) peut également se situer dans ce cadre.

2. Dans les pays développés et plus particulièrement au sein de l'Union Européenne, on préfère définir un seuil relatif de pauvreté. L'Union Européenne a lancé un programme de mesure de la pauvreté où le seuil de pauvreté y est défini par rapport à la moyenne ou la médiane de la distribution des revenus. Sera considéré comme pauvre tout individu touchant un revenu inférieur à 50% ou 60% de la moyenne des revenus de son pays. Il s'agit alors de pauvreté relative, ce qui nous rapproche de la notion de pauvreté ressentie.

2.5 Les indices de pauvreté

Il existe toute une série d'indices de pauvreté. Mais d'une certaine façon les indices les plus simples à comprendre et à manipuler sont les indices linéaires de Foster, Greer, and Thorbecke (1984). Ces indices sont basés sur des moyennes partielles construites à partir de la distribution des revenus. Si $F(.)$ est la distribution des revenus et z le seuil de pauvreté, alors pour un α donné cet indice s'écrit

$$P_\alpha = \int_0^z \left(\frac{z-x}{z}\right)^\alpha dF(x). \tag{16}$$

En faisant varier le paramètre α entre 0 et 2, on retrouve un certain nombre des mesures classiques de pauvreté.

Pour $\alpha = 0$, on tombe sur la mesure traditionnelle dite de headcount qui est une mesure de dénombrement:

$$P_0 = \int_0^z f(x)\,dx = F(z). \tag{17}$$

halshs-00277281, version 1 - 6 May 2008

Il s'agit d'une première mesure, intéressante en soi, elle permet de connaître le nombre de pauvres en multipliant simplement P_0 par la taille de la population. Toutefois cette mesure est insuffisante car elle ne distingue pas entre les pauvres et ne tient pas compte de leur niveau de pauvreté, c'est à dire qu'elle ne distingue pas entre les individus qui sont près de la ligne de pauvreté et ceux qui en sont loin.

Avec $\alpha = 1$, on tombe sur une mesure faisant intervenir le déficit de pauvreté ou poverty gap $z - x_i$ qui affecte chaque individu en dessous du seuil de pauvreté:

$$P_1 = \int_0^z (1 - x/z) f(x)\, dx. \tag{18}$$

Cet indice respecte le principe de transfert à l'inverse de la mesure de comptage P_0 qui ne le respecte pas. Cet indice est continu alors que le head count ne l'est pas. Mais il est insensible à certains types de transferts entre les pauvres.

Pour $\alpha = 2$, on arrive à une mesure qui est sensible à la distribution parmi les pauvres:

$$P_2 = \int_0^z (1 - x/z)^2 f(x)\, dx, \tag{19}$$

mais qui n'est pas très souvent employée. Atkinson (1987) examine simplement les propriétés d'une généralisation des deux premiers indices et leur relation avec la dominance stochastique restreinte, notion que l'on explicitera plus bas.

L'indice de Foster, Greer, and Thorbecke (1984) répond à la propriété de décomposabilité car il a une structure linéaire. Considérons par exemple une partition de la population entre urbain et rural. Si X représente l'ensemble des revenus, la partition de X se définira comme $X = X_U + X_R$. Appelons p la proportion de X_U dans X. Alors l'indice total de pauvreté se décomposera en

$$\begin{aligned} P_\alpha &= p \int_0^z \left(\frac{z - x}{z}\right)^\alpha dF(x_U) + (1 - p) \int_0^z \left(\frac{z - x}{z}\right)^\alpha dF(x_R) \\ &= p\, P_\alpha^U + (1 - p)\, P_\alpha^R. \end{aligned} \tag{20}$$

Une autre classe d'indices de pauvreté a été proposée à la suite de Sen (1976) pour tenir compte de l'inégalité entre les pauvres. L'indice original de Sen, P_S, combine une mesure de headcount P_0, une mesure du poverty gap, $zP_0 I_P = \int_0^z (z - x)) f(x)\, dx$ et un indice de Gini G_P calculé sur le segment $x < z$. Cet indice se note

$$P_S = P_0(I_P + (1 - I_P)G_P) = P_0 \left(1 - (1 - G_P)\frac{\mu_p}{z}\right) \tag{21}$$

où μ_p est la moyenne des revenus parmi les pauvres:

$$\mu_p = \int_0^z x f(x) dx / F(z) \qquad \text{et} \qquad I_P = 1 - \frac{\mu_p}{z}.$$

Quand il n'y a pas d'inégalité entre les pauvres ($G_P = 0$), alors $P_S = P_1$. Quand l'inégalité devient extrême ($G_P = 1$), on retombe sur la mesure de headcount, ce que traduit bien la factorisation

$$P_S = P_0 G_P + P_1(1 - G_P). \tag{22}$$

Mais tout comme l'indice de Gini, cet indice n'est pas décomposable. Il viole également l'axiome de transfert et de plus n'est pas continu. Shorrocks (1995) a proposé une modification de cet indice appelé aussi l'indice de Sen-Schorrocks-Thon qui résoud une partie de ces difficultés. Cet indice s'écrit par analogie avec l'indice de Sen comme

$$P_{SST} = (2 - P_0)P_0 I_P + P_0^2(1 - I_P)G_P.$$

2.6 Pauvreté et inégalité

La formulation initiale de la fonction de bien-être (4) implique qu'une augmentation du bien-être peut tout à fait s'accompagner d'un accroissement des inégalités. Comment inclure dans cette décomposition une attention plus particulière à la pauvreté? En d'autres termes, quelle forme doit-on considérer pour $W(x)$ si l'on veut maximiser le bien-être tout en insistant sur la pauvreté. Atkinson (1987) traite de cette question dans la section 3 de son papier en distinguant quatre options possibles.

La première option consiste à ne pas se soucier particulièrement de la pauvreté. On va simplement maximiser

$$W(x) = \mu(1 - I), \tag{23}$$

où I est un indice d'inégalité et μI mesure le coût de l'inégalité. Si l'on a choisi de manière adéquate la fonction de bien-être, on peut décomposer cet indice en distinguant le groupe des pauvres du reste de la population. On pourra donc mesurer l'évolution de la pauvreté sans avoir la réduction de la pauvreté comme objectif principal.

Dans une deuxième option, on va chercher à introduire un coût prioritaire pour la pauvreté $C_P = \mu P$ et laissant un rôle secondaire au coût de l'inégalité. Ceci peut se faire en adoptant une fonction de bien-être du type

$$W(x) = \mu - \mu P - \mu I. \tag{24}$$

Atkinson (1987) indique que dans ce cas il est logique d'utiliser une mesure de comptage pour P et une mesure satisfaisant le principe de transfert pour I.

La troisième option consiste à se focaliser uniquement sur la pauvreté. La fonction de bien-être à maximiser sera alors de la forme

$$W(x) = \mu - \mu P. \tag{25}$$

Enfin la dernière option consiste à utiliser un arbitrage entre inégalité et pauvreté. On aura toujours la fonction de bien-être donnée en (24)

$$W(x) = \mu - \mu I - \mu P. \tag{26}$$

halshs-00277281, version 1 - 6 May 2008

halshs-00277281, version 1 - 6 May 2008

Mais cette fois-ci des considérations de justice vont conduire à utiliser pour I un indice de Gini calculé sur toute la population et pour P un indice de pauvreté de Sen (1976) modifié.

Ces considérations montrent que la construction de la fonction de bien-être peut être relativement complexe quant à ses propriétés et à la façon dont sont agrégés les individus. La forme simple (4) était peut-être un peu trop simple.

2.7 La dominance stochastique

Les indices de pauvreté reposent sur la définition d'une ligne de pauvreté et celle-ci a un caractère arbitraire. On est donc en droit de se poser la question de la robustesse des indices par rapport à la définition retenue pour le niveau de pauvreté z. On voudrait avoir une mesure qui puisse permettre de porter des jugements soit pour tout z, soit pour z variant dans une certaine fourchette. On passe alors aux notions de dominance stochastique et de dominance stochastique restreinte.

La dominance stochastique est une notion mathématique qui permet de comparer des distributions entre elles. Elle vient de la théorie des probabilités (Blackwell 1953), a été utilisée pour résoudre les problèmes de décision dans l'incertain (Hanoch et Levy 1969), puis en finance pour caractériser des choix de portefeuille (Fishburn 1977) et enfin par Atkinson (1970) pour comparer des fonctions de bien-être.

Nous avons opté le plus souvent jusqu'à présent pour une présentation discrète, c'est à dire que l'on a considéré les individus et les ménages de manière isolée et utilisé des sommes discrètes pour calculer les indices en supposant que la population était finie. Même si l'on a parfois utilisé la présentation continue d'Atkinson (1970) qui suppose une population infinie. La notion de dominance stochastique est une notion qui, dans la plupart des cas, fait intervenir des distributions continues. On va comparer deux distributions continues de revenu $F(x)$ et $G(x)$ et dire quelle est la plus désirable dans un sens que nous allons préciser. On distinguera différents ordres de dominance, un deux ou trois.

La notion de dominance la plus immédiate est la dominance stochastique à l'ordre 1. Elle vise à comparer directement deux distributions point par point:

Definition 1 *La distribution de probabilité F domine stochastiquement la distribution G à l'ordre un si et seulement si*

$$F(x) - G(x) \leq 0 \qquad \forall x \in [0, +\infty[. \tag{27}$$

Cette définition signifie que la probabilité d'obtenir x n'est pas plus grande avec F qu'avec G, quelle que soit la valeur de x. Mais dès que les distributions se croisent, cette définition ne peut plus s'appliquer. On doit alors passer à la dominance d'ordre deux qui va comparer les surfaces sous les distributions de probabilité.

Definition 2 *La distribution de probabilité F domine stochastiquement la distribution G à l'ordre deux si et seulement si*

$$\int_0^x [F(t) - G(t)]dt \leq 0 \qquad \forall x \in [0, +\infty[. \tag{28}$$

halshs-00277281, version 1 - 6 May 2008

Pour comprendre ce qui se passe, il est utile de considérer la suite d'intégrales de la densité f

$$\begin{aligned} F_0(x) &= f(x) \\ F_1(x) &= \int_0^x F_0(t)dt \\ F_2(x) &= \int_0^x F_1(t)dt \end{aligned} \tag{29}$$

que l'on peut généraliser en

$$F_s(x) = \int_0^x F_{s-1}(t)dt = \frac{1}{(s-1)!}\int_0^x (x-t)^{s-1} f(t)dt. \tag{30}$$

Cette dernière écriture est particulièrement instructive dans la mesure où elle permet de faire le lien entre la dominance stochastique et les indices de pauvreté de Foster, Greer, and Thorbecke (1984). En effet si l'on fixe x égal à z le seuil de pauvreté, on se rend compte qu'il y a identité entre les fonctions de dominance $F_s(z)$ et $P_{s-1}(z)$ à un facteur multiplicatif près qui ne dépend que de s. La dominance stochastique correspond donc à la généralisation de ces indices quand on fait varier le seuil de pauvreté sur tout le segment $[0, +\infty[$. C'est le point de vue que développe Atkinson (1987). Notons au passage la notion de poverty deficit curve qui est la courbe obtenue en faisant varier z dans $zP_1(z)$.

Le lien avec les indices de pauvreté est encore plus apparent si l'on passe à une notion de dominance restreinte en ne considérant plus les inégalités pour tout x, mais pour un intervalle réduit $[z_*, z^*]$. On va donc considérer

$$F_s(z) = \frac{1}{(s-1)!}\int_0^z (z-t)^{s-1} f(t)dt \qquad \forall z \in [z_*, z^*]$$

Cette écriture permet de comparer deux distributions de revenu en faisant varier le seuil de pauvreté entre deux bornes. On aura donc une mesure de comparaison plus robuste. Si on appelle $z - x$ le poverty gap, c'est à dire la différence entre l'observation du revenu x et un seuil de pauvreté z, alors pour $s = 1$, on compte les pauvres, pour $s = 2$, on mesure la moyenne du poverty gap et pour $s = 3$, on s'intéresse à sa dispersion pour un intervalle donné a priori.

3 La mesure de la pauvreté en Europe

En 1990, Antony Atkinson a donné la Yrjö Jahnsson lecture sur le thème *La Pauvreté en Europe*. Sa première leçon était consacrée à la mesure de la pauvreté. Les données datent un peu maintenant, mais restent d'actualité pour leur aspect méthodologique. Atkinson relate des recherches menées sous l'égide de l'Union Européenne et de son programme de lutte contre la pauvreté. L'union Européenne estimait à 50 million le nombre de pauvres en 1990 dans l'Union en se basant sur des enquêtes sur les budgets familiaux.

Definition 3 *L'UE définit comme pauvre toute personne vivant avec moins de 50% de la dépense moyenne par habitant du pays dans lequel elle vit.*

Table 1: La pauvreté dans l'UE
(chiffres de 1988 extraits de Atkinson 1998)

Pays	Nbre de personnes (en millions)	Pourcentage
Portugal	2.5	25.5
Italy	12.1	21.1
Grèce	1.8	18.7
Espagne	6.5	16.9
Irlande	0.5	15.7
Royaume Uni	8.4	14.8
France	8.2	14.7
Luxembourg	0.041	11.1
Allemagne de l'Ouest	6.6	10.9
Belgique	0.7	7.4
Pays Bas	0.7	4.8
Danemark	0.2	3.9
Total	49.0	15.0

L'application de cette définition donnait en 1994 le classement suivant rapporté dans la Table 1. La pauvreté semble localisée dans les pays du sud, mais les 2/3 de ces 50 millions se situent en France, Allemagne, Italie et Royaume Uni. L'analyse de ces chiffres pose un certain nombre de problèmes que l'on va maintenant détailler avec Atkinson (1998).

3.1 Comparaison entre la France et la Grande Bretagne

Dans le tableau précédent dressé pour le compte de l'UE, la France et le Royaume Uni sont très semblables. Pourtant quand on prend des études nationales (effectuées dans chacun de ces deux pays), on peut trouver des chiffres très différents, même s'ils semblent basés sur la même définition apparente du seuil de pauvreté. On va comparer les résultats obtenus à partir de *l'enquête sur le budget des familles* (France) et le *Family expenditure survey* (UK). Ces études prennent le revenu comme base de départ et non la dépense comme cela est fait dans le premier tableau. Selon ces études, on arrive à 9.6% de pauvres pour la France et 4.1% pour le UK. On est donc loin des chiffres du premier tableau. Mais on peut retrouver les chiffres initiaux en effectuant les glissements suivants:

1. On peut choisir la médiane ou la moyenne pour mesurer la tendance centrale. Les deux sont identiques si la distribution est symétrique. Mais les deux distributions sont très différentes entre les deux pays quant à l'asymétrie.

2. On doit calculer le revenu moyen de chaque individu alors que l'on n'observe que les ménages. Comment pondérer chaque ménage? Par un, par le nombre de ses membres ou par le nombre d'équivalents adultes?

halshs-00277281, version 1 - 6 May 2008

halshs-00277281, version 1 - 6 May 2008

3. Il faut ensuite choisir une échelle d'équivalence. L'échelle de l'OCDE vaut 1, 0.7, 0.5. Celle utilisée au Royaume Uni est plus complexe car elle tient compte de l'age des enfants.

4. Enfin, on peut prendre le revenu avant ou après le coût du logement. Au Royaume Uni, ce coût varie très fort car les taux d'intérêt sont variables et donc les dépenses de logement ne résultent pas d'une décision des ménages, mais sont pour une bonne part exogènes.

On voit donc qu'en adoptant des choix particuliers de mesure on peut faire dire des choses très différentes à la simple définition *50% en dessous du revenu moyen.*

3.2 Pauvreté absolue ou pauvreté relative

Quand on veut mesurer l'évolution de la pauvreté dans le temps, que faut-il choisir comme seuil de pauvreté: un seuil absolu ou un seuil relatif? On a vu comment définir un seuil relatif. Si le revenu moyen du pays augmente, le seuil relatif va augmenter avec lui. Par contre si l'on choisi un seuil fixe à partir d'un panier de biens prédéfini, ce seuil ne va être ajusté dans le temps que pour l'inflation. Mais on ne tiendra pas compte de l'évolution des besoins et des désirs. En Italie, la pauvreté a diminué de 1980 à 1990 passant de 8.3% à 3.4% si l'on prend un seuil de pauvreté fixe (pouvoir d'achat), mais elle a augmenté sensiblement passant de 8.3% à 10.2% en prenant un seuil de pauvreté relatif. Au Royaume Uni, la pauvreté a beaucoup plus augmenté en considérant un seuil relatif qu'un seuil fixe de pouvoir d'achat.

Un seuil fixe a une logique de pouvoir d'achat. Un seuil relatif repose sur une logique de capacité de participation à la vie sociale. Il s'agit là de pauvreté ressentie. Ce n'est pas la même chose de simplement manger à sa faim en Afrique et en Europe.

Enfin, faut-il considérer un seuil unique pour tous les pays de l'Union? Pour le moment ce seuil a été calculé pays par pays. Mais on espère du fonctionnement de l'Union une convergence entre les pays, comme l'on peut espérer une convergence entre les régions d'un même pays. Il faudrait alors calculer le seuil de pauvreté par rapport à la moyenne des revenus de l'Union. D'ici là, on pourrait décider d'une moyenne pondérée pour le revenu moyen de référence

$$\tilde{x}_i = \bar{x}_{UE}^{\theta} \bar{x}_i^{1-\theta}$$

où θ est le paramètre de pondération que l'on doit fixer a priori.

3.3 Revenu ou dépense

Tous les pays mènent des enquêtes sur les budgets des ménages, mais ceux-ci contiennent des informations variables sur les revenus. On s'accorde souvent à dire que la qualité des données portant sur la dépense est meilleure que celle des données de revenus qui sont souvent sous évalués.

Sur le plan théorique, une approche de pouvoir d'achat (ligne fixe) voudrait que l'on se base sur la consommation. Tandis qu'une approche en terme de droits minimaux ou de capacité voudrait que l'on se base sur les revenus. Mais bien souvent, le choix s'opère seulement en fonction de la question de la qualité des observations.

halshs-00277281, version 1 - 6 May 2008

Ces choix ont une incidence sur la mesure de l'évolution de la pauvreté et sur la composition de la population des pauvres. L'UE a choisi de passer du revenu à la dépense, alors que le Royaume Uni a choisi de passer de la dépense au revenu, donc le choix inverse.

3.4 Famille ou ménage

Au Royaume Uni, on est passé de la famille au ménage. Cela a une incidence sur les mesure de pauvreté car le ménage est plus important que la famille. Il peut contenir plusieurs adultes, par exemple les enfants majeurs au chômage ou pas. Il y a donc agrégation de plusieurs revenus. Ceci a permis de faire passer la population avec un revenu inférieur à 50% de la moyenne de 11.1% à 8.1% en 1989. Plus généralement le coefficient de Gini sur la population totale est passé de 0.285 à 0.261. L'agrégation masque donc certains aspects de la pauvreté. Il faut désagréger les données pour étudier la pauvreté en détail. Par exemple, si l'on s'intéresse à la pauvreté des femmes, il faut regarder ce qui se passe à l'intérieur du ménage.

3.5 Les échelles d'équivalence

On observe des données sur des ménages qui ont des compositions différentes. Comment tenir compte de ces tailles différentes et surtout comment passer du bien-être du ménage au bien-être individuel? Les échelles d'équivalence permettent de passer de l'unité d'un ménage à celui d'un adulte équivalent. Si x_i est le revenu du ménage, le revenu de l'adulte équivalent va être obtenu en divisant le revenu du ménage par un nombre p_i

$$\tilde{x}_i = \frac{x_i}{p_i}$$

où p_i peut être calculé de différentes façons. On peut prendre $p_i = n_i$ où n_i est le nombre total de personnes dans le ménage. Mais en général, on va accorder un poids inférieur aux enfants par exemple. Le système fiscal français retient 1, 1, 0.5, 0.5 et 1 à partir du troisième enfant, ce qui signifie que chaque adulte compte pour une part, les deux premiers enfants pour une demi part, mais le troisième enfant pour une part entière. L'échelle OCDE donne 0.7 aux adultes non chef de ménage et 0.5 à tous les enfants. On a proposé dans la littérature

$$p_i = n_i^s$$

avec $s < 1$. Ce paramètre permet de résumer l'élasticité de p_i à la taille du ménage. Selon les échelles subjectives prises, ce paramètre peut varier entre 0.3 et 0.7. Ce choix de s permet de moduler la répartition des coûts à l'intérieur d'un ménage. Si les coûts fixes sont importants, on pourra prendre un s faible. Dans le cas contraire, le poids relatif d'une personne supplémentaire sera plus fort et alors un s plus grand se justifie.

En reprenant sa comparaison entre France et Royaume Uni, Atkinson montre qu'en faisant varier s on peut inverser le ranking entre des deux pays. Avec $s < 0.55$, le Royaume Uni a plus de pauvres. Avec $s > 0.55$, c'est la France qui a le plus de pauvres. En changeant la valeur de s, on change la composition de la population des pauvres. Avec $s = 0$, la majorité des pauvres se trouve dans les petits ménages. Avec s grand, ce sont les familles nombreuses qui constituent le gros des bataillons de pauvres.

halshs-00277281, version 1 - 6 May 2008

3.6 L'utilisation de la dominance stochastique

On a présenté jusqu'ici dans cette section des mesures ponctuelles de pauvreté. Le Tableau 1 fournit des indications sur la valeur de P_0 quand $z = \mu_i/2$. On peut maintenant montrer que les comparaisons tant internationales que temporelles peuvent être différentes si l'on change la valeur de z. Au lieu de considérer un simple indice P_0 comme un nombre fixe, on va considérer P_0 comme une fonction de z avec z variant à l'intérieur d'un intervalle fixé à l'avance. Pour réaliser une comparaison entre deux situations, on devra alors comparer non plus deux points, mais deux courbes. C'est la notion de dominance stochastique restreinte telle qu'illustrée dans Atkinson (1987). Ainsi, Atkinson (1987) montre que la pauvreté aux USA a augmenté entre 1974 et 1982 car la courbe de 1974 est bien en dessous de celle de 1982 pour $z \in [0.50z_o, 1.50z_o]$ si z_o est le seuil de pauvreté officiel américain. Par contre entre 1979 et 1982, les deux courbes se croisent, ce qui fait qu'il n'est pas possible de conclure dans un sens ou dans l'autre, du moins à l'ordre un. Dans ce cas précis l'adoption d'un seuil fixe de pauvreté aurait conduit à une conclusion dans un sens ou dans l'autre, sans rendre compte de l'ambiguité de la situation. Mais si l'on avait examiné les courbes de dominance à l'ordre deux, on aurait pu lever l'ambiguïté, car la dominance à l'ordre deux est un critère moins exigeant que la dominance à l'ordre un. Examinons maintenant les données britanniques avec Atkinson (1998). Sa Figure 1.8 montre que la courbe de dominance de 1979 est toujours inférieure à celle de 1994 pour $z \in [0.40\mu, 0.60\mu]$. Donc pour cet intervalle, il y a dominance stochastique à l'ordre 1, ce qui est un résultat relativement robuste quant à la valeur du seuil de pauvreté considéré.

4 Les problèmes de statistique mathématique

Les divers résultats présentés dans la section précédente démontrent déjà qu'il est difficile d'obtenir un résultat unique quant à la quantification de la pauvreté. Des choix de mesure différents, des définitions différentes conduisent à des résultats souvent opposés dans les comparaisons tant internationales que temporelles. Pourtant, l'approche suivie par la plupart des économistes et non des moindres, ignore le fait que les données que l'on traite proviennent de la réalisation de variables aléatoires. Quand on compare deux indices de Gini ou deux indices de pauvreté, il faut impérativement se poser la question de savoir s'il sont statistiquement différents. Ce qui veut dire que leur différence constatée, si elle n'est pas très importante, peut simplement être due à la réalisation particulière d'une variable aléatoire et qu'un autre tirage aurait conduit à des résultats légérement différents. Nous allons dans cette section détailler une série de questions qui ont trait à la pratique statistique de l'inférence et des tests.

4.1 L'utilisation des statistiques d'ordre

Les premières techniques d'estimation que l'on va présenter ici sont relativement simples. Elles reposent sur le fait que l'on puisse ordonner les observations et donc calculer ce que l'on appelle les statistiques d'ordre. Supposons donc que les observations de X soient classées par ordre croissant et que l'on note ce classement

$$x_{(1)} \leq x_{(2)} \leq \cdots \leq x_{(n)}. \tag{31}$$

$x_{(1)}$ représente la plus petite observation et $x_{(n)}$ la plus grande. Il devient facile dans ce cadre d'estimer de façon naturelle une distribution et ses quantiles. En effet, un distribution se définit comme $F(x) = \text{Prob}(X < x)$. Elle peut s'approximer au moyen de

$$\text{Prob}(X \leq x_{(i)}) \simeq i/n \tag{32}$$

quand on dispose de suffisamment d'observations. Le premier décile de cette distribution correspond à la valeur $x_{0.10}$ telle que $\text{Prob}(X \leq x_{0.10}) = 0.10$. Il suffira alors de trouver l'observation dont le rang i correspondra à peu près à $i/n = 0.10$ dans la suite ordonnée des X. D'une manière générale, si l'on note $Q(p)$ le quantile d'ordre p, celui-ci s'estime comme

$$Q(p) = x_{(s)} \qquad s - 1 \leq np \leq s. \tag{33}$$

L'estimation des quantiles permet par exemple de calculer une mesure de dispersion qui est l'écart interdécile $(x_{0.90} - x_{0.10})/x_{0.50}$.

A partir de ces mêmes statistiques d'ordre, on peut définir un estimateur pour la courbe de Lorenz généralisée en traçant

$$L(p = i/n) = \frac{1}{n} \sum_{j=1}^{i} x_{(j)}. \tag{34}$$

On a utilisé ici les sommes partielles des statistiques d'ordre. La courbe de Lorenz s'obtient en normalisant par la moyenne. Enfin, on peut estimer le coefficient de Gini au moyen d'une simple somme et non d'une double somme comme dans la définition originale (8):

$$\hat{I}_G = \frac{2}{n(n-1)\hat{\mu}} \sum_i i\, x_{(i)} - \frac{n+1}{n-1}. \tag{35}$$

Ce type de calcul peut également s'appliquer pour calculer l'indice de pauveté de Sen-Schorrocks-Thon:

$$\hat{I}_{SST} = \frac{1}{n^2} \sum_{i=1}^{q} (2n - 2i + 1) \frac{z - y_{(i)}}{z}.$$

où q correspond au rang de z dans la distribution de X.

4.2 L'estimation des densités

Le graphique d'une distribution est par nature une courbe relativement lisse pour laquelle on peut se contenter d'une approche simplifiée. Par contre avec les densités, on passe de la notion de probabilité cumulée $\text{Prob}(X \leq x)$ à la notion de probabilité d'un intervalle, par exemple $\text{Prob}(x_{(i-1)} < X < x_{(i)})$. Le graphique cherché ne sera donc plus lissé par l'augmentation du nombre des observations. On devra le lisser de manière artificielle en utilisant un estimateur non paramétrique à noyau. Si l'on veut par exemple estimer la distribution des revenus à partir

halshs-00277281, version 1 - 6 May 2008

d'un échantillon, on devra choisir un noyau K (K est une densité paramétrique normalisée et standardisée qui s'intègre à 1) et un paramètre de lissage h appelé également fenêtre. On aura alors

$$\hat{f}_n(x) = \frac{1}{nh} \sum_i K\left(\frac{x_i - x}{h}\right). \tag{36}$$

Le noyau[1] K est en général pris égal à la densité normale, ce qui fait que la hauteur de la densité f au point x est prise comme une moyenne pondérée des valeurs qu'aurait une densité normale dans le voisinage de x. Au plus h est grand, au plus le graphique obtenu sera lisse et proche d'une normale. Une grande fenêtre cachera donc les détails de la densités réelle. Par contre pour $h \to 0$, ce graphique deviendra très irrégulier et l'on ne se concentrera que sur les détails de la distribution.

Il faut en général beaucoup d'observations pour obtenir une bonne estimation de la densité car la vitesse de convergence des estimateurs non-paramétriques est très lente, de l'ordre de $n^{1/5}$. Pour obtenir une vitesse de convergence plus rapide de l'ordre de $n^{1/2}$, on doit passer à un estimateur paramétrique et proposer une famille paramétrique à estimer. Les deux distributions historiquement les plus utilisées ont été la distribution de Pareto et la distribution log normale. La distribution de Pareto s'ajuste particulièrement bien à la distribution des revenus des classes moyennes supérieures et supérieures. La distribution log normale représente bien la distribution des revenus moyens à faibles. Ces deux familles de distributions induisent des courbes de Lorenz qui ne peuvent pas se couper. Elles sont donc trop restrictives.

On a cherché dans la littérature des densités qui soient à la fois commodes à ajuster et dont les queues s'ajustent mieux à la distribution complète des revenus. On a ainsi des familles de distributions à trois ou quatre paramètres. Le distribution de Singh and Maddala (1976) à trois paramètres est la plus connue. Mais ces distributions sont toujours des distributions unimodales. Elles gomment donc certains aspects de la distribution des revenus, comme par exemple la montée du groupe des retraités pauvres au Royaume Uni, groupe qui se manifeste par un mode secondaire dans la distribution des bas revenus. Les études empiriques modernes sur la distribution des revenus emploient donc soit des approches non-paramétriques, voir par exemple Marron and Schmitz (1992), soit des approches semi-paramétriques au moyen de mélanges de distributions, voir Flachaire and Nunez (2007).

4.3 Ecart-types

Pendant très longtemps, on a supposé qu'il n'était pas utile d'associer un écart type aux mesures d'inégalité et de pauvreté. On pensait que ceux-ci étaient de toute façon très petits à cause de la grande taille des échantillons. Mais, comme l'on s'intéresse très souvent à des sous-groupes, par exemple celui des parents isolés, l'argument de la taille ne tient pas très longtemps, car les sous-classes sont intrinsèquement de petite taille. Ce qui est vrai par contre, c'est que ces écart-types sont souvent difficiles à calculer.

[1]Le noyau optimal est le noyau d'Epanechnikov. Mais la perte d'efficience est très minime si l'on choisit le noyau Gaussien. Le choix critique est celui de la fenêtre de lissage.

halshs-00277281, version 1 - 6 May 2008

halshs-00277281, version 1 - 6 May 2008

Les mesures de pauvreté que l'on veut estimer sont des fonctions complexes de l'échantillon, comme par exemple les quantiles, et l'on doit avoir recours à des techniques spéciales pour en évaluer les écart-types. On pourra par exemple consulter Berger and Skinner (2003) à ce sujet. On peut obtenir soit des formules analytiques, le plus souvent approchées par une méthode de linéarisation, soit utiliser des techniques de re-échantillonnage comme le bootstrap pour calculer alors un écart-type d'échantillonnage. Davidson (2007) compare plusieurs méthodes pour calculer la variance de l'indice de Gini et fournit une bonne méthode d'approximation. Il donne également une procédure applicable à l'indice de pauvreté de Sen et à l'indice de pauvreté de Sen-Shorrocks-Thon.

L'approche Bayésienne peut fournir une solution intéressante à cette question en évitant certains problèmes de biais attachés à l'emploi des techniques de bootstrap. Si l'on a estimé une distribution paramétrique pour les revenus par une technique de Monte Carlo, on dispose alors d'un ensemble de tirages des paramètres de cette distribution. Si la distribution estimée est suffisamment simple, on aura une expression analytique de certains indices, comme le Gini. Pour chaque valeur des tirages on calculera la valeur de ces indices et il sera alors trivial d'en calculer les écart-types d'échantillonnage. Si l'on désire ajuster une distribution plus riche, on passe alors à un mélange de distributions basé sur des membres appartenant à une famille simple pour laquelle on dispose de résultats analytiques. En général l'indice général sera obtenu comme une somme pondérée des indices calculés pour chaque membre du mélange.

4.4 Tester l'égalité de deux indices

Kakwani (1993) montre comment traiter ce problème de manière relativement simple. La simplicité tient au type d'indice considéré, à savoir les indices de pauvreté de Foster, Greer, and Thorbecke (1984) qui sont linéaires et décomposables. Ce qui fait que ces indices ne font intervenir que des sommes de variables aléatoires indépendantes et identiquement distribuées. Le théorème central limite s'applique alors directement. Ce ne serait pas le cas avec l'indice de Sen (1976) car il fait intervenir le rang des variables aléatoires.

L'indice de Foster, Greer, and Thorbecke (1984) se définit conditionnellement à une valeur du paramètre α comme

$$P_\alpha = \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x)dx \qquad \alpha \geq 0. \tag{37}$$

Cet indice s'estime assez facilement. Considérons un échantillon sur n ménages dont le revenu par adulte équivalent est de $x_1, \cdots, x_n$. Supposons que l'on ait ordonné les observations x_i par ordre croissant et que q soit le rang du dernier pauvre ($q = \text{Max}_i \; i \, \mathbb{1}(x_{(i)} \leq z)$. Un estimateur consistant de l'indice P_α est donné par

$$\widehat{P}_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-x_{(i)}}{z}\right)^\alpha \tag{38}$$

dans le cas général et par $\widehat{P}_0 = q/n$ quand $\alpha = 0$. L'application du théorème central limite permet tout d'abord de montrer la normalité asymptotique de cet estimateur

$$\sqrt{n}(\widehat{P}_\alpha - P_\alpha) \sim N(0, \sigma^2). \tag{39}$$

halshs-00277281, version 1 - 6 May 2008

La variance σ^2 se définit comme

$$\sigma^2 = \mathrm{E}(\widehat{P}_\alpha - P_\alpha)^2 = \int_0^z \left(\frac{x-z}{z}\right)^{2\alpha} - P_\alpha^2, \tag{40}$$

et cette quantité s'estime naturellement comme

$$\widehat{\sigma}^2 = \widehat{P}_{2\alpha} - \widehat{P}_\alpha^2. \tag{41}$$

L'écart-type de $\widehat{P}_\alpha$ s'estimera comme $\widehat{\sigma}/\sqrt{n}$ que l'on va noter $\hat{\sigma}_P$. Alors la variable aléatoire

$$t = \frac{\widehat{P}_\alpha - P_\alpha}{\hat{\sigma}_P} \tag{42}$$

est distribuée asymptotiquement selon une normale de moyenne nulle et de variance unité. Si on appelle $t_{0.05}$ la valeur critique de la normale au seuil de 5%, on peut former l'intervalle de confiance

$$\widehat{P}_\alpha - t_{0.05}\hat{\sigma} \leq P_\alpha \leq \widehat{P}_\alpha + t_{0.05}\hat{\sigma}. \tag{43}$$

Considérons maintenant deux échantillons indépendants de taille n_1 et n_2. On considère les distributions asymptotiques de $\sqrt{n_i}\widehat{P}_i$ de variance σ_i^2. On a omis l'indice α dans P_α pour la clarté des notations. Alors l'écart type de l'estimateur de la différence $\widehat{P}_1 - \widehat{P}_2$ est égal à

$$SE(\widehat{P}_1 - \widehat{P}_2) = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \tag{44}$$

et la statistique

$$\eta_y = \frac{\widehat{P}_1 - \widehat{P}_2}{SE(\widehat{P}_1 - \widehat{P}_2)} \tag{45}$$

est asymptotiquement normale de moyenne zero et de variance un. Cette statistique permet de tester le fait que deux indices de pauvreté sont statistiquement les mêmes contre le fait qu'il sont différents.

4.5 Tester la dominance stochastique

Ce problème est beaucoup plus complexe que le précédent. On ne considère plus un simple indice, mais une courbe complète, la courbe de dominance (30). Il va donc s'agir de comparer deux courbes entre elles, ce qui suppose une comparaison point par point. On doit donc traiter simultanément de la comparaison d'un ensemble de points. Comme l'on s'intéresse à la pauvreté, c'est le concept de dominance stochastique restreinte qui est opérant ici. Considérons donc un intervalle $[z_*, z^*]$ qui correspond à deux valeurs extrêmes pour le seuil de pauvreté et deux échantillons A et B pour lesquels on a calculé la courbe de dominance (30) à l'ordre s que l'on notera $F_s^A(x)$ et $F_s^B(x)$ pour les deux échantillons. Il est possible de définir trois hypothèses distinctes:

1. $H_0 : \delta_s(x) = F_s^A(x) - F_s^B(x) = 0 \qquad \forall x \in [z_*, z^*]$

halshs-00277281, version 1 - 6 May 2008

2. $H_1 : \delta_s(x) = F_s^A(x) - F_s^B(x) \geq 0 \qquad \forall x \in [z_*, z^*]$

3. H_2 : aucune restriction sur $D_s(x)$

La première hypothèse dit que les deux distributions des deux échantillons A et B sont équivalentes et qu'elles ne peuvent pas être distinguées. Avec la deuxième hypothèse, les deux distributions sont classées, la distribution B domine clairement la distribution A. La dernière hypothèse n'implique aucune restriction entre les deux distributions, aucune ne domine l'autre, il n'est pas possible de les classer. Si l'on était dans un cadre unidimensionnel, H_0 correspondrait à une hypothèse ponctuelle, H_1 permettrait de construire un test unilatéral et H_2 un test bilatéral. Mais ici ces hypothèses doivent être vérifiées soit sur la partie de l'échantillon contenue dans l'intervalle $[z_*, z^*]$, soit sur une grille de points équidistants couvrant ce même intervalle $[z_*, z^*]$. Choisissons ce dernier cas et considérons la grille suivante de K points équidistants

$$z = [z_k] = z_*, z_2, \cdots, z_{K-1}, z^*. \tag{46}$$

Davidson and Duclos (2000) dégagent un résultat fondamental pour la suite à savoir que

$$\sqrt{n}(\hat{F}_s(z) - F_s(z)) \sim \mathrm{N}(0, \Sigma), \tag{47}$$

et ils donnent la valeur de la matrice Σ pour toute une série de cas. Cette matrice exprime la structure de corrélation qui existe entre les différentes valeurs de $\hat{F}_s(.)$ calculées sur la grille. Les tests que nous allons maintenant examiner tiennent compte de cette structure de corrélation. Appellons $\hat{\delta}(z)$ l'estimation de la différence entre deux courbes de dominance calculées sur la grille $[z_k]$. Si les deux populations sont indépendantes, alors on aura

$$\sqrt{n}(\hat{\delta}(z) - \delta(z)) \sim \mathrm{N}(0, \Omega = \Sigma_A + \Sigma_B). \tag{48}$$

Le test le plus simple à mettre en place est le test de l'hypothèse H_0 contre l'alternative H_2. C'est une simple généralisation du test de Kakwani (1993). Il a été analysé dans Beach and Davidson (1983) et dans Dardanoni and Forcina (1999) pour tester l'égalité de deux courbes de Lorenz. Ce test s'effectue au moyen de la statistique de Wald:

$$T_{02} = \hat{\delta}(z)'\Omega^{-1}\hat{\delta}(z), \tag{49}$$

qui est distribuée selon une χ^2 à K degrés de liberté, K étant la taille de la grille.

Dardanoni and Forcina (1999) envisagent deux autres tests dont l'implémentation est plus complexe, bien que toujours basée sur le résultat de normalité asymptotique (48). Le premier test compare H_0 (l'égalité des deux courbes de dominance) avec H_1 qui représente la dominance de B sur A. L'hypothèse nulle est donc une hypothèse d'équivalence contre une alternative de dominance. Le second test compare H_1 avec H_2. Dans ce cas, l'hypothèse nulle devient la dominance et l'alternative la forme la plus générale de non-dominance. Pour décrire ces tests, il faut pouvoir décrire une hypothèse dans un espace multidimensionnel. Pour cela, Dardanoni and Forcina (1999) définissent une fonction de distance d'un vecteur par rapport à un espace. Ainsi

halshs-00277281, version 1 - 6 May 2008

par exemple la distance du vecteur δ par rapport à l'espace décrit par l'hypothèse nulle H_0 se définit comme

$$d(\delta, H_0, \Omega) = \min_{y \in H_0} (\delta - y)' \Omega^{-1} (\delta - y).$$

Calculer cette distance revient à résoudre un problème de programmation quadratique. Les statistiques de test cherchées vont être

$$T_{01} = n\hat{\delta}(z)'\Omega^{-1}\hat{\delta}(z) - \min_{y \geq 0} n(\hat{\delta}(z) - y)'\Omega^{-1}(\hat{\delta}(z) - y)$$

pour le test de l'égalité contre la dominance et

$$T_{12} = \min_{y \geq 0} n(\hat{\delta}(z) - y)'\Omega^{-1}(\hat{\delta}(z) - y).$$

pour le test de la dominance contre la non-dominance.

Le problème de ces tests, c'est qu'outre qu'ils demandent la résolution d'une problème de programmation quadratique, c'est qu'ils ont une distribution compliquée qui repose sur un mélange de χ^2. Toutefois, Dardanoni and Forcina (1999) font remarquer que le test de l'hypothèse nulle de la non-dominance (l'espace décrit par H_2 dans nos notations auquel on a retiré l'espace décrit par H_1) contre l'hypothèse alternative de la dominance (H_1 dans nos notations) conduit à un résultat de beaucoup simplifié. La simplification tient au fait que l'on ne perd aucune information en négligeant la structure de corrélation de Ω. On retombe alors sur le test proposé par Howes (1993) pour une grille et par Kaur, Prakasa Rao, and Singh (1994) pour tous les points de l'échantillon. Ce test consiste à calculer séparément les K valeurs de la statistique de Student et à en prendre le minimum. Donc la statistique de test cherchée est

$$T_{21} = \min_{z_i} \hat{\delta}(z_i)/\omega_{ii}.$$

Le gros avantage de cette statistique, c'est qu'elle est distribuée selon une $\mathrm{N}(0, 1)$ sous l'hypothèse nulle. Davidson and Duclos (2006) recommandent l'utilisation de ce test. Ils montrent toutefois qu'il ne donne des résultats cohérents que si l'on élimine les queues des distributions, régions dans lesquelles on ne dispose pas de suffisament d'observations. Cette opération de nettoyage s'opère naturellement quand on teste la dominance restreinte si les bornes z_* et z^* sont correctement choisies.

5 Les modèles explicatifs de la pauvreté

Nous nous sommes intéressés jusqu'à présent à la description de la distribution des revenus. On a vu comment comparer deux distributions, soit dans deux pays différents, soit entre deux points du temps pour un même pays. Mais tout cela est resté très descriptif et l'on n'a pas essayé de rendre compte de la formation de la distribution des revenus ou de l'explication du phénomène de pauvreté. Et en cela nous avons été fidèles à cette dichotomie qui existe dans la littérature entre la mesure des inégalités et de la pauvreté et la théorie de la formation des revenus. Le revenu des ménages se décompose en plusieurs parties à savoir les salaires (la plus grosse part), les revenus

halshs-00277281, version 1 - 6 May 2008

locatifs ou financiers, et les impôts et les transferts. Les économistes du travail se sont penchés dans les années 1980 sur l'explication de la dispersion des salaires individuels, mais n'ont jamais relié cette recherche à la mesure des inégalités de revenus des ménages. Le but de cette section n'est pas de combler cette lacune, on regardera pour cela par exemple Atkinson (2003). Mais on va par contre s'attacher à présenter certains des outils économétriques qui permettent d'expliquer la formation et l'évolution de la distribution des revenus et des phénomènes de pauvreté.

5.1 La décomposition des indices

Certains indices d'inégalité et de pauvreté peuvent facilement se décomposer comme on l'a vu avec les indices de Foster, Greer, and Thorbecke (1984). La décomposition se fait alors par groupes de la population. Mais dans cette procédure aucune explication n'est donnée en fonction des caractéristiques des sous groupes. Oaxaca (1973) a le premier tenté une explication des inégalités en utilisant une technique de régression.

Oaxaca (1973) s'intéresse aux inégalités de salaire entre hommes et femmes. Pour chaque groupe, on estime une équation de salaire

$$\log(W_i) = X_i\beta_i + u_i, \qquad i = h, f.$$

On regarde ensuite les différences salariales moyennes entre hommes et femmes. Une partie de cette différence s'explique par des différences de caractéristiques objectives mesurées par $X_h - X_f$, l'autre partie s'explique par des différences de rendements de ces mêmes caractéristiques, c'est à dire par la discrimination que le marché opère entre hommes et femmes. Comme dans une régression OLS $\log(\overline{W}_i) = \overline{X}_i\hat{\beta}_i$, on aura la décomposition suivante appelée décomposition de Oaxaca:

$$\log(\overline{W}_h) - \log(\overline{W}_f) = (\overline{X}_h - \overline{X}_f)\hat{\beta}_h + \overline{X}_f(\hat{\beta}_h - \bar{\beta}_f).$$

Ce type de décomposition a donné lieu à des développements importants dans la littérature. Par exemple Juhn, Murphy, and Pierce (1993) généralisent le résultat précédent aux différents quantiles de la distribution des résidus. Radchenko and Yun (2003) donnent une approche Bayésienne qui permet d'implémenter facilement des tests de significativité.

La décomposition de Oaxaca (1973) est basée sur une hypothèse de régression linéaire. Yun (2004) en donne une généralisation non-linéaire qui permet de proposer une décomposition des mesures de head count. Celles-ci étant assimilables à des proportions, on peut les relier à des variables explicatives au moyen d'un modèle probit. Une des applications dans Yun (2004) concerne justement les modèles probit. Une équation de régression par groupe permet d'expliquer le ratio entre la dépense y et le seuil de pauvreté z:

$$\log\left(\frac{y}{z}\right) = X\beta + e,$$

où X est une matrice de caractéristiques personnelles à k composantes pour n observations. Sous une hypothèse de normalité pour e, la probabilité d'être pauvre pour le groupe des n individus est égale au vecteur $\Phi(-X\beta/\sigma)$ où σ^2 est la variance des résidus. Quand n tend vers l'infini, la moyenne des $\Phi(-X\beta/\sigma)$ tend vers le head count ratio, c'est à dire P_0 dans nos notations. Alors, on aura la décomposition suivante de la différence entre deux indices de pauvreté correspondant à deux groupes distincts A et B:

$$\begin{aligned} P_A^0 - P_B^0 &= [\overline{\Phi(-X_A\beta_A/\sigma_A)} - \overline{\Phi(-X_B\beta_A/\sigma_A)}] \\ &\quad + [\overline{\Phi(-X_B\beta_A/\sigma_A)} - \overline{\Phi(-X_B\beta_B/\sigma_B)}], \end{aligned}$$

ce qui correspond à la différence entre les caractéristiques et la différence entre les coefficients. L'application de la procédure de Yun (2004) permet de pondérer cette décomposition en fonction du poids de chacune des k caractéristiques individuelles

$$\begin{aligned} P_A^0 - P_B^0 &= \textstyle\sum_{i=1}^k W_{\Delta X}^i [\overline{\Phi(-X_A\beta_A/\sigma_A)} - \overline{\Phi(-X_B\beta_A/\sigma_A)}] \\ &\quad + \textstyle\sum_{i=1}^k W_{\Delta\beta}^i [\overline{\Phi(-X_B\beta_A/\sigma_A)} - \overline{\Phi(-X_B\beta_B/\sigma_B)}] \end{aligned}$$

où les poids $W_{\Delta X}^i$ et $W_{\Delta\beta}^i$ sont donnés dans Bhaumik, Gang, and Yun (2006) sur un argument de linéarisation de Yun (2004).

5.2 Les régressions quantiles

Un modèle de régression donne le lien linéaire ou non-linéaire qui existe entre une variable endogène et une ou plusieurs variables explicatives. Dans le cas le plus simple, la ligne de régression représente une espérance conditionnelle linéaire. Une régression non-paramétrique va expliquer une espérance conditionelle de façon non-linéaire. Mais aucune de ces régressions ne se préoccupe d'expliquer tous les quantiles de la distribution de la variable endogène. La régression quantile, introduite en économétrie par Koenker and Basset (1978) fournit l'instrument recherché pour expliquer l'évolution complète de la variable endogène. Le principe est simple, mais sa réalisation numérique est plus complexe, en particulier pour calculer des écart-types.

Considérons un modèle linéaire de régression $y = x'\beta + e$ à propos duquel on ne fait pour le moment aucune hypothèse sur la distribution conditionelle de y. On va noter $q_\alpha(x)$ le quantile de niveau α qui se définit par

$$q_\alpha(x) = F^{-1}(\alpha)$$

On veut expliquer ce quantile par une combinaison linéaire des x

$$q_\alpha(x) = x'\beta$$

Il faut tout d'abord noter que si F est la distribution normale, nous n'apprendrons pas grand chose par ce modèle, car premièrement pour cette densité la moyenne et la médiane sont confondues et qu'ensuite les quantiles conditionnels sont des droites. Il faut s'écarter de ce cadre traditionnel pour que le modèle prenne tout son sens. Une première possibilité consiste à avoir des erreurs hétéroscédastiques, c'est à dire que si les erreurs sont normales de variance σ^2, ce paramètre va être fonction des variables exogènes. Une hypothèse plus extrême consiste à ne faire aucune

halshs-00277281, version 1 - 6 May 2008

halshs-00277281, version 1 - 6 May 2008

hypothèse paramétrique sur la distribution F et à travailler dans un cadre semi-paramétrique. Pour cela on va définir une fonction d'erreurs

$$\rho_\alpha(u) = \begin{cases} u\alpha & \text{si } u > 0 \\ u(\alpha - 1) & \text{sinon.} \end{cases}$$

On va chercher ensuite à minimiser en β la quantité $\sum \rho_\alpha(y_i - x_i'\beta)$, ce qui revient à résoudre un problème de programmation quadratique. En utilisant un cadre Bayésien, Yu and Moyeed (2001) montrent que l'estimation du quantile de y peut être rendue équivalente à l'estimation du paramètre de localisation d'une distribution asymétrique de Laplace. Ceci permet d'écrire une fonction de vraisemblance avec

$$L(\beta|y, x) \propto \alpha^n (1 - \alpha)^n \exp\{-\sum_i \rho_\alpha(y_i - x_i'\beta)\}$$

pour évaluer ensuite la densité a posteriori des paramètres. Il devient alors facile d'estimer des écart-types, ce qui n'est par contre pas du tout évident dans le cadre initial proposé par Koenker and Basset (1978).

La régression quantile a été très souvent employée pour rendre compte de l'évolution des inégalités salariales à partir d'une équation Mincerienne de salaire, ce qui permet de mesurer les différences de rendements de l'éducation et de l'expérience. Voir par exemple Machado and Mata (2005) pour une étude récente. La régression quantile a été utilisée pour rendre compte des phénomènes de pauvreté dans certains pays du tiers monde en expliquant l'écart de revenu entre les villes et les campagnes. Ainsi Binh T. Nguyen (2007) montrent que pour le Vietnam l'écart ville-campagne dans la distribution des revenus est dû à des différences d'éducation pour les quantiles les plus faibles, mais que pour le reste de la distribution, cet écart est principalement du à des différences dans le rendement des attributs. On peut déduire de ces résultats qu'une politique de réduction de la pauvreté dans les campagnes doit se centrer sur l'éducation.

5.3 La dynamique des états de pauvreté

La pauvreté est un phénomène qui évolue dans le temps. Même si un indice de pauvreté reste constant ou évolue peu, ce ne seront pas les mêmes personnes qui resteront dans l'état de pauvreté. Les ménages font l'expérience d'épisodes de pauvreté, avec une entrée, une durée et une sortie. Il s'agit d'expliquer ces mouvements. Pour cela, nous avons besoin de données d'enquêtes spécifiques où l'on suit un même ménage plusieurs années consécutives. Ce sont des panels dont on verra plusieurs exemples dans la section suivante. Jenkins (2000), après avoir dressé un tableau de la situation en Grande Bretagne entre 1991 et 1996, passe en revue les différents modèles qui peuvent être utilisés. Il cite en particulier les résultats et l'approche de Stevens (1999), dont on ne retiendra qu'un seul modèle. On va estimer les paramètres d'un modèle dynamique de séries temporelles qui explique l'évolution du logarithme du income-to-needs ratio en fonction d'un certain nombre de variables exogènes et d'erreurs composées. Les paramètres de ce modèle sont ensuite utilisés pour former des jugements sur la persistance de la pauvreté. Le income-to-needs ratio est une mesure couramment utilisée aux USA pour évaluer la situation économique d'un ménage. Il est calculé en prenant le revenu du ménage sans tenir compte des diverses allocations et en le divisant par le seuil de pauvreté officiel correspondant à la composition de ce ménage. Le modèle de base est le suivant:

$$\begin{aligned}\log(Y_{it}) &= Z_i\alpha + X_{it}\beta + \epsilon_{it} \\ \epsilon_{it} &= \delta_i + v_{it} \\ v_{it} &= \gamma v_{it-1} + \eta_{it}\end{aligned}$$

Le revenu est expliqué par des variables individuelles qui sont fixes au cours du temps comme le sexe, le niveau d'éducation, l'origine des parents et par des variables individuelles qui varient au cours du temps. Le paramètre δ_i mesure un effet d'hétérogénéité individuelle aléatoire de moyenne nulle et de variance σ_δ. Le paramètre γ modélise une permanence commune à tous les individus. On peut dire que les individus reçoivent des shocks v_{it} qui perdurent. Sous une hypothèse de normalité pour δ_i et η_{it}, Stevens (1999) simule ce modèle pour 20 années et calcule le temps moyen passé sous la ligne de pauvreté. Estimé sur les données du PSID, ce modèle indique que ce temps moyen est en général très légèrement supérieur si le chef de ménage est noir ou est une femme.

6 L'évolution de l'accès aux bases de données

L'analyse de la distribution des revenus et la mesure de la pauvreté suppose que l'on puisse avoir accès à des données individuelles d'enquête. Si ces données d'enquête existent depuis longtemps, leur champ s'est considérablement élargi depuis vingt ans. Ainsi le PSID américain (Panel Study of Income Dynamics), la base micro-économique la plus connue, est passé de 400 variables en 1968 à 3000 en 2005. Mais l'accès à ces bases de données pose le problème de la confidentialité et donc des procédures d'anonymisation. L'examen statistique des chiffres ne doit pas permettre de retrouver la personne interrogée; par exemple il ne doit pas permettre de publier la feuille d'impôt d'un personnage connu. Or il est évident qu'au plus l'enquête comporte de variables renseignées, au plus il sera difficile d'anonymiser les informations contenues dans cette enquête. C'est sur cet arbitrage que repose toute la problématique d'accès aux données et les restrictions imposées par les agences statistiques qui les produisent. On va essayer de faire un état des lieux et de son évolution dans différents pays.

6.1 La France

En France, l'INSEE est le principal pourvoyeur de données d'enquêtes. L'accès à ces données a toujours posé problème pour les chercheurs, si bien qu'en 1999 Claude Allègre, Ministre de la Recherche donne à Roxane Silberman une Lettre de Mission pour mettre en place les instruments nécessaires pour en faciliter l'accès . Le rapport Silberman de 1999 aboutira en 2001 à la création d'une unité mixte de service par le CNRS, le Centre Quetelet. Ce centre a effectué tout un travail

halshs-00277281, version 1 - 6 May 2008

de mise en forme de différentes enquêtes. En ce qui concerne la mesure de la pauvreté, il s'agit principalement des enquêtes successives de l'INSEE sur les budgets des familles.

De cette initiative, il ne reste aujourd'hui que le réseau Quetelet qui est un portail de recherche sur la disponibilité de données en sciences sociales: *http://www.centre.quetelet.cnrs.fr*. Il renvoie entre autres au Centre Maurice Halbwachs (ancien LASMAS) à Paris qui se charge maintenant de leur diffusion. Leur accès est règlementé, mais gratuit. On est donc revenu à la situation de départ. Le chercheur dépend de la politique de diffusion de l'INSEE, qui est devenu plus restrictive au fil du temps. Il y a une différentielle d'accès très marquée entre les chercheurs faisant partie d'un centre rattaché à l'INSEE et les autres. Il est par exemple très difficile, si ce n'est quasi impossible, d'obtenir le croisement de deux fichiers.

6.2 L'Europe

Le chercheur français peut se tourner vers le réseau Européen EQUALSOC (Economic Change, Quality of Life and Social Cohesion *http://www.equalsoc.org*) pour avoir un accès direct et gratuit au Panel Européen, après agrément. Le Panel Européen était une enquête européenne sur les conditions de vie des ménages mis en place par EUROSTAT voici une quinzaine d'années. En 2001, cette enquête a été remplacée par une enquête plus large sur les revenus et les conditions de vie dans l'Union Européenne (EU-SILC, "European Union-Statistics on Income and Living Conditions"). La fourniture d'informations statistiques a désormais force de loi pour l'ensemble des pays de l'Union alors que ce n'était pas le cas pour l'ancien panel. Cet outil devrait permettre de mieux analyser les phénomènes d'exclusion et l'impact des politiques sociales et fiscales sur la redistribution au niveau Européen.

Le Luxembourg Income Study (LIS, *http://www.lisproject.org*) correspond à un projet déjà ancien (1983). Il collecte des données d'enquête sur les ménages pour une trentaine de pays. Outre les pays Européens, participent également l'Australie, le Canada, Israel, le Mexique, Taiwan, la Russie et les USA. A partir des données d'enquêtes nationales, le LIS produit une base de données individuelles harmonisées portant sur les revenus et les autres variables du même champ. Il a servi de base à de nombreuses études de comparaisons internationales (voir Jenkins et Micklewright 2007 pour un survey). On notera le rôle pionnier du LIS dans l'accès aux données individuelles. Mais si cet accès est gratuit, il est toutefois limité. On ne peut accéder à ces données que sous la forme de l'exécution d'un programme SAS avec impossibilité de récupérer les données elles-mêmes en vue d'un traitement sur un autre logiciel.

6.3 Le Royaume Uni et les USA

La situation dans ces deux pays est relativement particulière et si l'on peut dire plus en accord avec la pratique des chercheurs. Dans les années 1970, les chercheurs britanniques n'avaient accès qu'à des données groupées. Ils peuvent maintenant accéder directement aux données du Family Expenditure Survey (FES) sur les trente dernières années en les téléchargeant sur *http://www.data-archive.ac.uk/findingData/fesTitles.asp*. Mais il faut être un utilisateur enreg-

halshs-00277281, version 1 - 6 May 2008

halshs-00277281, version 1 - 6 May 2008

istré et avoir un mot de passe. Apparemment seules les universités britanniques y ont accès.

Le Panel Study of Income Dynamics (PSID), a débuté en 1968 aux USA. Il s'agit de données de panel fournissant un échantillon représentatif des conditions de vie de la population des USA. Il concerne environ 8 000 ménages suivis sur une base annuelle. Ce sont toujours les mêmes ménages qui sont interrogés. On peut télécharger les données sur le site *http://psidonline.isr.umich.edu*. Il faut être un utilisateur enregistré, mais l'enregistrement ne semble pas très restrictif.

6.4 La Banque Mondiale et les pays en développement

Depuis 1980, la Banque Mondiale s'est lancée dans la constitution de bases de données micro-économiques les *Living Standards Measurement Surveys* (LSMS). Ces enquêtes concernent une quarantaine de pays, en très grande majorité des pays en voie de développement. On peut accéder à ces données sur le site de la Banque Mondiale: *http://www.worldbank.org/lsms*. Deaton (1997) donne un bon aperçu des études qui ont été faites à partir de ces données. On trouvera également sur ce site les codes sources de programmes Stata utilisés dans Deaton (1997).

7 Conclusion

La mesure des inégalités et de la pauvreté est devenu un domaine scientifique bien structuré, son axiomatique est maintenant bien établie. La question de la mesure empirique continue de poser des problèmes non-triviaux. Le premier des problèmes concerne l'accès aux données d'enquête. Si beaucoup d'études portent sur les pays en développement, c'est bien souvent que les problèmes de confidentialité et d'accès aux données y sont moins cruciaux. Par exemple le papier de Kyosuke and Kurosaki (2007) est basé sur des long panels portant sur les Philippines et la Thailande pour analyser la relation entre croissance, inégalité et pauvreté.

Les comparaisons internationales posent le problème du choix et de la définition des variables. On l'a bien vu à propos de la comparaison France-Royaume Uni. Le second problème concerne le traitement statistique de ces données. Trop souvent encore, les études, en se focalisant sur l'axiomatique, tendent à ignorer que les données proviennent de variables aléatoires et doivent donc impérativement donner lieu à l'estimation d'écart-type. Deux indices qui paraissent différents peuvent dans les faits ne pas être statistiquement différents l'un de l'autre. Enfin, la pauvreté n'est pas un problème statique. La situation des ménages, des groupes évolue dans le temps. Même si un indice de pauvreté reste le même au cours du temps, la composition du groupe des pauvres change, les ménages subissent des épisodes de pauvreté. Il s'agit alors pour mieux cibler les politiques publiques de déterminer les facteurs qui conduisent à entrer et à sortir de l'état de pauvreté. Pour cela on a besoin de deux choses: de modèles structurels, c'est à dire pleinement explicatifs et de données de panel.

halshs-00277281, version 1 - 6 May 2008

References

ATKINSON, A. (1970): "The measurement of inequality," *Journal of Economic Theory*, 2, 244–263.

——— (1987): "On the measurement of poverty," *Econometrica*, 55, 749–764.

——— (1998): *Poverty in Europe*. Blackwell, Oxford.

——— (2003): "Income Inequality in OECD Countries: Data and Explanations," Working Paper 49: 479–513, CESifo Economic Studies, CESifo.

BEACH, C. M., AND R. DAVIDSON (1983): "Distribution-Free Statistical Inference with Lorenz Curves and Income Shares," *The Review of Economic Studies*, 50, 723–735.

BERGER, Y. G., AND C. J. SKINNER (2003): "Variance Estimation of a Low-Income Proportion," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 52, 457–468.

BHAUMIK, S. K., I. N. GANG, AND M.-S. YUN (2006): "A Note on Decomposing Differences in Poverty Incidence Using Regression Estimates: Algorithm and Example.," Working Paper 2006-33, Department of Economics, Rutgers University, Rutgers University.

BINH T. NGUYEN, JAMES W. ALBRECHT, S. B. V. M. D. W. (2007): "A quantile regression decomposition of urban–rural inequality in Vietnam," *Journal of Development Economics*, 83(2), 466–490.

CARLSSON, F., D. DARUVALA, AND O. JOHANSSON-STENMAN (2005): "Are People Inequality-Averse, or Just Risk-Averse?," *Economica*, 72, 375–396.

DARDANONI, V., AND A. FORCINA (1999): "Inference for Lorenz curve orderings," *The Econometrics Journal*, 2(1), 49–75.

DAVIDSON, R. (2007): "Reliable Inference for the Gini Index," Working Paper 2007-23, GREQAM, Marseille.

DAVIDSON, R., AND J.-Y. DUCLOS (2000): "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica*, 68, 1435–1464.

——— (2006): "Testing for Restricted Stochastic Dominance," working paper 2006-36, ECINEQ.

DEATON, A. (1997): *The Analysis of Household Surveys*. The John Hopkins University Press, Baltimore and London.

FLACHAIRE, E., AND O. NUNEZ (2007): "Estimation of the income distribution and detection of subpopulations: An explanatory model," *Computational Statistics and Data Analysis*, 51(7), 3368–3380.

halshs-00277281, version 1 - 6 May 2008

FOSTER, J., J. GREER, AND E. THORBECKE (1984): "A class of decomposable poverty measures," *Econometrica*, 52, 761–765.

HOWES, S. (1993): "Asymptotic Properties of Four Fundamental Curves of Distributional Analysis," Unpublished paper, STICERD, London School of Economics.

JENKINS, S. P. (2000): "Modelling Household Income Dynamics," *Journal of Population Economics*, 13, 529–567.

JUHN, C., K. M. MURPHY, AND B. PIERCE (1993): "Wage Inequality and the Rise in Returns to Skill," *The Journal of Political Economy*, 101(3), 410–442.

KAKWANI, N. (1993): "Statistical inference in the measurement of poverty," *Review of Economics and Statistics*, 75, 632–639.

KAUR, A., B. PRAKASA RAO, AND H. SINGH (1994): "Testing for Second-Order Stochastic Dominance of Two Distributions," *Econometric Theory*, 10, 849–866.

KOENKER, R., AND G. BASSET (1978): "Regression quantiles," *Econometrica*, 46(1), 33–50.

KOLM, S. C. (1969): "The Optimal Production of Social Justice," in *Public Economics*, ed. by J. Margolis, and H. Guitton, p. 145–200. Macmillan, London.

KOLM, S.-C. (1976): "Unequal inequalities I," *Journal of Economic Theory*, 12, 416–442.

KYOSUKE, K., AND T. KUROSAKI (2007): "The Dynamics of Growth, Poverty, and Inequality: A Panel Analysis of Regional Data from the Philippines and Thailand," Discussion Paper 223, Institute of Economic Research, Hitotsubashi University.

MACHADO, J. A. F., AND J. MATA (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(4), 445 – 465.

MARRON, J., AND H. SCHMITZ (1992): "Simultaneous Density Estimation of Several Income Distributions," *Econometric-Theory*, 8(4), 476–488.

OAXACA, R. (1973): "Male-female wage differentials in urban labor markets," *International Economic Review*, 14, 693–709.

OKUN, A. (1975): *Equality and Efficiency: The Big Tradeoff*. The Brookings Institution, Washington (D.C.).

RADCHENKO, S. I., AND M.-S. YUN (2003): "A Bayesian approach to decomposing wage differentials," *Economics Letters*, 78(3), 431–436.

RAWLS, J. (1972): *A Theory of Justice*. Oxford University Press, London.

SEN, A. (1976): "Poverty: an Ordinal Approach to Measurement," *Econometrica*, 44, 219–231.

halshs-00277281, version 1 - 6 May 2008

SEN, A. K. (1997): *On Economic Inequality*. Clarendon Press, Oxford, expanded edition with a substantial annexe by j. e. foster and a.k. sen. edn.

SHORROCKS, A. F. (1995): "Revisitng the Sen poverty index," *Economerica*, 63(5), 1225–1230.

SINGH, S., AND G. MADDALA (1976): "A function for the size distribution of incomes," *Econometrica*, 44, 963–970.

STEVENS, A. H. (1999): "Climbing out of poverty, falling back in," *The Journal of Human Ressources*, 34(3), 557–588.

YU, K., AND R. A. MOYEED (2001): "Bayesian quantile regression," *Statistics and Probability Letters*, 54, 437–447.

YUN, M.-S. (2004): "Decomposing differences in the first moment," *Economics Letters*, 82(2), 275–280.