

GiniDecompLY

Abdessamad Ait Mbarek

Jan, 2024

GiniDecompLY R package provides a set of functions to estimate the effect of each income source on income inequalities based on the decomposition of Lerman and Yitzhaki (1985) [doi:10.2307/1928447](https://doi.org/10.2307/1928447). This R package offers tidy output, presenting results in tidy tibbles, making it easy for users to explore and utilize the results for further analysis and visualization.

You can access the package page on [CRAN](#) and you can install it using :

```
install.packages("GiniDecompLY")
```

The functions include :

- Calculating the Gini correlation index between two distributions.
- Estimation of the contribution of each income source on the overall income inequality.
- Computation of elasticity of Gini index associated with change in income sources.
- Growth-redistribution decomposition of the effects of income sources on social welfare function as defined by Amartya Sen (1970)

A sample data on income sources has been provided in the package :

```
dplyr::glimpse(sample_income_data)
```

Rows: 200

Columns: 6

[illegible]

The sample data contains 200 observations representing households and 6 columns :

- `region` column categorizes households based on whether they reside in urban or rural areas.
- `sample_wgt` column contains the sampling weights assigned to each observation.
- `wage`, `self_employment_rev`, `farming_rev`, and `other_rev` represent different sources of income for the household. Specifically, `wage` refers to wages and salaries earned, `self_employment_rev` denotes revenue from independent activities, `farming_rev` signifies income from farming, and `other_rev` encompasses income from other sources.

Gini Correlation Index

Gini correlation was first introduced in Schechtman and Yitzhaki (1987). It is a non-symmetric measure and can take the following two forms, depending on which variable is given in its actual values and which one is ranked:

If X and Y are random variables from continuous distribution functions F_X and G_Y respectively

$$\Gamma(X, Y) = \frac{Cov(X, G_Y(Y))}{Cov(X, F_X(X))}$$

and similarly,

$$\Gamma(Y, X) = \frac{Cov(Y, F_X(X))}{Cov(Y, G_Y(Y))}$$

The range of the gini correlation coefficient is $[-1, 1]$

`gini_corr` function takes as first argument the variable in its actual values, and as second argument the variable of the distribution with the rank information. An optional argument for sampling weights is also available.

If we want to calculate the gini correlation coefficient of salaries distribution ranked with the total income distribution, then we can call the function as follow:

```
# Calculate the gini correlation between the salary and total income distributions

Salary_distribution = sample_income_data$wage
Total_income_distribution = rowSums(sample_income_data[3:6])

gini_corr(Salary_distribution, Total_income_distribution)
```

```
[1] 0.6307609
```

Gini decomposition by income sources

The Lerman and Yitzhaki decomposition method aims to understand the contribution of different income sources to overall income inequality.

Suppose the total income of households comes from 4 income sources as described in `sample_income_data` dataset. The decomposition of Lerman and Yitzhaki provide estimation of the contribution of each income source to the income inequality as measured by Gini coefficient.

$$Gini(income) = 0.421$$

The final output of the decomposition is represented by the **Absolute Contribution** for each income source which sum up to the overall $Gini(income)$, and the **Relative Contribution** which sum up to

The steps of estimation include :

- 1- Calculation of the **Share** of each income source to the total income;
- 2- Calculation of **Gini** coefficients for each income source;
- 3- Calculation of **Gini_corr** (gini correlation coefficient) between the distribution each income source and the total income.

Absolute Contribution is calculated as follow:

$$\text{Absolute Contribution} = \text{Share} * \text{Gini} * \text{Gini_corr}$$

and then

$$\text{Relative Contribution} = \text{Absolute Contribution} / Gini(income)$$

`gini_decomp_source` provide a tibble containing all the components described above. It takes as a first argument the `.data` containing the income sources variables. And then we pass the columns names (or positions), separated by commas, indicating income sources.

```
sample_income_data |>
  gini_decomp_source(wage, self_employment_rev, farming_rev, other_rev)
```

The function provides two optional arguments:

- `.wgt` for sampling weights; - and `.by` for results disaggregation, in that case the output will be a grouped tibble.

We can also pass the income sources variables by their positions in the data frame:

```
gini_decomp_source(sample_income_data, 3:6, .by = region, .wgt = sample_wgt)
```

```
# A tibble: 8 x 7
```

```
# Groups:   region [2]
```

	region	income_source	Share	Gini	Gini_corr	Absolute_Contribution	Relative_Contribution
	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Rural	farming_rev	0.412	0.815	0.854	0.287	0.61
2	Rural	other_rev	0.301	0.402	0.395	0.0478	0.10
3	Rural	self_employment_rev	0.0674	0.907	0.471	0.0288	0.06
4	Rural	wage	0.219	0.706	0.488	0.0755	0.16
5	Urban	farming_rev	0.00189	0.989	0.737	0.00138	0.00
6	Urban	other_rev	0.373	0.618	0.634	0.146	0.31
7	Urban	self_employment_rev	0.165	0.827	0.327	0.0446	0.10
8	Urban	wage	0.460	0.648	0.665	0.198	0.50

Gini income elasticity

The decomposition method described above can be represented in the following equation:

$$Gini(Y) = \sum_k S_k \Gamma(Y_k, Y) Gini(Y_k)$$

Where Y_k and S_k are the distribution and the **Share** of income source k , respectively. And Y denotes the distribution of the total income.

Thus, **Absolute Contribution** for the k^{th} income source is the product $S_k \Gamma(Y_k, Y) Gini(Y_k)$

Stark and al. (1986) showed that this decomposition equation offers a simple method to assess the marginal effect on total income inequality following a marginal percentage variation in income from a given source, equal for all households. Thus, the impact of an increase in income from source k , for all households, in such a way that Y_k is multiplied by $(1 + e_k)$, where e_k is sufficiently small can be represented as follow :

$$\frac{\partial Gini(Y)}{\partial e_k} = S_k (\Gamma(Y_k, Y) Gini(Y_k) - Gini(Y))$$

This expression show a measure for the marginal contribution of source k to the overall inequality. It can be expressed differently to reveal the relative marginal contribution of inequality due to a marginal variation in the income from source k :

$$\frac{\partial Gini(Y)/\partial e_k}{Gini(Y)} = S_k \left(\frac{\Gamma(Y_k, Y) Gini(Y_k)}{Gini(Y)} - 1 \right) = S_k (\eta_k - 1)$$

Where

$$\eta_k = \frac{\Gamma(Y_k, Y)Gini(Y_k)}{Gini(Y)}$$

η_k will denote the elasticity of Gini index associated with a percentage change in the mean income (for each income source)

This definition clearly demonstrates that the marginal impact of a source of income depends on its income elasticity of the Gini index. Thus, a percentage increase in the income from a source with a lower η_k (higher) than 1, will decrease (increase), the overall income inequality. When the η_k is close to 1, it means that the variation in this source does not affect overall inequality.

`gini_income_elasticity` function calculate the Elasticity along with the `Marginal_Impact` of a change in the mean of each income source on the overall Gini index.

The function has the same set of arguments as the previous function

```
sample_income_data |>
  gini_income_elasticity(wage, self_employment_rev, farming_rev, other_rev,
    .by = region,
    .wgt = sample_wgt)
```

```
# A tibble: 8 x 7
# Groups:   region [2]
  region income_source      Share  Gini Gini_corr Elasticity Marginal_Impact
  <chr>   <chr>          <dbl> <dbl>   <dbl>      <dbl>      <dbl>
1 Rural  farming_rev      0.412  0.815   0.854      1.59      0.241
2 Rural  other_rev        0.301  0.402   0.395      0.362     -0.192
3 Rural  self_employment_rev 0.0674 0.907   0.471      0.973     -0.00182
4 Rural  wage            0.219  0.706   0.488      0.784     -0.0474
5 Urban  farming_rev      0.00189 0.989   0.737      1.87      0.00164
6 Urban  other_rev        0.373  0.618   0.634      1.00      0.000976
7 Urban  self_employment_rev 0.165  0.827   0.327      0.693     -0.0506
8 Urban  wage            0.460  0.648   0.665      1.10      0.0480
```

Growth-redistribution impacts on social welfare function

In terms of social welfare, if individuals or households assess their level of well-being, on one hand, in absolute terms (that is, the income they have), and, on the other hand, in relative terms (how much they have compared to others), the level of social welfare can be represented as a function combining the level and inequality of well-being (Wodon and Yitzhaki, 2002;

Essama-Nssah, 2000 & Sen, 1997). In other words, such a function combines both the social preference for more income and that for more equality. It is an increasing function of the mean and decreasing function of inequality (Essama-Nssah, 2000).

Sen (1997) showed, under the assumption that individual well-being is approximated by individual income and that the well-being of any pair of individuals equals that of the poorest, then the following expression can be considered as a function of social welfare underlying the Gini coefficient.

$$SW = \bar{Y}(1 - G)$$

Where \bar{Y} denotes average income and G the Gini index.

Using the decomposition of G , the impact of a variation in an income source on the Social Welfare function is estimated according to the following formulation:

$$\frac{\Delta SW}{SW} |_{Y_k} = \frac{S_k}{1 - Gini(Y)} - \frac{S_k \Gamma(Y_k, Y) Gini(Y_k)}{1 - Gini(Y)}$$

`social_welfare_impact` function provides a Growth-redistribution decomposition of the impact of a variation in each income source on the Social Welfare function.

The output of the function is three columns (added to the decomposition components described before)

- `Total_Variation` = $\frac{\Delta SW}{SW} |_{Y_k}$;
- `Growth_Effect` = $\frac{S_k}{1 - Gini(Y)}$;
- `Redistribution_Effect` = $-\frac{S_k \Gamma(Y_k, Y) Gini(Y_k)}{1 - Gini(Y)}$

```
sample_income_data |>
  social_welfare_impact(wage, self_employment_rev, farming_rev, other_rev,
    .wgt = sample_wgt)
```

A tibble: 4 x 7

	income_source	Share	Gini	Gini_corr	Growth_Effect	Redistribution_Effect	Total_Variation
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	farming_rev	0.125	0.926	0.620	0.215	-0.123	0.092
2	other_rev	0.352	0.574	0.600	0.607	-0.209	0.398
3	self_employment_rev	0.136	0.864	0.424	0.234	-0.0858	0.148
4	wage	0.388	0.697	0.659	0.669	-0.308	0.361

References

- Essama-Nssah B. (2000), *Inégalité, pauvreté et bien-être social, fondements analytiques et normatifs*, De Boeck Université, Bruxelles
- Handcock, M. (2016), *Relative Distribution Methods in the Social Sciences*, Springer-Verlag, Inc., New York, 1999 ISBN 0387987789
- Lerman, R. I., & Yitzhaki, S. (1985). *Income Inequality Effects by Income Source: A New Approach and Applications to the United States*. The Review of Economics and Statistics, 67(1), 151–156. <https://doi.org/10.2307/1928447>
- Rawls J. (1971), *A Theory of Justice*, The Belknap Press of Harvard Univers
- Schechtman, E., and Yitzhaki, S. (1987). *A Measure of Association Based on Gini's Mean Difference*. *Communications in Statistics: Theory and Methods* A16:207–31.
- Schechtman, E., and Yitzhaki, S. (1999) *On the proper bounds of the Gini correlation*, Economics Letters, Volume 63, Issue 2, p. 133-138, ISSN 0165-1765
- Sen A. (1970) *Collective Choice and Social Welfare*, Holden-Day, ISBN:978-0-444-85127-7
- Soudi K. (20)
- Stark, O., J. Taylor, and S. Yitzhaki. (1986). *Remittances and Inequality*. Economic Journal 96(383):722–40.
- Wodon, Quentin and Yitzhaki, Shlomo (2002): *Inequality and Social Welfare*. Published in: A Source book for Poverty Reduction Strategies , Vol. 1, (April 2002): pp. 75-104.