E C O N O M I C S   B U L L E T I N

A note on the decomposition of the coefficient of variation squared: comparing entropy and Dagum's methods

Celestin Chameni Nembua
University of YaoundéII, Cameroon

Abstract

The aim of this paper is to propose a new decomposition of the coefficient of variation squared. The approach is similar to the Dagum's method when decomposing the Gini index. We compare the new method to the former entropy decomposition of this coefficient. An empirical study is elaborated. This concerns a subgroup decomposition of the non food expenditure cameronian households inequality.

Citation: Chameni Nembua, Celestin, (2006) "A note on the decomposition of the coefficient of variation squared: comparing entropy and Dagum's methods." *Economics Bulletin,* Vol. 4, No. 8 pp. 1−8
Submitted: November 10, 2005. **Accepted:** February 2, 2006.
URL: http://www.economicsbulletin.com/2006/volume4/EB−05D60003A.pdf

1- Introduction

This paper proposes a new subgroup decomposition of the coefficient of variation squared. It is equal to twice the Hirschman-Herfindahl index which belongs to the class of generalized entropy indices obtained with the parameter $\beta = 1$ or C=2. The approach here is similar to the one used by Dagum (1997a; 1997b) when decomposing the Gini index and it is based on the interpersonal expression of the coefficient of variation squared; so, we shall consider this as Dagum's method. The comparison between this method and the one issued from the generalised entropy indices decomposition (Cowell, 1980) is operated. The theoretical results are applied to analyse the decomposition of household non food expenditure inequality in Cameroon. The ECAMII-2001(a household survey carried out by Cameroon's National Institute of Statistics) data base is used and the empirical study is carried out using the free program provided on the web site http//www.Lameta.univ-montp1.fr/online/gini.html for the entropy method in one hand, and using a computer program we have designed for applications of our method in another hand.

The remaining text is subdivided into four sections in addition to the present introduction. In Section 2, we establish the interpersonal expression of the coefficient of variation squared (section 2.1) and we present his Dagum subgroups decomposition. Section 3 is devoted to a rapid overview of the entropy decomposition of the coefficient of variation squared. As to section 4, the preceding results are applied to decompose the inequalities in non food expenditure of Cameroonian households. The results obtained here lead to the comparison of the two decomposition methods and we motivate our preference to the Dagum decomposition by giving important reasons for that. Finally the paper is concluded in section 5.

2- The Dagum subgroup decomposition of coefficient of variation squared

Lets consider a population P with n income units x_1, x_2, $x_3 \ldots x_i \ldots x_n$ where CV^2, Var and μ are respectively the square of coefficient of variation, the variance and the mean on P. We assume that P is partitioned into K subpopulations $P_1, P_2, P_3, \ldots, P_h, \ldots, P_K$ with respectively $n_1, n_2, n_3, \ldots, n_h, \ldots n_K$, members; $x_{h1}, x_{h2}, ..., x_{hn_k}$ CV_h^2 and μ_h are respectively the income units, the square of coefficient of variation and the mean on P_h.

2.1 The interpersonal expression of the coefficient of variation squared

By definition,

$$CV^2 = \frac{Var}{\mu^2} = \frac{\frac{1}{n}\sum_{i=1}^{n} x_i^2 - \mu^2}{\mu^2} = \frac{1}{2n\mu^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{2n\mu^2}\sum_{j=1}^{n} x_j^2 - \frac{1}{\mu^2}\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)\left(\frac{1}{n}\sum_{j=1}^{n} x_j\right) \quad (1)$$

$$= \frac{1}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i^2 + \frac{1}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_j^2 - \frac{2}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j \quad (2)$$

$$= \frac{1}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_i^2 + x_j^2 - 2x_i x_j\right) = \frac{1}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(x_i - x_j\right)^2 \quad (3)$$

$$= \frac{1}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|^2 \qquad (4)$$

2.2 The expression of the decomposition

We set, for h=1,2,…K : $f_h = \frac{n_h}{n}$ and $s_h = \frac{n_h}{n}\left(\frac{\mu_h}{\mu}\right)^2$.

a) Decomposition into two components

Using (4), we have:

$$CV^2 = \frac{1}{2n^2\mu^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|^2 = \frac{1}{2n^2\mu^2}\sum_{h=1}^{K}\sum_{k=1}^{K}\sum_{i=1}^{n_h}\sum_{j=1}^{n_k}\left|x_{hi} - x_{kj}\right|^2 \qquad (5)$$

The mean of the differences of order 2 between the subpopulations P_h and P_k is defined by:

$$\Delta_{hk}(2) = E\left|X_h - X_k\right|^2 = \frac{1}{n_h n_k}\sum_{i=1}^{n_h}\sum_{j=1}^{n_k}\left|x_{hi} - x_{kj}\right|^2 \quad \text{and it follows that :} \qquad (6)$$

$$CV^2 = \frac{1}{2n^2\mu^\alpha}\sum_{h=1}^{K}\sum_{k=1}^{K} n_h n_k \Delta_{hk}(2) = \frac{1}{2n^2\mu^\alpha}\sum_{h=1}^{K}\sum_{k=1}^{K}\frac{n_h n_k \Delta_{hk}(2)}{(\mu_h^\alpha + \mu_k^\alpha)}(\mu_h^\alpha + \mu_k^\alpha) \qquad (7)$$

Let's introduce the index of inequality between the subpopulations P_h and P_k :

$$G_{hk} = \frac{\Delta_{hk}(2)}{\mu_h^2 + \mu_k^2} \qquad (8)$$

We have in particular: $G_{hh} = \frac{\Delta_{hh}(2)}{2\mu_h^2} = \frac{1}{2n_h^2\mu_h^2}\sum_{i=1}^{n_h}\sum_{j=1}^{n_h}\left|x_{hi} - x_{hj}\right|^2 = CV_h^2$ (9)

$$CV^2 = \frac{1}{2n^2\mu^2}\sum_{h=1}^{K}\sum_{k=1}^{K} G_{hk} n_h n_k (\mu_h^2 + \mu_k^2) \qquad (10)$$

$$= \sum_{h=1}^{K}\left(\frac{n_h}{n}\right)^2\left(\frac{\mu_h}{\mu}\right)^2 CV_h^2 + \frac{1}{2n^2\mu^2}\sum_{1=h\neq}^{K}\sum_{k=1}^{K} G_{hk} n_h n_k (\mu_h^2 + \mu_k^2) \qquad (11)$$

$$= \sum_{h=1}^{K} f_h s_h CV_h^2 + \sum_{h=2}^{K}\sum_{k=1}^{h-1}\left(f_k s_h + f_h s_k\right)G_{hk} = CV_W^2 + CV_B^2 \qquad (12)$$

Where CV_W^2 is the within group component and CV_B^2 is the gross between group component.

b) Decomposition into three components

The gross economic wealth noted d_{hk}, is defined between two subpopulations P_h and P_k such that $\mu_h > \mu_k$. d_{hk} is the mean of the difference $(x_{hi} - x_{kj})$ for each income x_{hi} of a member in P_h greater than income x_{kj} of a member in P_k :

$$d_{hk} = \int_0^{+\infty} dF_h(y)\int_0^{y}\left|y - x\right| \, dF_k(x) = \frac{1}{n_h n_k}\sum_{\substack{i=1 \\ x_{hi} > x_{kj}}}^{n_h}\sum_{j=1}^{n_k}\left|x_{hi} - x_{kj}\right| \le \Delta_{hk}(1) \qquad (13)$$

Where $\Delta_{hk}(1) = E\left|X_h - X_k\right| = \frac{1}{n_h n_k}\sum_{i=1}^{n_h}\sum_{j=1}^{n_k}\left|x_{hi} - x_{kj}\right|$ (14)

Following Dagum, we set $p_{hk} = \Delta_{hk}(1) - d_{hk}$ if $\mu_h > \mu_k$.
p_{hk} corresponds to the transvariational[1] component.
The net economic wealth between two subpopulations P_h and P_k such that $\mu_h > \mu_k$, is defined by the difference $d_{hk} - p_{hk} > 0$
and the relative economic difference between two such subpopulations is given by:

$$D_{hk} = \frac{d_{hk} - p_{hk}}{\Delta_{hk}(1)} \tag{15}$$

It is clear that, $\Delta_{hk}(1), \Delta_{hk}(2), G_{hk}$ and D_{hk} define symmetric matrices and it is well known (Dagum 1997b) that D_{hk} is a distance on the set of distributions X_h which is null if and only if there is perfect overlapping between distributions and $0 \le D_{hk} \le 1$.

According to (12) and inserting (15) in CV_B^2,

$$CV^2 = \sum_{h=1}^{K} f_h s_h CV_h^2 + \sum_{h=2}^{K}\sum_{k=1}^{h-1} \left(f_k s_h + f_h s_k\right)\left(D_{hk} + 1 - D_{hk}\right)G_{hk} \tag{16}$$

$$CV^2 = \sum_{h=1}^{K} f_h s_h CV_h^2 + \sum_{h=2}^{K}\sum_{k=1}^{h-1} D_{hk}\left(f_k s_h + f_h s_k\right)G_{hk} + \sum_{h=2}^{K}\sum_{k=1}^{h-1}\left(1 - D_{hk}\right)\left(f_k s_h + f_h s_k\right)G_{hk} \tag{17}$$

$$CV^2 = CV_W^2 + CV_{NB}^2 + CV_{TB}^2 \text{ where} \tag{18}$$

$CV_W^2 = \sum_{h=1}^{K} f_h s_h CV_h^2$ is the contribution of the within subgroup inequality to the overall CV^2.

$CV_{NB}^2 = \sum_{h=2}^{K}\sum_{k=1}^{h-1} D_{hk}\left(f_k s_h + f_h s_k\right)G_{hk}$ is the net contribution of the between subgroups inequality to the overall CV^2.

$CV_{BT}^2 = \sum_{h=2}^{K}\sum_{k=1}^{h-1}\left(1 - D_{hk}\right)\left(f_k s_h + f_h s_k\right)G_{hk}$ measures the contribution to the overall CV^2, of the inequality coming from the transvariation between the subgroup pairs.

c) Contribution of each group to the gross between group component

The classical decomposition of the variance implies:

$$CV^2 = \sum_{h=1}^{K} s_h CV_h^2 + \sum_{h=1}^{K} \frac{n_h}{n}\left(\frac{\mu_h}{\mu} - 1\right)^2 \tag{19}$$

By equating (12) and (19) we obtain:

$$CV_B^2 = \sum_{h=1}^{K} s_h\left(1 - f_h\right)CV_h^2 + \sum_{h=1}^{K} \frac{n_h}{n}\left(1 - \frac{\mu_h}{\mu}\right)^2. \tag{20}$$

What permit to gauge the contribution of P_h to CV_B^2 :

[1] transvariation come from ' tranvariazione' which is the term used by C.Gini in 1916.

$$CV_B^2(P_h)=\frac{n_h}{n}\left(1-\frac{\mu_h}{\mu}\right)^2+s_h(1-f_h)CV_h^2 \qquad (21)$$

Formula (20) reveals in particular that the gross between groups component, and consequently the total CV^2 index, are increasing functions of the within groups indices; which means that this decomposition satisfies the Shorrocks (1994) subgroup consistency property.

3- The entropy subgroup decomposition of coefficient of variation squared

It is obtained as particular case of generalized entropy ratio which is express by:

$$I_\beta=\frac{1}{\beta(\beta+1)n}\sum_{h=1}^{K}\sum_{i=1}^{n_h}\frac{x_{hi}}{\mu}\left[\frac{x_{hi}}{\mu}-1\right]^\beta \quad \beta \text{ real.} \qquad (22)$$

The generalised entropy can be decomposed (Cowell 1980) into two components, the within group component and the between group component such as :

$$I_\beta=\sum_{h=1}^{K}\frac{n_h\mu_h}{n\mu}\left(\frac{\mu_h}{\mu}\right)^\beta I_{\beta h}+\frac{1}{\beta(\beta+1)}\sum_{h=1}^{K}\frac{n_h}{n}\frac{\mu_h}{\mu}\left[\left(\frac{\mu_h}{\mu}\right)^\beta-1\right] \qquad (23)$$

$$=I_{\beta W}+I_{\beta B}$$

The coefficient of variation squared equal $2I_\beta$ with $\beta=2$. We deduce from (23) that:

$$CV^2=\sum_{h=1}^{K}\frac{n_h}{n}\left(\frac{\mu_h}{\mu}\right)^2 CV_h^2+\sum_{h=1}^{K}\frac{n_h\mu_h}{n\mu}\left(\frac{\mu_h}{\mu}-1\right) \qquad (24)$$

$$CV^2=\sum_{h=1}^{K}s_hCV_h^2+\sum_{h=1}^{K}s_h\left(1-\frac{\mu}{\mu_h}\right)=I_W+I_B \qquad (25)$$

4- Application

We study non food expenditure household inequality in Cameroon. The ECAMII-2001 database is used and it includes 10992 households subdivided according to their residential areas: Urban (group 1, n_1=4975) Semi-urban (group 2, n_2=2137) and Rural (group 3, n_3=3880). The two decomposition methods introduced above allow one to know if the inequalities are generated by the expenditure gaps within the three residential areas or if inequalities come from the expenditure gaps between the three groups. The computation of the entropy decomposition is provided by the free program on the web site http//www.Lameta.univ-montp1.fr/online/gini.html, while the results on the Dagum decomposition are obtained by a program package that we have designed for the circumstances.

Table1 illustrates these results in giving the contribution of each component of the two decompositions to the global inequality. The entropy method shows that the contribution within the subpopulation represent 93.38% and the differences between the three areas represent only 6.62% of the global inequality; whereas the Dagum method grants a little difference between the within groups element (40.69%) and the between group element (59.31%) with , in contrary, the predominance of the between group component.

Only the Dagum method can provide the intensity of net between group component and the intensity of transvariation which is the part of the between groups disparities issued from the overlap between the three distributions. The results obtained here shows that, a considerable part (59.03%) of the global inequality comes from overlapping between the distributions of the three groups and the transvariational between group component represents 99.52% (that is almost the totality) of the gross between group component. In the other hand, the Dagum method permits to gauge the contribution of each pair of groups to the net and the transvariational between groups component.

Table4 indicates that the pair (group 1; group3) , (group1; group2) and (group2; group3) explain respectively 61.83% , 33.42% and 4.75% of the transvariational component; which here indicates a significant overlapping of the amount of non food expenditure distributions in group3 and group1, as well as relative likeness between a considerable fraction of these distributions. The pairs (group 1; group3) and (group2; group3) contribute nothing to the net between group component, what confirms a perfect overlapping of their corresponding distributions.

Table2 reveals that, using the entropy method, the group3 (Rural area) contributes negatively (-0.86%) to the global inequality while all the contributions of the three groups are positive in the Dagum method. This phenomenon of negative contribution when using entropy method, are reinforced in table3 where the contribution of group2 (-4.63%) and group3 (-37.02%) to the between groups component are negative again. This implies an uncomfortable situation and shows that, the entropy decomposition of the coefficient of variation squared (or the Hirschman-Herfindahl index) should be used with a lot of precautions for, it between groups component have negative terms which may lead to nonsense interpretations.

Although the two decompositions method unveil together that the group1 (Urban area) plays a central role in generating inequality (the contribution of this group to the overall inequality (cf. table2) is higher in the two methods and equal respectively to 95.68%, 89.37% in the entropy and Dagum method) , the difference of the results between these two methods are important. So it seems indispensable to direct the choice of the users of the decomposition of the coefficient of variation squared (or the Hirschman-Herfindahl index).

We incite to privilege the Dagum decomposition for many reasons:

1) The Dagum method is based on an interpersonal expression of the coefficient of variation squared and thus integrate the criteria of the interpersonal utility comparison like the Gini index (for more details, see Dagum, 1980 or Mussard and alii., 2003).
2) The Dagum method is built on a better between group specification; not only its between groups component is an effective inequalities between the subgroups, but also its transvariational component constitutes an enrichment which permit to gauge inequalities coming from overlapping between the income distributions of various subgroups.

   While the entropy decomposition of the coefficient of variation squared has a between groups component which is a simple difference in mean. Moreover, this between groups contributions are obtained like a residual ($I_B = CV^2 - I_W$) that generates negative terms which may lead to nonsense interpretation as seen above and as Dagum has noticed in 1997.
3) The Dagum decomposition of the coefficient of variation squared satisfies the Shorrocks (1994) subgroup consistency property. This means that changes (increase or decrease) in one of the within group index implies changes (in the same direction) in the overall inequality index.

5- Conclusion

We have presented another way to decompose the coefficient of variation squared similar to the one used by Dagum when decomposing the Gini index. The comparison of the new method to the former entropy decomposition of the coefficient of variation squared has been done through an empirical study concerning the decomposition of the inequality in the non food expenditure Cameroonian households. The results obtained lead directly to the preference of the new method of decomposition. Since our choice has been motivated, we incite the users to give privilege to the new method.

REFERENCES

Cowell F.A. (1980a) Generalized entropy and the Measurement of distributional Change, *European Economics Review*, vol 13: 147-159.

Cowell F.A. (1980b) On the Structure of additive Inequality Measures, *Review of Economics Studies*, vol 47, 521-531.

Dagum C. (1997a) A new Approach to the decomposition of the Gini Income Inequality Ratio, *Empirical Economics*, 22(4),p515-531.

Dagum C. (1988) Inequality Measures between Income Distribution with Application, *Econometrica* 1791-1803

Dagum C. (1997b) Decomposition and Interpretation of Gini and the Generalized Entropy Inequality Measures, *Proceedings of the American Statistical Association, Business and Economic Statistics Section, 157 th Meeting*, p. 200-205.

Gini C.(1916), II concetto di transvariazione e le sue prime applicationzioni, Giornale

Musard S., Seyte F., Terraza M. (2002b), *Programme pour la decomposition de l'indice de Gini de C.Dagum*, site internet http://www.lameta.univ-montp1.fr/online/gini.html

Musard S., Seyte F., Terraza M. (2003), Decomposition of Gini and the Generalized entropy Inequality Measures, *Economics Bulletin*, vol.4 n°7,1-6.

Shorrocks A.F. (1984), Inequality Decomposition by factors Components and by Population Subgroup, *Econometrica*, vol.53,1369-1386.

Table 1: Contribution of each element of the two decompositions to the overall inequality

Decomposition Method ↓	Within group component		Between group component		Net between group component		Transvariational component	
	Absolute value	% value	Absolute value	% value	Absolute value	% value	Absolute value	% value
Entropy	3.252574	93.38	0.230556	6.62	NA*	NA*	NA*	NA*
Dagum	1.41727	40.69	2.06586	59.31	0.00983	0.28	2.05603	59.03

*NA: Non available for this method

Table 2: Contribution of each group to the global inequality

Groups ↓	Entropy Method		Dagum Method	
	Absolute value	**% value**	**Absolute value**	**% value**
Group 1	3,33268	95,68	3,11286	89,37
Group 2	0,180348	5,18	0,19169	5,50
Group 3	-0,0299	-0,86	0,17858	5,13
Global	3,48313	100,00	3,48313	100,00

Table 3: Contribution of each group to the within and between group component

Groups ↓	Entropy Method				Dagum Method			
	Within group component		Between group component		Within group component		Gross Between group component	
	Absolute value	% value	Absolute value	% value	Absolute value	% value	Absolute value	% value
Group 1	3,00609	92,42	0,3265	141,65	1,36057	96,00	1,75229	84,82
Group 2	0,19103	5,87	-0,0106	-4,63	0,03714	2,62	0,15455	7,48
Group 3	0,05544	1,70	-0,0853	-37,02	0,01957	1,38	0,15901	7,70
Global	3,25257	100	0,23055	100	1,41728	100	2,06585	100

Table4: Pair wise groups contribution to the net and transvariational between groups component

	Net between groups component			Transvariational between groups component		
	Group1	**Group2**	**Group3**	**Group1**	**Group2**	**Group3**
Group1	-			-		
Group2	100	-		33,42	-	
Group3	0	0	-	61,83	4,75	-
Global	100			100		

NB: This table is not available for the entropy method