Factor decomposition of spatial income inequality: a revision

Juan Antonio Duro[*]

Universitat Autònoma de Barcelona

Abstract

Duro and Esteban (1998) proposed an additive decomposition of Theil population-weighted index by four income multiplicative factors (in spatial contexts). This note makes some additional methodological points: first, it argues that interaction effects are taken into account in the factoral indexes although only in a fairly restrictive way. As a consequence, we suggest to rewrite the decomposition formula as a sum of strict Theil indexes plus the interactive terms; second, it might be instructive to aggregate some of the initial factors; third, this decomposition can be immediately extended to the between- and within-group components.

Key words: Theil index; inequality decomposition.

JEL Classification: C10; F00

1. Introduction

---

[*] Juan Antonio Duro, Departament d'Economia Aplicada, Facultat de Ciències Econòmiques i Empresarials, Universitat Autònoma de Barcelona, Bellaterra, 08193, Barcelona, Spain.

Phone: 34-93-5806612. E-mail address: juanantonio.duro@uab.es

The use of Theil population-weighted index has been largely appreciated by the literature[1]. This is because its appealing properties and, specially, its capacity for being decomposed by parts[2]. Recently, Duro and Esteban (1998) have suggested that per capita income inequality (when it is measured through this index) can be additively decomposed by multiplicative income factors. In this note we revise this methodology and point out that: (i) interactions among factors are included although in a fairly unsatisfactory way; (ii) it seems useful to aggregate employment and activity rates in a single factor and, in some contexts, to remove the demographic factor; (iii) this decomposition can be readily applied to between- and within-group components.

2. Basic methodology and correlation effects

Duro and Esteban (1998) point out that if we (i) use Theil population-weighted index as the summary inequality measure, T(y); (ii) decompose per capita incomes in multiplicative factors like productivity (x), employment rate (e), activity rate (a) and working-age ratio (w); (iii) define four fictitious incomes by letting factor values to differ from the average by only one at time we obtain

$$T(y) = I\left(y^{x}\right) + I\left(y^{e}\right) + I\left(y^{a}\right) + I\left(y^{w}\right) \qquad (1)$$

---

[1] See Bourguignon (1979) and Ram (1992).

[2] Conventionally, this inequality measure has been decomposed by income sources (Theil (1979)), by subgroups of populations (Shorrocks (1980)) and by population and income changes (Theil and Sorooshian (1979)).

where y is the per capita income; y^r is the per capita income when only the factor r differs from the average (r = x,e,a and w) and $I(y^r)$ are inequality indices for each ficticious income, where the average income is used as a reference rather than the corresponding mean of the relevant ficticious income[3].

Also, it is readily verified that[4]

$$I(y^r) = T(y^r) + \log\left(\frac{\mu}{\mu^r}\right) \qquad \text{for r=x,e,a} \quad (2)$$

where $T(y^r)$ is the proper Theil index applied to ficticious income y^r and μ^r is the real mean of the ficticious income y^r.

What is the meaning of the second element in (2)? It can be demonstrated that this term is influenced by factoral covariances. In particular, we might express each conventional factoral index like[5]:

[3] Goerlich (2001) demonstrates that this decomposition also holds for Theil income-weighted index. This is true but we believe that the population-weighted version seems a more preferable measure. First, if our objective is to make a comparison of well being of populations inequality measures based on population-weights would be preferable. Second, Theil population-weighted index is more sensitive to income changes lower down the scale, which can be interesting for some researchers. Finally, this measure is strictly decomposable by groups whereas Theil income-weighted is only weakly decomposable (see Shorrocks (1980)).

[4] For r = w we have $I(y^w) = T(y^w)$.

[5] Details can be consulted in the appendix.

$$I\left(y^{x}\right)=T\left(y^{x}\right)+\log\left(1+\frac{\sigma_{x,eaw}}{\mu^{x}}\right)$$

$$I\left(y^{e}\right)=T\left(y^{e}\right)+\log\left(1+\frac{x\sigma_{e,aw}}{\mu^{e}}\right) \qquad (3)$$

$$I\left(y^{a}\right)=T\left(y^{a}\right)+\log\left(1+\frac{xe\sigma_{a,w}}{\mu^{a}}\right)$$

where $\sigma_{x,eaw}$ is the (weighted) covariance between productivity and occupation rate in a wide sense (employment/working-age population); $\sigma_{e,aw}$ denotes the (weighted) covariance between occupation rate and activity rate in a wide sense (active/total population) and $\sigma_{a,w}$ is the (weighted) covariance between activity and working-age ratio.

Note that interactive effects are not homogeneously included into the different indexes. Thus, for the case of employment rates its correlation with productivity is omitted; in case of activity rates only its interaction with demographical structures is considered; and for the workink-age ratio any correlation is taken into account. Surely, we can find a variety of reasons to believe that these exclusions should be considered.

In these circumstances, we propose a less restrictive solution which would consist in rewriting aggregate inequality as a sum of proper Theils by factor and, separately, the interaction components[6]. That is

---

[6] Note that this decomposition holds for every variable which can be broken down in multiplicative factors.

$$T(y)=T(y^{x})+T(y^{e})+T(y^{a})+T(y^{w})+\log\left(1+\frac{\sigma_{x,eaw}}{\mu^{x}}\right)+\log\left(1+\frac{x\sigma_{e,aw}}{\mu^{e}}\right)+\log\left(1+\frac{xe\sigma_{a,w}}{\mu^{a}}\right) \quad (4)^{7}$$

Observe that under the condition of independence of factors the aggregate inequality would be elegantly expressed as a sum of individual factor indexes.

3. Aggregating factors

An additional aspect is related with the possibility of aggregating factors. Immediate candidates would be the employment and activity rates. First, we have reasons to believe that employment and activity are correlated and its joint consideration can be perceived as a global indicator about the role made by labour markets. Second, occupation rates would be less informative about the position of labour markets due to the non-inclusion of the non-employed, that is the potential labour force which leave markets. Finally, this aggregation can help to solve the possible variation in unemployed definitions which hamper international comparisons. Then, the proposed decomposition formula would take the following form:

$$T(y)=T(y^{x})+T(y^{ea})+T(y^{w})+\log\left(1+\frac{\sigma_{x,eaw}}{\mu^{x}}\right)+\log\left(1+\frac{x\sigma_{ea,w}}{\mu^{ea}}\right) \quad (5)$$

[7] Note that, in case of small inequality values, this formula would be approximately equal to:

$$T(y)\approx T(y^{x})+T(y^{e})+T(y^{a})+T(y^{w})+\frac{\sigma_{x,eaw}}{\mu^{x}}+\frac{x\sigma_{e,aw}}{\mu^{e}}+\frac{xe\sigma_{a,w}}{\mu^{a}}$$

Also, this aggregation might be extended, in some cases, to the working-age ratio, because its limited explanative role (i.e., OECD-countries) and, therefore, we might consider a simple two-factor decomposition:

$$T(y) = T(y^x) + T(y^{eaw}) + \log\left(1 + \frac{\sigma_{x,eaw}}{\mu^x}\right) \qquad (6)$$

4. Extensions

A well-known inequality decomposition exercise breaks down total inequality in two synthetic components: between- and within-group terms. Its algebraic expression is:

$$T(y) = \sum_{g=1}^{G} p_g T(y)_g + \sum_{g=1}^{G} p_g * \ln\left(\frac{\mu}{y_g}\right) \qquad (7)$$

where p_g is the relative population of group g, T_g denotes the internal inequality present in group g and, finally, y_g represents the mean income in group g.

Observe that the first term, the within-group component, is a weighted average of internal Theil indexes, which can be immediately decomposed by our multiplicative factors. On the other hand, the between-group component is purely a Theil population-weighted index and also the implementation of our decomposition is trivial.

ACKNOWLEDGEMENTS

Research partially supported by Project SGR2001-160 from Direcció General de Recerca, Departament d'Universitats, Recerca i Societat de la Informació, Generalitat de Catalunya. The author wishes to acknowledge helpful discussions with Joan Esteban. Usual disclaimer applies.

APPENDIX

a) Four-factor decomposition

a1) $\sigma_{x,eaw} = \sum_{i=1}^{n} p_i\left[(x_i - x)(eaw_i - eaw)\right] = \sum_{i=1}^{n} p_i\left[x_i eaw_i - x_i eaw - xeaw_i + xeaw\right] =$

$$= \sum_{i=1}^{n} p_i x_i eaw_i - \sum_{i=1}^{n} p_i x_i eaw - \sum_{i=1}^{n} p_i xeaw_i + \sum_{i=1}^{n} p_i xeaw = \mu - \mu^{x} - \mu^{eaw} + \mu = \mu - \mu^{x}$$

if we divide all the terms by μ^{x} and manipulate we get:

$$\frac{\mu}{\mu^{x}} = \frac{\sigma_{x,eaw}}{\mu^{x}} + 1$$

and if we apply logarithms

$$\ln\left(\frac{\mu}{\mu^{x}}\right) = \ln\left(\frac{\sigma_{x,eaw}}{\mu^{x}} + 1\right)$$

a2) $\sigma_{e,aw} = \sum_{i=1}^{n} p_i\left[(e_i - e)(aw_i - aw)\right] = \sum_{i=1}^{n} p_i\left[e_i aw_i - e_i aw - eaw_i + eaw\right] =$

$$= \sum_{i=1}^{n} p_i e_i aw_i - \sum_{i=1}^{n} p_i e_i aw - \sum_{i=1}^{n} p_i eaw_i + \sum_{i=1}^{n} p_i eaw = eaw - \sum_{i=1}^{n} p_i e_i aw - eaw + eaw = eaw - aw\sum_{i=1}^{n} p_i e_i$$

if we multiply all by x we will have

$$x\sigma_{e,aw} = xeaw - xaw\sum_{i=1}^{n} p_i e_i = \mu - \mu^{e}$$

and we divide all the terms by μ^{e} and manipulate them

$$\frac{\mu}{\mu^e} = \frac{x\sigma_{e,aw}}{\mu^e} + 1$$

Finally, we apply logarithms and we will get

$$\ln\left(\frac{\mu}{\mu^e}\right) = \ln\left(\frac{x\sigma_{e,aw}}{\mu^e} + 1\right)$$

a3) $\sigma_{a,w} = \sum_{i=1}^{n} p_i\left[(a_i - a)(w_i - w)\right] = \sum_{i=1}^{n} p_i\left[a_i w_i - a_i w - aw_i + aw\right] =$

$$= \sum_{i=1}^{n} p_i a_i w_i - \sum_{i=1}^{n} p_i a_i w - \sum_{i=1}^{n} p_i aw_i + \sum_{i=1}^{n} p_i aw = aw - \sum_{i=1}^{n} p_i a_i w - aw + aw = aw - w\sum_{i=1}^{n} p_i a_i$$

if we multiply all by xe we will get

$$xe\sigma_{a,w} = xeaw - xew\sum_{i=1}^{n} p_i a_i = \mu - \mu^a$$

and divide all the terms by μ^a and manipulate them

$$\frac{\mu}{\mu^a} = \frac{xe\sigma_{a,w}}{\mu^a} + 1$$

Finally, we apply logarithms and we will get

$$\ln\left(\frac{\mu}{\mu^a}\right) = \ln\left(\frac{xe\sigma_{a,w}}{\mu^a} + 1\right)$$

b) Three-factor decomposition

In this case we will have two adjustment terms. The first one would be equivalent to a1) and the second one would be extracted through the following derivation:

b1) $\sigma_{ea,w} = \sum_{i=1}^{n} p_i\left[(ea_i - ea)(w_i - w)\right] = \sum_{i=1}^{n} p_i\left[ea_i w_i - ea_i w - eaw_i + eaw\right] =$

$$= \sum_{i=1}^{n} p_i ea_i w_i - \sum_{i=1}^{n} p_i ea_i w - \sum_{i=1}^{n} p_i eaw_i + \sum_{i=1}^{n} p_i eaw = eaw - \sum_{i=1}^{n} p_i ea_i w - eaw + eaw = eaw - w\sum_{i=1}^{n} p_i ea_i$$

if we multiply all by x we will get

$$x\sigma_{ea,w} = xeaw - xw\sum_{i=1}^{n} p_i ea_i = \mu - \mu^{ea}$$

and divide all the terms by μ^{ea} and manipulate them

$$\frac{\mu}{\mu^{ea}} = \frac{x\sigma_{ea,w}}{\mu^{ea}} + 1$$

Finally, we apply logarithms and we will get

$$\ln\left(\frac{\mu}{\mu^{ea}}\right) = \ln\left(\frac{x\sigma_{ea,w}}{\mu^{ea}} + 1\right)$$

c) Two-Factor decomposition

We only would have one adjustment term, which would coicide with a1).

REFERENCES

Bourguignon, F., 1979, Decomposable income inequality measures, Econometrica, 47, 901-920.

Duro, J.A. and J. Esteban, 1998, Factor decomposition of cross-country income inequality, 1960-1990, Economics Letters, 60, 269-275.

Goerlich, F., 2001, On factor decomposition of cross-country income inequality: some extensions and qualification, Economics Letters, 70, 303-309.

Ram, R., 1992, International inequalities in human development and real income, Economics Letters, 38, 351-354.

Shorrocks, A., 1980, The Class of Additively Decomposable Inequality Measures, Econometrica, 48, 613-625.

Theil, H., 1979, The measurement of inequality by components of income, Economics Letters, 2, 197-199.

Theil, H. and Sorooshian, 1979, Components of the change in regional inequality, Economics Letters, 4, 191-193.