Solutions Problem Set 1 Social Welfare

Professor: Marcelo Neri TA: Tiago Bonomo

March 29, 2017

Exercise 1

*** 1.1

A population is divided into four groups, each one with four individuals. The individual incomes are:

$x_1 = [1, 1, 2, 4]\,; \quad x_2 = [1, 1, 2, 4]\,; \quad x_3 = [2, 2, 4, 8]\,; \quad x_4 = [4, 4, 8, 16]$

Calculate the two Theil measures of inequality, verifying that the between and the within group components are the same for both measures.

Solution

For the two Theil measures, we use the formulas $T = \sum_i (y_i \ln N y_i)$ and $L = -\sum_i \left(\frac{1}{N} \ln N y_i\right)$, where N is the size of the population and y_i is the share of individual i's income in total income, that is, $y_i = \frac{x_i}{\sum_i x_i} = \frac{x_i}{N\mu}$. Using individual incomes, we have that $T = \frac{1}{N\mu} \sum_i x_i \ln \frac{x_i}{\mu}$ e $L = -\frac{1}{N} \sum_i \ln \frac{x_i}{\mu}$.

For the total population, we have that $N = 16$ and

$\mu = (1+1+2+4+1+1+2+4+2+2+4+8+4+4+8+16)/16 = 4$, therefore $N\mu = 64$.

The Theil-T index for the total population is

$$
\begin{aligned}
T &= +\frac{1}{64}\left(1\times\ln\frac{1}{4}+1\times\ln\frac{1}{4}+2\times\ln\frac{2}{4}+4\times\ln\frac{4}{4}\right)\\
&\quad+\frac{1}{64}\left(1\times\ln\frac{1}{4}+1\times\ln\frac{1}{4}+2\times\ln\frac{2}{4}+4\times\ln\frac{4}{4}\right)\\
&\quad+\frac{1}{64}\left(2\times\ln\frac{2}{4}+2\times\ln\frac{2}{4}+4\times\ln\frac{4}{4}+8\times\ln\frac{8}{4}\right)\\
&\quad+\frac{1}{64}\left(4\times\ln\frac{4}{4}+4\times\ln\frac{4}{4}+8\times\ln\frac{8}{4}+16\times\ln\frac{16}{4}\right)\\
&= +\frac{1}{64}\left[4\left(1\times\ln\frac{1}{4}\right)+4\left(2\times\ln\frac{2}{4}\right)+5\left(4\times\ln\frac{4}{4}\right)+2\left(8\times\ln\frac{8}{4}\right)+16\times\ln\frac{16}{4}\right]\\
&= +\frac{1}{64}\left[4\left(1\times\ln\frac{1}{2^2}\right)+4\left(2\times\ln\frac{1}{2}\right)+5\left(4\times 0\right)+2\left(8\times\ln 2\right)+16\times\ln 2^2\right]\\
&= +\frac{1}{64}\left[4\left(-2\times\ln 2\right)+4\left(-2\times\ln 2\right)+2\times 8\times\ln 2+16\times 2\times\ln 2\right]\\
&= +\frac{1}{64}\left[\left(-8-8+16+32\right)\times\ln 2\right]=\frac{1}{64}\left(32\times\ln 2\right)=\frac{1}{2}\times\ln 2\\
T &= 0.34657
\end{aligned}
$$

The Theil-L for the total population is

$$
\begin{aligned}
L &= -\frac{1}{16}\left(\ln\frac{1}{4}+\ln\frac{1}{4}+\ln\frac{2}{4}+\ln\frac{4}{4}\right)\\
&\quad-\frac{1}{16}\left(\ln\frac{1}{4}+\ln\frac{1}{4}+\ln\frac{2}{4}+\ln\frac{4}{4}\right)\\
&\quad-\frac{1}{16}\left(\ln\frac{2}{4}+\ln\frac{2}{4}+\ln\frac{4}{4}+\ln\frac{8}{4}\right)\\
&\quad-\frac{1}{16}\left(\ln\frac{4}{4}+\ln\frac{4}{4}+\ln\frac{8}{4}+\ln\frac{16}{4}\right)\\
&= -\frac{1}{16}\left(4\times\ln\frac{1}{4}+4\times\ln\frac{2}{4}+5\times\ln\frac{4}{4}+2\times\ln\frac{8}{4}+\ln\frac{16}{4}\right)\\
&= -\frac{1}{16}\left(-2\times 4\times\ln 2-4\times\ln 2+5\times 0+2\times\ln 2+2\times\ln 2\right)\\
&= -\frac{1}{16}\left[\left(-8-4+2+2\right)\times\ln 2\right]=-\frac{1}{16}\left(-8\times\ln 2\right)=\frac{\ln 2}{2}\\
L &= 0.34657
\end{aligned}
$$

Now let's decompose each index and calculate their within and between groups components. We denote the size of group h by n_h and the share of group h in total population by $\pi_h = \frac{n_h}{N}$. The share of total income for individual i from group h is $y_{hi} = \frac{x_{hi}}{N\mu}$ and the fraction of total income for group h is $Y_h = \sum_i y_{hi}$.

We have that $h \in \{1,2,3,4\}$ and $n_h = 4$ for each h. Therefore, $\pi_h = 0.25$ for each h.

We also have that

$$
\begin{aligned}
Y_1 &= \sum_i y_{1i} = \frac{1}{64}\sum_i x_{1i} = \frac{1}{64}(1+1+2+4) = \frac{1}{8} \\
Y_2 &= \sum_i y_{2i} = \frac{1}{64}\sum_i x_{2i} = \frac{1}{64}(1+1+2+4) = \frac{1}{8} \\
Y_3 &= \sum_i y_{3i} = \frac{1}{64}\sum_i x_{3i} = \frac{1}{64}(2+2+4+8) = \frac{2}{8} \\
Y_4 &= 1-(Y_1+Y_2+Y_3) = 1-\frac{4}{8} = \frac{4}{8}
\end{aligned}
$$

For the Theil-T index, we note that is possible to decompose it as follows: $T = T_b + \sum_{h=1}^{4} Y_h T_h$, where $T_b = \sum_{h=1}^{4} Y_h \ln \frac{Y_h}{\pi_h}$ e $T_h = \sum_{i=1}^{4} \frac{y_{hi}}{Y_h} \ln\left(n_h \frac{y_{hi}}{Y_h}\right)$. Let's calculate its between groups component, T_b.

$$
\begin{aligned}
T_b &= \frac{1}{8}\ln\frac{1}{8\times 0,25} + \frac{1}{8}\ln\frac{1}{8\times 0,25} + \frac{2}{8}\ln\frac{2}{8\times 0,25} + \frac{4}{8}\ln\frac{4}{8\times 0,25} \\
&= \frac{1}{8}\left(\ln\frac{1}{2} + \ln\frac{1}{2} + 2\times\ln\frac{2}{2} + 4\times\ln\frac{4}{2}\right) \\
&= \frac{1}{8}(-\ln 2 - \ln 2 + 2\times 0 + 4\times\ln 2) = \frac{1}{8}\times 2\times\ln 2 = \frac{\ln 2}{4} \\
T_b &= 0.17329
\end{aligned}
$$

Note that if we know T and T_b, it is possible to obtain the within groups component:

$$
\begin{aligned}
\sum_{h=1}^{4} Y_h T_h &= T - T_b = \frac{\ln 2}{2} - \frac{\ln 2}{4} = \frac{\ln 2}{4} \\
&= 0.17329
\end{aligned}
$$

Note that, in this specific case, inequality between groups and within groups are the same.

For the Theil-L index, the decomposition is as follows: $L = L_b + \sum_{h=1}^{4} \pi_h L_h$, where $L_b = \sum_{h=1}^{4} \pi_h \ln\frac{\pi_h}{Y_h}$ and $L_h = \frac{1}{n_h}\sum_{i=1}^{4} \ln\frac{Y_h}{n_h y_{hi}}$. To calculate the within groups component, let's find $\{L_h\}_{h=1}^{4}$ first:

$$
\begin{aligned}
L_1 &= \frac{1}{4}\left(\ln\frac{1/8}{4\times 1/64}+\ln\frac{1/8}{4\times 1/64}+\ln\frac{1/8}{4\times 2/64}+\ln\frac{1/8}{4\times 4/64}\right)=\frac{1}{4}\left(\ln 2+\ln 2+\ln 1-\ln 2\right)=\frac{\ln 2}{4}\\
L_2 &= \frac{\ln 2}{4}\\
L_3 &= \frac{1}{4}\left(\ln\frac{2/8}{4\times 2/64}+\ln\frac{2/8}{4\times 2/64}+\ln\frac{2/8}{4\times 4/64}+\ln\frac{2/8}{4\times 8/64}\right)=\frac{1}{4}\left(\ln 2+\ln 2+\ln 1-\ln 2\right)=\frac{\ln 2}{4}\\
L_4 &= \frac{1}{4}\left(\ln\frac{4/8}{4\times 4/64}+\ln\frac{4/8}{4\times 4/64}+\ln\frac{4/8}{4\times 8/64}+\ln\frac{4/8}{4\times 16/64}\right)=\frac{1}{4}\left(\ln 2+\ln 2+\ln 1-\ln 2\right)=\frac{\ln 2}{4}
\end{aligned}
$$

Note that $L_1 = L_2 = L_3 = L_4$, that is, inequalities within the groups are the same. This happened because the Theil-L index (as well as the Theil-T and the Gini) attends the income homogeneity condition, also known as scale independence. Observe that $x_4 = 2x_3 = 4x_2 = 4x_1$, that is, the income distribution of any given group can be written as the income distribution of any other group multiplied by a constant.

The Theil-L within groups component is

$$\sum_{h=1}^{4}\pi_h L_h = 0,25\times 4\times\frac{\ln 2}{4}=\frac{\ln 2}{4}=0.17329$$

Therefore, the between groups component is

$$L_b = L-\sum_{h=1}^{4}\pi_h L_h=\frac{\ln 2}{2}-\frac{\ln 2}{4}=\frac{\ln 2}{4}=0.17329$$

We also verify for the Theil-L index that inequality between and within groups are the same.

**1.2

The individual incomes for three groups are given. In group 1, there are six individuals with incomes

$x_{11} = x_{12} = 0.5$; $x_{13} = x_{14} = x_{15} = 1$; $x_{16} = 8$

In group 2, there are five individuals with incomes

$x_{21} = x_{22} = x_{23} = x_{24} = 1$; $x_{25} = 16$

Group 3 has only three individuals and their incomes are

$x_{31} = x_{32} = x_{33} = 16$

The three groups together constitute a total population of 14 individuals. Calculate the mean, median, mode, amplitude and variance of the income taking into account the 14 individuals. Calculate the Theil-T index related to the inequality in each group, the index related to global inequality and its within and between groups componentes. Do the same for the Theil-L index. Do the same for the Gini index.

Solution

First, let's calculate the measures for the total population. We have that $N = 14$ and total income is

$\sum_h \sum_i x_{hi} = 2 \times 0,5 + 7 \times 1 + 8 + 4 \times 16 = 80$

The mean is $\mu = {}^{80}/_{14} = {}^{40}/_{7}$.

To find the median, let's rank the distribution from the lowest to the highest income and see what is the central term of the distribution.

We have that

$x = [{}^{1}/_{2}, {}^{1}/_{2}, 1, 1, 1, 1, 1, 1, 1, 8, 16, 16, 16, 16]$.

The median is 1 and is the mean between the 7^o and the 8^o terms in the distribution.

We have therefore that half of the population have incomes less than or equal to 1, while the other half have incomes bigger than or equal to 1.

The mode is the most repeated value in the distribution and is also 1.

The amplitude is the differente between the extreme incomes of the distribution and therefore is $16 - \frac{1}{2} = 15.5$.

We denote the variance by σ^2. We have that

$\sigma^2 = \frac{\sum(x_i - \mu)^2}{N} = \frac{\sum(x_i - {}^{40}/_{7})}{14} = 45.59694$

The Theil-T index for the total distribution is

$$
\begin{aligned}
T &= \frac{1}{N\mu} \sum_i x_i \ln \frac{x_i}{\mu} \\
&= \frac{1}{80} \left[2 \times \frac{1}{2} \ln \left(\frac{{}^{1}/_{2}}{{}^{40}/_{7}} \right) + 7 \times 1 \times \ln \left(\frac{1}{{}^{40}/_{7}} \right) + 8 \times \ln \left(\frac{8}{{}^{40}/_{7}} \right) + 4 \times 16 \times \ln \left(\frac{16}{{}^{40}/_{7}} \right) \right] \\
T &= 0.67438
\end{aligned}
$$

The Theil-L is

$$
\begin{aligned}
L &= -\frac{1}{N} \sum_i \ln \frac{x_i}{\mu} \\
&= -\frac{1}{14} \left(2 \times \ln \frac{0,5}{{}^{40}/_{7}} + 7 \times \ln \frac{1}{{}^{40}/_{7}} + \ln \frac{8}{{}^{40}/_{7}} + 4 \times \ln \frac{16}{{}^{40}/_{7}} \right) \\
L &= 0.90129
\end{aligned}
$$

Populations in each group are

$n_1 = 6$; $n_2 = 5$; $n_3 = 3$

Shares of population by each groups are

$\pi_1 = \frac{6}{14}$; $\pi_2 = \frac{5}{14}$; $\pi_3 = \frac{3}{14}$

Mean incomes in each group are

$\mu_1 = \frac{(2 \times 0,5 + 3 \times 1 + 8)}{6} = 2$; $\mu_2 = \frac{(4 \times 1 + 16)}{5} = 4$; $\mu_3 = \frac{3 \times 16}{3} = 16$

Shares in total income by each group are

$Y_1 = \frac{2 \times 6}{80} = \frac{12}{80} = \frac{3}{20}$; $Y_2 = \frac{4 \times 5}{80} = \frac{20}{80} = \frac{5}{20}$; $Y_3 = \frac{16 \times 3}{80} = \frac{48}{80} = \frac{12}{20}$

Let's calculate the Theil-T indexes within each group. We have that

$$T_1 = \sum_{i=1}^{n_1} \frac{x_{1i}}{n_1\mu_1} \ln \frac{x_{1i}}{\mu_1} = 2 \times \left(\frac{0,5}{12} \times \ln \frac{0,5}{2} \right) + 3 \times \left(\frac{1}{12} \times \ln \frac{1}{2} \right) + \frac{8}{12} \ln \frac{8}{2} = \frac{11}{12} \times \ln 2 = 063538$$

$$T_2 = \sum_{i=1}^{n_2} \frac{x_{2i}}{n_2\mu_2} \ln \frac{x_{2i}}{\mu_2} = 4 \times \left(\frac{1}{20} \times \ln \frac{1}{4} \right) + \frac{16}{20} \times \ln \frac{16}{4} = \frac{6}{5} \times \ln 2 = 0.83178$$

$$T_3 = \sum_{i=1}^{n_3} \frac{x_{3i}}{n_3\mu_3} \ln \frac{x_{3i}}{\mu_3} = 3 \times \left(\frac{16}{48} \times \ln \frac{16}{16} \right) = 0$$

The within groups component is therefore

$$\sum_h Y_h T_h = \frac{3}{20} \times T_1 + \frac{5}{20} T_2 + \frac{12}{20} \times T_3 = 0.30325$$

The between groups component is

$$T_b = T - \sum_h Y_h T_h = 0,67438 - 0.30325$$

$$T_b = 0,37113$$

The Theil-L indexes within each group are

$$L_1 = -\frac{1}{n_1} \sum_i \ln \frac{x_{1i}}{\mu_1} = -\frac{1}{6} \left(2 \times \ln \frac{0,5}{2} + 3 \times \ln \frac{1}{2} + \ln \frac{8}{2} \right) = 0.57762$$

$$L_2 = -\frac{1}{n_2} \sum_i \ln \frac{x_{2i}}{\mu_2} = -\frac{1}{5} \left(4 \times \ln \frac{1}{2} + \ln \frac{16}{2} \right) = 0.83178$$

$$L_3 = -\frac{1}{n_3} \sum_i \ln \frac{x_{3i}}{\mu_3} = -\frac{1}{3} \left(3 \times \ln \frac{16}{16} \right) = 0$$

The Theil-L within groups component is therefore

$$\sum_{h=1}^{3} \pi_h L_h = \frac{6}{14} L_1 + \frac{5}{14} L_2 + \frac{3}{14} L_3 = 0.54461$$

The between groups component is

$$L_b = L - \sum_{h=1}^{3} \pi_h L_h = 0,90129 - 0,54461 = 0.35668$$

For the global distribution, we have that the Gini is

$$
\begin{aligned}
G &= \frac{2}{n^2\mu}\sum_i ix_i - \left(1+\frac{1}{n_h}\right) \\
&= \frac{2}{14^2 \times {}^{40}\!/_7}\left[(1+2)\times 0,5 + (3+4+5+6+7+8+9)\times 1 + 10\times 8 + (11+12+13+14)\times 16)\right] - \\
G &= 0.57768
\end{aligned}
$$

Now let's decompose it in the within and between groups components. Let's consider the groups ranked from the lowes to the highest share in total income and, within each group, we rank the individuals from the lowest to the highest income.

We denote group h's share in total income as $\Phi_h = \frac{1}{N\mu}\sum_{j=1}^{h}\mu_j n_j$ and the share of group h's income of individual $i$as $\Phi_{hi} = \frac{1}{n_h\mu_h}\sum_{j=1}^{i} x_{hj}$. We can decompose the Gini index as follows:

$G = G_b + \sum_h \pi_h Y_h G_h + G_s$, where $G_b = 1 - \sum_h (\Phi_h + \Phi_{h-1})\,\pi_h$ and $G_h = \frac{2}{n_h^2\mu_h}\sum_i ix_{hi} - \left(1+\frac{1}{n_h}\right)$

We can make use of the following table:

h	$n_h\mu_h$	π_h	$N\mu\Phi_h$	Φ_h
1	12	$6/14$	12	$12/80$
2	20	$5/14$	32	$32/80$
3	48	$3/14$	80	1

We have that

$$
\begin{aligned}
G_b &= 1 - \sum_h (\Phi_h + \Phi_{h-1})\,\pi_h \\
&= 1 - \left[\left(\frac{12+0}{80}\right)\times\frac{6}{14} + \left(\frac{32+12}{80}\right)\times\frac{5}{14} + \left(\frac{80+32}{80}\right)\times\frac{3}{14}\right] \\
G_b &= 0.43929
\end{aligned}
$$

The Gini for each group is

$$
\begin{aligned}
G_1 &= \frac{2}{n_1^2\mu_1}\sum_i ix_{1i} - \left(1-\frac{1}{n_1}\right) = \frac{2}{6^2\times 2}\left[0,5\times(1+2) + 1\times(3+4+5) + 8\times 6\right] - \left(1+\frac{1}{6}\right) \\
&= 0.54167 \\
G_2 &= \frac{2}{n_2^2\mu_2}\sum_i ix_{2i} - \left(1-\frac{1}{n_2}\right) = \frac{2}{5^2\times 4}\left[1\times(1+2+3+4) + 16\times 5\right] - \left(1+\frac{1}{5}\right) \\
&= 0.60000 \\
G_3 &= = \frac{2}{n_3^2\mu_3}\sum_i ix_{3i} - \left(1-\frac{1}{n_3}\right) = \frac{2}{3^2\times 16}\left[16\times(1+2+3)\right] - \left(1+\frac{1}{3}\right) \\
&= 0
\end{aligned}
$$

The within groups component is therefore

$$\sum_h \pi_h Y_h G_h = \frac{6}{14} \times \frac{3}{20} \times G_1 + \frac{5}{14} \times \frac{5}{20} \times G_2$$
$$= 0.08839$$

****1.3**

Consider two populations divided into three stratum each.

In population A, the 40% poorest have 10% of total income, the 40% of the middle have 40% and the 20% richest have 50%.

In population B, the three stratum (40% poorest, 40% of the middle and 20% richest) have 20%, 20% and 60% of total income, respectively.

We suppose there is no inequality within each stratum.

Calculate the Gini index for each one of the two populations. Do the same for the Theil-T index and for the Theil-L indexes. Based on these results, verify in each one of the two populations the income distribution is more unequal.

Comment the results taking into account the Lorenz curve for each population.

Solution

We can calculate the Gini using the formula of the between and within groups decomposition, noting that the within groups component is zero in this case. Therefore, the Gini for each population is

$$G_1 = 1 - \sum_h (\Phi_h + \Phi_{h-1}) \pi_h = 1 - [(0,1-0) \times 0,4 + (0,5+0,1) \times 0,4 + (1+0,5) \times 0,2)] = 0.42000$$
$$G_2 = 1 - \sum_h (\Phi_h + \Phi_{h-1}) \pi_h = 1 - [(0,2-0) \times 0,4 + (0,4+0,2) \times 0,4 + (1+0,4) \times 0,2)] = 0.40000$$

The Theil-T for each population is

$$T_1 = \sum_i \left(y_h \ln \frac{y_h}{\pi_h} \right) = 0,1 \times \ln \frac{0,1}{0,4} + 0,4 \times \ln \frac{0,4}{0,4} + 0,5 \times \ln \frac{0,5}{0,2} = 0.31952$$
$$T_2 = \sum_i \left(y_h \ln \frac{y_h}{\pi_h} \right) = 0,2 \times \ln \frac{0,2}{0,4} + 0,2 \times \ln \frac{0,2}{0,4} + 0,6 \times \ln \frac{0,6}{0,2} = 0.38191$$

The Theil-L is

$$L_1 = \sum_h \pi_h \ln \frac{\pi_h}{Y_h} = 0,4 \times \ln \frac{0,4}{0,1} + 0,4 \times \ln \frac{0,4}{0,4} + 0,2 \times \ln \frac{0,2}{0,5} = 0.37126$$
$$L_2 = \sum_h \pi_h \ln \frac{\pi_h}{Y_h} = 0,4 \times \ln \frac{0,4}{0,2} + 0,4 \times \ln \frac{0,4}{0,2} + 0,2 \times \ln \frac{0,2}{0,6} = 0.33480$$

We can't say unequivocally which of the two income distributions is more unequal. If we use the Gini or the Theil-L, we have that population A is more unequal but if we use the Theil–T we have the converse result. The results are different because the Lorenz curves cross each other. The three measures of inequality are consistent with Lorenz, that is, when the Lorenz curves for two distributions don't cross, we can tell which distribution is more unequal. The distribution closer to the line of perfect equality will be the less unequal using the three indexes.

1.4

Answer true or false and comment. The dual permits to compare different measures of inequality in the same scales and study the average sensibility of inequality to income transfers.

Solution

The sentence is true. The dual of the Theil convert the indexes in a measure that take values from 0 to 1 and is dimensionless, so we can compare it to the Gini index. Besides that, as the dual permit us to compare populations with different numbers of null incomes, we can therefore infer the impact of income transfers on the distribution.

1.5

IBGE (National Bureau of Statistics and Geography) has recently lauched the dual of the Theil index taking into account the working population with positive income using data from PNADs 2002 and 2003. Calculate the evolution (percentual) of inequality using the dual of the Theil-T index taking into account the active age population (therefore PIA – 15 to 65 years old).

	2002	2003	Variation
Dual of the Theil-T	0.563	0.555	-1.42%
% of PIA with zero income	0.507	0.519	2.43%

Calculate the Theil-T index for PNAD 2003

Solution

Finding the dual for the PIA consist in finding the dual of a new distribution obtained adding a share of the population with null income. For 2002, the proportion $\phi_{02} = 0.507$ of the new distribtuion will be added with null income. The new dual will be:

$U_{02\prime} = \phi + U_{02}\,(1 - \phi_{02}) = 0.507 + 0.563\,(1 - 0.507) = 0.78456$

For 2003 we have that $\phi_{03} = 0,519$. Therefore

$U_{03\prime} = \phi + U_{03}\,(1 - \phi_{03}) = 0.519 + 0.555\,(1 - 0.519) = 0.78596$

Inequality in the PIA as measured by the dual of the Theil-T has evolved as follows

$\frac{U_{03\prime} - U_{02\prime}}{U_{02\prime}} = 0.17793\%.$

Inequality measured by the Theil-T can be obtained from its dual using the formula

$U_T = 1 - \exp(-T)$

Therefore, for 2003 we have that

$$\begin{aligned} U\prime_{03} &= 1 - \exp(-T_{2003}) \\ 0,78596 &= 1 - \exp(-T_{2003}) \\ \exp(-T_{2003}) &= 0.21405 \\ -T_{2003} &= \ln 0.21405 \\ T_{2003} &= -\ln 0.21405 = 1.54155 \end{aligned}$$

1.6

We present the mean and the inequality of per capita income using the Gini index of a hipothetical country before and after a socialist revolution. Calculate the evolution of the well being in this society taking into account the function proposed by Sen.

- Before: Mean Income 300 and Gini 0.6
- After: Mean Income 250 and Gini 0.5

Solution

Sen's social welfare function is $W = \mu(1 - G)$, where μ is the mean and G is the Gini index of the distribution. We have that,

$$\begin{aligned} W_o &= 300 \times (1 - 0.6) = 120 \\ W_1 &= 250 \times (1 - 0.5) = 125 \end{aligned}$$

Social welfare, as measured by the Sen's function, increased from 120 to 125, that is, 4.2%.

1.7

Answer true or false and comment. The extension of temporal variability of observed income always influence inequality of annual incomes keeping constant the presente value of the income earned during the life cycle.

Solution

False. If the temporal variability of observed income doesn't change the permanent income, it will only affect the distribution of current income, once a negative shock in one period will be compensated by a positive shock in a future one.

1.8

Write down the formula and discuss possible problems of the following indicators:

- Sen's Social Welfare Function
- Variance of the logs as a measure of inequality

Solution

- Sen's Social Welfare Function

We have that

$W = \mu (1 - G)$,

where μ is the mean and G is the Gini index of the distribution.

The inequality measure used in Sen's Social Welfare Function is the Gini index, measure that has a complex decomposition in between and within groups components. Except in the cases where thera are no intersections between the income brackets in the different groups, we observe a residual that is very hard to interpret in the decomposition. Besides that, the funcion considers that efficient and equality has the same weights, being a particular case of the Graaff's Social Welfare Function, given by $W = \mu (1 - G)^{\sigma}$. In the Sen's case, we have $\sigma = 1$.

- Variance of the logs

We have that

$V = \frac{1}{N \sum [E[log(y)] - log(y)]^2}$

The main advantages are it is scale-invariant and decomposable. The decomposition follows because of the properties of the logarithmic function, specially additivity. However, it is now defined for null incomes and it is not very sensitive to changes in the top of the distribution, once the logarithmic function "smoothes" the distribution.

1.9

Write down the two alternative formulas and explain the logic and the intution behind the Gini index.

Solution

- $G = \frac{1}{\mu N(N-1)} \sum_{i>j}^{N} \sum_{j}^{N} \mid x_i - x_j \mid$

The Gini corresponds to the ratio between the average of the absoluto deviations of the incomes of all the people in the sample and twice the average. Remember that there are $\frac{N(N-1)}{2}$ distinct pairs. Note that in the case of perfect equality ($x_i = \mu$ for all i), we have that the sum is equal to zero and the Gini is also zero. On the other hand, in the case of perfect inequality ($x_i = N\mu$for one individual and $x_i = 0$ for the other ones), we have that the Gini is equal to one.

- $G = \frac{N+1}{N-1} - \frac{2}{N(N-1)\mu} \sum_{i=1}^{N} \rho_i x_i$

where ρ_i is the ranking in the distribution of income from the highest to the lowest.

It is clear that poorer individuals receive bigger weights: the richest in the distribution has weight 1, while the poorest has weight N. The first term of the expression aprroaches 1 as N approaches infinity. The advantages of this formula is that it is simpler to use and it is insensitive to scale. The disadvantage is that it is not fully decomposable.

- $G = 1 - \frac{1}{n} \sum_{i=1}^{n} (\Phi_i + \Phi_{i-1})$

The Gini is 2 times the area between the Lorenz curve and the perfect equality line. Alternatively, as expressed by the above formula, the Gini is equal to 1 minos 2 times the are between the Lorenz curve and the perfect unequality line. For a discrete distribution, the area between the Lorenz curve and the perfect unequality line consist in trapeziums with bases Φ_i (bigger) and Φ_{i-1} (smaller) and height $1/n$.

1.10

According to the empirical evidence seen in class, the influence of the attribute in the decomposition of the Theil-T index is bigger when we measure:

i) The race attribute using household per capita income or individual labor income?

ii) The gender attribute using total individual labor income or labor income normalized per hour?

Solution

i) The influence of the race attribute probably is higher when we use household per capita income because an assortative matching effect impacts positively the probability that most discriminated individuals (like blacks, for example) marry individuals that are similar, the same happening for the most privileged. The consequence is that the contribution of the race attribute should be high for the inequality between households.

ii) In Brazil, we have that the women work more than men despite earning less. That means that if we use labor income normalized per hour, we will have that the influence of the gender attribute will be higher than if we use total individual labor income.

Exercise 2

2.1

Calculate the inequality indexes seen in the course (Theil-T, Theil-L, Gini and the duals) according to the following sample of incomes:

$x = [1, 1, 2, 6, 30]$

If we add one person with zero income to the sample, how do the indexes change?

Solution

In the above distribution, we have $n = 5$, $\mu = {}^{(1+1+2+6+30)}/_{5} = 8$ and therefore $n\mu = 40$.

- Theil-T

$$
\begin{aligned}
T &= \sum_{i=1}^{5} y_i \ln ny_i = \frac{1}{n\mu} \sum_{i=1}^{5} x_i \ln \frac{x_i}{\mu} \\
&= \frac{1}{40} \left(2 \times 1 \times \ln \frac{1}{8} + 2 \times \ln \frac{2}{8} + 6 \times \ln \frac{6}{8} + 30 \times \ln \frac{30}{8} \right) \\
&= 0.77488
\end{aligned}
$$

The dual is

$$U_T = 1 - \exp(-T) = 1 - \exp(-0.77488) = 0.53924$$

Adding an individual with zero income represents a share of $\phi = \frac{1}{6}$ with zero income in the new distribution. The dual of the new distribution is

$$U_T' = \phi + (1 - \phi) U_T = \frac{1}{6} + \frac{5}{6} \times 0.53924 = 0.61078$$

The Theil-T of the new distribution is therefore

$$T' = -\ln(1 - U_T') = -\ln(1 - 0,61078) = 0.94361$$

- Theil-L

$$
\begin{aligned}
L &= -\frac{1}{n} \sum_{i=1}^{5} \ln \frac{x_i}{\mu} = -\frac{1}{5} \left(2 \times \ln \frac{1}{8} + \ln \frac{2}{8} + \ln \frac{6}{8} + \ln \frac{30}{8} \right) \\
&= 0,90222
\end{aligned}
$$

It is not possible to calculate the changes in the Theil-L when we add a person with null income because the index is not defined for zero incomes.

- Gini

$$
\begin{aligned}
G &= \frac{2}{n^2 \mu} \sum_i i x_i - \left(1 + \frac{1}{n} \right) \\
&= \frac{2}{5^2 \times 8} (1 \times 1 + 1 \times 2 + 2 \times 3 + 6 \times 4 + 30 \times 5) - \left(1 + \frac{1}{5} \right) \\
&= 0.63000
\end{aligned}
$$

The dual is the Gini itsfelf. If we add an individual with null income, we will have that

$$G' = \phi + (1 - \phi)\, G = \frac{1}{6} + \frac{5}{6} \times 0.63 = 0.69167$$

2.2

Suppose that per capita income of household A, composed of only one individual, is 8. Suppose also that there is only another household in the economy, with incomes $\{1,\ 1,\ 2,\ 6,\ 30\}$. Calculate the level of inequality according to the following concepts:

i) Household per capita income between households

ii) Household per capita income between individuals

iii) Calculate the inequality component of individual income between groups of households (i.e., A and B). Assume now that income of household A is 7. Recalculate it.

iv) Suppose there is no socialization of incomes inside the households. How much of total income inequality is going to be underestimated taking into account both scenarios?

Solution

Let's use the Theil-T index to calculate the level of inequality.

House per capita income for household B is $x_2 = \frac{x_{21}+x_{22}+x_{23}+x_{24}+x_{25}}{n_2} = \frac{1+1+2+6+30}{5} = 8$.

The share in total population for each household is $\pi_1 = \frac{1}{6}$ and $\pi_2 = \frac{5}{6}$.

The share in total income for each household is $Y_1 = \frac{8}{48} = \frac{1}{6}$ and $Y_2 = \frac{40}{48} = \frac{5}{6}$.

The Theil-T between households is, therefore

$$T_b = \sum_h Y_h \ln \frac{Y_h}{\pi_h} = \frac{1}{6} \ln \frac{1/6}{1/6} + \frac{5}{6} \ln \frac{5/6}{5/6} = 0$$

Because all individuals have household per capita income of 8, the Theil-T index between individuals will also be 0.

Considering now $x_1 = 7$, we have that $Y_1 = \frac{7}{47}$ and $Y_2 = \frac{40}{47}$.

The Theil-T between households is

$$T_b' = \frac{7}{47} \ln \frac{7/47}{1/6} + \frac{40}{47} \ln \frac{40/47}{5/6} = 0.00117$$

Note that, in this case, $T = T_b$, because we still have that the within group components are zero since we are considering household per capita income.

If we consider individual income, we have that

$$
\begin{aligned}
T &= \frac{1}{n\mu}\sum x_i \ln\frac{x_i}{\mu} = \frac{1}{48}\left(8\ln\frac{8}{8}+\ln\frac{1}{8}+\ln\frac{1}{8}+6\ln\frac{6}{8}+30\ln\frac{30}{8}\right) = 0.70349 \\
T' &= \frac{1}{n\mu'}\sum x_i \ln\frac{x_i}{\mu'} = \frac{1}{47}\left(7\ln\frac{7}{47/6}+\ln\frac{1}{47/6}+\ln\frac{1}{47/6}+6\ln\frac{6}{47/6}+30\ln\frac{30}{47/6}\right) = 0.71873
\end{aligned}
$$

We verify that, in the case with no socialization of incomes within the households, inequality would be substantially higher. In the first case, for example, the Theil-T index would rise from $T = 0$ to $T = 0.70349$.

2.3

Write down the formulas and compare advantages and disadvantages of the Theil-T and Gini inequality indexes.

Solution

- Theil-T

$T = \frac{1}{N\mu}\sum_i x_i \ln\frac{x_i}{\mu}$

- Gini

$G = \frac{2}{n^2\mu}\sum_i ix_i - \left(1+\frac{1}{n_h}\right)$

The Gini index has the advantage of being easy to interpret because it takes values from 0 to 1, while the Theil-T can take values from 0 to $\ln N$.

However, we cannot decompose it into between and within group components as we can do with the Theil-T.

2.4

What is the meaning and the importance of the Principle of Transfers (Pigou-Dalton) in the specification of a Social Welfare Funciton?

Solution

The Pigou-Dalton Principle of Transfers is one of the desirable properties of a social welfare function and it reveals a preference for equality. Ignoring the effects of incentives and allocation restrictions, we have that the social welfare function is higher the less unequal is the income distribution, conditional on having the same mean. This principle says that any progressive income transfer that keeps constant the ranking in the distribution must reduce inequality and therefore increases the social welfare function.

2.5

Define and illustrate the concept of Lorenz dominance.

Solution

Lorenz dominance is a criterion that permits us to compare two different distribution and say unequivocally which one is more unequal. The dominance is verified when a Lorenz curve is always above another one, that is, when they don't cross. If they cross, there is no Lorenz dominance and it might be the case that different inequality indexes imply different results in terms of which distribution is more unequal. Therefore, we cannot say unequivocally which one is more unequal. If is the case of Lorenz dominance, all the most relevant inequality (including the Theil-T, Theil-L and the Gini) indexes will point in the same direction.

Exercise 3

3.1 - Empirical Estimates

For the model $ln(Y_i) = \alpha + \beta \times X_i + u_i$, we have the following estimate

$$ln(Y_i) = \underset{(0.01768)}{0.8972} + \underset{(0.0497)}{0.1543} \times X_i$$

$$R^2 = 0.4456$$

where
Y_i = income from the main activity
X_i = years of schooling
The numbers in brackets are the standard errors of the estimates.

Interpret the slope coefficient (give the formula), its significance and the R^2 of the regression.

Solution

We have that the slope coefficient is given by

$$\hat{\beta} = \frac{Cov(X_i, lnY_i)}{Var(X_i)}$$

Note that the t statistic is

$$t = \frac{\hat{\beta}}{ep(\hat{\beta})} = \frac{0.1543}{0.0497} \approx 3.1$$

so that the estimative is statistically significant.

The interpretation is that each additional year of schooling is associated on average with an increase in the wages of approximately 15.4%.

The R^2 of the regression indicates that approximately 45% of the variation in the wages is explained by the variation in the years of schooling.

3.2

Using the regression above and the Theil-T index, discuss and explain the logic of the role of education in the determination of the labor income inequality in Brazil.

Solution

The estimatives of the regression above, considering that the model is a good one, show us that education has an important role in determing the wages of the individuals. Actually, 15.43% is a bery big number for the returns to education, much bigger than in most of the developed countries. Even with this big educational premium, we still have in Brazil very low educational levels, so there is a big room for improvement. We will also have that education must have an important role in determing the Theil-T index for income inequality, as wages represent the most important component in individual or even household income.

Exercise 4

4.1

Taking into account the following social welfare function, discuss how to incorporate the Principle of Transfers in the measures of inequality. What would be the case for the Gini and Atkinson's (with $\varepsilon = 1$) measures?

$W = u(x^*) = \int_0 u(x)w(x)f(x)dx$

Solution

The Principle of Transfer can be incorporated if we consider bigger weights for the poorest individuals, that is, with lower x. In the case of the Gini, we have that $w(x) = 2[1 - F(x)]$. Note then that the poorest individual, for whom $F(x) = 0$, has weight $w(x) = 2$, while the richest individual, for whom $F(x) = 1$, has weight $w(x) = 0$. Another possibility is to consider individual utility functions with decreasing marginal utility. In this case, an income transfer from a relative rich individual to a relative poor onde increase welfare, onde the increase in the utilty of the poorest of the two individuals is bigger than the decrease in the utilty of the richest. This is the case of the utility considered in Atkinson's Welfare Function, in which $u(x) = ln(x)$.

4.2

Calculate the Theil-T index between groups by gender for per capita individual income using the following data. Interpret.

- Male: Per Capita Income 806.54 and Population 91,507,992
- Female: Per Capita Income 468.31 and Population 96,686,391

- Total: Per Capita Income 630.25 and Population 188,194,383

Solution

Note that we are considering two representative individuals (one for the male and one for the female). Then, we only have to do the same procedure as we did in exercise 2.2, noting that the within groups component will be equal to zero (only onde representative individual in each group). Therefore, we will have that the Theil-T index will be equal to the between groups component. Considering men as group 1 and women as group 2, we have that $\pi_1 = \frac{91{,}707{,}992}{(91{,}707{,}992+96{,}686{,}391)} = 0.486$ and $\pi_2 = \frac{96{,}686{,}391}{(91{,}707{,}992+96{,}686{,}391)} = 0.514$.

Also, $Y_1 = \frac{806.54}{(806.54+468.31)} = 0.633$ and $Y_2 = \frac{468.31}{(806.54+468.31)} = 0.367$. Therefore, we have that

$$T = T_b = \sum_{h=1}^{2} Y_h \ln \frac{Y_h}{\pi_h} = (0.633)\ln(\frac{0.633}{0.486}) + (0.367)\ln(\frac{0.367}{0.514}) = 0.344 - 0.027 = 0.317$$

Note that for the male we have that their share in total income is higher than their share in the population, the converse happening for the female.

4.3

Explain how the following dimensions affect the measurment of inequality and its decomposition between and within groups:

a) use disaggregated income between individuals from the same household and calculate the Theil and Gini indexes between households versus between individuals.

Solution

a) If we calculate the Theil and Gini indexes between households taking into account dissaggregated income between individuals, we will have that the differences in income between individuals will tend to be reduced when we add them all in the household. That is, the within groups component will tend to be smaller because we are implicitly considering that individuals "socialize" income inside the households. In terms of the between groups component, it will depend in the heterogeneity of the households. The more different they are in terms of per capita income, the higher inequality between groups will be. If otherwise we calculate the indexes between individuals, we will have that the within and between group components will be the same, as we are considering disaggregated income.

4.4 - Empiric

Consider the labor decomposition of individual income taking into account different sources:

a. What is the rate of unemployment in the PIA (Active Age Population - 15 to 65 years old)?

b. What is the fraction of the growth of the mean labor income in the PIA that is explained for the rise in occupation?

c. If we assume a 0.5% per year growth of the PIA as a result of the recent demographic transition, what should be the growth of income from all sources?

d. Compare the impacts in total income of the demographic bonus with the impacts of the rise in average years of schooling of the occupied (educational bonus).

Solution

For this exercise, see handout 11.

a. The rate of unemployment in the Economically Active Population (therefore PEA) for 2009 was $1 - \frac{ocup}{PEA} = 1 - 0.833 = 0.167$. We also know that paricitpation rate in the labor market corresponds to $\frac{PEA}{PIA} = 0.739$. Therefore, we can calculate the rate of unemployment in the PIA as following

$$\frac{unemp}{PIA} = \frac{unemp}{PEA} \times \frac{PEA}{PIA} = 0.167 \times 0.739 = 0.123$$

That is, the rate of unemployment in the PIA for 2009 was 12.3%.

b. The rise in occupation in PIA was

$$\frac{(\frac{ocup}{PIA})_{2009} - (\frac{ocup}{PIA})_{2003}}{(\frac{ocup}{PIA})_{2003}} = \frac{(\frac{ocup}{PEA})_{2009} \times (\frac{PEA}{PIA})_{2009} - (\frac{ocup}{PEA})_{2003} \times (\frac{PEA}{PIA})_{2003}}{(\frac{ocup}{PEA})_{2003} \times (\frac{PEA}{PIA})_{2003}}$$

$$\Rightarrow \frac{(\frac{ocup}{PIA})_{2009} - (\frac{ocup}{PIA})_{2003}}{(\frac{ocup}{PIA})_{2003}} = \frac{(0.833) \times (0.739) - (0.803) \times (0.721)}{(0.803) \times (0.721)} = 6.32\%$$

We have that

$$totincome_{2009} = 806.56 \text{ and } (\frac{totincome}{laborincome})_{2009} = 1.1703$$

Therefore,

$$laborincome_{2009} = (\frac{806.56}{1.1703}) = 689.2$$

Doing the same for 2003, we have that

$$laborincome_{2009} = (\frac{642.65}{1.1874}) = 541.2$$

Therefore, the rise in labor income was

$$\frac{laborincome_{2009} - laborincome_{2003}}{laborincome_{2003}} = \frac{689.2 - 541.2}{541.2} = 27.34\%$$

We have then that the fraction of the growth in labor income in the PIA explained by the rise in occupation is $\frac{6.32}{27.34} = 23.13\%$.

c. Let's remember the following relation

$$totincome = \frac{totincome}{laborincome} \times hourlywage \times educ \times worktime \times \frac{ocup}{PEA} \times \frac{PEA}{PIA} \times \frac{PIA}{pop}$$

Taking logs and taking the derivative with respect to time, we have that the growth rate of total income is the sum of the growth rate of each component above, as we have on the table. Considering a growth rate of 0.5% per year in the PIA, we have that growth rate in total income will be approximately 0.5% higher per year, that is, it will be

$$3.86\% + 0.5\% = 4.36\%$$

d. The impact of the demographic bonus is $\frac{0.5\%}{4.36\%} = 11.47\%$. The impact of the rise in average years of schooling of the occupied is $\frac{2.12\%}{4.36\%} = 48.62\%$. Therefore, the impact of the educational bonus is more than 4 times the impact of the demographic bonus.