Information Theory based Indexes (Theil T and Theil L) 01.17 details from 01.17 to 01.183

Concept of the Theil-T index assess how much a given income distribution (each person receive y_i of total income) is away of a perfect uniform distribution (each person receive 1/n of total income), or the redundancy degree. The Theil L is the other way around. (see 01.172)

$$T = \frac{1}{N}\sum_{i=1}^{N}\frac{x_i}{\mu}\ln\left(\frac{x_i}{\mu}\right) \quad \text{e} \quad L = -\frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{x_i}{\mu}\right),$$

Where xi is individual ***i income***, ***N is*** population size and μ is mean income.

B. Dual: $U_2 = \phi + (1-\phi)U_1$ allows to compare different inequality measures in the same 0 to 1 scale **The Dual of the Gini Index is the Gini Index $G^* = G(1-\%) + \%$, % are** new 0s a way to proceed with maximum inequality (G=1) so is adding top incomes. One can use this formula for introducing both ends of income distribution. As the dual of any inequality measure since its dual transformation measures in the Gini scale. Applying this formula $U_2 = \phi + (1-\phi)U_1$ to the to the Theil –T we get $T2 = T1 - \ln(1-\phi)$. A fully decomposable overall measure of social welfare inspired on Sen (1973) is $SW = mean.(1-U_{T1})$. Since the Theil L does not admit null values, it also does not admit a Dual measure.

D. Intra and Inter Groups Decomposition of Theil T (Theil L allows a similar formula)

$T = T_e + \sum_{h=1}^{K} Y_h T_h$ Where, $T_e = \sum_{h=1}^{k} Y_h \log \frac{Y_h}{\pi_h}$ is the Theil T between groups and $T_h = \sum_{i=1}^{n_h} \frac{y_{hi}}{Y_h} \log n_h \frac{y_{hi}}{Y_h}$ is the Theil intra groups. Therefore $\sum_{h=1}^{K} Y_h T_h$ is the weighted average of intra-groups Theil Ts. **Te / T is the Contribution of a certain characteristic to inequality (say how much schooling (or gender) explains exactly total inequality?). Alternative to mincerian regressions based decompositions.**

$$\boxed{\text{Theil Between groups}} \longleftarrow \boxed{\text{Theil Total}} \longrightarrow \boxed{\text{Theil Within groups}}$$

Other application: Does per capita Household Income underestimates true inequality?

Theil Between Families (per capita)	← Theil Total →	Theil Within Families

***Applying Decomposition to Inequality & Temporal Variability (Mobility, Risk or measurement error)**

Dispersion of Mean Income (earned over time) between people	← Theil Total →	Mean variability of income in relation to the mean across time for each person

We have used the micro-longitudinal aspect of PME/IBGE to the Real Plan Stabilization. The main result here is that the fall of month-to-month inequality measures observed after the fall of inflation in 1994 drastically overestimates the fall of inequality when one compares it with mean earnings over four months. The greater fall of traditional inequality measures on a monthly basis in comparison with measures on a four-month basis is explained by the fall of the individual volatility measures following the sharp decline in inflation rates observed in this period. In sum, stabilization produced more stable earnings trajectories (i.e., lower temporal inequality (in fact, volatility) of individual earnings). On the other hand, the observed fall of inequality *stricto sensu* was much smaller than inequality measures based on monthly measures would have suggested. In sum, the post-stabilization fall in inequality for the group of population is much higher on a monthly basis (as traditionally used in Brazil) than when one uses mean earnings over four months. The fall of Theils is around 4 times higher when one uses the former concept.

Still another application: Understanding Inequality of Opportunities in Brazil

Drawing on the distinction between variables of 'circumstance' (not in control of the individual) and 'effort' (in control of the individual) in John Roemer's work on equality of opportunity, their approach is to simulate the reduction in earnings inequality which would attain if differences in circumstance variables were eliminated.

The five observed circumstances (father's and mother's education; father's occupation; race; and region of birth) are found to account for between 10% and 37% of the Theil index, when accounting for possible biases. Parental education is the most important circumstance affecting earnings, but the occupation of the father and race also play a role. On average, some 60% of the effect of these circumstances operates directly through earnings, while the remaining 40% or so operate by affecting the level of efforts expended by individuals. The decomposition is applied to the distribution of male earnings in urban Brazil in 1996.

J-divergence (see ***01.19**) measure is the simple sum of Theil T and Theil L indexes **(J = *T+L*):**

$$J = \frac{1}{N\mu}\sum_{i=1}^{N}(x_i - \mu)\ln\left(\frac{x_i}{\mu}\right).$$

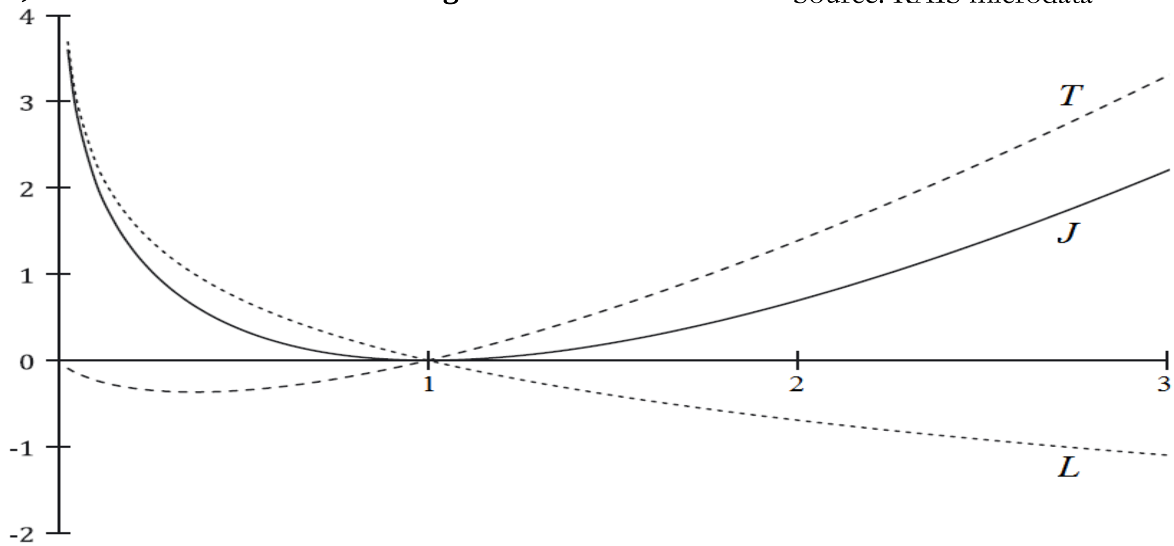
J-Divergence can also be expressed in terms of its within and between groups components, in terms of the sum of Theil-T and Theil-L respective components:

$$\mathrm{J} = \mathrm{T} + \mathrm{L} = \mathrm{T}e + \mathrm{L}e + \sum_{\mathrm{h}=1}^{k} \mathrm{Y}h\,\mathrm{T}h + \sum_{\mathrm{h}=1}^{k} \pi h\,\mathrm{L}h$$

$Gross\ Contribution\ Level\ J = J_{et}/\ J_t$ & $Change\ = \Delta(J_{et})/\ \Delta(J_t)$

	Gross	Contribution to Inequality J-Divergence					
	Education*	Gender	Age	Firm size	Sector	Type of Firm	Individual Firm
Levell 2015	**32,81%**	**0,96%**	**10,82%**	**13,62%**	**8,63%**	**8,15%**	**64,7%**
Change 2001 to 2015	**33,33%**	**0,82%**	**10,82%**	**7,65%**	**13,47%**	**-2,61%**	**75,86%**

***T*, *L* & *J* Contributions according to income level** Source: RAIS microdata



Key point: J-divergence implies in shares (Always non negative) for each income-bracket and individuals. It allows to move from variables (education for example) to specific groups (higher education, top 1% incomes, etc).

	Gross	Contribution to Inequality J-Divergence			CATEGORIES
Level 2015	**Education: Completed College**	**Firm Size: >1000 Employees**	**Income Brackets: Top 0,1%**	**Income Brackets: Top 1%**	**Income Brackets:Top 5%**
	48,7%	**34,8%**	**7,13%**	**27,57%**	**52,2%**

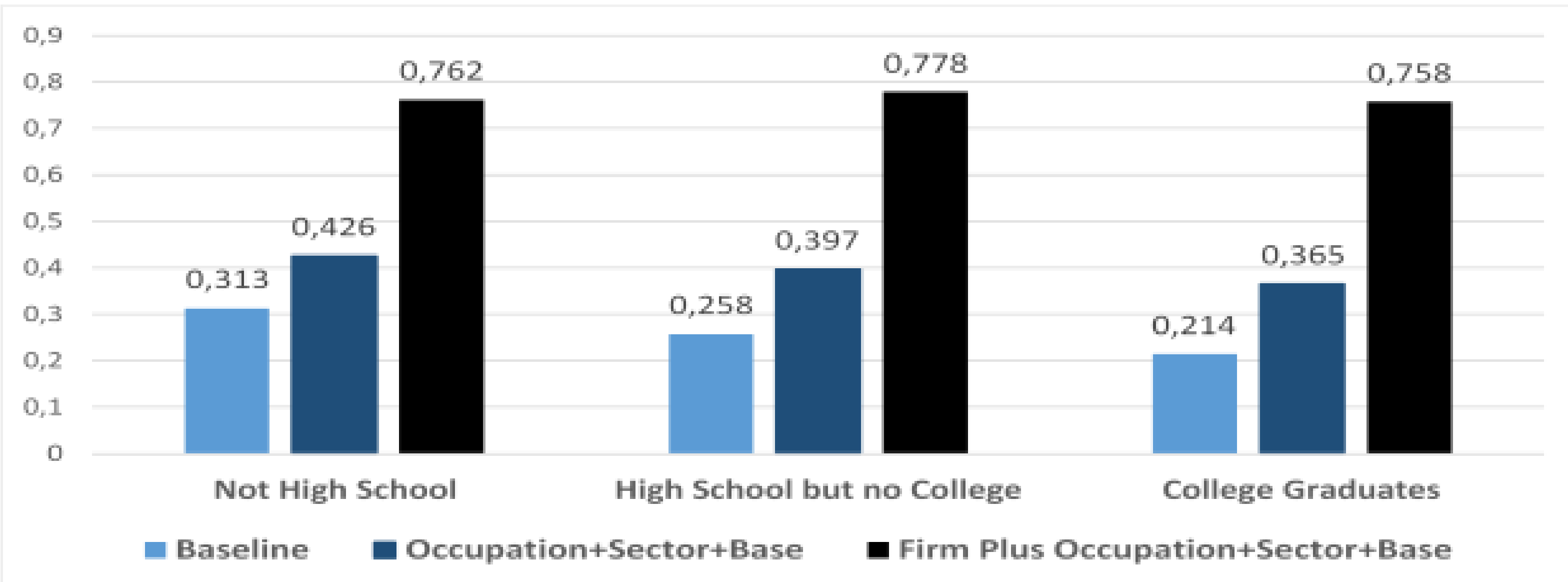
Source: RAIS microdata

Mincerian Model: (Mincer 1974; Lemieux 2006, Card 2001) *01.20

$$y_i = \ln(Y_i) = \alpha + \beta S_i + x_i'\gamma + \varepsilon_i$$

where Y_i is the labour income of individual i (we change this metric below) , S_i is the level of education of individual i measured by years of schooling, x_i is a vector of controls and ε_i is an error term.

Net Contribution to Income Inequality (%) * % Difference of R^2 without specific Variable wrt full regression R^2



Demographics

Source: Rais microdata 1994 to 2015

Mincerian Regression: **Individual earnings inequality within educational groups**
How much do variables explain? Variance of Logs – share of inequality explained
Firms fixed effects are key!