GRÉDI 

Groupe de Recherche en Économie et Développement International

Cahier de recherche / Working Paper
06-05

La décomposition des mesures d'inégalité en sources de revenu : l'indice de Gini et les généralisations

Stéphane Mussard

UNIVERSITÉ DE SHERBROOKE

La décomposition des mesures d'inégalité en sources de revenu : l'indice de Gini et les généralisations

Stéphane MUSSARD *

* GREDI, Département d'économique
Université de Sherbrooke
Québec, Canada
GEREM, Département des Sciences Economiques
Université de Perpignan,
Avenue Paul Alduy
66100 Perpignan – France.
E-mail: smussard@adm.usherbrooke.ca

La décomposition des mesures d'inégalité en sources de revenu : l'indice de Gini et les généralisations

RESUME. – La littérature montre que la mesure des inégalités de revenu n'est pas suffisante pour appréhender les déterminants des inégalités. Les sources de rémunération, leurs corrélations au revenu global, leurs parts dans le revenu global, rentrent en compte dans la décomposition de la mesure de Gini, faisant aussi apparaître la mesure du pseudo-Gini. Les dernières méthodes en date sont aussi exposées afin de mettre en évidence les possibilités de généralisation dans ce domaine.

The Decomposition of the Inequality Measures by Income Sources: The Gini Ratio and the Generalisations

ABSTRACT. – The literature points out that the income inequality measures are not sufficient to apprehend the determinants of the inequalities. The factors of remuneration, their correlations to the overall income, their shares in the mean income, yield the Gini decomposition by income sources and the so-called pseudo-Gini index. The last techniques are also exposed in order to bring out the generalisations of the income source decomposition.

I. Introduction

Les inégalités dans la répartition des revenus et celle des fortunes doivent-elles être réduites ? Plusieurs économistes l'ont soutenues. Pareto et Keynes par exemple. En 1896, Pareto préconise un lissage de la distribution des revenus au détriment des plus élevés et en faveur des plus faibles. En 1936, dans le dernier chapitre de la Théorie générale, Keynes n'hésite pas à conclure que sa conception du fonctionnement des économies de marché justifie une réduction sensible des inégalités arbitraires et inéquitables qui caractérisent la distribution des revenus et la répartition des fortunes au sein du monde dans lequel il vit. Ces prises de position permettent de comprendre que la réflexion sur la mesure des inégalités ait pu constituer et constitue l'un des thèmes de recherche privilégié des économistes.

En ce domaine, les travaux de Lorenz (1905) et ceux de Gini (1912, 1914, 1916) peuvent être considérés comme précurseurs. Ils constituent les points de départ des recherches développées ultérieurement par Kolm en 1966, Atkinson en 1970 et Sen en 1973. Toutefois ce sont les résultats obtenus en 1969 par Rao qui vont retenir notre attention.

Dés 1969, Rao essaie de comprendre la détermination des mesures d'inégalité du revenu en privilégiant soit une décomposition en sous-groupes soit une décomposition en sources de revenus. La première est consacrée à la détermination des inégalités à l'intérieur des groupes et entre les groupes. La seconde, celle qui fait l'objet de notre réflexion, appelée aussi décomposition en facteurs, s'intéresse au poids des différents revenus (revenus du travail, revenu du capital, prestations sociales, taxes, etc.) dans la mesure de l'inégalité globale. En 1982, Shorrocks évoque des doutes quant à la crédibilité de la décomposition de la mesure de Gini en facteurs. Mais le nombre de publications privilégiant cette seconde piste de recherche témoigne de l'engouement qu'elle a suscité. Parmi elles, citons celles qui font le lien avec la croissance et le développement économique [Fei, Ranis et Kuo (1978)], la consommation [Garner (1993)] ou encore l'analyse économétrique [Murdoch et Sicular (2002)]. Parmi ces publications, celles dont l'apport nous paraît le plus important sont dues à Chantreuil et Trannoy (1999), et à Shorrocks (1999). Ces recherches donnent l'occasion d'associer la Valeur Shapley, concept central de la théorie des jeux coopératifs, et la généralisation de la décomposition en sources de revenu.

II. La genèse de la décomposition en facteurs : Rao (1969)

La paternité de la décomposition de la mesure de Gini en sources de revenu peut être attribuée à Rao (1969). Après avoir, spécifié une décomposition matricielle de l'indice de Gini en sous-groupes, l'auteur exprime dans la deuxième partie de sa réflexion une méthode permettant de séparer le coefficient de Gini en contributions représentant les différentes sources de rémunération.

Soit une population composée de n individus ($\forall i = 1,\ldots, n$). Les revenus des individus comprennent q sources de revenu indexées par $m = 1,\ldots, q$. Soit la fonction $R_m(i)$. Elle exprime le rang du $i^{\text{ième}}$ individu lorsque les sources de nature m sont ordonnées de manière croissante. Parallèlement, on peut définir le rang de l'individu i lorsque les revenus de la distribution globale sont ordonnés de manière croissante : $R_t(i)$. Avant les années 1980, les inégalités de revenu sont étudiées en classes. Nous verrons par la suite la conséquence principale de ces découpages en classes. On considère que les distributions de chaque source de revenu m sont étudiées sur k classes S_{jm} définies par le classement croissant $R_m(i)$ (on a donc k quantiles : $j = 1,\ldots, k$ comprenant n/k individus). La distribution du revenu total est aussi divisée en k classes S_{jt} comprenant chacune n/k individus et où les individus sont classés selon la répartition $R_t(i)$. Soit ensuite la somme cumulée Q_{jm} (sur les classes S_{jt}, non S_{jm}) des proportions des sources m (dans le revenu total) inhérentes à chaque classe S_{jt}. Soit la somme cumulée Q_{jt} (sur les classes S_{jt}, non S_{jm}) des proportions de revenu total inhérentes à chaque classe S_{jt}. Par conséquent, $\sum_m Q_{jm} = Q_{jt}$. Soit le terme $Q_{\cdot m}$ désignant la proportion de revenu total issue du facteur m et P_{jt} la somme cumulée des proportions d'individus sur les classes S_{jt}. On obtient donc le coefficient de concentration[1] :

$$c = \sum_{j=1}^{k-1} \left(P_{jt} \times Q_{(j+1)t} - P_{(j+1)t} \times Q_{jt} \right). \tag{1}$$

Soit le ratio $Q^*_{jm} = Q_{jm} / Q_{\cdot m}$. En appliquant la formule (1) aux variables P_{jt} et Q^*_{jm} on obtient l'indicateur de concentration suivant :

$$c^*_m = \sum_{j=1}^{k-1} \left(P_{jt} \times Q^*_{(j+1)m} - P_{(j+1)t} \times Q^*_{jm} \right). \tag{2}$$

[1] Rao exprime sa décomposition en fonction du coefficient de concentration. L'indice de Gini est un cas particulier de l'indice de concentration [cf. Pyatt, Chen et Fei (1980) et section V ci-après]. Ce dernier peut être négatif, notamment lorsque les classes sont ordonnées de manière décroissante.

Par rapport à la mesure c_m qui désigne le coefficient de concentration mesuré sur la source de revenu m, l'indicateur c^*_m est atypique. Le calcul de c^*_m s'effectue en prenant en compte le rang de la population globale et non celui de la source m [c'est-à-dire les classes S_{jt} dépendantes du classement $R_t(i)$]. En effet, c^*_m est obtenu à partir de S_{jt} alors que c_m est obtenu à partir de S_{jm} :

$$c_m = \sum_{j=1}^{k-1} \left(P_{jm} \times Q_{(j+1)m} - P_{(j+1)m} \times Q_{jm}\right). \tag{3}$$

La décomposition de l'indicateur de concentration en sources de revenu s'exprime alors sous la forme suivante :

$$c = \left(Q_{\cdot m}\right)' \left(c_m^*\right), \tag{4}$$

où $(Q_{\cdot m})'$ est le vecteur colonne transposé des proportions des sources m dans le revenu global ; et où (c^*_m) est le vecteur colonne des q indicateurs de concentration correspondant aux q sources de revenu. Cependant, étant donné que la décomposition est basée sur l'indice c^* (qui n'est pas véritablement l'indice de concentration), Rao reformule l'expression précédente afin de proposer une décomposition exacte et conforme au coefficient c_m :

$$c = \left(Q_{\cdot m}\right)' \left(c_m\right) - \left(Q_{\cdot m}\right)' \left(c_m f_m\right), \tag{5}$$

où (c_m) est le vecteur colonne des coefficients de concentration pour les sources m, où $f_m \in [0,2]$ est défini par $f_m = 1 - c_m^*/c_m$ et où $(c_m f_m)$ est le vecteur colonne des produits scalaires $c_m.f_m$. Lorsque les classes S_{jt} et S_{jm} son identiques, $c^*_m = c_m$. De ce fait, $f_m = 0$ et : $c = \left(Q_{\cdot m}\right)' \left(c_m\right).$ (6)

Outre la difficulté technique soulevée par cette décomposition, la méthode reste particulièrement intéressante. L'indice global est une moyenne pondérée des indicateurs de concentration par source et de la proportion des ces sources dans la distribution globale. Une autre condition plus forte permet d'atteindre ce résultat : $R_m(i) = R_t(i), \forall m$. Autrement dit, les individus doivent avoir le même rang dans la distribution de revenu total et dans les distributions des sources de revenu. Il s'agit d'une condition nécessaire mais non suffisante pour réunir la condition $S_{jt} = S_{jm}$ et atteindre la relation (6). Empiriquement, ce résultat est très difficile à obtenir. On a donc recours à la relation (5). Elle indique que si l'on accède à la parfaite répartition au sein des distributions des q sources de revenu, alors l'inégalité globale est nulle. Le premier produit $\left(Q_{\cdot m}\right)' \left(c_m\right)$ est considéré comme la valeur maximale que c peut atteindre. Le deuxième $\left(Q_{\cdot m}\right)' \left(c_m f_m\right)$

mesure l'intensité avec laquelle les inégalités issues des sources peuvent s'annuler. En effet, $c = 0$ lorsque c_m est égal à zéro ou bien lorsque le second terme annule l'effet du premier. Les interprétations peuvent être classifiées par rapport au terme f_m.

- Si $f_m = 0$, le deuxième élément (les inégalités provenant des sources autres que m) ne permet pas d'atténuer les inégalités dérivées de la source m. Il s'agit d'une inégalité de « supplément ».
- Si $f_m = 2$, les inégalités impulsées par les sources (exceptées m) viennent contrer l'effet des inégalités issues du facteur m.
- Si $f_m = 1$, l'effet des deux éléments est neutre.

En définitive, l'approche de Rao (1969) est fondamentale. En premier lieu, elle repose sur une approche matricielle. Silber (1989) estime, à ce propos, que les approches décomposées sont souvent basées sur des algorithmes intéressants mais encore difficiles à évaluer, même en utilisant l'outil informatique. Il propose, comme l'a fait Rao, de simplifier la décomposition en sources de revenu en recourant à une approche matricielle [cf. aussi Silber (1993)]. Deuxièmement, la méthode de Rao procure deux types de décomposition : l'une est basée sur l'indice de concentration c_m inhérent à chaque source de rémunération, l'autre est un indice de concentration hybride, que la littérature nommera pseudo-Gini.

III. Le pseudo-Gini de Fei, Ranis et Kuo (1978) : le lien avec l'économie du développement

Fei, Ranis et Kuo (que nous noterons désormais FRK) construisent la décomposition de la mesure de Gini à des fins spécifiques : celles de l'étude de la croissance économique, du développement économique, et de la détermination des sources d'inégalité qui séparent le comportement des familles riches et celui des familles pauvres. La démarche des auteurs peut donc être qualifiée d'approche statistico-économique, alors que le modèle de Rao (1969) est purement statistique.

Kuznets (1955) établit un lien étroit unissant les inégalités de revenu et la croissance économique. FRK étendent cette idée. Comment la croissance économique s'est-elle aussi rapidement accrue ? Pour répondre à cette question, l'analyse de l'histoire économique d'un pays est une bonne piste. Une alternative consisterait à appréhender les déterminants de la distribution des revenus. On considère pour cela une distribution constituée de n ménages $(x_1, \ldots, x_i, \ldots, x_n)$, où chaque ménage i possède un nombre fini de sources de revenu telles que : $x_i = x_i^1 + ... + x_i^m + ... + x_i^q$. On note G, l'indice de Gini global et G_m l'indice de Gini inhérent à la source de revenu m. De manière intuitive, l'indicateur global peut se décomposer à l'aide des indices G_m. Il suffit d'envisager une moyenne pondérée telle que : $G = \varphi_m G_m$. Le but des auteurs est de fournir une décomposition exacte et cohérente, en s'appuyant sur l'approximation suivante :

$$G = \hat{G} + \theta , \tag{7}$$

où, $$\hat{G} = \sum_{m=1}^{q_1} \varphi_m G_m + \sum_{m=1}^{q_2} \varphi_m G_m - \sum_{m=1}^{q_3} \varphi_m G_m , \text{ et } \sum \varphi_m = 1. \tag{8}$$

Les facteurs de revenu sont regroupés en trois types. Les sources « type 1 » (ces sources sont indexées par q_1) sont les facteurs détenus en plus grande quantité (de manière absolue) par les ménages riches et qui, proportionnellement, sont beaucoup plus importants chez les familles riches que chez les familles pauvres. Ces revenus sont liés à la propriété privée (les biens matériels, fonciers, etc.). Les facteurs « type 2 » (indexés par q_2) représentent les sources qui sont absolument plus importantes chez les familles riches mais proportionnellement moins importantes que chez les familles pauvres. Il s'agit du salaire. Enfin, les facteurs « type 3 » associés au signe (−) décroissent de manière absolue pour les familles qui sont les plus dotées en bien-être. Ces sources sont les revenus de transfert.

Cette classification est basée sur trois types de sources de revenu qui permettent d'élaborer un modèle général estimé sur un panel de Taiwan (1964, 1966, 1968, 1970, 1971 et 1972). Le principal résultat de l'estimation économétrique de l'équation (7) est la faiblesse des sources de type 3. Ce modèle est cependant beaucoup plus apte à rendre compte du lien qui prévaut entre les inégalités et la croissance économique. Considérons une économie composée d'une seule branche et deux facteurs de revenu (q = 2) : le capital (K) et le travail (L). Par conséquent, l'indice de Gini global peut se réécrire comme suit :

$$G = \varphi_k G_k + \varphi_l G_l , \tag{9}$$

où φ_k et φ_l sont respectivement les parts de capital et de travail.[2] La différentielle de l'expression précédente permet de déterminer deux impacts marginaux :

$$\frac{dG}{dt} = D + I , \tag{10}$$

[2] Ces parts sont calculées sur la base du revenu : source moyenne / revenu moyen.

où $\quad I=\varphi_l\left(\frac{dG_l}{dt}\right)+\varphi_k\left(\frac{dG_k}{dt}\right)$ et $D=\left(G_k-G_l\right)\frac{d\varphi_l}{dt}$. (11)

La statique comparative permet d'expliquer le changement du Gini total (G) en fonction d'une période considérée. Le premier effet (D) est un effet de distribution fonctionnelle. Il décrit le changement de l'inégalité totale inhérente aux variations des parts de chaque source de revenu : capital et travail. Le second (I) est l'effet des facteurs de Gini. Il représente les changements de G au cours du temps par l'intermédiaire des impacts favorables ou défavorables des indices de Gini pour chaque facteur. Lorsque les salaires sont répartis de manière plus égalitaire que les capitaux ($G_l < G_k$), et lorsque les salaires sont favorisés (φ_l augmente), alors l'effet de distribution fonctionnelle va améliorer les distributions de revenu des familles.

Afin d'évaluer l'impact de chaque facteur, FRK introduisent le concept de pseudo-Gini sur la base de la courbe de Lorenz. Le coefficient de Gini est égal à $G = 1\!-\!2A$ où A est l'aire située entre la courbe de Lorenz et l'axe des abscisses. L'aire peut se formuler de la manière suivante :

$$A=1-\left(\sum_{i=1}^{n} iQ_i/n \;+\; (n)/2\right), \tag{12}$$

où Q_i renvoie à la proportion de revenu totale détenue par $i: Q_i = x_i/\sum_{i=1}^{n} x_i$. Cette expression repose sur une transformation monotone non décroissante des poids (i) et des proportions de revenus (Q_i). L'indicateur se réécrit alors :

$$G=(2/n)\;u_x-(n+1)/n, \tag{13}$$

où $u_x=\sum_{i=1}^{n} iQ_i$. Soit μ^m la moyenne arithmétique de la source de m et μ celle du revenu global. Ensuite, définissons q_i^m comme étant la part de la source m totale détenue par l'individu i, et φ_m comme étant la part moyenne de la source m dans le revenu total (μ^m/μ). Si le terme $\bar{u}_m$ est un pseudo-indicateur de u_x pour la source m tel que $\bar{u}_m=\sum_{i=1}^{n} iq_i^m$, alors l'indice de Gini peut s'exprimer par :

$$\begin{aligned} G &= \sum_{m=1}^{q}\varphi_m\left[(2/n)\bar{u}_m-(n+1)/n\right] \\ &= \sum_{m=1}^{q}\varphi_m\,\bar{G}_m. \end{aligned} \tag{14}$$

Pourquoi trouve-t-on des pseudo-Gini $\overline{G}_m$? En fait, la condition de monotonie est perdue. On imagine que les éléments q_i^m du terme $\overline{u}_m$ ne sont pas classés par ordre croissant. Ils sont ordonnés selon le classement croissant des revenus : $x_1 \leq x_2 \leq \ldots \leq x_n$. Dans ces conditions, les éléments q_i^m ne respectent pas systématiquement cette même condition de monotonie. On retrouve alors la discussion de Rao (1969) sur la différence des rangs qui prévaut entre le vecteur de revenu global et les vecteurs de chaque source de revenu, et les conditions qu'il faut réunir afin que le pseudo-Gini de la source m ($\overline{G}_m$) soit égal au réel Gini de la source m (G_m).

En définitive, cette approche est équivalente à celle de Rao (1969), mais néanmoins plus complète, puisque FRK mettent en exergue une décomposition à des fins économiques comme celles de la détermination des sources d'inégalité expliquant le fossé qui règne entre les pauvres et les non-pauvres.

IV. Une application aux ménages colombiens : Fields (1979)

La méthodologie adoptée par Fields provient de FRK (1978). L'indice de Gini global calculé sur un ensemble de sources de rémunération s'exprime sous la forme d'une moyenne pondérée des pseudo-Gini. L'auteur étudie les sources de revenu de 2949 ménages colombiens de 1967 et 1968. Le revenu total de chaque ménage est constitué de quatre facteurs principaux. La première source concerne les revenus salariaux. Il s'agit des salaires, du partage des bénéfices et des primes. Le deuxième facteur concerne les revenus indépendants issus de services rendus ou de travaux professionnels. Le troisième facteur concerne les revenus du capital : intérêts, dividendes et rentes. La dernière source représente les revenus des transferts. L'indice de Gini global s'exprime alors sous la forme d'une moyenne pondérée de quatre pseudo-Gini :

$$G = \varphi_1 \overline{G}_1 + \varphi_2 \overline{G}_2 + \varphi_3 \overline{G}_3 + \varphi_4 \overline{G}_4 . \qquad (15)$$

D'après la décomposition de FRK (1978), le pseudo-Gini d'une source m ($\overline{G}_m$) est une fonction de l'indice de Gini associé à la source m (G_m) :

$$\overline{G}_m = G_m \, \Omega_m , \qquad (16)$$

où Ω_m est un coefficient de corrélation relatif. La population comprend n ménages ($\forall i = 1,\ldots, n$). Soit x^m le vecteur des n éléments de la $m^{\text{ième}}$ source de revenu. On note R le vecteur des positions (rangs) des revenus globaux de chaque ménage[3] et R_m le vecteur des rangs des sources m de chaque ménage. Par conséquent, le coefficient Ω_m est défini par :

$$\Omega_m = \frac{\mathrm{Cov}(x^m, R)}{\mathrm{Cov}(x^m, R^m)}, \tag{17}$$

où $\mathrm{Cov}(x^m, R)$ est la covariance entre les salaires et la position du revenu global de chaque ménage dans la distribution. Par exemple, $\mathrm{Cov}(x^4, R^4)$ représente la covariance entre les transferts des n ménages et le rang de ces transferts lorsqu'ils sont ordonnés de manière croissante. Par conséquent, la relation qui définit l'indice de Gini global comme une somme pondérée des pseudo-Gini par source se réécrit de la manière suivante :

$$G = \varphi_1 \Omega_1 G_1 + \varphi_2 \Omega_2 G_2 + \varphi_3 \Omega_3 G_3 + \varphi_4 \Omega_4 G_4 \tag{18}$$

Outre, cette brève analyse des corrélations (qui sera reprise et développé par Lerman et Yitzhaki (1985), cf. section VII), l'auteur s'intéresse au poids de chaque source de revenu dans la mesure du Gini total :

$$100\% = \varphi_1 \Omega_1 \frac{G_1}{G} + \varphi_2 \Omega_2 \frac{G_2}{G} + \varphi_3 \Omega_3 \frac{G_3}{G} + \varphi_4 \Omega_4 \frac{G_4}{G}. \tag{19}$$

Cette réécriture de la décomposition de FRK procure la contribution des quatre sources de revenu colombiennes. Les facteurs issus du travail [salaires et revenus indépendants (sources 1 et 2)] représentent respectivement 27,02% et 42,08% de l'inégalité totale. Les revenus provenant du capital génèrent 26,47% de l'inégalité contre 4,42% pour les revenus de transfert.

Fields soulève ensuite une question capitale. Existe-t-il une différence significative entre les estimations faites à partir de données individuelles et celles menées sur les données groupées ?

Les revenus sont répartis en 21 classes. Le tableau 1 expose, pour chaque méthode d'estimation, les contributions de chaque source à l'inégalité totale.

Tableau 1 : Contributions des sources à l'inégalité totale

Contributions (%)	*Données groupées*	*Données individuelles*
S^1	27,65	27,02
S^2	41,87	42,08
S^3	25,46	26,47
S^4	4,48	4,42

[3] Tous les rangs sont obtenus en classant les éléments de la distribution par ordre croissant.

Les écarts entre les contributions sont très faibles. Ce n'est pas le cas pour les indicateurs de Gini par source (G_m) et les corrélations Ω_m.

En somme, l'auteur apporte, par l'intermédiaire de son application menée sur les revenus colombiens, une confirmation des résultats précédemment obtenus par FRK sur les revenus de Taiwan et du Pakistan. Les revenus du travail sont répartis de manière très inégalitaire. Ils accroissent les inégalités globales. Cette conclusion, autant valable pour les données individuelles que pour les données groupées, est en partie due à l'importance du poids des sources 1 et 2 dans le revenu total. Pyatt, Chen, et Fei (1980) vont pourtant démontrer qu'une différence fondamentale prévaut entre les estimations menées à l'aide de données individuelles et celles provenant de données groupées.

V. La différence existant entre les données individuelles et les données groupées : Pyatt, Chen et Fei (1980)

Avant de mettre en évidence les difficultés concernant les données individuelles et les données groupées, les auteurs proposent de démontrer que la mesure de Gini est un cas particulier du coefficient de concentration.

Le rang des ménages dépend de la règle de classification t. Lorsque t_i est le plus faible élément, on a $R_{(ti)} = 1$ et lorsque t_i est le plus fort on note $R_{(ti)} = n$ (le nombre de ménages). Un problème qui n'avait pas été évoqué jusqu'ici est soulevé. Il peut en effet exister plusieurs ménages possédant le même rang dans la distribution. Il faut donc leur attribuer un rang identique. On leur octroie la moyenne arithmétique de leur rang. De cette manière, il est possible de définir le rang moyen de la distribution :

$$\overline{R} = (1/n)\sum_{i=1}^{n} R_{(ti)} = (n+1)/2 . \tag{20}$$

On considère que x_i est une variable inconnue, mais on exige que la moyenne de cette variable μ soit non-négative, bien que les x_i puissent être négatifs. On définit ensuite q_i comme étant la proportion de revenu total détenue par le ménage i : $q_i = x_i / n\mu$. Soit la variable Q_i définissant la proportion de revenu cumulé $Q_i = \sum_{j=1}^{i} q_j$ et P_i celle représentant la proportion de rang cumulé : $P_i = \sum_{j=1}^{i} R_{(tj)} / n$. En liant les points issus des couples (P_i, Q_i) dans le plan Euclidien à deux dimensions, on obtient une courbe de concentration. Celle-ci n'est pas forcément monotone. Le coefficient de concentration peut être négatif ou supérieur à 1 si certaines valeurs de x_i sont négatives. En prenant la somme des aires des trapèzes situés en dessous de la courbe de concentration, on obtient l'aire suivante :

$$A=(1/2n)\sum_{i=1}^{n} q_i \left[1+2(n-i)\right]. \tag{21}$$

L'indice de concentration est donc :

$$c(x/t)=1-(1/n)\sum_{i=1}^{n} q_i \left[1+2(n-i)\right] . \tag{22}$$

Après quelques transformations algébriques, le coefficient de concentration peut se réécrire comme suit :

$$c(x/t)=2\,\mathrm{Cov}\,(q(x), R_{(ti)})=(2/n\mu)\mathrm{Cov}\,(x, R_{(ti)}), \tag{23}$$

où Cov désigne l'opérateur de covariance. Cette expression montre que le coefficient de concentration est calculé à partir de la variable x et du critère de classification t. Un lien s'établit alors avec la courbe de Lorenz. Il faut considérer que x_i est le revenu du ménage i et que le rang du ménage correspond au rang du revenu dans la distribution : $R(_{ti}) = R(_x)$. Etant donné que $t_i = 1$ pour la plus faible valeur de t_i, alors $R(_x) = 1$ pour le plus faible revenu et $R(_x) = n$ pour le plus fort. Les auteurs proposent alors :

$$G=c(x/x)=2\,\mathrm{Cov}\,(q(x), R_{(x)})=(2/n\mu)\mathrm{Cov}\,(x, R_{(x)}). \tag{24}$$

On comprend alors que la courbe de Lorenz est un cas particulier de la courbe de concentration et que l'indice de Gini est un cas particulier du coefficient de concentration. Ce cas particulier est construit sur le fait que le rang attribué à chaque famille correspond au rang de son revenu dans la distribution. Le revenu de chaque ménage étant la résultante de q facteurs additifs, le coefficient de Gini s'écrit alors :

$$G=(2/n\mu)\mathrm{Cov}\,(x, R_{(x)})=(2/n\mu)\mathrm{Cov}\left(\sum_{m=1}^{q} x_i^m, R_{(x)}\right)=\sum_{m=1}^{q}\varphi_m\, c(x^m/x). \tag{25}$$

Il s'agit de l'expression de Rao (1969).[4] Le pseudo-Gini de la source m s'exprime par :

$$\overline{G}_m=c(x^m/x)=(2/n\mu)\mathrm{Cov}\,(x^m, R_{(x)}). \tag{26}$$

A contrario, l'indice de Gini mesuré sur la source m est :

$$G_m=c(x^m/x^m)=(2/n\mu)\mathrm{Cov}\,(x^m, R_{(x^m)}). \tag{27}$$

Un autre lien unit l'indice de Gini et son pseudo-Gini. Le $i^{\text{ième}}$ point sur la courbe de concentration de l'indice de Gini de la source m regroupe tous les revenus situés en dessous de i. En prenant le pseudo-Gini, la somme des revenus en dessous de ce point i ne peut pas excéder celle obtenue à partir du Gini standard. Par conséquent, la courbe de

[4] Les articles antérieurs à Pyatt, Chen et Fei attribuent la paternité de la décomposition de l'indicateur de Gini en sources de revenu à Fei, Ranis et Kuo (1978).

concentration du pseudo-Gini procure un indice moins important. D'où la relation suivante :

$$c(x^m/x) \leq G_m \Leftrightarrow c(x^m/x)/G_m = \Omega_m(x, x^m) \leq 1 . \qquad (28)$$

De manière générale, les indices de concentration $c(x/t)$ ne peuvent procurer une valeur supérieure à celle de leur coefficient de Gini.

Il est ensuite possible de calculer le coefficient de Gini à partir des données groupées. Les ménages sont maintenant regroupés en k classes ($j = 1,\ldots, k$). On note $\underline{\mu}_j$ la moyenne des revenus de la classe j et $\underline{\mu}$ le vecteur des moyennes de chaque classe. L'indice de Gini mesuré sur les revenus des ménages regroupés en j classes est :

$$G(\underline{\mu}) = c(\underline{\mu}/\underline{\mu}) = (2/n\mu)\mathrm{Cov}(\underline{\mu}, R_{(\underline{\mu})}) . \qquad (29)$$

D'autre part, la relation suivante est toujours vérifiée :

$$\mathrm{Cov}(\underline{\mu}, R_{(\underline{\mu})}) = \mathrm{Cov}(x, R_{(\underline{\mu}_j)}) . \qquad (30)$$

On sait aussi, d'après la remarque précédente, que les indices formés à partir d'une corrélation avec un rang diffèrent de celui qui prévaut pour la distribution globale sous-estime l'indice de Gini : $\mathrm{Cov}(x, R_{(\underline{\mu}_j)}) \leq \mathrm{Cov}(x, R_{(x)})$. On a donc :

$$\mathrm{Cov}(x, R_{(\underline{\mu}_j)}) = \mathrm{Cov}(\underline{\mu}, R_{(\underline{\mu})}) \leq \mathrm{Cov}(x, R_{(x)}) \;\Rightarrow\; G(\underline{\mu}) \leq G(x) . \qquad (31)$$

En définitive, l'indice de Gini estimé à partir de données en classes sous-estime en permanence l'indice de Gini mesuré à partir de données individuelles.[5] Il est donc primordial de mener à bien les décompositions avec des données individuelles. Les auteurs étudient néanmoins les conditions dans lesquelles l'erreur peut revêtir une forme particulière selon les rangs considérés. Une difficulté d'interprétation subsiste car la décomposition de l'indicateur de Gini dépend de l'indice de corrélation des rangs Ω_m :

$$G = \textstyle\sum_{m=1}^{q} \varphi_m \, \Omega_m(x, x^m) \, G(x^m) . \qquad (32)$$

Selon FRK, Ω_m reste constant dans le temps. A contrario, Pyatt, Chen et Fei estiment que ce coefficient est dépendant d'un aspect dynamique du développement, comme le départ des agriculteurs vers des empois salariés. Cette idée d'application au domaine agricole sera ensuite reprise par Adams (1994) sur 734 ménages pakistanais.

[5] Les années 70 et 80 sont marquées par d'abondantes applications menées sur des données groupées. Les auteurs tentent donc de justifier leur approche par une étude comparative. Fields (1979) compare de manière empirique les estimations. Pour Pyatt, Chen et Fei, l'étude est accompagnée par une justification mathématique. La différence entre les données groupées et individuelles fut aussi étudiée par Gastwirth (1972).

En définitive, les approches précédentes montrent que la science, et en particulier les sciences économiques, forme une matrice disciplinaire cumulative dans le sens où les auteurs ont chacun apporté leur contribution afin d'étendre la décomposition de la mesure de Gini en relation avec l'indice de concentration. Une césure va pourtant s'effectuer au sein de la littérature puisqu'en 1982, Shorrocks va exprimer des doutes concernant la décomposabilité de la mesure de Gini.

VI. L'apport axiomatique de Shorrocks (1982)

En 1982, Shorrocks propose de faire une synthèse des travaux effectués dans le domaine de la décomposition des mesures d'inégalité en sources de revenu. Son approche est qualifiée d'analyse axiomatique. Il refuse cependant cette appellation. Il introduit différentes règles de décomposition sous forme d'hypothèses, puis propose en 1983 une application sur les revenus américains.

Hypothèse 1 : Symétrie, Continuité, et Normalisation.

La mesure d'inégalité est une fonction continue et symétrique de ses arguments.[6] La normalisation est considérée comme une borne inférieure. Si tous les revenus sont égaux à un même réel, comme la moyenne μ, alors : $I(\mu,\ldots,\mu) = 0$.

Hypothèse 2 : (a) Continuité et (b) Traitement symétrique des facteurs.

(a) La continuité est définie par : $S^m(x\,;q)$ est continue en x^m, où S^m est la contribution de la source m à l'inégalité totale, x le vecteur des revenus, et q le nombre de facteurs.

(b) Le traitement symétrique des facteurs est : $S^m(x^1, x^2,\ldots, x^m\,;q) = S^{\pi\, m}(x^{\pi 1}, x^{\pi\, 2},\ldots, x^{\pi\, m}\,;q)$, où πm est une permutation des éléments du vecteur des sources m. Cela signifie, que le mélange des éléments à l'intérieur de la distribution x^m ne modifie pas la valeur de la contribution $S^m(x\,;q)$.

Hypothèse 3 : Indépendance par rapport au niveau de désagrégation.

La contribution $S^m(x\,;q)$ ne doit pas dépendre de la manière dont les revenus sont distingués ou décomposés. Par exemple, que le revenu soit divisé en salaires, primes, rentes et transferts ou seulement en salaires et autres revenus, la contribution du facteur salaire à l'inégalité totale ne doit pas être modifiée.[7] Autrement dit, si la contribution de

[6] En effet, lorsqu'une perturbation est introduite dans la distribution, la mesure d'inégalité reste inchangée. La symétrie équivaut aussi au principe qui permet de conserver l'anonymat des individus.

[7] Lorsqu'une décomposition en facteurs est réalisée par la valeur Shapley (1953), on ne respecte pas cette propriété (cf. la section X sur les méthodes permettant de généraliser la décomposition). Il est nécessaire de recourir à d'autres valeurs comme celle d'Owen (1977).

la source 1 est évaluée conjointement avec quatre autres sources ou par rapport à un ensemble qui regroupe les quatre facteurs, alors la contribution reste inchangée :

$$S^1(x^1, x^2,\ldots, x^4 ; q) = S^1(x^1, x-x^1 ; 2) = S^1(x^1, x).$$

Hypothèse 4 : Décomposition cohérente.

Elle implique trivialement que la somme des contributions $S^m(x ; q)$ procure la valeur globale de l'inégalité $I(x)$. Elle est définie par :

$$\sum_{m=1}^{q} S^m\left(x^1, x^2,\ldots, x^m,\ldots, x^q ; q\right) = \sum_{m=1}^{q} S^m\left(x^m ; q\right) = I(x). \tag{33}$$

Ces quatre hypothèses de base permettent de spécifier un théorème dont le résultat montre qu'une mesure d'inégalité se réécrit comme une moyenne pondérée des sources de revenu.

$$I(x) = \sum_{m=1}^{q} S^m\left(x^m, x\right) = a.\sum_{m=1}^{q} x^m = ax, \tag{34}$$

De plus, la contribution de la source m (S^m) à l'inégalité totale se formule avec les mêmes poids. Par exemple, pour l'indicateur de Gini, on a :[8]

$$S^m\left(x^m, x\right) = \frac{2}{n^2\mu}\sum_{i=1}^{n}\left(i - \frac{n+1}{2}\right)x_i^m = \frac{\mu^m}{\mu}\overline{G}\left(x^m\right), \tag{35}$$

où : $$a_i(x) = 2\left(i - \frac{n+1}{2}\right)\Big/\mu n^2 . \tag{36}$$

Hypothèse 5 : (a) Symétrie de la population et (b) Normalisation des distributions de facteurs équivalents.

(a) Les contributions de chaque facteur à l'inégalité totale ne doivent pas dépendre de la manière dont les individus sont classés : $S^m(x^mP, xP) = S^m(x^m, x)$. En introduisant une perturbation (par une matrice de permutation) sur la distribution globale et sur la distribution de la source m, on change le rang des individus. Cette perturbation ne doit pas affecter la valeur de la contribution. On constate immédiatement que cette restriction invalide la décomposition de l'indice de Gini, car le pseudo-Gini dépend du rang de la distribution globale.

(b) La normalisation pour des distributions de sources équivalentes permet d'attribuer la valeur nulle à la contribution de la source m lorsque le vecteur de la source m est constitué d'un même réel. Par exemple, si le vecteur est uniquement constitué de la valeur moyenne μ^m, alors la contribution de la source m à l'inégalité totale est nulle : $S^m(\mu^m e\ , x) = 0,\ \forall \mu^m$, où e est le vecteur « unité » (1,…,1) de taille n. Cette hypothèse est nécessaire mais non suffisante. Elle permet de définir une décomposition unique pour chaque indicateur. Le problème est que l'unicité est seulement valable pour une population de deux personnes. Les décompositions ne sont donc pas naturelles.[9] Il faut alors introduire une hypothèse encore plus restrictive.

Hypothèse 6 : Symétrie de deux facteurs.

Il s'agit d'une restriction de l'hypothèse 2 (b) pour deux facteurs : x^1 et x^2. Supposons que x^2 soit une permutation de x^1 : $x^2 = x^1P$, où P est une matrice de permutation ($n\times n$). Cette hypothèse permet de définir l'égalité entre la contribution issue de x^1 et son permuté x^2. Lorsqu'il existe plusieurs variables, la corrélation entre x^1 et $x–x^1$ n'est pas la même que celle qui prévaut entre x^2 et $x–x^2$. Ces différences de corrélation donnent des contributions différentes à l'inégalité totale. En revanche, lorsqu'il existe seulement deux facteurs, ces contributions doivent être égales : $S^1(x^1, x^1 + x^1P) = S^2(x^1P, x^1 + x^1P)$.

Théorème. *Les hypothèses* 1-6 *impliquent* :

$$s^m(I) = \frac{S^m\left(x^m, x\right)}{I(x)} = \frac{\mathrm{Cov}\left(x^m, x\right)}{\sigma^2(x)}. \tag{37}$$

Ce théorème fournit trois résultats décisifs. Premièrement, la règle de décomposition est unique. Deuxièmement, la contribution relative de chaque source à l'inégalité totale est indépendante du choix de la mesure. Cela signifie que pour l'indice de Gini, s'il existe une équation qui satisfait les hypothèses 1-6, la contribution relative pourrait alors se mesurer avec la covariance et la variance. Troisièmement, la variance et le coefficient de variation vérifient les hypothèses 1 à 6. Ceci permet d'invalider, en partie, toutes les recherches qui ont été menées jusqu'à présent et qui tente de démontrer la justesse de la décomposition de l'indicateur de Gini. D'autres chercheurs vont néanmoins démontrer que la décomposition de la mesure de Gini est « désirable ».

VII. La décomposition de l'indice de Gini étendu : Lerman et Yitzhaki (1985)

En partant de la formule de l'indice de Gini basée sur la covariance, notamment développée et généralisée par Pyatt, Chen et Fei (1980), Lerman et Yitzhaki (1985) présentent la décomposition de l'indice de Gini étendu. Il s'agit de l'indicateur de Gini

---

[8] Cette formule de l'indicateur de Gini est valable lorsque les revenus sont ordonnés de manière croissante, et où i représente le rang de l'individu i.

[9] Le terme naturel renvoie à l'unicité de la décomposition.

qui tient compte de l'aversion à l'inégalité. En d'autres termes, il intègre la préférence relative pour l'égalité des revenus.

On appelle F la fonction de réparation de la distribution globale et F^m la fonction de répartition du facteur m. Les distributions sont uniformément réparties entre [0,1] de sorte que la moyenne soit de 1/2. L'indicateur de Gini se formule par $G = 2\mathrm{Cov}(x, F)/\mu$. Il peut se réécrire en tenant compte de l'ensemble des sources de revenu :

$$G = \frac{2\sum_{m=1}^{q}\mathrm{Cov}\left(x^m, F\right)}{\mu}. \tag{38}$$

Afin de faire apparaître l'indice de Gini de la source m, il est possible de multiplier et de diviser l'expression précédente par des termes qui se neutralisent :

$$G = \frac{2\sum_{m=1}^{q}\mathrm{Cov}\left(x^m, F\right)}{\mu} = \sum_{m=1}^{q}\left[\frac{\mathrm{Cov}\left(x^m, F\right)}{\mathrm{Cov}\left(x^m, F^m\right)}\right]\times\left[\frac{2\,\mathrm{Cov}\left(x^m, F^m\right)}{\mu^m}\right]\times\left[\frac{\mu^m}{\mu}\right]. \tag{39}$$

On obtient alors la relation suivante :

$$G = \sum_{m=1}^{q} R^m G^m \varphi^m, \tag{40}$$

où $$R^m = \left[\frac{\mathrm{Cov}\left(x^m, F\right)}{\mathrm{Cov}\left(x^m, F^m\right)}\right],\ G^m = \left[\frac{2\,\mathrm{Cov}\left(x^m, F^m\right)}{\mu^m}\right],\ \varphi^m = \left[\frac{\mu^m}{\mu}\right]. \tag{41}$$

On appelle R^m la corrélation de Gini entre la source m et le revenu global [équivalent à Ω_m de l'équation (17) introduit par FRK (1978) puis par Fields (1979)]. Le terme G^m est l'indice de Gini associé à la source m. L'élément φ^m est la proportion du facteur m dans le revenu moyen. La corrélation de Gini est similaire au coefficient de corrélation de Pearson.[10] Le ratio est borné dans l'intervalle [-1,1].

- Si $R^m = 1$, la source m est une fonction croissante du revenu total.
- Si $R^m = -1$, le facteur m est une fonction décroissante du revenu total.
- Si $R^m = 0$, la source de revenu m est également distribuée et sa contribution à l'inégalité totale est nulle.

Ceci nous conduit à l'évaluation de l'impact marginal de la source m sur le coefficient global. Cette mesure peut être effectuée par une simple dérivée partielle. Supposons que chaque personne voit le montant de sa source m varié de $e.x^m$, où e tends vers la valeur 1. Il est donc possible de jauger la dérivée partielle du Gini global compte tenu d'une petite variation en pourcentage de la source m (représentée par e^m) :

$$\frac{\partial G}{\partial e^m} G = \varphi^m \left(R^m G^m - G \right). \qquad (42)$$

D'après Yitzhaki (1983), il est possible de construire un indice de Gini qui dépend de l'intensité de la préférence pour l'égalité. Si cette intensité est captée par le paramètre v, alors :

$$G(v) = 1 - v\,(v-1) \int_0^1 (1-F)^{v-2} L(F)\, dF\ , \qquad (43)$$

où $G(v)$ est l'indice de Gini étendu (ou généralisé), et où $L(F)$ est la courbe de Lorenz. Le paramètre de préférence pour l'égalité est compris entre 0 et l'infini. Lorsque :

- $v \in [0,1[$, l'indice reflète la préférence pour l'inégalité ;
- $v = 1$, l'indice reflète l'indifférence à l'égalité ;
- $v = 2$, le coefficient de Gini généralisé est égal à l'indice de Gini standard ;
- $v \to \infty$, on atteint le critère de Rawls pour lequel la parfaite égalité est préférée.

La décomposition en sources de revenu est analogue à la précédente :

$$G = \sum_{m=1}^{q} R^m(v)\, G^m(v)\, \varphi^m\ , \qquad (44)$$

où $$R^m(v) = \left[\frac{\mathrm{Cov}\left(x^m, \left(1-F^0\right)^{v-1}\right)}{\mathrm{Cov}\left(x^m, \left(1-F^m\right)^{v-1}\right)} \right],\ G^m(v) = \frac{-v\ \mathrm{Cov}\left(x^m, \left(1-F^m\right)^{v-1}\right)}{\mu^m}. \qquad (45)$$

Lerman et Yitzhaki s'interrogent sur la cohérence de leur décomposition. Ils s'appuient pour cela sur les hypothèses introduites par Shorrocks (1982). La décomposition de Lerman et Yitzhaki tombe dans la catégorie des multiples décompositions de l'indicateur de Gini, dont la séparation n'est pas unique. Selon Shorrocks, l'idée de choisir l'indice de Gini est « non acceptable ». Selon Lerman et Yitzhaki, leur décomposition est « désirable ». Ils font ici allusion au fait que l'indice de Gini permet d'atteindre le concept de dominance stochastique. Cette décomposition est ensuite beaucoup plus attrayante du fait qu'elle intègre trois indices : l'indice de Gini du facteur, la part dans le revenu, et la corrélation de la source avec le revenu global. L'approche de Shorrocks est une analyse mathématique très intéressante, mais le fait que les poids (a_i) procurent de multiples décompositions n'est pas, selon les auteurs, explicité de manière économique. Enfin, la dernière idée qui incite à privilégier cette décomposition est le calcul des contributions marginales. Il s'agit de mesurer la variation d'une source précise m et d'évaluer son impact sur le coefficient global.

[10] Le coefficient de Pearson possède le même numérateur que R^m, mais il est basé sur le produit des écart-types.

L'application de cette méthode révèle, par exemple, que la contribution marginale des revenus féminins (dans un ménage) excède celle issue du revenu du capital.[11]

On remarquera, finalement, que cette nouvelle décomposition de la mesure de Gini permet de se détacher du concept de pseudo-Gini.[12]

VIII. Une application à la consommation : Garner (1993)

L'approche de Lerman et Yitzhaki (1985) est intéressante, car l'analyse des contributions marginales constitue un apport fondamental. Ces contributions marginales peuvent être évaluées par de petites variations inhérentes aux valeurs d'une source de revenu. Supposons qu'il existe un changement dans les dépenses de consommation. Ce changement concerne une source *m*. L'intensité du changement est mesurée par le pourcentage e^m. On considère ensuite que cette variation concerne l'ensemble des individus d'une population sur laquelle sont mesurées les inégalités. Ce changement marginal peut être évalué en proportion de la valeur de l'indice de Gini global. Il est appelé contribution marginale relative :

$$\frac{\partial G / \partial e^m}{G} = \frac{R^m G^m \varphi^m}{G} - \varphi^m . \tag{46}$$

Cette formule est très attrayante. Contrairement à Lerman et Yitzhaki (1985), Garner mesure le poids relatif de l'impact marginal. Ceci aboutit à une définition claire. Il s'agit de la contribution en pourcentage de la source *m* à laquelle on retranche la contribution moyenne de la dépense de consommation du facteur *m*. Cela permet d'envisager et de mesurer l'effet d'une taxe imposée sur une source particulière. Lorsque la contribution marginale de la source *m* est positive, une taxe sur ce facteur permet de réduire l'inégalité totale. Cette taxe doit être progressive car elle affecte marginalement les riches plus qu'elle n'affecte les pauvres. A contrario, lorsque la contribution marginale relative est négative, une taxe sur le bien de consommation a pour conséquence d'augmenter l'inégalité totale.

D'autre part, d'après Yitzhaki (1990), il est possible de calculer l'élasticité de la source *m* par rapport à la consommation globale :

[11] Notons que Shorrocks (1983) atteint le même résultat sans utiliser la contribution marginale. En 1989, Lerman et Yitzhaki utilisent cette méthode pour étudier les inégalités de sources de revenu des ménages américains de 1983.

[12] Lerman et Yitzhaki dérivent leur mesure sans mentionner la mesure du pseudo-Gini. Néanmoins, il est possible de remarquer l'étroite relation avec l'approche de FRK (1978) et celle de Fields (1979).

$$\eta^m = \frac{R^m G^m}{G}. \tag{47}$$

Cette élasticité peut se réécrire de la manière suivante :

$$\eta^m = \frac{\text{Cov}\left(x^m, F\right)}{\text{Cov}\left(x^m, F^m\right)} \frac{\mu}{\mu^m} = b^m \frac{\mu}{\mu^m}, \tag{48}$$

où b^m peut être considéré comme un paramètre de régression ou comme la propension moyenne à consommer le bien m. Les valeurs des élasticités renvoient à des biens de consommation particuliers :

- une élasticité supérieure à 1 représente un bien de luxe ;
- une élasticité appartenant à l'intervalle [0,1] représente un bien de nécessité ;
- une élasticité négative représente un bien de consommation inférieur.

Une application est menée sur les dépenses de consommation américaines de 1987. L'indice de Gini global de 0,33 est décomposé.[13] Les élasticités sont notamment très élevées pour les dépenses vestimentaires et les dépenses de loisir. Une conséquence de ces résultats est que la taxation de ces biens augmenterait la progressivité du système fiscal.

IX. Une application à la redistribution : Podder (1993)

Une autre élasticité est ici considérée. Elle mesure l'effet de l'augmentation de 1% de la moyenne de la source m sur l'inégalité totale.

Théorème. *Soit une variation de la moyenne de la source m (μ^m) sans changement de sa courbe de concentration. L'élasticité du coefficient de Gini par rapport à sa source m est donnée par :*

$$\eta^m = \frac{1}{G}\left[\frac{\mu}{\mu^m}\left(G^m - G\right)\right]. \tag{49}$$

Il est possible de démontrer que cette élasticité est égale à :

$$\eta^m = \frac{dG/G}{d\mu^m/\mu}, \text{ où } \sum_{m=1}^{q} \eta^m = 0 \tag{50}$$

Les implications ne sont pas fiscales. Elles concernent les effets de la redistribution. En effet, la redistribution d'une source particulière (notamment une augmentation de x% de la source m) peut affecter la répartition globale des revenus, et par conséquent modifier les inégalités globales. Il faut donc planifier les redistributions avec un effet neutralisant

[13] Pour les lecteurs qui s'intéressent à la valeur particulière G=0,33 ; confère Milanovic (1997).

les inégalités de revenu. Pour cela, il est nécessaire de modifier l'indice de Gini de la source *m* de manière à ce qu'il soit égal à l'indicateur de Gini global : $G = G^m$.

On constate alors que l'approche de Lerman et Yitzhaki est fondamentale. Elle a permis de créer une approche en terme d'élasticité de Gini, utilisée à la fois en fiscalité, en redistribution, et permet d'apprécier l'impact du travail des femmes sur l'évolution du marché du travail et les distributions de revenu [cf. Cancian et Reed (1998)].[14]

X. Les tentatives de généralisation

L'approche de Lerman et Yitzhaki (1985) est la méthode de décomposition de l'indice de Gini en sources de revenu la plus élaborée. Les recherches récentes s'attachent donc à généraliser la technique de décomposition en sources de revenu. Cette généralisation prend forme autour de deux axes : l'analyse économétrique et la Valeur de Shapley (1953).

X.1. L'approche économétrique : Morduch et Sicular (2002)

Contrairement aux sections précédentes, où l'on est contraint par la structure additive des sources de revenu, une solution alternative consiste à privilégier l'approche économétrique. Elle paraît à première vue générale, car les facteurs peuvent ne pas être liés par une structure particulière. Il est par conséquent concevable d'utiliser des facteurs multiplicatifs qui ne sont pas des sources de revenu.

Les auteurs critiquent les approches traditionnelles de décomposition en dénonçant un manque de contrôle vis à vis de l'endogénéité des variables. En effet, les sources sont issues des revenus, qui sont eux-mêmes déterminés par d'autres variables.

Ces problèmes ont incité les chercheurs à examiner la piste économétrique. Plus précisément, ils introduisent la régression dans la décomposition. Ce recours permet notamment de contrôler l'endogénéité et d'intégrer la continuité des variables. Leur méthode permet de calculer : (*i*) l'exacte contribution d'une source particulière à l'inégalité totale, (*ii*) son écart-type, (*iii*) son intervalle de confiance. Pour cela, deux définitions de base sont proposées.

[14] Confère aussi Flückiger et Silber (1995) dont la méthode permet d'expliquer la différence entre l'indicateur de Gini des hommes et celui des femmes par l'intermédiaire des sources de revenu.

Définition 1 : Additions uniformes. *Un indice d'inégalité* $I(x)$ *décroît lorsque les individus reçoivent un même transfert positif t :*

$$I(\underline{x} + t\underline{e}) \leq I(x), \tag{51}$$

où $\underline{x}$ *est le vecteur des revenus* $n{\times}1$, *et où* $\underline{e}$ *est le vecteur unitaire* $n{\times}1$ *dont les éléments sont uniquement égaux à la valeur* 1.

Il s'agit d'une propriété issue de l'axiome de transfert de Pigou-Dalton, où la mesure d'inégalité décroît lorsqu'une personne riche transfère une partie de son revenu à une personne pauvre. En notant $\underline{x}^m$ le vecteur représentant la $m^{\text{ième}}$ source de revenu des n individus, il est alors possible de modifier la définition 1.

Définition 2. *Une méthode de décomposition qui attribue à chaque facteur une proportion d'inégalité* s^m, *satisfait la propriété d'additions uniformes si pour tout vecteur* $\underline{x}_m = t\,\underline{e}$, *alors* $s^m < 0$.

Par conséquent, toute méthode de décomposition qui accorde une contribution relative négative à une source (pour des distributions de facteur égalitaires) satisfait le principe d'additions uniformes (néanmoins la définition 1 n'implique pas la définition 2). Par exemple, l'indice de Gini satisfait la définition 1. Mais sa décomposition en pseudo-Gini ne permet pas de vérifier la définition 2. De même, la séparation du coefficient de variation en sources de revenu procure un rejet de la définition 2 (cependant sous une écriture différente il peut l'inclure).

Soit un modèle de régression généralisé de type :

$$x = Y\beta + \varepsilon\,. \tag{52}$$

La matrice Y comprend en première colonne le vecteur e, puis les autres colonnes sont constituées de la matrice X. La matrice Y est donc de taille $n{\times}(q{+}1)$. L'estimation est faite par moindres carrés (régression quantile) en tenant compte des différentes corrections (endogénéité, etc.). Cette méthode est donc plus générale. Elle peut en effet être utilisée pour des sources de revenu ou bien des dimensions (attributs) tels que l'éducation, l'âge, la santé, etc. Par conséquent, le revenu estimé en fonction des attributs est :

$$\hat{x} = Y\hat{\beta}\,. \tag{53}$$

De même, les revenus issus de l'attribut m sont :

$$\hat{x}^m = Y\hat{\beta}^m\,. \tag{54}$$

Le revenu global de l'individu i est :

$$x_i = \sum_{m=1}^{q+1} \hat{x}_i^m , \qquad (55)$$

où $\hat{x}_i^m = \hat{\varepsilon}_i$, $\forall m = q+1$. On rappelle que, d'après Shorrocks (1982), les contributions relatives de chaque source se formulent par :

$$s^m = \frac{\sum_{i=1}^{n} a_i \left(x_i^m\right)}{I(x)} . \qquad (56)$$

Ainsi, la contribution relative d'un facteur à l'inégalité totale devient :

$$s^m = \hat{\beta}^m \left(\frac{\sum_{i=1}^{n} a_i \left(x_i^m\right)}{I(x)} \right) . \qquad (57)$$

Il est alors permis de décomposer tous les indices d'inégalité qui peuvent s'écrire comme une moyenne pondérée des revenus. D'autre part, peu d'études mettent l'accent sur les erreurs liées à l'estimation des inégalités. Pourtant, étant donné la linéarité dans l'estimation des paramètres, on a :

$$\sigma\left(s^m\right) = \sigma\left(\hat{\beta}^m\right) \left(\frac{\sum_{i=1}^{n} a_i \left(x_i^m\right)}{I(x)} \right) . \qquad (58)$$

L'écart-type des erreurs est donc :

$$\sigma\left(s^{\varepsilon}\right) = \left\{ \sigma_{\varepsilon}^2 \sum_{i=1}^{n} \left(\frac{a_i(x)}{I(x)} \right)^2 \right\}^{1/2} . \qquad (59)$$

La relation (59) permet ainsi le calcul des intervalles de confiance pour les contributions relatives (sous l'hypothèse d'homoscédasticité).

En définitive, on remarque que cette approche remet en cause les décompositions analytiques (non paramétriques) des inégalités. La première remise en cause s'effectue par les définitions 1 et 2. La seconde attire l'attention du chercheur sur la contrainte : la somme des facteurs doit être égale au revenu. L'approche des auteurs permet une certaine généralisation. On peut à la fois utiliser la méthode pour des sources de revenu ou pour des attributs. Il faut tout de même noter que cette généralisation implique certaines restrictions. Les variables doivent être indépendantes, le résidu doit suivre une loi normale, et le modèle doit être de qualité. D'autre part, les auteurs s'appuient uniquement sur la propriété des additions uniformes. Elle permet de privilégier la mesure de Theil. Néanmoins, cette propriété ne procure pas une règle unanime.

X.2. La valeur de Shapley (1953) : un outil de décomposition et de généralisation

En 1992, une nouvelle approche originale est apparue dans la littérature. Auvray et Trannoy (1992) proposent de décomposer les mesures d'inégalité par la valeur de Shapley (1953 ; valeur Shapley en abrégé). La généralisation provient du fait que cette technique peut être intégrée à une grande majorité d'indicateurs (les inégalités, la pauvreté, le R^2, etc.). Il s'agit donc d'une avancée radicale au sein de la littérature et certains auteurs en sont convaincus [cf. Silber (2004)]. Il s'agit d'une synthèse entre la théorie des jeux coopératifs et l'économie des inégalités. L'utilisation de la valeur Shapley suggère que l'inégalité globale peut être partagée au même titre que le problème de partage d'une somme d'argent que l'on rencontre en théorie des jeux coopératifs.

Au regard des principes de décomposition en sources de revenu introduits par Shorrocks (1982), Chantreuil et Trannoy (1999) vérifient que la décomposition par Shapley ignore deux règles cruciales : l'indépendance par rapport au niveau de désagrégation et la normalisation des distributions de facteurs équivalents. Afin de surmonter la difficulté liée à l'indépendance par rapport au niveau de désagrégation, Chantreuil et Trannoy recommandent l'utilisation de la méthode « Nested Shapley ». Shorrocks (1999) va lui-même s'intéresser à cet outil de décomposition très intéressant et va proposer le recours à la valeur d'Owen..

Shorrocks identifie tout d'abord un problème de symétrie. Selon lui, la procédure de Shapley envisage l'ensemble de toutes les éliminations possibles des variables considérées. En éliminant les facteurs selon une règle précise, on met en évidence l'impact marginal de chaque facteur. En prenant la moyenne des ces impacts marginaux, on mesure la contribution du facteur à l'indicateur global. Mais dans ce processus, il existe un problème de dépendance, car les facteurs ne sont pas traités de manière symétrique. Par conséquent, un algorithme est proposé afin d'équilibrer les impacts marginaux.

Supposons qu'un indicateur I (Gini, pauvreté, R^2, etc.) soit déterminé par q facteurs x^m, $m \in K = \{1, 2, \ldots, q\}$. La fonction $F(S)$ procure la valeur de l'indice I lorsque le facteur x^m ($m \notin S$) a été éliminé de l'ensemble des facteurs. On appelle s le nombre de facteurs restant après la séquence d'élimination des x^m : $s = |S|$. La fonction F, est définie par :

$$F : \{S \mid S \subseteq K\} \to \mathbb{R} \qquad (60)$$

Les q! éliminations possibles des q facteurs permettent de mesurer la contribution du facteur m à l'inégalité totale. On mesure donc la contribution du facteur considéré quand la séquence d'élimination est faite au hasard. Les facteurs sont traités de manière symétrique. En conséquence, la contribution du facteur m à la valeur totale de l'indicateur I est donnée par :

$$C_m^s(K,F)=\sum_{s=0}^{q-1}\sum_{S\subseteq K\setminus\{m\}}\frac{(q-1-s)!\,s!}{q!}\Delta_m F(S)\,, \tag{61}$$

où $\Delta_m F(S)$ est l'impact marginal du facteur m lorsque ce dernier a été retiré de l'ensemble des facteurs S :

$$\Delta_m F(S)=F(S\cup\{m\})-F(S) \text{ et } F(\emptyset)=0\,. \tag{62}$$

La contribution du facteur m correspond à l'impact marginal du facteur m lorsque le processus d'élimination des variables intègre toutes les combinaisons possibles.

Pour démontrer la portée de la méthode de Shapley, Shorrocks (1999) propose une application au modèle de Datt et Ravallion (1992). Les auteurs suggèrent un modèle selon lequel les effets de la croissance économique et de la redistribution expliquent les changements de la pauvreté au cours du temps. Le modèle s'exprime par :

$$\Delta P = F(G,R), \tag{63}$$

où ΔP représente la variation de la pauvreté, G la croissance économique et R la redistribution.

Pour mesurer les effets respectifs de G et R sur la variation du niveau de la pauvreté, il est possible de mesurer les contributions de chaque facteur par la valeur Shapley. Les contributions de la croissance et de la redistribution sont données par :

$$C_G^s=\frac{1}{2}\left(F(G,R)-F(R)+F(G)\right) \tag{64}$$

$$C_R^s=\frac{1}{2}\left(F(G,R)-F(G)+F(R)\right). \tag{65}$$

Etant donné que la somme de ces deux composantes est égale à la valeur de la variation de la pauvreté, on peut exactement déterminer quel est le rôle (en pourcentage) de chaque facteur. Shorrocks généralise ensuite ce résultat aux mesures de pauvreté, notamment à la classe des mesures de Foster-Greer-Thorbecke (FGT, 1984).

Contrairement à ce que préconisent Chantreuil et Trannoy (1999), Shorrocks (1999) s'interroge sur la possibilité de recourir à la valeur d'Owen (1977). Elle permettrait de traiter les structures hiérarchiques dans les sources de revenu. La première partition des facteurs est l'ensemble A. Cet ensemble de q facteurs primaires varie avec le processus d'élimination des variables. Lorsqu'une ou plusieurs variables sont éliminées, l'ensemble A devient l'ensemble T (où t est le nombre de facteurs primaires restants après les différentes éliminations). La première phase consiste à définir, par la valeur Shapley, la contribution du facteur L ($L \in A$) :

$$C_L^O(K, A, F) = \sum_{t=0}^{q-1} \sum_{T \subseteq A \setminus \{L\}} \frac{(q-1-t)!\,t!}{q!} \Delta_L F(T)\text{, où} \tag{66}$$

$$\Delta_L F(T) = F^A(T \cup \{L\}) - F(T). \tag{67}$$

La fonction F^A représente l'indicateur d'inégalité (par exemple Gini) mesuré sur la première partition.

La deuxième phase consiste à faire la décomposition par Shapley sur un ensemble L de la première partition. On suppose que L est déterminé par z facteurs de la deuxième partition. Lorsqu'un ou plusieurs facteurs sont éliminés, l'ensemble L est défini par l'ensemble S (s étant le nombre de facteurs restants après les éliminations). Par conséquent, la contribution du facteur m à la contribution $C_L^O(K, A, F)$ préalablement définie, est :

$$C_m^O(K, A, F) = \sum_{s=0}^{z-1} \sum_{S \subseteq L \setminus \{m\}} \frac{(z-1-s)!\,z!}{z!} \Delta_m F(S). \tag{68}$$

Si les facteurs de l'ensemble L sont eux-mêmes fonctions d'une troisième partition de variables, la procédure se poursuit jusqu'à ce qu'il n'y ait plus de partitions. Cette technique en multi-niveaux respecte toujours la cohérence agrégative (quel que soit le nombre de partitions).

Le problème que l'on rencontre avec la valeur Shapley est le suivant. Si l'on applique la décomposition par la valeur Shapley au dernier niveau de partitions (autrement dit à tous les facteurs) puis que l'on somme certains facteurs pour mesurer la contribution d'un facteur de la première partition, le résultat sera différent de la décomposition par Shapley appliquée au premier niveau de la partition. C'est pourquoi, la procédure d'Owen en plusieurs niveaux est recommandée. La méthode Nested Shapley est aussi appropriée [cf. Chantreuil et Trannoy (1999)].

En somme, les techniques utilisées en théorie des jeux coopératifs ouvrent des perspectives de généralisation intéressantes.

XI. Conclusion

La décomposition de la mesure de Gini en facteurs permet de comprendre qu'une source particulière peut être à l'origine des inégalités de revenu. Cette décomposition autorise aussi une vision descriptive des inégalités. Cependant, les sources de revenu autorisent la mise en place de liens plus étroits avec la théorie économique. Par exemple, Fei, Ranis et Kuo (1978) établissent des relations avec la croissance économique ou encore l'économie du développement. La détermination du rôle des sources de revenu dans l'inégalité globale permet aussi d'analyser la distance qui sépare les familles riches des familles pauvres. L'approche de FRK (1978), basée sur les travaux précurseurs de Rao (1969), notamment fondée sur le concept de pseudo-Gini, est le point de départ d'une multitude d'études empiriques. On recense par exemple, des applications sur les sources de revenu colombiennes [Fields (1979)], américaines [Shorrocks (1983)] ou encore pakistanaises [cf. Adams (1994)]. Shorrocks (1982) va ensuite généraliser cette décomposition en facteurs en essayant de comprendre pourquoi certains indicateurs ne sont pas parfaitement décomposables au même titre que la variance et le coefficient de variation. Au total, une mesure est décomposable en facteurs si elle respecte six propriétés fondamentales, sinon, la décomposition est jugée « non acceptable ». Néanmoins, Lerman et Yitzhaki (1985) refusent l'intransigeance de ces règles et montrent que l'indice de Gini satisfait un procédé de décomposition « désirable » puisqu'il fait apparaître trois indices : la mesure de Gini pour chaque facteur, la part du facteur dans le revenu global, et un coefficient de corrélation entre le facteur et le revenu global. La littérature se désunit donc en partie des six règles introduites par Shorrocks (1982), puisque les décompositions en facteurs s'attachent désormais aux généralisation, notamment par les techniques économétriques [cf. Morduch et Sicular (2002)] ou encore par l'intérimaire de la valeur Shapley [cf. Chantreuil et Trannoy (1999), Shorrocks (1999)].

En définitive, outre Paréto et Keynes, la lecture de la littérature indique que de nombreux chercheurs soutiennent la réduction des inégalités dans la répartition des revenus et celle des fortunes, et que le procédé de décomposition en sources de revenu de la mesure de Gini continue d'exercer l'esprit de nombreux auteurs, notamment quand il s'agit d'unifier cette méthode à celle de la décomposition en sous-groupes [cf. par exemple Mussard (2004)].

BIBLIOGRAPHIE

Adams R. (1994), « Non-Farm Income and Inequality in Rural Pakistan: A Decomposition Analysis », *Journal of Development Studies*, Vol. 31(1) : p. 110-133.

Atkinson A. (1970), « On the Measurement of Inequality », *Journal of Economic Theory*, Vol. 55 : p. 244-263.

Auvray C. et Trannoy A. (1992), « Décomposition par source de l'inégalité des revenus à l'aide de la Valeur Shapley », *Journées de Microéconomie Appliquée*, Sfax, Tunisie.

Cancian M. et Reed D. (1998), « Assessing the Effects of Wives' Earnings on Family Income Inequality », *Review of Economics and Statistics*, Vol. 80(1) : p. 73-79.

Chantreuil F. et Trannoy A. (1999), « Inequality Decomposition Values : The Trade-off Between Marginality and Consistency », DP 9924 THEMA.

Dalton H. (1920), « The Measurement of Inequality of Incomes », *Economic Journal*, Vol. 30 : p. 348-361.

Datt G. et Ravallion M. (1992), « Growth and Redistribution Components of Changes in Poverty Measures - A Decomposition with Applications to Brazil and India in the 1980s », *Journal of Development Economics*, Vol. 38 : p. 275-296.

Fei J.C.H., Ranis G. et Kuo S.W.Y (1978), *Growth with Equity – The Taiwan Case*, Oxford : Oxford University Press.

FGT - Foster J. E., Greer J. et Thorbecke E. (1984), « Notes and Comments. A Class of Decomposable Poverty Measures », *Econometrica*, Vol. 52 : p. 761-766.

Fields G. (1979), « Income Inequality in Urban Columbia : A decomposition Analysis », *Review of Income and Wealth*, Vol. 25(3) : p. 327-341.

Flückiger Y. et Silber J. (1995), « Income Inequality by Income Source and the Breakdown of Inequality Differences Between Two Population Subgroups », *Swiss Journal of Economics and Statistics*, Vol. 131(4/1) : p. 599-615.

Garner T. (1993), « Consumer Expenditures and Inequality: An Analysis Based on Decomposition of the Gini Coefficient », *Review of Economics and Statistics*, Vol. 75(1) : p. 134-138.

Gastwirth J. L. (1972), « The Estimation of the Lorenz Curve and the Gini Index », *Review of Economics and Statistics*, Vol. LIV : p. 306-316.

Gini C. (1912), « Variabilità e mutabilità », *Memori di Metodologia Statistica*, Vol. 1, *Variabilità e Concentrazione*. Libreria Eredi Virgilio Veschi, Rome, p. 211-382.

Gini C. (1914), *L'Ammontare e la composizione della ricchezza delle nazione*, Bocca, Torino.

Gini C. (1916), « Il concetto di transvariazione e le sue prime applicazioni », *Giornale degli Economisti e Rivista di Statistica*, dans Gini C. (ed.) (1959) : p.21-44.

Keynes J. M. (1936), « La théorie générale de l'emploi, de l'intérêt et de la monnaie », traduit de l'anglais par Jean de Largentaye, Payot, Paris.

Kolm S-C. (1966), « The Optimal Production of Social Justice », *Colloques Internationaux du CNRS*, Biarritz, 2-9 septembre 1966.

Kuznets S. (1955), « Economic growth and income inequality », *American Economic Review*, Vol. 45 : p. 1-28.

Lerman R. et Yitzhaki S. (1985), « Income Inequalities Effects by Income Source : A New Approach and Applications to United States », *Review of Economics and Statistics*, Vol. 67 : p.151-156.

Lerman R. et Yitzhaki S. (1989), « Income Sources and Income Inequality : Measurement from Three US Income Surveys », *Journal of Economic and Social Measurement*, Vol. 15(2) : p. 167-179.

Lorenz M. O. (1905), « Methods for Measuring Concentration of Wealth », *Journal of the American Statistical Association*, Vol. 70 : p. 209-219.

Milanovic B., (1997), « A Simple Way to Calculate the Gini Coefficient, and Some Implications », *Economics Letters*, Vol. 56 : p. 45-49.

Morduch J. et Sicular T. (2002), « Rethinking Inequality Decomposition, with Evidence from Rural China », *Economic Journal*, Vol. 112(476) : p. 93-106.

Mussard S. (2004), *Décompositions multidimensionnelles du rapport moyen de Gini. Applications aux revenus italiens de 1989 et 2000*, Thèse, Université de Montpellier I.

Owen G. (1977), *Values of Games with Priori Unions*, dans Heim R. et Moesclin O. (ed.) Essays in Mathematical Economics and Game Theory (Springer-Verlag).

Pareto V. (1896), « *Ecrits sur la courbe de la répartition de la richesse*. Œuvres complètes de Vilfredo Pareto publiées sous la direction de Giovanni Busino. Genève : Librairie Droz, 1965.

Pyatt G., Chen C-N, Fei J. (1980), « The Distribution of Income by Factor Component », *Quarterly Journal of Economics*, Vol. 95(3) : p. 451-473.

Podder N. (1993), « The Disaggregation of the Gini Coefficient by Factor Components and Its Applications to Australia », *Review of Income and Wealth*, Vol. 39(1) : p. 51-61.

Rao V.M. (1969), « Two Decompositions of Concentration Ratio », *Journal of the Royal Statistical Society*, Séries A 132 : p. 418-425.

Sastre M. et Trannoy A. (2002), « Shapley inequality decomposition by factor components: Some methodological issues », in P. Moyes, C. Seidl and A.F. Shorrocks (Eds.), *Inequalities: Theory, Experiments and Applications*, *Journal of Economics*, Supplément 9 : p. 51-90.

Sen A. K. (1973), *On Economic Inequality*, Clarendon Press, Oxford.

Shapley L. (1953), « A value for n-person games », in: H. W. Kuhn and A. W. Tucker, eds., *Contributions to the Theory of Games, Vol. 2* (Princeton University Press).

Shorrocks A. F. (1982), « Inequality Decomposition by Factor Component », *Econometrica*, Vol. 50 : p. 193-211.

Shorrocks A. F. (1983), « The Impact of Income Component on the Distribution of Family Incomes », *Quarterly Journal of Economics*, p. 311-326.

Shorrocks A. F. (1999) “Decomposition Procedures for Distributional Analysis: A Unified Framework Based on the Shapley Value”, *Mimeo*, University of Essex.

Silber J. (1989), « Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality », *Review of Economics and Statistics*, Vol. 71 : p. 107-115.

Silber J. (1993), « Inequality Decomposition by Income Source : a Note », *Review of Economics and Statistics*, Vol. 75 n° 3 : p. 545-547.

Silber J. (2004), « Inequalities : theory, experiments and applications », *European Journal of Political Economy*, Vol. 20 : p. 813-820.

Theil H. (1967), *Economics and Information Theory*, North-Holland Publishing Company, Amsterdam.

Tsui K. (1999), « Multidimensional Inequality and Multidimensional Generalized Entropy Measures : An axiomatic Derivation », *Social Choice and Welfare*, Vol. 16 : p. 145-157.

Yitzhaki S. (1983), « On an Extension of the Gini Inequality Index », *International Economic Review*, Vol. 24(3) : p. 617-28.

Yitzhaki S. (1990), « On the Effect of Subsidies to Basic Commodities on Inequality in Egypt », *Oxford Economic Papers*, Vol. 42 : p. 772-792.