CENTRAL EUROPEAN REVIEW
OF ECONOMICS & FINANCE
Vol. 2, No. 1 (2012) pp. 109-116

Bruno Bracalente*, **Cristiano Perugini***

FACTOR DECOMPOSITION OF CROSS-COUNTRY INCOME INEQUALITY WITH INTERACTION EFFECTS

Abstract

In this paper we describe a decomposition of the Theil measures of per capita income inequality which accounts for interaction effects between its multiplicative factors. Our theoretical evidence, supported by an empirical application referring to EU-27 countries in the year 2010, suggest that neglecting these effects may strongly bias the relative importance of some factors, with consequent misleading policy implications.

JEL classification code: C10; D63; O10.

Keywords: Inequality, Decomposition, Interaction Effects.

Introduction

Per capita income may be expressed as the product of many factors. The basic decomposition is into the two classical determinants of the *wealth of nations*: the share of population employed, and labour productivity. In turn, each of these can be multiplicatively decomposed into more specific factors.

Irrespective of their number, an interesting question concerns how to measure their contribution to inequality in cross-country (or region) per capita income. For this aim, Duro and Esteban (1998) proposed an additive decomposition based on the *second* Theil inequality measure. This approach was criticised by Cheng and Li (2006) who developed a method, noted by the same Duro and Esteban, in which a residual term emerges and is interpreted as an interaction effect between the components.

* Department of Economics, Finance and Statistics, University of Perugia, Via Pascoli, 06123 Perugia, Italy.

In this paper, moving from Duro and Padilla (2006), we describe a possible decomposition of the *first* and *second* Theil inequality measures (Theil, 1967; Bourguignon, 1979) with interaction components and explain the reasons why this method should be considered preferable to the above-mentioned ones. An application referring to EU-27 countries for the year 2010 is provided to corroborate the theoretical evidence.

1. Decomposition of per capita GDP inequality with interactions

Let X_i, E_i, P_i $(i=1, \ldots, N)$ be country i GDP, employment and population, respectively, and X, E, P the corresponding total values (i.e., $X = \sum_i X_i$, etc.). Let x_i be country per capita GDP, with weighted mean $\mu(x) = X/P$, and p_i the country share of population on total ($p_i = P_i/P$).

Both x_i and $\mu(x)$ may be expressed as the product of two factors:

$$x_i = \frac{X_i}{P_i} = \frac{X_i}{E_i} \cdot \frac{E_i}{P_i} = y_i \cdot e_i$$

$$\mu(x) = \frac{X}{P} = \frac{X}{E} \cdot \frac{E}{P} = \mu(y) \cdot \mu(e)$$

where y_i and e_i are country labour productivity and employment rate on total population, respectively, and $\mu(y)$ and $\mu(e)$ their weighted means.

The corresponding population-weighted Theil inequality index (so-called second measure) may be decomposed into two additive components, as follows:

$$T(x,p) = \sum_i p_i \cdot \ln \frac{\mu(x)}{x_i} = \sum_i p_i \cdot \ln \frac{\mu(y) \cdot \mu(e)}{y_i \cdot e_i} =$$
$$= \sum_i p_i \cdot \ln \frac{\mu(y)}{y_i} + \sum_i p_i \cdot \ln \frac{\mu(e)}{e_i} = T(y,p) + T(e,p) \tag{1}$$

This decomposition is provided, via a more complex procedure, by Duro and Esteban (1998) who show that each additive term represents the contribution to GDP per capita inequality of each initial multiplicative factor.

As noted by Cheng and Li (2006), this approach does not consider explicitly the interaction effect deriving from the correlation between the components of per capita GDP. In order to account for this effect, Cheng and Li (2006) proposed another method (reported by Duro and Esteban in a footnote) which they developed using an unweighted version of the Theil inequality index:

$$T(x)=\frac{1}{N}\cdot\sum_i \ln\frac{\mu'(x)}{x_i}=\frac{1}{N}\cdot\sum_i \ln\frac{\mu'(y)\cdot\mu'(e)}{y_i\cdot e_i}\cdot\frac{\mu'(x)}{\mu'(y)\cdot\mu'(e)}=$$

$$=T(y)+T(e)+\ln\frac{\mu'(x)}{\mu'(y)\cdot\mu'(e)} \quad (2)$$

where $\mu'(x)$, $\mu'(y)$ and $\mu'(e)$ are the un-weighted means of x_i, y_i and e_i, respectively.

The residual $\ln\frac{\mu'(x)}{\mu'(y)\cdot\mu'(e)}$ is interpreted by Cheng and Li as an interaction effect which reflects the correlation between y and e.

However, it should be noted that this residual term only survives if the un-weighted Theil measure and means are used. This choice implies assigning to each country (or region) per capita income, irrespective of their economic or demographic size, equal importance in determining inequality. When weighted means are used, since $\mu(x) = \mu(y) \cdot \mu(e)$, the residual term of Cheng and Li becomes zero.

Following Duro and Padilla (2006), to properly consider the interaction effect between components, we must go back to decomposition (1), where the second component, $T(e, p)$, is a proper Theil index (second measure) and can correctly be interpreted as the share of inter-country per capita income inequality attributable to the employment factor. However, the first component, $T(y, p)$, is not a proper Theil index, since the weighting factor here should be the country employment share ($h_i = E_i/E$), rather than the population share (p_i). Thus, the proper Theil index for the first component is:

$$T(y,h)=\sum_i h_i\cdot\ln\frac{\mu(y)}{y_i}$$

Consequently, decomposition (1) becomes:

$$T(x,p)=\sum_i h_i\cdot\ln\frac{\mu(y)}{y_i}+\sum_i p_i\cdot\ln\frac{\mu(e)}{e_i}+\sum_i (p_i-h_i)\cdot\ln\frac{\mu(y)}{y_i} \quad (3)$$

The third addend (residual term) is easily rearranged as follows:

$$\sum_i (p_i - h_i) \cdot \ln \frac{\mu(y)}{y_i} = \sum_i (h_i - p_i) \cdot \ln \frac{y_i}{\mu(y)} =$$

$$= \sum_i \frac{P_i}{E} \cdot \left(\frac{E_i}{P_i} - \frac{E}{P}\right) \cdot \left[\ln y_i - \ln \mu_g(y) + \ln \frac{\mu_g(y)}{\mu(y)}\right] = \tag{4}$$

$$= \frac{1}{\mu(e)} \cdot \sum_i p_i \cdot [e_i - \mu(e)] \cdot [\ln y_i - \ln \mu_g(y)]$$

where μ_g is the geometric mean of the variable.

Net to the scalar $1/\mu(e)$, this residual term is the co-variance, weighted by p_i, between e_i and $\ln y_i$, and may this be interpreted as an interaction effect, depending on the correlations between these two variables.

Denoting this interaction effect by $\Delta_{y,e}$, equation (3) becomes:

$$T(x, p) = T(y, h) + T(e, p) + \Delta_{y,e} \tag{5}$$

Since $T(y, p) = T(y, h) + \Delta_{y,e}$, in equation (1) the interaction factor is inside $T(y, p)$, which therefore cannot be interpreted as the share of inequality attributable only to the productivity factor of per capita GDP. This clarifies a point considered puzzling by Cheng and Li, i.e, why one factor of (1) could contribute negatively to inequality. The two factors may indeed be of opposite sign when the contribution of one of them is partially (or totally) offset by the other one. For the sake of clarity, let us suppose that the countries examined have the same per capita GDP, so that the Theil index is zero, but both labour productivities and employment rates show some variability. Using decomposition (1), we would necessarily obtain components of opposite signs (and strength). However, the true reason why a component may be negative lies in the fact that it includes the interaction effect, which may be negative (if the correlation between the interacting variables is negative) and strong enough to affect the sign of the improperly measured component.

Using the same approach, we can of course decompose the income-weighted Theil index (first measure) as follows:

$$T(x, q) = \sum_i q_i \cdot \ln \frac{x_i}{\mu(x)} = \sum_i q_i \cdot \ln \frac{y_i}{\mu(y)} + \sum_i q_i \cdot \ln \frac{e_i}{\mu(e)} = \tag{6}$$

$$= T(y, q) + T(e, q)$$

Now the second component, $T(e, q)$, is not a proper Theil index, in which the weighting factor should be h_i:

$$T(e,h) = \sum_i h_i \cdot \ln \frac{e_i}{\mu(e)}$$

In this case decomposition (6) becomes:

$$T(x,q) = \sum_i q_i \cdot \ln \frac{y_i}{\mu(y)} + \sum_i h_i \cdot \ln \frac{e_i}{\mu(e)} + \sum_i (q_i - h_i) \cdot \ln \frac{e_i}{\mu(e)} \tag{7}$$

The residual term is now:

$$\sum_i (q_i - h_i) \cdot \ln \frac{e_i}{\mu(e)} == \frac{1}{\mu(y)} \cdot \sum_i h_i \cdot [y_i - \mu(y)] \cdot [\ln e_i - \mu_g(e)] \tag{8}$$

Now the residual term measures the interaction effect deriving from the correlation between y_i and ln e_i. Denoting this interaction effect as $\Delta'_{y,e}$, equation (7) becomes:

$$T(x,q) = T(y,q) + T(e,h) + \Delta'_{y,e} \tag{9}$$

Again, since $T(e, q) = T(e, h) + \Delta'_{y,e}$, in equation (6) the interaction component is inside $T(e, q)$, which therefore cannot be interpreted as the share of inequality attributable only to employment differentials.

The proposed approach can obviously be used for more complex decompositions of per capita GDP, for example into four factors:

$$x_i = \frac{X_i}{P_i} = \frac{X_i}{L_i} \cdot \frac{L_i}{E_i} \cdot \frac{E_i}{Pl_i} \cdot \frac{Pl_i}{P_i} = y'_i \cdot c_i \cdot e'_i \cdot d_i$$

where the new notations L_i and Pl_i are country internal employment and working-age population, respectively. The corresponding weighted mean $\mu(x)$ is then:

$$\mu = \frac{X}{P} = \frac{X}{L} \cdot \frac{L}{E} \cdot \frac{E}{Pl} \cdot \frac{Pl}{P} = \mu(y') \cdot \mu(c) \cdot \mu(e') \cdot \mu(d)$$

where $L = \sum_i L_i$, and $Pl = \sum_i Pl_i$.

This longer decomposition allows us: (i) to measure productivity and employment rate more correctly by means of two different measures of employment[1]; (ii) to measure the employment rate on the working age population; and (iii) to take into account the age structure of the population.

Going back to equation (5), the decomposition of $T(x, p)$ into four components is easily obtained by applying the decomposition in two components to both $T(y, h)$ and $T(e, p)$:

$$T(y,h) = \sum_i h_i \cdot \ln\frac{\mu(y)}{y_i} = \sum_i l_i \cdot \ln\frac{\mu(y')}{y'_i} + \sum_i h_i \cdot \ln\frac{\mu(c)}{c_i} +$$

$$+\sum_i (h_i - l_i) \cdot \ln\frac{\mu(y')}{y'_i} = T(y', l) + T(c,h) + \Delta_{y',c} \qquad (10)$$

where $l_i = L_i/L$ and:

$$T(e,p) = \sum_i p_i \cdot \ln\frac{\mu(e)}{e_i} = \sum_i w_i \cdot \ln\frac{\mu(e')}{e'_i} + \sum_i p_i \cdot \ln\frac{\mu(d)}{d_i} +$$

$$+\sum_i (p_i - w_i) \cdot \ln\frac{\mu(e')}{e'_i} = T(e', w) + T(d,p) + \Delta_{e',d} \qquad (11)$$

where $w_i = Pl_i/Pl$.

Lastly, the decomposition of $T(x, p)$ (second measure) into four components is:

$$T(x,p) = T(y',l) + T(c,h) + T(e',w) + T(d,p) + \Delta_{y,e} + \Delta_{y',c} + \Delta_{e',d} \qquad (12)$$

Correspondingly, moving from equation (9) and decomposing the two components $T(y, q)$, $T(e, q)$, we obtain the following Theil inequality (first measure) index, broken down into four components:

$$T(x,q) = T(y',q) + T(c,l) + T(e',h) + T(d,w) + \Delta'_{y,e} + \Delta'_{y',c} + \Delta'_{e',d} \qquad (13)$$

2. Decomposition of per capita GDP inequality in the EU-27 countries

We employ the approach to decompose into four components cross-country per capita GDP (market prices, millions PPPs) of the 27 EU members. Data are

[1] *Internal* employment (L_i) is used for productivity, since GDP is measured on „internal bases"; *residential* employment (E_i) is used for the employment rate, since working age population is, of course, resident population. The two measures may differ due to commuting flows, presence of foreign non-resident workers, and statistical errors.

from Eurostat and refer to 2010, the most recent year available at the time of writing. We aim here at providing one example of the differences obtained by: (i) using or not using the interaction terms; (ii) using the two Theil measures, i.e, weighting with population versus income shares.

Table 1 lists results for the first Theil measure, with and without interaction terms. Apart from the productivity component, which obviously remains unchanged, the data reveal that a great deal of the impact assigned to the L/E factor (7.8%) is due to the interaction effect between this component and productivity (5.5%); similarly, the true role of the employment rate (15.6%) is less than half that emerging without interaction terms (31.7%). Interestingly, the properly measured demographic component contributes around 1.4% to total inequality, whereas in the absence of interactions it seemed to act as a factor strongly reducing inequality (-18.1%). However, this negative sign is the effect of the negative correlation between ln e' and d, as highlighted by the sign of their interaction term.

Table. 1. Decomposition of per capita GDP in 2010 without and with interactions (Theil 1, income weighted, EU-27)

Components		%	Components		%
T(y',q)	0.02359	78.551	T(y',l)	0.02359	78.551
T(c,q)	0.00233	7.765	T(c,h)	0.00069	2.298
			$\Delta'_{y',c}$	0.00164	5.467
T(e',q)	0.00954	31.750	T(e',w)	0.00468	15.596
T(d,q)	-0.00543	-18.066	T(d,p)	0.00043	1.445
			$\Delta'_{y,e}$	-0.00003	-0.094
			$\Delta'_{e',d}$	-0.00098	-3.263
T(x,q)	*0.03004*	*100.000*	*T(x,p)*	*0.03004*	*100.000*

Similar comments may be provided with reference to the outcomes obtained using the second Theil measure (Table 2). The comparison between the population-weighted and properly weighted decompositions reveals, in the first case, an overestimation of the role of productivity differences (89% instead of 80%), biased by the interaction factors between y and e and *y'* and *c*. Conversely, the importance of employment rates is underestimated if the interaction between *e'* and *d* (negative) is not considered.

The comparison between the left panels of the two tables shows that, if the interaction terms are not taken into account, the decomposition may lead to very different outcomes in identifying the components of inequality, depending on the choice of the first or second Theil measure. However, these huge differences tend to shrink greatly if the interaction terms are accounted for (right panels).

Table. 2. Decomposition of per capita GDP in 2010 without and with interactions (Theil 2, population weighted, EU-27)

Components		%	Components		%
T(y',p)	0.03053	89.177	T(y',l)	0.02738	79.974
T(c,p)	-0.00043	-1.262	T(c,h)	0.00067	1.956
			$\Delta_{y,e}$	0.00037	1.066
			$\Delta_{y',c}$	0.00168	4.918
T(e',p)	0.00371	10.830	T(e',w)	0.00466	13.609
			$\Delta_{e',d}$	-0.00095	-2.778
T(d,p)	0.00043	1.255	T(d,p)	0.00043	1.255
T(x,p)	*0.03423*	*100.000*	*T(x,p)*	*0.03423*	*100.000*

Conclusions

This paper describes a decomposition of the Theil (first and second) measures of per capita income inequality which allows distinguishing the role of the multiplicative components of per capita income from their interaction effects. We show that correctly isolating these effects allows addressing otherwise unresolved points, e.g., the negative sign of one or more components when interactions are not considered. Our empirical exercise for the 27 EU countries (2010) shows that the explicit consideration of interaction terms drastically reduces the remarkable differences obtained using the first or second Theil measure. Conversely, if not explicitly considered, interaction effects may strongly bias the relative importance of the multiplicative factors of per capita GDP, with consequent misleading policy implications.

References

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

Cheng, Yuk-shing, Li, Sung-ko, 2006. Income Inequality and inefficiency: a decomposition approach and applications to China. Economics Letters 91, 8-14.

Duro, Juan Antonio, Esteban, Joan, 1998. Factor decomposition of cross-country income inequality, 1960-1990. Economics Letters 60, 269-275.

Duro, Juan Antonio, Padilla Emilio, 2006. International inequalities in per capita CO_2 emissions: a decomposition methodology by Kaya factors. Energy Economics 28, 170-187.

Theil, Henri, 1967. Economics and Information Theory. North-Holland, Amsterdam.