Groupe de Recherche en Économie et Développement International

Cahier de recherche / Working Paper
05-04

Une réconciliation entre la décomposition en sous-groupe et la décomposition en sources de revenu de l'indice de Gini.
La multi-décomposition de l'indicateur de Gini

Stéphane Mussard

UNIVERSITÉ DE SHERBROOKE

Une réconciliation entre la décomposition en sous-groupes et la décomposition en sources de revenu de l'indice de Gini
La multi-décomposition de l'indicateur de Gini

Stéphane Mussard*†

Résumé

Nous proposons une nouvelle méthode de décomposition. La multi-décomposition de l'indicateur de Gini combine deux méthodes de décomposition : la décomposition en sous-groupes et la décomposition en sources de revenu. La technique retenue est mise en oeuvre dans le cadre d'une étude concernant les revenus italiens des années 1989 et 2000. Elle permet de mettre en évidence la faiblesse du rôle des transferts et des pensions dans la réduction des inégalités intragroupes et intergroupes.

Résumé

This research provides a new approach of decomposition. The multi-decomposition of the Gini index allows one to combine the two methods of decomposition : the decomposition by subgroup and the decomposition by income source. Thanks to the Italian income earners in 1989 and 2000, we illustrate this technique in order to show the weakness of the transfers and the pensions to reduce the within-group and the between-group inequalities.

Classification JEL : D63, D31.

Mots-clés : Décomposition en sources de revenu, Décomposition en sous-groupes, Multi-décomposition.

* *GREDI Université de Sherbrooke et LAMETA Université de Montpellier I, UFR Sciences Economiques, Avenue de la Mer - Site Richter - CS 79606, F-34960 Montpellier Cedex 2, France. E-mail : s-mussard@lameta.univ-montp1.fr*

† L'auteur remercie Camilo Dagum et Jacques Silber pour leurs précieux conseils sur les méthodes de décomposition de l'indicateur de Gini. Il tient aussi à remercier les arbitres anonymes des Annales et Bernard Philippe. Selon la formule consacrée, les éventuelles erreurs relèvent de l'entière responsabilité de l'auteur.

1 Introduction

Shorrocks (1980, 1982) propose des règles pour décomposer les mesures d'inégalité du revenu soit par sous-groupe (sous-population) soit par source de revenu (facteur). La première méthode se borne à décomposer la structure de l'indice en une mesure intragroupe et une mesure intergroupe. Lorsqu'une population est divisée en plusieurs groupes (par exemple hommes et femmes), le coefficient intragroupe représente l'intensité des inégalités de revenu qui prévaut à l'intérieur des groupes. La mesure intergroupe symbolise les inégalités qui existent entre les groupes de la population. La deuxième méthode de décomposition, la décomposition en facteurs, met en exergue le poids de chaque source de revenu (par exemple le revenu du travail, le revenu du capital, les taxes, etc.) au sein de l'inégalité globale. On peut ainsi déterminer les sources qui contribuent à accroître les inégalités dans la répartition du revenu.

Shorrocks (1999) reprend l'idée de Auvray et Trannoy (1992) (voir aussi Chantreil-Trannoy (1999) et Sastre-Trannoy (2002)) selon laquelle la valeur de Shapley (1953), outil de la théorie des jeux coopératifs, autorise la séparation des indicateurs d'inégalité et de pauvreté en sous-groupes ou en sources de revenu. Il mentionne aussi certaines possibilités de fusion entre les deux techniques de décomposition. Notamment celles qu'offre le recours aux indices de pauvreté additivement décomposables.

Un an avant les travaux de Shorrocks (1999), Chakravarty *et alii* (1998) avaient introduit, sans utiliser la valeur de Shapley, la classe des mesures de pauvreté multidimensionnelles simultanément décomposables en sous-groupes et en attributs. Le principal apport de leur recherche est le calcul des couples "attribut/groupe" où un attribut, dans l'approche multidimensionnelle, caractérise un niveau de santé, un niveau d'éducation, etc. La technique est attrayante. Elle permet de cerner tous les couples qui tendent à augmenter le degré de pauvreté total. A ce titre, elle peut servir aux décideurs dont le but est de réduire la pauvreté par des politiques socio-économiques de redistribution.

Jusqu'à ce jour, le domaine des mesures d'inégalité du revenu ne permettait pas aux chercheurs de recenser une méthode unifiant la décomposition en sous-groupes et la décomposition en sources de revenu. La principale difficulté de la réalisation de cette unification, en comparaison avec les mesures de pauvreté, réside dans le fait que les mesures d'inégalité décomposées en sous-groupes procurent plusieurs éléments. La greffe de la décomposition en sources de revenu sur la décomposition en sous-groupes paraît donc plus délicate. Pourtant la lecture de la littérature laisse penser que la synthèse est possible. Flückiger et Silber (1995) proposent une idée voisine en privilégiant l'indicateur de Gini. La différence des indices de Gini des hommes et des femmes peut être expliquée par les sources de revenu. Cet écart est défini par la décomposition en sources de revenu proposée par Silber (1989, 1993). Le modèle est donc proche de la synthèse, mais en définitive, ne permet pas d'atteindre une décomposition simultanée en sous-groupes et en sources de revenu, puisque la décomposition en sous-groupes n'est pas envisagée. En conséquence, ce modèle ne permet pas de connaître l'influence des inégalités féminines et masculines (et des facteurs sous-jacents) sur la valeur de l'inégalité globale. Néanmoins, les auteurs montrent que l'indicateur de Gini possède une structure favorable à l'utilisation des techniques de décomposition.

Le but de cet article est d'aboutir à la synthèse des deux techniques de décomposition par l'intermédiaire de la mesure de Gini. La multi-décomposition du coefficient de Gini est alors proposée. Elle permet le calcul de la contribution des sources de revenu aux mesures intragroupes et intergroupes. On peut ainsi évaluer les couples "source/intragroupe" et "source/intergroupe" qui déterminent le montant global de l'inégalité.

Le plan adopté pour présenter notre réflexion est le suivant. La section 2 décrit la décomposition en sous-groupes de l'indicateur de Gini réalisée par Dagum (1997a). Il s'agit de la méthode la plus récente. Elle spécifie de manière claire toutes les étapes de la décomposition. La section 3 introduit une décomposition de la mesure de Gini en sources de revenu. La section 4 expose l'unification des deux techniques. Enfin, avant de conclure en section 6, la section 5 est consacrée à un exemple d'application sur les revenus italiens des années 1989 et 2000.

2 La décomposition de l'indice de Gini en sous-groupes

En 1967, Bhattacharya et Mahalanobis fournissent l'une des premières décompositions en sous-populations[1]. Ensuite, Rao (1969), Pyatt (1976), Mookherjee et Shorrocks (1982), Silber (1989), Lerman et Yitzhaki (1991), Dagum (1997a, 1997b), font parti des nombreux auteurs qui ont contribué de manière significative à l'élaboration de la décomposition. Dans les lignes qui suivent, nous nous intéressons à celle qui a été présentée en dernière. La méthodologie de Dagum repose sur l'identification des éléments de la décomposition par l'intermédiaire d'indicateurs statistiques usuels.

Soit une population Q, composée de n individus dont les revenus sont notés $x_{Q,i}$ $(i = 1, \dots, n)$. Cette population globale comprend k sous-groupes Q_j $(j, h = 1, \dots, k)$ comprenant chacun n_j individus $(i, r = 1, \dots, n_j)$. On note μ la moyenne arithmétique des revenus de Q et μ_j la moyenne arithmétique des revenus de Q_j. Le coefficient de Gini associé à Q est mesuré par :

$$G = \frac{\sum_{i=1}^{n} \sum_{r=1}^{n} |x_{Q,i} - x_{Q,r}|}{2n^2\mu}. \tag{1}$$

L'indicateur de Gini permet de répertorier au total n^2 différences binaires de revenu. En rassemblant ces différences, il est possible de mettre en évidence

[1]La première peut être attribuée à Soltow (1960).

les inégalités intragroupes et les inégalités intergroupes brutes :

$$\begin{aligned}
G &= \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\sum_{r=1}^{n_j}|x_{j,i}-x_{j,r}|}{2n^2\mu} + \frac{2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|}{2n^2\mu} \\
&= G_w + G_{gb}. \qquad (2)
\end{aligned}$$

Le terme $x_{j,i}$ correspond au revenu de l'individu i appartenant au groupe Q_j. Cette méthode procure donc : (a) l'indice d'inégalité de Gini intragroupe (G_w) qui représente la contribution des inégalités issues de chaque groupe à l'inégalité totale ; (b) la contribution brute de l'indice de Gini intergroupe (G_{gb}) qui permet de jauger les écarts de revenu entre chaque paire de sous-groupes. Comme l'indiquait Rao (1969), cette composante intergroupe est très inhabituelle. En effet, elle ne révèle pas les différences moyennes entre les groupes (d'où l'expression de Dagum "contribution brute"). Elle paraît cependant plus complète puisque, par rapport aux inégalités moyennes, elle fournit davantage d'informations (par l'intermédiaire du calcul de toutes les différences binaires de revenu et de la comparaison entre chaque paire de sous-groupes).

Dagum montre que le coefficient de Gini peut être décomposé en trois indices lorsque la contribution brute intergroupe est séparée en deux composantes. La première est l'indice d'inégalité de Gini net intergroupe (G_{nb}), qui mesure les différences de revenu moyennes entre les groupes. La deuxième mesure l'intensité de transvariation[2] entre les groupes (G_t), qui est, l'inégalité provenant du chevauchement entre les distributions. Par conséquent, quatre composantes sont disponibles : G_w, G_{gb}, G_{nb} et G_t. Leur spécification conduit Dagum à introduire deux sortes d'indicateurs de Gini. Le coefficient de Gini associé à la sous-population Q_j (G_{jj}) et l'indicateur de Gini associé aux sous-population Q_j et Q_h (G_{jh}) sont respectivement donnés par (*Cf.* Dagum (1987)) :

$$G_{jj} = \frac{\sum_{i=1}^{n_j}\sum_{r=1}^{n_j}|x_{j,i}-x_{j,r}|}{2n_j^2\mu_j}\ ;\ G_{jh} = \frac{\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|}{n_j n_h(\mu_j+\mu_h)}. \qquad (3)$$

Ce sont des indicateurs de Gini au sens propre du terme. Le premier est le même que l'indice de Gini standard, que l'on applique à la sous-population Q_j. Le deuxième est différent car il concerne deux groupes, mais l'interprétation reste identique. Lorsque l'indice de Gini entre les groupes Q_j et Q_h tend vers la valeur 1, la répartition des revenus entre ces deux groupes est inégalitaire. A contrario, lorsque le coefficient est égal à zéro, la répartition est égalitaire.

[2]Depuis les articles de Silber (1989, 1993), Deutsch et Silber (1997) et Dagum (1997a), le troisième élément de la décomposition G_t a été clairement spécifié (voir aussi l'approche en terme de stratification de Lerman et Yitzhaki (1991)). La définition de Dagum (1997a) permet d'adopter l'expression d'intensité de transvariation issue du concept "transvariazione" de Gini (1916).

Lemme 2.1 : Dagum (1997a). *La distance directionnelle brute d_{jh} est une moyenne pondérée des différences de revenu $x_{j,i}-x_{h,r}$ pour chaque revenu $x_{j,i}$ d'un membre de Q_j supérieur au revenu $x_{h,r}$ d'un membre de Q_h ($x_{j,i} > x_{h,r}$) lorsque le groupe Q_j est en moyenne plus riche que le groupe Q_h ($\mu_j > \mu_h$) :*

$$d_{jh} = \int_0^\infty dF_j(x) \int_0^x (x-y)dF_h(y). \tag{4}$$

Les fonctions $F_j(x)$ et $F_h(y)$ sont respectivement les fonctions de répartition des sous-groupes Q_j et Q_h.

Lemme 2.2 : Dagum (1997a). *Le moment d'ordre 1 de transvariation,*

$$p_{jh} = \int_0^\infty dF_h(x) \int_0^x (x-y)dF_j(y), \tag{5}$$

est la moyenne pondérée des différences binaires de revenu $x_{h,r} - x_{j,i}$ telles que $x_{j,i} < x_{h,r}$ et $\mu_j > \mu_h$.

Les deux lemmes précédents permettent la mesure de la distance économique relative et directionnelle :

$$D_{jh} = \frac{d_{jh} - p_{jh}}{d_{jh} + p_{jh}}. \tag{6}$$

D_{jh}, élément de l'intervalle fermé [0,1], indique la distance existante entre les sous-groupes Q_j et Q_h (*Cf.* Dagum (1980)). En d'autres termes, c'est la proportion des différences binaires de revenu liée à la zone de non-chevauchement entre les distributions j et h. Par conséquent, l'expression $G_{jh}.D_{jh}$ permet d'évaluer les inégalités intergroupes nettes : les inégalités moyennes entre les sous-groupes Q_j et Q_h. En effet, lorsque les moyennes sont égales, $\mu_j = \mu_h$, la distance économique entre Q_j et Q_h est nulle, ce qui implique la nullité de l'inégalité intergroupe nette de Gini. A contrario, l'expression $G_{jh}(1 - D_{jh})$ donne l'intensité de transvariation entre les sous-populations Q_j et Q_h. Cette intensité est une fonction croissante de l'écart qui existe entre les revenus élevés du groupe qui est le plus pauvre en moyenne et les revenus faibles du groupe le plus riche en moyenne (soit $x_{h,r} - x_{i,j} > 0$). Par définition, la transvariation (Gini (1916), Dagum (1960)) désigne les différences de revenu qui sont de signe opposé à celui de la différence des moyennes des sous-groupes correspondants (*Cf.* Dagum (1997a)). Le pourcentage d'individus appartenant au sous-groupe Q_j (p_j) et leur part de revenu dans le revenu de la population totale (s_j) sont donnés par :

$$p_j = \frac{n_j}{n} \ ; \ s_j = \frac{n_j \mu_j}{n\mu}. \tag{7}$$

Ces pondérations permettent d'introduire les indices de la décomposition en sous-groupes. L'indicateur d'inégalité de Gini *intragroupe* est :

$$G_w = \sum_{j=1}^{k} G_{jj} p_j s_j. \tag{8}$$

L'indicateur de Gini *intergroupe net* entre les sous-populations est mesuré par :

$$G_{nb} = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} D_{jh} (p_j s_h + p_h s_j). \tag{9}$$

L'intensité de *transvariation intergroupe* l'est par :

$$G_t = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} (1 - D_{jh}) (p_j s_h + p_h s_j). \tag{10}$$

Le coefficient de Gini *intergroupe brut* est :

$$G_{gb} = G_{nb} + G_t = \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} (p_j s_h + p_h s_j). \tag{11}$$

En définitive, le résultat fondamental de Dagum peut être exprimé sous la forme du théorème suivant.

Théorème 2.3 : Dagum (1997a). *De* (8), (9), (10) et (11), *on déduit que la décomposition de la mesure de Gini en sous-groupes est donnée par :*

$$\begin{aligned} G &= G_w + G_{nb} + G_t \\ &= G_w + G_{gb}. \end{aligned} \tag{12}$$

3 La décomposition de l'indice de Gini en sources de revenu

Rao (1969), Fei *et alii* (1978), Shorrocks (1982) ont échafaudé la méthode de décomposition en facteurs du coefficient de Gini. Ces approches sont basées sur le pseudo-Gini[3]. Nous introduisons une méthode différente permettant de mesurer la contribution de chaque source de revenu à l'inégalité totale. Nous partons de la formule bien connue suivante :

$$|x_{Q,i} - x_{Q,r}| = x_{Q,i} + x_{Q,r} - 2min\{x_{Q,i}, x_{Q,r}\}. \tag{13}$$

[3]Il s'agit de l'indicateur de Gini mesuré sur une source particulière où les individus sont ordonnés conformément à leur rang dans la distribution de revenu globale (et où les revenus sont ordonnés de manière croissante).

L'indice de Gini global, mesuré sur la population Q, est tel que :

$$G = \frac{\sum_{i=1}^{n}\sum_{r=1}^{n}(x_{Q,i} + x_{Q,r} - 2min\{x_{Q,i}, x_{Q,r}\})}{2n^2\mu}. \tag{14}$$

Supposons que les revenus de chaque individu soient divisés en q sources x^m ($m = 1, \ldots, q$). Le revenu du i^{me} individu de la population Q est alors additivement séparé :

$$x_{Q,i} = \sum_{m=1}^{q} x_{Q,i}^{m}. \tag{15}$$

L'indice de Gini devient :

$$G = \frac{\sum_{i=1}^{n}\sum_{r=1}^{n}(\sum_{m=1}^{q} x_{Q,i}^{m} + \sum_{m=1}^{q} x_{Q,r}^{m} - 2min\{x_{Q,i}, x_{Q,r}\})}{2n^2\mu}. \tag{16}$$

Pour séparer l'expression $2min\{x_{Q,i}, x_{Q,r}\}$ en facteurs, on utilise la notation :

$$\sum_{m=1}^{q} 2x_{Q,ir}^{*m} := 2min\{x_{Q,i}, x_{Q,r}\}. \tag{17}$$

Exemple 3.1 *Soit* $x_{Q,i} = 2$ *avec* $x_{Q,i} = x_{Q,i}^{1} + x_{Q,i}^{2} = 0 + 2$ *et* $x_{Q,r} = 4$ *avec* $x_{Q,r} = x_{Q,r}^{1} + x_{Q,r}^{2} = 2 + 2$. *On obtient :* $2min\{x_{Q,i}, x_{Q,r}\} = \sum_{m=1}^{q} 2x_{Q,ir}^{*m} = 2x_{Q,i}^{1} + 2x_{Q,i}^{2} = 2(0) + 2(2)$.

Ceci nous permet d'énoncer la proposition suivante.

Proposition 3.2 *Une décomposition de l'indicateur de Gini peut s'écrire :*

$$G = \sum_{m=1}^{q} \frac{\sum_{i=1}^{n}\sum_{r=1}^{n}(x_{Q,i}^{m} + x_{Q,r}^{m} - 2x_{Q,ir}^{*m})}{2n^2\mu}, \tag{18}$$

où S_m *peut être assimilée à la contribution du facteur* m *à la mesure du Gini global.*

Preuve : *Cf.* Annexe. □

Cette proposition signifie que l'indice de Gini est séparé en q composantes dont chacune représente la contribution absolue de chaque facteur S_m au montant de l'inégalité totale. Il est ainsi possible d'estimer la contribution d'une source de revenu précise (comme les salaires nets, les primes, les prestations sociales, etc.) à la valeur du Gini total. La proposition admet donc le corollaire suivant.

Corollaire 3.3 *Si les* q *distributions de facteur sont* q *réplications, correspondant à* q *variables i.i.d., alors :*

$$S^1 = \ldots = S^m = \ldots = S^q \Longrightarrow G = qS^m. \tag{19}$$

Preuve : La démonstration est évidente. Prenons un exemple. Soit une population de 3 individus dont les revenus, $x = \{2, 4, 6\}$, sont constitués de deux sources. Les distributions des facteurs sont : $x^1 = \{1, 2, 3\}$ et $x^2 = \{1, 2, 3\}$. On a alors :

$$S^1 = S^2 = \frac{2(1 + 2 - 2(1)) + 2(1 + 3 - 2(1)) + 2(2 + 3 - 2(2))}{2.4.3^2} = \frac{8}{72}. \quad (20)$$

Le coefficient de Gini se détermine par :

$$G = qS^m = 2\frac{8}{72} = \frac{2}{9}. \square \quad (21)$$

Ce corollaire montre que la mesure S^m accorde la même valeur aux sources identiquement distribuées. On trouve le même résultat lorsque les sources sont distribuées de manière égalitaire.

Corollaire 3.4 *Si les q distributions de facteur sont q distributions également distribuées, alors :*

$$S^1 = \ldots = S^m = \ldots = S^q \Longrightarrow G = 0. \quad (22)$$

Preuve : Dès lors qu'une distribution est répartie de manière égalitaire, les différences binaires de revenu sont nulles, ce qui implique la nullité des contributions S^m. Prenons un exemple. Si $x^1 = \{1, 1, 1\}$, $x^2 = \{3, 3, 3\}$ et $x = \{4, 4, 4\}$, on a :

$$\begin{aligned} S^1 &= \frac{2(1 + 1 - 2(1)) + 2(1 + 1 - 2(1)) + 2(1 + 1 - 2(1))}{2.4.3^2} = 0 = S^2 \\ &\Longrightarrow G = 0. \square \end{aligned} \quad (23)$$

Dans le cadre des deux résultats précédents, l'indicateur de Gini réagit de la même façon que la mesure S^m. Cependant S^m reste différente puisqu'elle peut conduire à des valeurs négatives.

Corollaire 3.5 *La contribution de la source m à l'inégalité totale peut être négative.*

Preuve : Soit une distribution globale $x = \{1, 2, 3\}$ où $x^m = \{-1, -2, -3\}$. La mesure S^m est négative :

$$\begin{aligned} S^m &= \frac{2(-1 - 2 - 2(-1)) + 2(-1 - 3 - 2(-1)) + 2(-2 - 3 - 2(-2))}{180} \\ &= \frac{-2}{45}. \square \end{aligned} \quad (24)$$

La mesure S^m indique que certaines sources, comme les transferts, peuvent réduire les inégalités. Une autre propriété va différencier S^m et G.

Corollaire 3.6 *Supposons que certaines distributions de facteur soient proportionnelles à la distribution x^m. Les contributions de ces sources de revenu à l'inégalité totale seront donc proportionnelles à la contribution de la source m.*

Preuve : Soit une distribution composée de deux sources de revenu telles que : $x^1 = \{a, b, c\}$ et $x^2 = \lambda x^1 = \{d, e, f\}$, où $\lambda > 0$. On a alors :

$$\begin{aligned} S^1 = & \frac{2(a+b-2min\{a,b\}) + 2(a+c-2min\{a,c\})}{2n^2\mu} \\ & + \frac{2(b+c-2min\{b,c\})}{2n^2\mu}. \end{aligned} \tag{25}$$

On peut ensuite déterminer la contribution de la source 2 par rapport à celle de la source 1 :

$$\begin{aligned} S^2 & = \frac{2(d+e-2min\{d,e\}) + 2(d+f-2min\{d,f\}) + 2(e+f-2min\{e,f\})}{2n^2\mu} \\ & = \frac{2(\lambda a+\lambda b-2min\{\lambda a,\lambda b\}) + 2(\lambda a+\lambda c-2min\{\lambda a,\lambda c\})}{2n^2\mu} \\ & + \frac{2(\lambda b+\lambda c-2min\{\lambda b,\lambda c\})}{2n^2\mu} \\ & = \lambda\frac{2(a+b-2min\{a,b\} + 2(a+c-2min\{a,c\}) + 2(b+c-2min\{b,c\})}{2n^2\mu} \\ & = \lambda S^1. \ \square \end{aligned} \tag{26}$$

On constate que les contributions des facteurs sont proportionnellement liées, avec la même intensité que le lien unissant les distributions de facteur. Par exemple, si les salaires sont cinq fois plus importants que les primes (ceci pour tous les individus), alors la contribution des salaires à l'inégalité totale est cinq fois plus élevée que celle des primes.

Les quatre corollaires montrent que l'indicateur S^m paraît heuristique. Néanmoins, le calcul des contributions S^m exprime une dépendance entre les facteurs et les revenus[4]. On retrouve ce genre de dépendance dans le calcul du pseudo-Gini. Mais le recours à S^m est préférable dans la conception de la multi-décomposition.

[4]Les termes x^* montrent une dépendance entre la source de revenu et le revenu global des individus.

4 La multi-décomposition de l'indicateur de Gini

A partir de la décomposition en facteurs (18), il est possible de décomposer l'indice de Gini mesuré sur la sous-population Q_j :

$$G_{jj} = \sum_{m=1}^{q} \frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_j} (x_{j,i}^m + x_{j,r}^m - 2x_{j,ir}^{*m})}{2n_j^2 \mu_j} = \sum_{m=1}^{q} S_j^m, \tag{27}$$

où S_j^m est la contribution de la source m à G_{jj}. En conséquence, la désagrégation de l'indice de Gini intragroupe (G_w) est donnée par :

$$G_w = \sum_{j=1}^{k} G_{jj} p_j s_j = \sum_{m=1}^{q} \sum_{j=1}^{k} S_j^m p_j s_j. \tag{28}$$

Les inégalités à l'intérieur des groupes sont expliquées par les q sources de revenu. On peut donc estimer l'ensemble des couples "source m/groupe Q_j" qui composent l'indicateur d'inégalité de Gini intragroupe (G_w).

La seconde étape de la multi-décomposition concerne la séparation des mesures intergroupes. On passe tout d'abord par la phase de décomposition de la distance économique relative et directionnelle D_{jh} et celle de sa contrepartie $P_{jh} := 1 - D_{jh}$. La mesure P_{jh} est un élément inclus dans l'intervalle fermé [0,1]. Elle mesure le pourcentage des différences binaires de revenu issues du chevauchement entre les distributions des sous-groupes Q_j et Q_h. Il s'agit donc d'un ratio de chevauchement.

Proposition 4.1 *La distance économique relative et directionnelle D_{jh} et le ratio de chevauchement P_{jh} sont décomposables en facteurs :*

$$D_{jh} = \sum_{m=1}^{q} D_{jh}^m \ ,$$

$$P_{jh} = \sum_{m=1}^{q} P_{jh}^m \ , \tag{29}$$

où D_{jh}^m et P_{jh}^m sont respectivement les contributions de la source m aux mesures D_{jh} et P_{jh}.

Preuve : *Cf.* Annexe. □

La décomposition des mesures D_{jh} et P_{jh} constitue la pierre angulaire des décompositions en facteurs des éléments G_{nb}, G_t, et G_{gb}. On associe ensuite les décompositions en sources de revenu de D_{jh} et P_{jh} aux pondérations de population (p_j) et de revenu (s_j) afin de réduire les expressions des inégalités nettes et de transvariation :

$$w_{nb,jh}^m := D_{jh}^m (p_j s_h + p_h s_j) \ , \tag{30}$$

$$w_{t,jh}^m := P_{jh}^m (p_j s_h + p_h s_j) \ . \tag{31}$$

En considérant les équations (9), (10), (11), et les pondérations (30) et (31), on obtient la décomposition en facteurs de l'indice de Gini *net intergroupe*, de l'intensité de *transvariation intergroupe*, et de l'indice de Gini *brut intergroupe* :

$$G_{nb} = \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} w_{nb,jh}^{m} \ , \tag{32}$$

$$G_{t} = \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} w_{t,jh}^{m} \ , \tag{33}$$

$$G_{gb} = \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} (w_{nb,jh}^{m} + w_{t,jh}^{m}) \ . \tag{34}$$

Proposition 4.2 *La multi-décomposition de l'indicateur de Gini en deux composantes correspond à la séparation de la décomposition de Dagum (2) en sources de revenu :*

$$\begin{aligned} G &= G_w + G_{gb} \\ &= \sum_{m=1}^{q} \sum_{j=1}^{k} S_j^m p_j s_j + \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} (w_{nb,jh}^{m} + w_{t,jh}^{m}). \end{aligned} \tag{35}$$

Elle est égale à la multi-décomposition exacte suivante :

$$\begin{aligned} G = & \sum_{m=1}^{q} \frac{\sum_{j=1}^{k} \sum_{i=1}^{n_j} \sum_{r=1}^{n_j} (x_{j,i}^m + x_{j,r}^m - 2x_{j,ir}^{*m})}{2n^2\mu} \\ & + \sum_{m=1}^{q} \frac{2 \sum_{j=2}^{k} \sum_{h=1}^{j-1} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} (x_{j,i}^m + x_{h,r}^m - 2x_{jh,ir}^{*m})}{2n^2\mu}. \end{aligned} \tag{36}$$

Preuve : *Cf.* Annexe. □

L'équation (35) présente un indice de Gini avec une structure naturelle de multi-décomposition. L'équation (36) indique que cette décomposition naturelle permet le calcul de toutes les contributions possibles (sources, sous-groupes, sources et sous-groupes). Cette seconde formulation révèle aussi qu'il n'existe pas de termes redondants : la multi-décomposition est exacte. Autrement dit, aucun terme multiplicatif d'interaction ne vient contraindre le calcul d'une contribution particulière[5]. En revanche, la première formulation équivalente (35), signale qu'il n'est pas nécessaire de décomposer les indices de Gini entre deux groupes (G_{jh}). Il suffit en effet de décomposer la distance économique et le ratio de chevauchement pour que la multi-décomposition soit parfaite.

[5]On pourrait d'ailleurs démontrer que l'entropie généralisée n'est pas multi-décomposable. Elle fait apparaître des termes redondants.

Comment interpréter cette multi-décomposition en deux éléments ? Le premier terme représente les inégalités intragroupes expliquées par les sources de revenu. Par exemple, si la population est divisée en deux régions $\mathcal{A}$ et $\mathcal{B}$, on va pouvoir mesurer l'influence des facteurs (salaires nets, primes, taxes, etc.) sur les inégalités de la région $\mathcal{A}$ et sur celle de la région $\mathcal{B}$. Le second terme de la multi-décomposition caractérise les inégalités entre les paires de groupes (ici les régions $\mathcal{A}$ et $\mathcal{B}$). Les différences de revenu entre les régions $\mathcal{A}$ et $\mathcal{B}$ vont donc être expliquées par les q facteurs. Au final, il est possible d'isoler tous les couples "source m/intragroupe Q_j" et "source m/entre Q_j et Q_h" dont la somme fournit l'indicateur de Gini global.

Proposition 4.3 *La multi-décomposition de l'indicateur de Gini en trois éléments permet de séparer parfaitement le montant global des inégalités simultanément par sous-groupe et par source de revenu, en faisant la distinction entre l'indice de Gini net intergroupe et l'intensité de transvariation entre les groupes :*

$$G = \sum_{m=1}^{q} \sum_{j=1}^{k} S_j^m p_j s_j + \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} w_{nb,jh}^m + \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} w_{t,jh}^m. \tag{37}$$

Preuve : *Cf.* Annexe. □

Cette proposition prouve que l'indice de Gini intragroupe (G_w), l'indice de Gini net intergroupe (G_{nb}), et l'intensité de transvariation intergroupe (G_t) sont décomposables en sources de revenu. De plus, on montre que cette décomposition est parfaite (*Cf.* la preuve en annexe). Comment interpréter cette multi-décomposition en trois composantes ? Les inégalités nettes intergroupes entre les régions $\mathcal{A}$ et $\mathcal{B}$, autrement dit les inégalités moyennes entre les régions, sont expliquées par les salaires, les primes, etc. D'autre part, les inégalités de transvariation entre les régions $\mathcal{A}$ et $\mathcal{B}$ sont aussi expliquées par les q facteurs. Supposons que la région $\mathcal{B}$ soit, en moyenne, plus riche que la région $\mathcal{A}$. L'intensité de transvariation signifie que la différence entre les hauts revenus de $\mathcal{A}$ et les faibles revenus de $\mathcal{B}$ contribue à creuser l'écart de revenu global. L'intensité de transvariation entre les deux régions est expliquée par les q sources de revenu.

Les deux propositions précédentes montrent que le coefficient de Gini est issu de l'intersection de deux domaines : les mesures d'inégalité décomposables en facteurs et les mesures d'inégalité décomposables en sous-groupes. En conséquence, il existe une propriété de multi-décomposition qui s'exprime par :

$$I(x) = \sum_{m=1}^{q} S_W^m + \sum_{m=1}^{q} S_B^m. \tag{MD}$$

L'indicateur S_W^m est la contribution de la source m de l'inégalité intragroupe à l'inégalité totale $I(x)$. Le coefficient S_B^m représente la contribution du facteur

m des inégalités intergroupes (brutes) à $I(x)$. La propriété (MD) permet donc de spécifier une extension de la propriété de décomposition cohérente de Shorrocks (1982) :

$$I(x) = \sum_{m=1}^{q} S^m, \tag{CD}$$

où S^m est la contribution de la source m à l'inégalité totale.

Mussard (2004) montre que l'indice GMR (Gini Mean Ratio) vérifie aussi cette propriété de décomposition combinatoire (MD). Cependant, la complexité de cette synthèse n'a donné lieu à aucune tentative de généralisation. Il est néanmoins possible de généraliser la multi-décomposition de la mesure de Gini à des facteurs dont la portée dépasse les simples sources de revenu.

Corollaire 4.4 *Lorsque les revenus des individus sont linéairement régressés sur q variables explicatives, on obtient une multi-décomposition paramétrique.*

Preuve : Supposons que la détermination des revenus des individus x_i puisse être représentée par un modèle linéaire ou par un modèle linéairement transformable par anamorphose :

$$x_i = \alpha_0 + \sum_{m=1}^{q} \alpha_m y_i^m + \epsilon_i. \tag{38}$$

On pose : $x_i^0 = \alpha_0$; $x_i^m = \alpha_m y_i^m$; $x_i^{q+1} = \epsilon_i$. On retrouve la relation additive :

$$x_i = \sum_{m=0}^{q+1} x_i^m, \tag{39}$$

qui procure une multi-décomposition exacte. Les indices de Gini G_w, G_{nb} et G_t sont ainsi expliqués par les q variables explicatives, par la qualité d'ajustement (ϵ) et par ce que le modèle est incapable d'expliquer (α_0). $\square$

5 Un exemple d'application sur les revenus italiens de 1989 et 2000

Les données proviennent du "*Survey of Households' Income and Wealth*" de la Banque d'Italie. L'échantillon comprend 13809 individus en 1989 (5538 femmes et 8271 hommes) et 14254 individus en 2000 (6343 femmes et 7911 hommes). Les revenus annuels sont fractionnés en sept facteurs :
(a) salaires nets ;
(b) primes ;
(c) pensions (et arriérés) ;
(d) autres transferts (scolarité, pensions alimentaires) ;

(e) revenus entrepreneuriaux ;
(f) revenus immobiliers ;
(g) et revenus financiers.

La décomposition en sous-groupes de Dagum fournit les résultats suivants :

$$G_{1989} = G_w + G_{nb} + G_t = 0,1718 + 0,1212 + 0,0475 = 0,3405; \quad (40)$$
$$G_{2000} = G_w + G_{nb} + G_t = 0,1883 + 0,121 + 0,0762 = 0,3855. \quad (41)$$

L'indice de Gini augmente de 13,22% entre 1989 et 2000. Les inégalités moyennes, autrement dit, l'indice de Gini net intergroupe (G_{nb}) reste quasi-constant. En conséquence, l'augmentation des inégalités est imputable à la croissance des inégalités intragroupes et des inégalités de transvariation. En 1989, l'indice de Gini masculin ($G_{jj} = 0,3199$) est supérieur à celui des femmes ($G_{jj} = 0,3016$; *Cf.* Tableau 1).

Tableau 1 : Les composantes intragroupes en 1989

Indices	Hommes	Femmes	Total
Individus n_i	8271	5538	13809
Revenus Moyens μ_j (euros)	12771,33	7408,33	10620,53
Ecart-types (euros)	9779,24	4473,72	8497,98
Gini G_{jj}	0,3199	0,3016	**0,3405**
Contribution de la source a	0,0391	0,0229	**0,0620**
Contribution de la source b	0,0011	0	**0,0011**
Contribution de la source c	-0,0037	-0,0027	**-0,0064**
Contribution de la source d	-0,0003	0,0003	**0**
Contribution de la source e	0,0472	0,0076	**0,0548**
Contribution de la source f	0,038	0,0043	**0,0423**
Contribution de la source g	0,0167	0,0013	**0,018**
Total	$G_{wHommes}$ = **0,1381**	$G_{wFemmes}$ = **0,0337**	G_w = **0,1718**

Source : Survey of Households' Income and Wealth 1989, Banque d'Italie

En 2000, ces indices subissent une forte augmentation. De plus, l'indice de Gini féminin ($G_{jj} = 0,37$) devient plus important que celui du groupe masculin ($G_{jj} = 0,3597$; *Cf.* Tableau 4). L'indice de Gini intragroupe devient donc plus intense. Cependant, comme en 1989, la contribution des hommes à l'inégalité totale reste supérieure ($G_{wHommes} = 0,1349$ et $G_{wFemmes} = 0,0534$) car les pondérations de taille et de revenu sont plus importantes pour le groupe masculin. L'intensité de transvariation représente les inégalités issues des différences entre les hauts revenus féminins et les faibles revenus masculins. Ces inégalités sont de plus en plus importantes ($G_{t1989} = 0,0475$ *Cf.* Tableau 2 et $G_{t2000} = 0,0762$ *Cf.* Tableau 5). Elles sont intéressantes dans la mesure où elles reflètent un rapprochement entre les distributions de revenu masculines et féminines (*Cf.* aussi les tableaux 3 et 6 où le ratio de chevauchement P_{jh} est de plus en plus important).

Tableau 2 : Les composantes intergroupes en 1989

Sources	Gini Net Intergroupe	Transvariation Intergroupe	Total
Contribution de la source a	0,04	0,0278	**0,0678**
Contribution de la source b	0,0009	0,0001	**0,001**
Contribution de la source c	-0,0071	-0,0012	**-0,0083**
Contribution de la source d	-0,0005	0,0007	**0,0002**
Contribution de la source e	0,0358	0,0137	**0,0495**
Contribution de la source f	0,0394	0,0037	**0,0431**
Contribution de la source g	0,0127	0,0027	**0,0154**
Total	$G_{nb} =$ **0,1212**	$G_t =$ **0,0475**	$G_{gb} =$ **0,1687**
Gini G_{jh}			0,3695

Source : Survey of Households' Income and Wealth 1989, Banque d'Italie

Tableau 3 : Distances et Ratios de Chevauchement en 1989

Sources	D_{jh}	P_{jh}	Total
Contribution de la source a	0,2374	0,1646	**0,402**
Contribution de la source b	0,0054	0,0006	**0,006**
Contribution de la source c	-0,0419	-0,0073	**-0,0492**
Contribution de la source d	-0,0033	0,0042	**0,0009**
Contribution de la source e	0,2125	0,0812	**0,2937**
Contribution de la source f	0,2336	0,0217	**0,2553**
Contribution de la source g	0,0754	0,0159	**0,0913**
Total	$D_{jh} = 0{,}7191$	$P_{jh} = 0{,}2809$	**1**

Source : Survey of Households' Income and Wealth 1989, Banque d'Italie

Quelles sont les causes de l'augmentation de la transvariation ? En 1989, la principale explication de G_t (*Cf.* Tableau 2) réside dans les salaires nets (a) et les revenus entrepreneuriaux (e). En 2000 (*Cf.* Tableau 5), les salaires nets (a) et les revenus financiers (g) représentent respectivement 37,3% et 27,2% de G_t. En 1989, les pensions et arriérés (c) permettaient de réduire l'intensité de la transvariation à hauteur de 2,53% de G_t, alors qu'en 2000 elles augmentent les inégalités globales à hauteur de 11,87% de G_t. La même remarque peut être faite concernant les contributions des distances économiques et des ratios de chevauchement (*Cf.* Tableau 3 et Tableau 6).

Tableau 4 : Les composantes intragroupes en 2000

Indices	Hommes	Femmes	Total
Individus n_i	7911	6343	14254
Revenus Moyens μ_j (euros)	18687,09	11168,56	15341,36
Ecart-types (euros)	16292,72	9923,88	14321,76
Gini G_{jj}	0,3597	0,37	**0,3855**
Contribution de la source a	0,0367	0,0243	**0,061**
Contribution de la source b	0,0009	0,0002	**0,0011**
Contribution de la source c	0,0131	0,0062	**0,0193**
Contribution de la source d	-0,0002	0,0001	**-0,0001**
Contribution de la source e	0,0355	0,0079	**0,0434**
Contribution de la source f	0,0387	0,0116	**0,0503**
Contribution de la source g	0,0102	0,0031	**0,0131**
Total	$G_{wHommes} =$ **0,1349**	$G_{wFemmes} =$ **0,0534**	$G_w =$ **0,1883**

Source : Survey of Households' Income and Wealth 2000, Banque d'Italie

Tableau 5 : Les composantes intergroupes en 2000

Sources	Gini Net Intergroupe	Transvariation Intergroupe	Total
Contribution de la source a	0,0394	0,0284	**0,0678**
Contribution de la source b	0,0008	0,0004	**0,0012**
Contribution de la source c	0,0106	0,0094	**0,02**
Contribution de la source d	-0,0003	0,0001	**-0,0002**
Contribution de la source e	0,0344	0,0104	**0,0448**
Contribution de la source f	0,0306	0,0207	**0,0513**
Contribution de la source g	0,0055	0,0068	**0,0123**
Total	$G_{nb} =$ **0,121**	$G_t =$ **0,0762**	$G_{gb} =$ **0,1972**
Gini G_{jh}			0,4104

Source : Survey of Households' Income and Wealth 2000, Banque d'Italie

Que représente le poids des transferts et des pensions ? En 2000, la réduction de l'indice de Gini intragroupe est imputable aux pensions et aux transferts (excepté les transferts féminins, *Cf.* Tableau 2). De manière surprenante, les pensions n'ont désormais plus vocation à réduire les inégalités. En effet, en 2000, la contribution de la source (c) est toujours positive. D'autre part, les transferts réduisent les inégalités entre les hommes avec exactement la même amplitude qu'en 1989, soit 0,052% de l'inégalité globale en 2000 (-0,0002, *Cf.* Tableau 4). Les transferts réduisent les inégalités moyennes entre les groupes (G_{nb}). On remarque qu'en 1989 cette diminution était plus importante (-0.0005 en 1989 et -0.0003 en 2000). Pourtant l'indice de Gini net intergroupe est le même en 1989 et 2000.

Tableau 6 : Distances et Ratios de Chevauchement en 2000

Sources	D_{jh}	P_{jh}	Total
Contribution de la source a	0,1998	0,1437	**0,3435**
Contribution de la source b	0,0042	0,0018	**0,006**
Contribution de la source c	0,0537	0,0478	**0,1015**
Contribution de la source d	-0,0014	0,0007	**-0,0007**
Contribution de la source e	0,1743	0,0528	**0,2271**
Contribution de la source f	0,1550	0,1051	**0,2601**
Contribution de la source g	0,0281	0,0344	**0,0625**
Total	D_{jh} = 0,6137	P_{jh} = 0,3863	**1**

Source : Survey of Households' Income and Wealth 2000, Banque d'Italie

Le système fiscal italien est-il vraiment efficace pour réduire les inégalités entre les hommes et les femmes ? De manière globale la réponse est positive, car en 2000, les contributions globales des transferts sont négatives (voir les colonnes Total des Tableaux 4, 5 et 6). En revanche, la contribution des transferts à la réduction de l'inégalité totale reste infinitésimale.

6 Conclusion

Cet article explore quatre points fondamentaux. Il permet premièrement de revoir la décomposition de Dagum (1997a, 1997b) selon laquelle le coefficient de Gini global est séparé en trois indices de Gini particuliers. Ensuite, une nouvelle approche de décomposition en sources de revenu est offerte. Elle attribue une part proportionnelle à chaque facteur. Troisièmement, cette décomposition en facteurs possède une configuration autorisant une multi-décomposition du coefficient de Gini (équations (36) et (37)). Cette méthode particulière est issue de l'intersection entre le domaine des mesures d'inégalité décomposables en sources de revenu et celui des mesures décomposables en sous-groupes. On s'aperçoit de plus que l'introduction d'un modèle économétrique linéaire permet un contrôle des variables qui agissent sur l'indice de Gini intragroupe (G_w) et sur les indices de Gini intergroupes (G_{nb}, G_{gb}, et G_t). Cette multi-décomposition paramétrique offre, de plus, des perspectives beaucoup plus larges que les simples sources de revenu, puisque toute sorte de dimension peut être envisagée. Quatrièmement, elle permet de faire un constat surprenant : les inégalités de revenu italiennes ont été accrues par les transferts en 1989 et par les pensions en 2000.

Au-delà de ce résultat, cette multi-décomposition permet d'obtenir d'autres conclusions. Les déterminants des inégalités sont très nombreux. La multi-décomposition permet alors de détecter précisément le rôle de chaque déterminant, son influence sur l'inégalité intragroupe, l'inégalité intergroupe et l'inégalité totale. Cette détection permet ensuite de calculer les contributions des combinaisons "facteur/intragroupe" et "facteur/intergroupe". Or, les contributions de ces combinaisons ne peuvent être évaluées lorsqu'on se contente des décompositions marginales (la décomposition en sous-groupes ou la décomposition en facteurs). La multi-décomposition permet donc d'envisager des politiques économiques redistributives en se focalisant sur les couples qui tendent à augmenter les inégalités ou qui ne réduisent pas assez les inégalités. Dans le cadre des distributions de revenu italiennes, les preneurs de décision doivent consolider leur système de redistribution notamment afin de réduire les inégalités moyennes entre les hommes et les femmes.

En définitive, la multi-décomposition peut permettre de tester les effets de toutes les politiques fiscales. Par exemple, lorsque les décideurs souhaitent modifier les taux d'imposition, la source transfert se modifie, et sa contribution aux inégalités intragroupes et intergroupes varie. Par conséquent, il est possible de simuler l'impact des changements du système fiscal sur les inégalités intragroupes et intergroupes. On peut par ailleurs simuler ces effets par rapport à toutes décisions visant à changer les structures du revenu comme les revenus financiers, les pensions, etc. En définitive, on peut raisonnablement espérer que la multi-décomposition paramétrique permette aussi l'étude des liens unissant les politiques économiques et les inégalités lorsque les facteurs explicatifs du revenu sont l'âge, l'éducation et la santé.

7 Références bibliographiques

AUVRAY C., TRANNOY A. (1992). - "Décomposition par source de l'inégalité des revenus à l'aide de la Valeur Shapley", *Journées de Microéconomie Appliquée*, Sfax.

BHATTACHARYA N., MAHALANOBIS B. (1967). - "Regional Disparities in Household Consumption in India", *Journal of the American Statistical Association*, Vol 62, pp. 143-161.

CHAKRAVARTY S.R., MUKHERJEE D., RANADE R.R. (1998). - "On the Family of Subgroup and Factor Decomposable Measures of Multidimensional Poverty", *Research on Economic Inequality*, Vol 8, pp. 175-194.

CHANTREUIL F., TRANNOY A. (1999). - "Inequality Decomposition Values : The Trade-off Between Marginality and Consistency", DP N° 9924 THEMA.

DAGUM C. (1960). - "Teoria de la transvariacion, sus aplicaciones a la economia", *Metron*, Vol XX, pp. 1-206.

DAGUM C. (1980). - "Inequality Measures Between Income Distributions with Applications", *Econometrica*, Vol 48, N° 7, pp. 1791-1803.

DAGUM C. (1987). - "Measuring the Economic Affluence Between Populations of Income Receivers", *Journal of Business and Economic Statistics*, Vol 5, N° 1, pp. 5-12.

DAGUM C. (1997a). - "A New Approach to the Decomposition of the Gini Income Inequality Ratio", *Empirical Economics*, Vol 22, pp. 515-531.

DAGUM C. (1997b). - "Decomposition and Interpretation of Gini and the Generalized Entropy Inequality Measures", *American Statistical Association*, Section 'Business and Economic Statistics', pp. 200-205.

DEUTSCH J., SILBER J. (1997). - "Gini's 'Transvariazione' and the Measurement of Distance Between Distributions", *Empirical Economics*, Vol 22, pp. 547-554.

FEI J. C. H., RANIS G., KUO S. W. Y. (1978). - "Growth and the Family Distribution of Income by Factor Components", *Quarterly Journal of Economics*, Vol 92, pp. 17-53.

FLÜCKIGER Y., 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, N° 4/1, pp. 599-615.

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

LERMAN R., YITZHAKI S. (1991). - "Income Stratification and Income Inequality", *Review of Income and Wealth*, Vol 37, N° 3, pp. 313-329.

MOOKHERJEE D., SHORROCKS A. F. (1982). - "A Decomposition Analysis of the Trend in UK Income Inequality", *Economic Journal*, Vol 92, pp. 886-902.

MUSSARD S. (2004). - *Décompositions multidimensionnelles du Rapport Moyen de Gini. Applications aux revenus italiens de 1989 et 2000*, Thèse de Doctorat, Université de Montpellier I.

PYATT G. (1976). - "On the Interpretation and Disaggregation of Gini Coefficients", *Economic Journal*, Vol 86, pp. 243-25.

RAO V.M. (1969). - "Two Decompositions of Concentration Ratio", *Journal of the Royal Statistical Society*, Séries A 132, pp. 418-425.

SASTRE M., TRANNOY A. (2002). - "Shapley Inequality Decomposition by Factor Components : Some Methodological Issues", dans Moyes P., Seidl C., Shorrocks A. F. (ed.), Inequalities : Theory, Experiments and Applications, *Journal of Economics*, Supplément N° 9, pp. 51-90.

SHAPLEY L. (1953). - *A Value for n-Person Games*, dans Kuhn H. W., Tucker A. W. (ed.), Contributions to the Theory of Games, Vol 2, Princeton University Press.

SHORROCKS A. F. (1980). - "The Class of Additively Decomposable Inequality Measures", *Econometrica*, Vol 48, pp. 613-625.

SHORROCKS A. F. (1982). - "Inequality Decomposition by Factor Component", *Econometrica*, Vol 50, pp. 193-211.

SHORROCKS A. F. (1999). - "Decomposition Procedures for Distributional Analysis : A Unified Framework Based on the Shapley Value", Mimeo, Université d'Essex.

SILBER J. (1989). - "Factor Components, Population Subgroups and the Computation of the Gini Index of Inequality", *Review of Economics and Statistics*, Vol 71, pp. 107-115.

SILBER J. (1993). - "Inequality Decomposition by Income Source : a Note", *Review of Economics and Statistics*, Vol 75, N° 3, pp. 545-547.

SOLTOW L. (1960). - "The Distribution of Income Related to Changes in the Distribution of Education, Age and Occupation", *Review of Economics and Statistics*, Vol 42, pp. 450-453.

8 Annexe

Preuve de la proposition 3.2.
En substituant,

$$2min\{x_{Q,i}, x_{Q,r}\} = \sum_{m=1}^{q} 2x_{Q,ir}^{*m},$$

dans l'équation suivante,

$$G = \frac{\sum_{i=1}^{n} \sum_{r=1}^{n} (\sum_{m=1}^{q} x_{Q,i}^{m} + \sum_{m=1}^{q} x_{Q,r}^{m} - 2min\{x_{Q,i}, x_{Q,r}\})}{2n^2\mu},$$

on obtient :

$$G = \frac{\sum_{i=1}^{n} \sum_{r=1}^{n} (\sum_{m=1}^{q} x_{Q,i}^{m} + \sum_{m=1}^{q} x_{Q,r}^{m} - 2\sum_{m=1}^{q} x_{Q,ir}^{*m})}{2n^2\mu}.$$

Etant donné que l'addition est associative et commutative, on peut rassembler les sources de même type entre elles, et extraire le symbole $\sum_m$:

$$G = \sum_{m=1}^{q} \frac{\sum_{i=1}^{n} \sum_{r=1}^{n} (x_{Q,i}^{m} + x_{Q,r}^{m} - 2x_{Q,ir}^{*m})}{2n^2\mu}.$$

Preuve de la proposition 4.1.
La distance économique relative et directionnelle peut être décomposée en facteurs de la manière suivante :

$$\begin{aligned} D_{jh} &= \frac{\frac{\left[\sum_{x_{j,i}>x_{h,r}} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} (x_{j,i}-x_{h,r})\right]}{n_j n_h} - \frac{\left[\sum_{x_{j,i}<x_{h,r}} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} (x_{h,r}-x_{j,i})\right]}{n_j n_h}}{\frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |x_{j,i}-x_{h,r}|}{n_j n_h}} \\ &= \sum_{m=1}^{q} \frac{\left[\sum_{x_{j,i}>x_{h,r}} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} (x_{j,i}^{m} - x_{h,r}^{m})\right] - \left[\sum_{x_{j,i}<x_{h,r}} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} (x_{h,r}^{m} - x_{j,i}^{m})\right]}{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |x_{j,i} - x_{h,r}|}. \end{aligned}$$

La décomposition en facteurs peut se réécrire :

$$D_{jh} = \sum_{m=1}^{q} D_{jh}^{m} \ , \forall \mu_j > \mu_h.$$

De manière analogue, on obtient la décomposition en sources de revenu du ratio de chevauchement :

$$\begin{aligned} P_{jh} = 1 - D_{jh} &= \sum_{m=1}^{q} \frac{2 \sum_{x_{j,i}<x_{h,r}} \sum_{i=1}^{n_j} \sum_{r=1}^{n_h} (x_{h,r}^{m} - x_{j,i}^{m})}{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |x_{j,i} - x_{h,r}|} \\ &= \sum_{m=1}^{q} P_{jh}^{m} \ , \forall \mu_j > \mu_h. \end{aligned}$$

Preuve de la proposition 4.2.
Le développement de l'expression,

$$G = \sum_{m=1}^{q} \sum_{j=1}^{k} S_j^m p_j s_j + \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} G_{jh} (w_{nb,jh}^{m} + w_{t,jh}^{m}),$$

donne :

$$\begin{aligned} G = &\sum_{j=1}^{k} \frac{n_j}{n} . \frac{n_j \mu_j}{n\mu} \sum_{m=1}^{q} \frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_j} (x_{j,i}^{m} + x_{j,r}^{m} - 2x_{j,ir}^{*m})}{2\mu_j n_j^2} \\ &+ \sum_{m=1}^{q} \sum_{j=2}^{k} \sum_{h=1}^{j-1} \frac{\sum_{i=1}^{n_j} \sum_{r=1}^{n_h} |x_{j,i} - x_{h,r}|}{n_j n_h (\mu_j + \mu_h)} \\ &\times \left(D_{jh}^{m} \left(\frac{n_j}{n} . \frac{n_h \mu_h}{n\mu} + \frac{n_h}{n} . \frac{n_j \mu_j}{n\mu} \right) + P_{jh}^{m} \left(\frac{n_j}{n} . \frac{n_h \mu_h}{n\mu} + \frac{n_h}{n} . \frac{n_j \mu_j}{n\mu} \right) \right). \end{aligned}$$

En supprimant les pondérations $\mu_j n_j^2$ et $n_j n_h(\mu_j+\mu_h)$, on a :

$$\begin{aligned}
G = &\sum_{m=1}^{q} \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\sum_{r=1}^{n_j}(x_{j,i}^m + x_{j,r}^m - 2x_{j,ir}^{*m})}{2n^2\mu} \\
&+ \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} \frac{\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|}{n^2\mu}\left(D_{jh}^m + P_{jh}^m\right).
\end{aligned}$$

On sait par ailleurs que $\sum_{m=1}^{q}(D_{jh}^m + P_{jh}^m) = 1$. En effet :

$$\begin{aligned}
\sum_{m=1}^{q}(D_{jh}^m + P_{jh}^m) &= \sum_{m=1}^{q} \frac{\frac{\left[\sum_{x_{j,i}>x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x_{j,i}^m - x_{h,r}^m)\right]}{n_j n_h} + \frac{\left[\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x_{h,r}^m - x_{j,i}^m)\right]}{n_j n_h}}{\frac{\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|}{n_j n_h}} \\
&= \frac{\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|}{\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|} \\
&= 1.
\end{aligned}$$

En insérant cette somme dans l'expression de l'indicateur de Gini, on a :

$$\begin{aligned}
G = &\sum_{m=1}^{q} \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\sum_{r=1}^{n_j}(x_{j,i}^m + x_{j,r}^m - 2x_{j,ir}^{*m})}{2n^2\mu} \\
&+ \sum_{m=1}^{q} \frac{\sum_{j=2}^{k}\sum_{h=1}^{j-1}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x_{j,i}^m + x_{h,r}^m - 2x_{jh,ir}^{*m})}{n^2\mu}.
\end{aligned}$$

La preuve est complète lorsque le second terme est multiplié et divisé par 2 :

$$\begin{aligned}
G = &\sum_{m=1}^{q} \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\sum_{r=1}^{n_j}(x_{j,i}^m + x_{j,r}^m - 2x_{j,ir}^{*m})}{2n^2\mu} \\
&+ \sum_{m=1}^{q} \frac{2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x_{j,i}^m + x_{h,r}^m - 2x_{jh,ir}^{*m})}{2n^2\mu}.
\end{aligned}$$

Exemple de la proposition 4.2.
Soit la distribution $x = \{4, 5, 8, 11\}$, de taille $n = 4$, et de moyenne $\mu = 7$. Deux groupes sont recensés ($j = 1, 2$). Les distributions et les tailles de chaque groupe sont : $x_1 = \{4, 8\}$ et $n_1 = 2$; $x_2 = \{5, 11\}$ et $n_2 = 2$. Les revenus sont constitués de deux sources ($m = 1, 2$). Les distributions deviennent donc : $x_1 = \{(2+2), (2+6)\}$ et $x_2 = \{(2+3), (5+6)\}$. L'estimation de la

multi-décomposition en deux éléments (36) donne :

$$\begin{aligned}
G =& \frac{(2+2-2(2)+2+2-2(2))+(2+5-2(2)+5+2-2(2))}{2.7.4^2} \\
&+ \frac{(2+6-2(2)+6+2-2(2))+(3+6-2(3)+6+3-2(3))}{2.7.4^2} \\
&+ 2\left(\frac{(2+2-2(2)+2+5-2(2))+(2+2-2(2)+2+5-2(2))}{2.7.4^2}\right) \\
&+ 2\left(\frac{(2+3-2(2)+2+6-2(2))+(6+3-2(3)+6+6-2(6))}{2.7.4^2}\right)
\end{aligned}$$

On a alors :

$$\begin{aligned}
G &= (0+0,0268)+(0,0357+0,0268)+2(0,0268)+2(0,0357) \\
&= 0,0893+0,125 \\
&= G_w + G_{gb}.
\end{aligned}$$

Cette technique permet donc de déterminer les contributions de chaque source à l'indice intragroupe et à l'indice brut intergroupe. L'indice de Gini global ($G = 0,2143$) est décomposé de la sorte :

- la contribution de la source 1 du groupe 1 est 0% (0) ;
- la contribution de la source 1 du groupe 2 est 12,51% (0,0268) ;
- la contribution de la source 2 du groupe 1 est 16,66% (0,0357) ;
- la contribution de la source 2 du groupe 2 est 12,51% (0,0268) ;
- la contribution de la source 1 à G_{gb} est 25% (0,0536) ;
- la contribution de la source 2 à G_{gb} est 33,32% (0,0714).

Preuve de la proposition 4.3.
Rappelons la forme de la multi-décomposition en trois éléments :

$$G = \sum_{m=1}^{q}\sum_{j=1}^{k} S_j^m p_j s_j + \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} G_{jh} w_{nb,jh}^m + \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} G_{jh} w_{t,jh}^m.$$

Pour montrer que la multi-décomposition est parfaite, autrement dit, qu'il n'existe pas de termes redondants, on développe l'expression de manière à extraire toutes les contributions "source/intragroupe" et "source/intergroupe".

L'indice de Gini net intergroupe se reformule par :

$$
\begin{aligned}
G_{nb} &= \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} G_{jh} w^m_{nb,jh} \\
&= \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} \frac{\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}|x_{j,i}-x_{h,r}|}{n_j n_h(\mu_j+\mu_h)} D^m_{jh}\left(\frac{n_j}{n}.\frac{n_h\mu_h}{n\mu}+\frac{n_h}{n}.\frac{n_j\mu_j}{n\mu}\right) \\
&= \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} \frac{\left(\sum_{x_{j,i}>x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{j,i}+x^m_{h,r}-2x^{*m}_{jh,ir})\right)}{\mu n^2} \\
&- \sum_{m=1}^{q}\sum_{j=2}^{k}\sum_{h=1}^{j-1} \frac{\left(\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{h,r}+x^m_{j,i}-2x^{*m}_{jh,ir})\right)}{\mu n^2}.
\end{aligned}
$$

Ensuite, en multipliant et en divisant l'expression par 2 ($\forall \mu_j > \mu_h$), on a :

$$
\begin{aligned}
G_{nb} = \sum_{m=1}^{q} & \frac{2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}>x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{j,i}-x^m_{h,r})\right)}{2\mu n^2} \\
& - \sum_{m=1}^{q} \frac{2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{h,r}-x^m_{j,i})\right)}{2\mu n^2}.
\end{aligned}
$$

De la même manière, il est possible de développer l'intensité de transvariation intergroupe ($\mu_j > \mu_h$) :

$$
\begin{aligned}
G_t &= \sum_{m=1}^{q} \frac{4\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{j,i}+x^m_{h,r}-2x^{*m}_{jh,ir})\right)}{2\mu n^2} \\
&= \sum_{m=1}^{q} \frac{4\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{h,r}-x^m_{j,i})\right)}{2\mu n^2}.
\end{aligned}
$$

En conséquence, la multi-décomposition en trois éléments s'exprime par :

$$
\begin{aligned}
G = \sum_{m=1}^{q} & \frac{\sum_{j=1}^{k}\sum_{i=1}^{n_j}\sum_{r=1}^{n_j}(x^m_{j,i}+x^m_{j,r}-2x^{*m}_{j,ir})}{2n^2\mu} \\
& + \sum_{m=1}^{q} \frac{2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}>x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{j,i}-x^m_{h,r})\right)}{2\mu n^2} \\
& - \sum_{m=1}^{q} \frac{2\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{h,r}-x^m_{j,i})\right)}{2\mu n^2} \\
& + \sum_{m=1}^{q} \frac{4\sum_{j=2}^{k}\sum_{h=1}^{j-1}\left(\sum_{x_{j,i}<x_{h,r}}\sum_{i=1}^{n_j}\sum_{r=1}^{n_h}(x^m_{h,r}-x^m_{j,i})\right)}{2\mu n^2}.
\end{aligned}
$$

En définitive, les trois indices G_w, G_{nb} et G_t, sont décomposés à la fois en sous-groupes et en sources de revenu. Excepté le dénominateur, les expressions ne comportent pas de termes multiplicatifs. La multi-décomposition en trois indices est donc parfaite. $\square$