Decomposition of the Gini coefficient using Stata

Alejandro López-Feldman
Universidad de Guanajuato
(Moving to CIDE in June)
Motivation

• The Gini coefficient is widely used to measure inequality in the distribution of income, wealth, expenditures, etc.

• By decomposing this measure you can better understand the determinants of inequality.

• Today I will show the use of a user written command `descogini` to decompose the Gini by income source.
Lerman and Yitzhaki (1985) show that the Gini coefficient for total income, $G$, can be represented as:

$$G = \sum_{k=1}^{K} S_k G_k R_k$$

- $S_k = \text{share of source } k \text{ in total income}$
- $G_k = \text{the source Gini}$
- $R_k = \text{Gini correlation of income from source } k \text{ with the distribution of total income}$
Background

• Stark, Taylor and Yitzhaki (1986) note that the influence of any income component upon total income inequality depends on:
  • how important the income source is with respect to total income;
  • how equally or unequally distributed the income source is; and
  • how the income source and the distribution of total income are correlated.

\[ G = \sum_{k=1}^{K} S_k G_k R_k \]
By using this method of Gini decomposition you can estimate the effect that a 1% change in income from source $k$ will have on total income inequality. This effect is given by:

$$\frac{S_k G_k R_k}{G} - S_k$$
Program syntax

• Syntax
   descogini varlist [ if ] [ in ] [, d(#) bar]

• Options
   d(#) allows the user to specify the number of decimal places to be reported in the table of results. The default is d(4)

   bar specifies that a vertical bar should be included before each output field.
Program syntax

• Remarks
  The first variable in varlist should always be the variable that captures total income. The order in which the income sources are included in the syntax does not matter.
Examples

• Example 1
  • Use data from Chiapas Mexico.
  • 3 income sources (family income, remittances and government transfers).
descogini output

```
: descogini income fam_inc rem_inc trans_inc, d(3)

Gini Decomposition by Income Source

Total Income Variable: income

<table>
<thead>
<tr>
<th>Source</th>
<th>Sk</th>
<th>Gk</th>
<th>Rk</th>
<th>Share</th>
<th>% Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>fam_inc</td>
<td>0.744</td>
<td>0.433</td>
<td>0.960</td>
<td>0.893</td>
<td>0.149</td>
</tr>
<tr>
<td>rem_inc</td>
<td>0.011</td>
<td>0.981</td>
<td>0.202</td>
<td>0.006</td>
<td>-0.005</td>
</tr>
<tr>
<td>trans_inc</td>
<td>0.245</td>
<td>0.341</td>
<td>0.418</td>
<td>0.101</td>
<td>-0.144</td>
</tr>
<tr>
<td>Total income</td>
<td>0.245</td>
<td>0.346</td>
<td>0.418</td>
<td>0.101</td>
<td>-0.144</td>
</tr>
</tbody>
</table>
```
Examples

• Example 2
  • Use data from rural Mexico (ENHRUM).
  • 3 income sources (family income, remittances and government transfers).
Saved Results

• The % changes are saved as a vector $e(b)$

• The estimates of $S_k, G_k, R_k, G$ are available as scalars.

• bootstrap can be easily applied to obtain standard errors and confidence intervals for any of these estimates.
bootstrap and descogini
References


Gracias!

lopezfeldman@gmail.com