Variations on Gini

Philippe Van Kerm

University of Luxembourg & Luxembourg Institute of Socio-Economic Research

2021 Belgian Stata Conference

e-Antwerp, June 7 2021

Applications of Gini and concentration coefficients in Social Sciences

- Income and wealth inequality
- Health inequality and income gradients in health
- Tax progressivity
- Poverty
- Income mobility
- Polarization
- Targeting and predictive performance (e.g., in credit scoring, in some Machine Learning models)
- Robust regression analysis
- ...

No shortage of material... (`findit gini`)

inequal7 igini1 lorenz igini

yadap progres

svylorenz dasp fastgini descogini

inequal

sgini giniinc survgini

egen_inequal ineqdecgini moremata

adgini ineqdeco conindex

anogi dsginideco inequaly bipolar

somersd ginireg

... and a very selective talk

http://www.vankerm.net/stata/manuals/sgini.pdf

(The benefit of “maturity”?)

```
*! v2.0.0, 2020-04-21, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   implementation of alternative calculation of ranks -- speeding up calculations
*   implementation of genfracrankvar() (was not previously activated)
* v1.1.5, 2014-04-29, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   minor bug fix (r(coeffs) size without source)
* v1.1.4, 2011-05-10, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   modify r(coeff) in the -sourcedeocmposition-
*   edit out put column label
*   add return matrices with contribution and relative contributions
* v1.1.3, 2010-05-20, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   add option welfare (synonymous to aggregate -- for backward compatibility)
* v1.1.2, 2010-03-12, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   add saved results for total Gini when using the sourcedecomp option
* v1.1.1, 2010-03-09, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   Change default format
* v1.1.0, 2010-02-05, Philippe Van Kerm, Generalized Gini/concentration coefficients
*   Add -fracrankvar- option and allow time-series operators in -varlist- or -sortvar-
*   computations for -sourcedecomposition- speed up considerably with use of -genp()-/-pvar()-
* v1.0.2, 2009-09-29, Philippe Van Kerm, Generalized Gini/concentration coefficients
* v1.0.1, 2009-09-11, Philippe Van Kerm, Generalized Gini/concentration coefficients
* v1.0.0, 2007-11-19, Philippe Van Kerm, Generalized Gini/concentration coefficients
* syntax varlist [if] [in] [fweight aweight] [ , Param(real 2.0) Sortvar(varname) SOURCEdecomposition ]
* (this version is based on _sgini.ado * v2.3.0, 2007-02-07)
```

... and a very selective talk

http://www.vankerm.net/stata/manuals/sgini.pdf

Overview

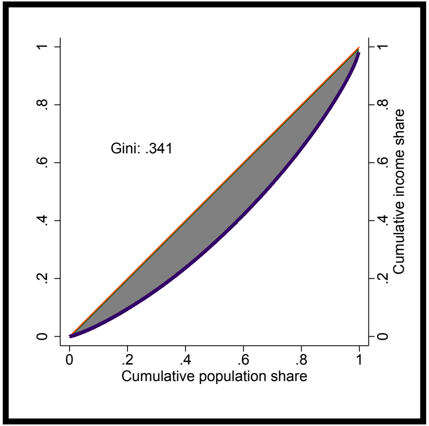
1. Definitions —The Gini coefficient and its nuclear family
2. Estimation
3. `sgini` —A Gini pocket calculator
4. Applications —The extended Gini family

Definitions —The Gini coefficient and its nuclear family

The Gini coefficient and the Lorenz curve

Twice the area between the 45 degree line and the Lorenz curve:

$$\mathrm{GINI}(X) = 1 - 2\int L_X(p)\,dp$$



Main properties

- Pigou-Dalton principle of transfer (transfer from rich to poor reduces inequality) – Lorenz consistency
- Scale invariant – homogenous of degree zero
- Population and permutation invariant
- Ranges between 0 (min inequality) and 1 (max inequality)
- Practically relatively robust to outliers
- Defined in presence of non-positive Y (but no more [0,1], nor PD consistent)

Many formulations ("More than a dozen ways to spell Gini", Yitzhaki, 1998)

Gini's mean difference:

Average of all pairwise absolute differences

$$\mathrm{GINI}(X) = \frac{1}{2N^2\mu} \sum_{i=1}^{N} \sum_{j=1}^{N} |x_i - x_j|$$

LISER LUXEMBOURG INSTITUTE OF SOCIO-ECONOMIC RESEARCH
uni.lu UNIVERSITÉ DU LUXEMBOURG

Many formulations ("More than a dozen ways to spell Gini", Yitzhaki, 1998)

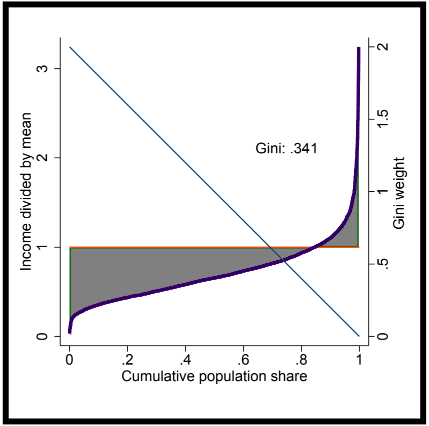
Gini weighted mean:

$$\text{GINI}(X) = 1 - \int 2(1-p)\frac{x(p)}{\mu}dp$$

where $x(p)$ is the quantile function

One minus weighted average of X with weight linear in rank

Leads to simple covariance expression:
$\text{GINI}(X) = -2 \ \text{Cov}\left(\frac{X}{\mu}, \ (1-F(X))\right)$



Single-parameter generalizations and linear inequality measures

A generalized Gini coefficient (a.k.a. the S-Gini, or extended Gini coefficient)

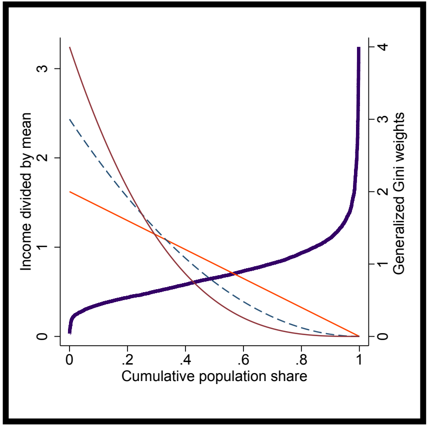
$$\mathrm{GINI}(X;\upsilon) = 1 - \int w(p;\upsilon)\frac{x(p)}{\mu}dp$$

Weighted average of X with weight non-linear in rank

$$w(p;\upsilon) = \upsilon(1-p)^{\upsilon-1}$$

The standard Gini corresponds to $\upsilon = 2$.

(Donaldson and Weymark, 1980, 1983, Yitzhaki, 1983)



Single-parameter generalizations and linear inequality measures

A generalized Gini coefficient (a.k.a. the S-Gini, or extended Gini coefficient)

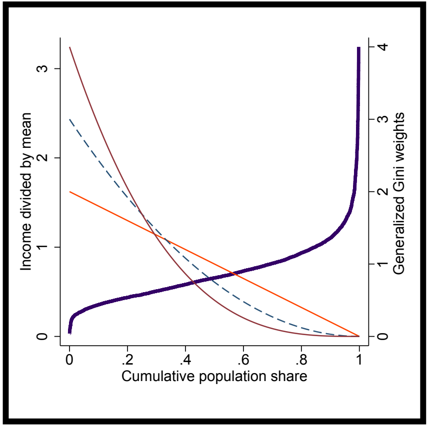
$$\text{GINI}(X;\upsilon) = 1 - \int w(p;\upsilon)\frac{x(p)}{\mu}dp$$

Weighted average of X with weight non-linear in rank

$$w(p;\upsilon) = \upsilon(1-p)^{\upsilon-1}$$

The standard Gini corresponds to $\upsilon = 2$.

(Donaldson and Weymark, 1980, 1983, Yitzhaki, 1983)



Concentration coefficient

The Concentration coefficient measures the association between *two* random variables.

- Weighted Gini means

$$\mathrm{CONC}(X, Y; \upsilon) = 1 - \frac{1}{N} \sum_i w(G(y_i); \upsilon) \frac{x_i}{\mu(X)}$$

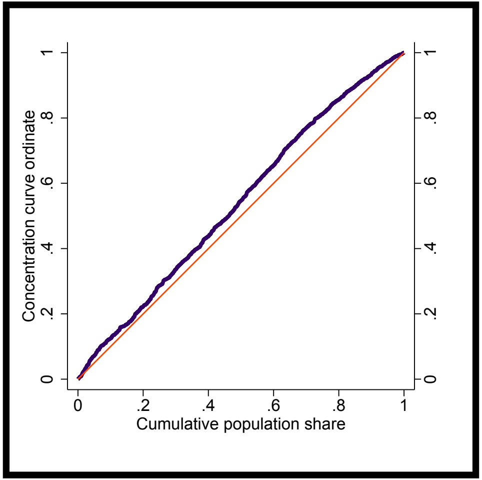
- Covariance

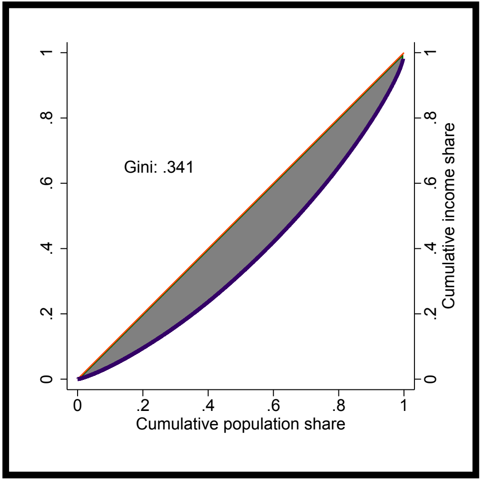
$$\mathrm{CONC}(X, Y; \upsilon) = -\upsilon \ \ \mathrm{Cov}\left(\frac{X}{\mu(X)}, \ (1 - G(Y)^{\upsilon - 1}) \right)$$

where $G(Y)$ is the cumulative distribution function of Y.

$\mathrm{CONC}(X, Y; \upsilon)$ reflects how much X is concentrated on observations with high ranks in Y (see, e.g., Kakwani, 1977*a*).

The Gini coefficient and the Lorenz curve





Extensions: translation invariance and degree 1 homogeneity

- Gini means – degree 1 homogeneity (Gini as 'cost of inequality'):

$$\begin{aligned} \mathrm{AGGCONC}(X, Y; \upsilon) &= \int w(p;\upsilon) Q(p) dp \\ &= \mu(X) \ (1 - \mathrm{CONC}(X, Y; \upsilon)) \end{aligned}$$

- Translation invariant measure:

$$\begin{aligned} \mathrm{ABSCONC}(X, Y; \upsilon) &= \mu(X) - \mathrm{AGGCONC}(X, Y; \upsilon) \\ &= \mu(X) \ \mathrm{CONC}(X, Y; \upsilon). \end{aligned}$$

- Scale invariant measure:

$$\begin{aligned} \mathrm{CONC}(X, Y; \upsilon) &= 1 - \frac{\mathrm{AGGCONC}(X, Y; \upsilon)}{\mu(X)} \\ &= \frac{\mathrm{ABSCONC}(X, Y; \upsilon)}{\mu(X)} \end{aligned}$$

Extensions: translation invariance and degree 1 homogeneity

- Gini means – degree 1 homogeneity (Gini as 'cost of inequality'):

$$\begin{aligned}\mathrm{AGGCONC}(X,Y;\upsilon) &= \int w(p;\upsilon)Q(p)\,dp \\ &= \mu(X)\ \left(1-\mathrm{CONC}(X,Y;\upsilon)\right)\end{aligned}$$

- Translation invariant measure:

$$\begin{aligned}\mathrm{ABSCONC}(X,Y;\upsilon) &= \mu(X)-\mathrm{AGGCONC}(X,Y;\upsilon) \\ &= \mu(X)\ \mathrm{CONC}(X,Y;\upsilon).\end{aligned}$$

- Scale invariant measure:

$$\begin{aligned}\mathrm{CONC}(X,Y;\upsilon) &= 1-\frac{\mathrm{AGGCONC}(X,Y;\upsilon)}{\mu(X)} \\ &= \frac{\mathrm{ABSCONC}(X,Y;\upsilon)}{\mu(X)}\end{aligned}$$

Extensions: translation invariance and degree 1 homogeneity

- Gini means – degree 1 homogeneity (Gini as 'cost of inequality'):

$$
\begin{aligned}
\mathrm{AGGCONC}(X, Y; \upsilon) &= \int w(p;\upsilon) Q(p) dp \\
&= \mu(X) \ \left(1 - \mathrm{CONC}(X, Y; \upsilon)\right)
\end{aligned}
$$

- Translation invariant measure:

$$
\begin{aligned}
\mathrm{ABSCONC}(X, Y; \upsilon) &= \mu(X) - \mathrm{AGGCONC}(X, Y; \upsilon) \\
&= \mu(X) \ \mathrm{CONC}(X, Y; \upsilon).
\end{aligned}
$$

- Scale invariant measure:

$$
\begin{aligned}
\mathrm{CONC}(X, Y; \upsilon) &= 1 - \frac{\mathrm{AGGCONC}(X, Y; \upsilon)}{\mu(X)} \\
&= \frac{\mathrm{ABSCONC}(X, Y; \upsilon)}{\mu(X)}
\end{aligned}
$$

LISER Luxembourg Institute of Socio-Economic Research

uni.lu Université du Luxembourg

Extensions: translation invariance and degree 1 homogeneity

- Gini means – degree 1 homogeneity (Gini as 'cost of inequality'):

$$\begin{aligned} \mathrm{AGGCONC}(X, Y; \upsilon) &= \int w(p; \upsilon) Q(p) dp \\ &= \mu(X) \; (1 - \mathrm{CONC}(X, Y; \upsilon)) \end{aligned}$$

- Translation invariant measure:

$$\begin{aligned} \mathrm{ABSCONC}(X, Y; \upsilon) &= \mu(X) - \mathrm{AGGCONC}(X, Y; \upsilon) \\ &= \mu(X) \; \mathrm{CONC}(X, Y; \upsilon). \end{aligned}$$

- Scale invariant measure:

$$\begin{aligned} \mathrm{CONC}(X, Y; \upsilon) &= 1 - \frac{\mathrm{AGGCONC}(X, Y; \upsilon)}{\mu(X)} \\ &= \frac{\mathrm{ABSCONC}(X, Y; \upsilon)}{\mu(X)} \end{aligned}$$

Gini correlation

Concentration coefficients capture the association between two random variables and leads to measures of 'Gini correlations' (Schechtman and Yitzhaki, 1987, 1999):

$$R(X, Y; \upsilon) = \frac{\mathrm{Cov}\left(X,\ (1 - G(Y))^{\upsilon - 1}\right)}{\mathrm{Cov}\left(X,\ (1 - F(X))^{\upsilon - 1}\right)} = \frac{\mathrm{CONC}(X, Y; \upsilon)}{\mathrm{GINI}(X; \upsilon)}.$$

Mixture of Pearson's and Spearman's correlations! (Equal to Somers' D if X is bivariate.)

The Gini nuclear family

- Gini coefficient
- Single-parameter extensions
- Concentration coefficient
- Gini correlation
- Gini means
- Absolute Gini coefficients

... ingredients of many more dishes

Estimation

Point estimation

Covariance-based expressions for the (generalized) Gini and Concentration coefficients are convenient for calculations from unit-record data.

$$\hat{\text{Cov}}(x,y) = \left(\frac{\sum_{i=1}^{N} w_i}{(\sum_{i=1}^{N} w_i)^2 - \sum_{i=1}^{N} w_i^2} \right) \sum_{i=1}^{N} w_i (x_i - \mu_x)(y_i - \mu_y)$$

so

$$\hat{\text{Gini}} = -2 \frac{\hat{\text{Cov}}(y,r)}{\hat{\mu}}$$

The only point of importance is the calculation of ranks – esp. in the presence of ties (ordinal data)

Fractional ranks with ties and/or weights

N observations on variable Y with associated sampling weights: $\{(y_i, w_i)\}_{i=1}^N$.

Let K distinct values observed on Y, denoted $y_1^* < y_2^* < \ldots < y_K^*$, and denote by π_k^* the corresponding weighted sample proportions:

$$\pi_k^* = \frac{\sum_{i=1}^N w_i \mathbf{1}(y_i = y_k^*)}{\sum_{i=1}^N w_i}$$

($\mathbf{1}(\mathit{condition})$ is 1 if $\mathit{condition}$ is true, 0 otherwise). (If all observations in Y are distinct and no sample weight are used, $\pi_k^* = 1/N$.)

The fractional rank attached to each y_k^* is then given by

$$F_k^* = \sum_{j=0}^{k-1} \pi_j^* + 0.5\pi_{j+1}^*$$

where $\pi_0^* = 0$

Fractional ranks with ties and/or weights (ctd.)

Each observation is then given the fractional rank

$$F_i = \sum_{k=1}^{K} F_k^* \mathbf{1}(y_i = y_k^*).$$

- Tied observations are associated to identical fractional ranks (no dependence on data order)
- The sample mean of the fractional ranks is equal to 0.5 (irrespective of sample size)

$\Longrightarrow$ Needed to guarantee population invariance and anonymity

$\{(F_i, y_i, w_i)\}_{i=1}^{N}$ is then plugged in covariance formula.

(See Yitzhaki and Schechtman (2005), Berger (2008), Chen and Roy (2009), and Davidson (2009).)

LISER LUXEMBOURG INSTITUTE OF SOCIO-ECONOMIC RESEARCH
uni.lu UNIVERSITÉ DU LUXEMBOURG

Two main approaches to variance estimation

Two main approaches for variance estimation, construction of confidence intervals, tests

- analytic, linearization approaches
- empirical, resampling-based approaches (jackknife and bootstrap)

Variance estimation with the influence function

An asymptotic approximation of the variance of θ is given by (Hampel, 1974)

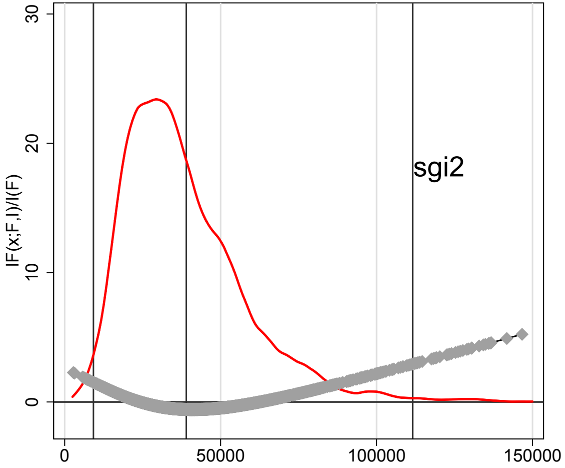
$$V(\theta) \approx \int \mathrm{IF}(y; \theta, F)^2 dF(y)$$

where IF is the influence function.

The IF for the Gini and concentration coefficients are relatively lengthy expressions (because of sampling variability of estimated ranks) but otherwise simple approach (and valid for complex survey design) (Van Kerm, 2015, 2017)

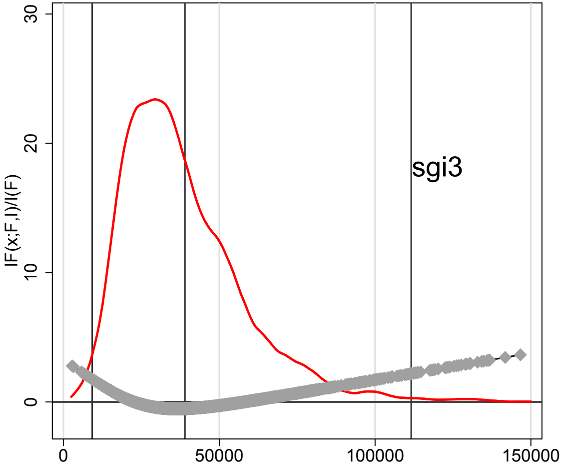
The influence function of Gini coefficients

The IF illustrates the relative robustness of Gini indices to extremes



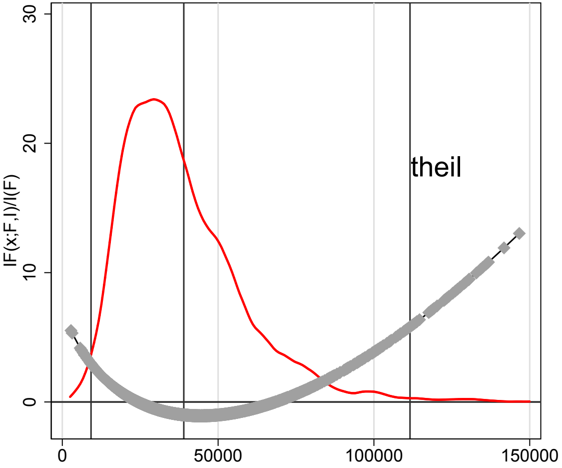
The influence function of Gini coefficients

The IF illustrates the relative robustness of Gini indices to extremes



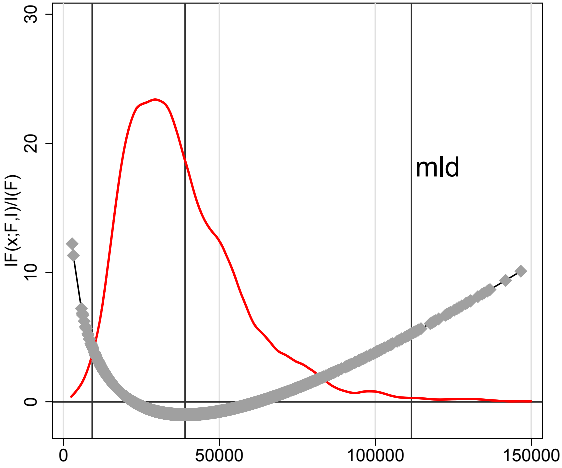
The influence function of Gini coefficients

The IF illustrates the relative robustness of Gini indices to extremes



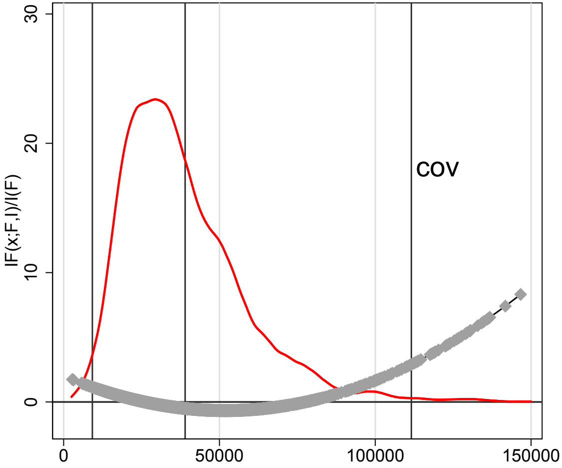
The influence function of Gini coefficients

The IF illustrates the relative robustness of Gini indices to extremes



The influence function of Gini coefficients

The IF illustrates the relative robustness of Gini indices to extremes



`sgini` —A Gini pocket calculator

sgini package

sgini is a small no-frills command for calculating Gini and the nuclear family

Syntax

sgini *varlist* [*if*] [*in*] [*weight*] [, parameters(*numlist*) sortvar(*varname*) fracrankvar(*varname*) sourcedecomposition absolute aggregate welfare format(*%fmt*)]

It has been optimized to be fast (also see fastgini). Point estimation but easily bootstrapped or jackknifed.[1]

[1] See, e.g., net install yadap , from(http://www.vankerm.net/stata for Gini coeffs with linearized variance (Van Kerm, 2017).

An illustrative example

```
. use http://www.stata-press.com/data/r9/nlswork , clear
(National Longitudinal Survey.  Young Women 14-26 years of age in 1968)

. xtset idcode year
       panel variable:  idcode (unbalanced)
        time variable:  year, 68 to 88, but with gaps
                delta:  1 unit

. gen w = exp(ln_wage)

. ssc install sgini
```

```
. sgini w
```

Gini coefficient for w

Variable	v=2
w	0.2732

```
. sgini w , parameters(1.5(.5)4) absolute
```

Generalized Gini coefficient for w

Variable	v=1.5	v=2	v=2.5	v=3	v=3.5	v=4
w	1.1207	1.6522	1.9812	2.2113	2.3844	2.5213

```
. sgini w L.w L2.w  , sortvar(w) param(2 3)

Generalized Concentration coefficient for w, L.w, L2.w against w

────────────────────────────────
    Variable │       v=2       v=3
─────────────┼──────────────────
           w │    0.2023    0.2836
         L.w │    0.1581    0.2198
        L2.w │    0.1387    0.1921
────────────────────────────────

. return list

scalars:
              r(sum_w) =  3481
                  r(N) =  3481
              r(coeff) =  .2023171625445915
```

(output omitted)

```
. matrix list r(coeffs)

r(coeffs)[1,6]
          param1:     param1:     param1:     param2:     param2:     param2:
                           L.         L2.                      L.         L2.
               w           w           w           w           w           w
Coeff  .20231716   .1581236  .13867147  .28359513  .21983376  .19212248
```

```
. bootstrap G=r(coeff) , reps(250) nodots : sgini w if !mi(w) & year==88

Warning:  Because sgini is not an estimation command or does not set e(sample),
          bootstrap has no way to determine which observations are used in calculating
          the statistics and so assumes that all observations are used.  This means that
          no observations will be excluded from the resampling because of missing values
          or other reasons.

          If the assumption is not true, press Break, save the data, and drop the
          observations that are to be excluded.  Be sure that the dataset in memory
          contains only the relevant data.

Bootstrap results                               Number of obs     =      2,272
                                                Replications      =        250

      command:  sgini w
            G:  r(coeff)

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
             |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           G |   .3552972   .0095474    37.21   0.000     .3365847    .3740097
------------------------------------------------------------------------------

. jackknife G=r(coeff) , rclass nodots: sgini w if !mi(w) & year==88
```

Two companion commands

Two companion commands

- `fracrank` for generating fractional ranks
- `sginicorr` for calculating generalized Gini *correlation* coefficients.

Applications —The extended Gini family

Applications of Gini and concentration coefficients

Variations on Gini and concentration coefficient are used in numerous areas (sometimes under different names)

- Health inequality and income gradients in health
- Factor decompositions
- Tax progressivity
- Predictive performance (e.g., in credit scoring) –cf.ROC curves
- Income mobility
- Poverty
- Polarization
- Gini Regression analysis
- ...

Applications of Gini and concentration coefficients

Variations on Gini and concentration coefficient are used in numerous areas (sometimes under different names)

- Health inequality and income gradients in health
- Factor decompositions
- Tax progressivity
- Predictive performance (e.g., in credit scoring) –cf.ROC curves
- Income mobility
- Poverty
- Polarization
- Gini Regression analysis
- ...

Factor decomposition

Total family income as sum of factors: earnings, capital income, transfer income, etc.

$$\text{GINI}(Y;\upsilon) = \sum_{k=1}^{K} \frac{\mu(Y^k)}{\mu(Y)} \times \text{CONC}(Y^k, Y;\upsilon)$$

where $\text{CONC}(Y^k, Y;\upsilon)$ is the (generalized) CC of factor k against total income (Fei et al., 1978, Lerman and Yitzhaki, 1985).

Furthermore,

$$\text{CONC}(Y^k, Y;\upsilon) = \text{GINI}(Y^k;\upsilon) \times R(Y^k, Y;\upsilon)$$

where $R(Y^k, Y;\upsilon)$ is the Gini correlation (Lerman and Yitzhaki, 1985, López-Feldman, 2006)

```
. sgini w L.w L2.w  if year==73 , sourcedecomposition
```

Gini coefficient for w, L.w, L2.w

Variable	v=2
w	0.2150
L.w	0.2077
L2.w	0.2043

Decomposition by source:
  TOTAL = w + L.w + L2.w

Parameter: v=2

Variable	Share s	Coeff. g	Corr. r	Conc. c=g*r	Contri. s*g*r	%Contri. s*g*r/G	Elasticity s*g*r/G-s
w	0.3492	0.2150	0.9427	0.2027	0.0708	0.3634	0.0142
L.w	0.3350	0.2077	0.9521	0.1978	0.0663	0.3402	0.0052
L2.w	0.3158	0.2043	0.8950	0.1828	0.0577	0.2964	-0.0194
TOTAL	1.0000	0.1948	1.0000	0.1948	0.1948	1.0000	0.0000

Tax progressivity and horizontal equity

How 'progressive' is a tax schedule? How much inequality is reduced after application of the tax?

$$\Pi^{RS} = GINI(X^{pre}) - GINI(X^{post})$$

where X^{pre} and X^{post} are pre- and post-tax incomes.

The Kakwani measure of progressivity (Kakwani, 1977*b*):

$$\Pi^{K} = CONC(T, X^{pre}) - GINI(X^{pre})$$

where T is the tax paid: $T = X^{pre} - X^{post}$.

Tax progressivity and horizontal equity

How ‘progressive’ is a tax schedule? How much inequality is reduced after application of the tax?

$$\Pi^{RS} = \text{GINI}(X^{\text{pre}}) - \text{GINI}(X^{\text{post}})$$

where X^{pre} and X^{post} are pre- and post-tax incomes.

The Kakwani measure of progressivity (Kakwani, 1977*b*):

$$\Pi^{K} = \text{CONC}(T, X^{\text{pre}}) - \text{GINI}(X^{\text{pre}})$$

where T is the tax paid: $T = X^{\text{pre}} - X^{\text{post}}$.

Tax progressivity and horizontal equity

Combining the progressivity measure with a component capturing the re-ranking induced by the tax schedule leads to a decomposition of Π^{RS} as

$$\Pi^{RS} = \frac{g}{1-g}\Pi^{K} - R$$

where $R = (\mathrm{CONC}(X^{\mathrm{post}}, X^{\mathrm{pre}}) - \mathrm{GINI}(X^{\mathrm{post}}))$ captures the effect of re-ranking on the net reduction in the Gini coefficient, and g is the average tax rate.[2]

[2] `progres` available on SSC (Peichl and Van Kerm, 2007).

Income mobility and pro-poor growth

Jenkins and Van Kerm (2006) relate the change in income inequality *over time* to the progressivity of individual income growth—a measure of the 'pro-poorness' of economic growth—and mobility in the form of re-ranking[3]

$$\Delta(\upsilon) = R(\upsilon) - P(\upsilon)$$

where

$$P(\upsilon) = \mathrm{GINI}(X^0; \upsilon) - \mathrm{CONC}(X^1, X^0 \upsilon)$$

and

$$R(\upsilon) = \mathrm{GINI}(X^1; \upsilon) - \mathrm{CONC}(X^1, X^0; \upsilon)$$

O'Neill and Van Kerm (2008) have interpreted $\Delta(\upsilon)$ as a measure of 'σ-convergence' and $P(\upsilon)$ as a measure of 'β-convergence' in analysis of cross-country (or regional) convergence in GDP.

[3] `dsginideco` SSC archive (Jenkins and Van Kerm, 2009).

Income mobility and pro-poor growth (ctd.)

Jenkins and Van Kerm (2016) propose to assess 'pro-poorness' of growth by 'progressivity-adjusted' individual income growth

$$M1(\upsilon) = \text{AGGCONC}(Z, X^0; \upsilon)$$

where Z is a measure of individual (or household-level) income change, the simplest of which is $Z = (Y_1 - Y_0)$, or $Z = (\ln(Y_1) - \ln(Y_0))$. Sensitivity to progressivity controlled by υ

Demuynck and Van de gaer (2012) advocate instead

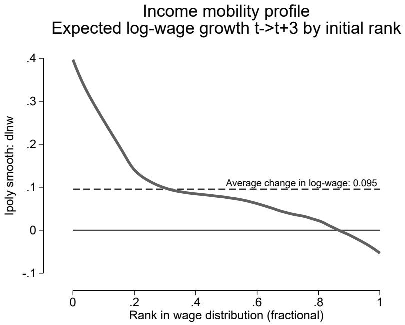
$$M2(\upsilon) = \text{AGGCONC}(Z, Z; \upsilon)$$

$M1(\upsilon)$ and $M2(\upsilon)$ differ in how individuals are ranked – and therefore the implicit weight of obs in the aggregation

```
. generate dlnw = ln(F3.w) - ln(w)

. sgini dlnw , parameter(1 2 3) sortvar(w) aggregate

Generalized Concentration coefficient for dlnw against w

Note: dlnw has 3687 negative observations (used in calculations).
```

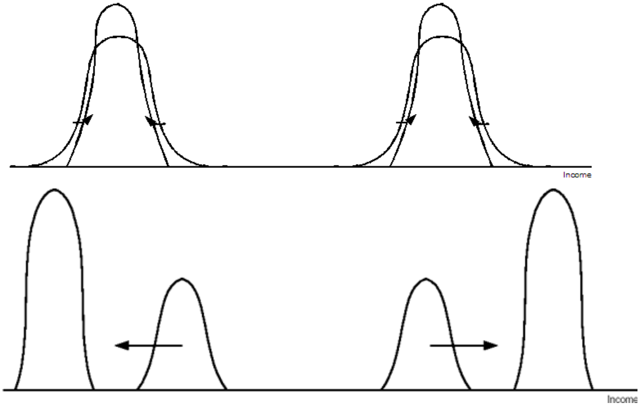
Variable	v=1	v=2	v=3
dlnw	0.0951	0.1526	0.1902

```
. fracrank w , gen(prank)

. range atp 0 1 100
(28,434 missing values generated)

. label variable atp "Rank in wage distribution (fractional)"

. lpoly dlnw prank , bw(0.08) gen(profile) at(atp) nograph
```



Polarization

- Concentration of the distribution around income 'poles'
- Connected to ideas of 'conflict': between-group 'alienation' vs. within-group 'identity'



(Duclos and Taptué, 2015)

Income bipolarization

Arrange population in increasing order of income and divide it in two equal-sized groups: the ‘poor’ and the ‘rich’.

Bipolarization can be expressed in terms of ‘within-group Gini’ and ‘between-group Gini’:

$$\mathrm{Within}(Y^P, Y^R) = \frac{1}{4}\left(\frac{\mu(Y^P)}{\mu(Y)}\mathrm{GINI}(Y^P) + \frac{\mu(Y^R)}{\mu(Y)}\mathrm{GINI}(Y^R)\right)$$

and

$$\mathrm{Between}(Y^P, Y^R) = \frac{1}{4}\left(\frac{\mu(Y^R)}{\mu(Y)} - \frac{\mu(Y^P)}{\mu(Y)}\right).$$

The ‘between-group Gini’ is equivalent to estimating the Gini coefficient of mean income in the two groups, that is $\mathrm{GINI}((\mu(Y^P), \mu(Y^R))')$.

Income bipolarization

The bipolarization index suggested by Silber et al. (2007) is defined as[4]

$$P_1 = \frac{\text{Between}(Y^P, Y^R) - \text{Within}(Y^P, Y^R)}{\text{GINI}(Y)},$$

the index of Wolfson (1994) as

$$P_2 = \left(\text{Between}(Y^P, Y^R) - \text{Within}(Y^P, Y^R)\right)\frac{\mu(Y)}{\text{Med}(Y)},$$

and the index proposed by Zhang and Kanbur (2001) as

$$P_3 = \frac{\text{Between}(Y^P, Y^R)}{\text{Within}(Y^P, Y^R)}.$$

[4] `bipolar` available on the SSC archive (Fusco and Van Kerm, 2020).

```
. prog define mypolar , rclass
  qui summarize w  , detail
      local mean = r(mean)
      local med  = r(p50)
  qui sgini w
      local sgini = r(coeff)
  su w if w<`med' , meanonly
      local mup = r(mean)
  su w if w>=`med' , meanonly
      local mur = r(mean)
  qui sgini w if w<`med'
      local sginip = r(coeff)
  qui sgini w if w>=`med'
      local sginir = r(coeff)
  return scalar Within = 0.25 * (1/`mean') * (`mup'*`sginip'  +  `mur'*`sginir')
  return scalar Between = 0.25 * (1/`mean') * (`mur'  -  `mup')
  return scalar P1 = ( return(Between) - return(Within) ) / `sgini'
  return scalar P2 = ( return(Between) - return(Within) ) * `mean' / `med'
  return scalar P3 =  return(Between) / return(Within)
  di "Within half-populations inequality: "  _col(42) %4.3f return(Within)
  di "Between half-populations inequality: " _col(42) %4.3f return(Between)
  di "Bipolarization index 1 (Silber et al.): " _col(42) %4.3f return(P1)
  di "Bipolarizarion index 2 (Wolfson): " _col(42) %4.3f return(P2)
  di "Bipolarization index 3 (Kanbur & Zhang): "_col(42) %4.2f return(P3)
end

. mypolar w
Within half-populations inequality:      0.121
Between half-populations inequality:     0.234
Bipolarization index 1 (Silber et al.):  0.318
Bipolarizarion index 2 (Wolfson):        0.145
Bipolarization index 3 (Kanbur & Zhang): 1.93

. bootstrap P1=r(P1) P2=r(P2) p3=r(P3) : mypolar w
```

```
. bipolar w
```

Bi-polarization measures	value
Deutsch Hanoka Silber (2007)	0.316
Foster Wolfson (1992, 2010)	0.145
Zhang Kanbur (2001)	1.925
Overall Gini index	0.355
Population share in low income group	0.498
Within group inequality	0.121
Between group inequality	0.234

No shortage of material... (`findit gini`)

inequal7 igini1 lorenz igini yadap progres svylorenz dasp fastgini descogini inequal sgini giniinc survgini ineqdecgini moremata egen_inequal adgini ineqdeco conindex bipolar anogi dsginideco inequaly somersd ginireg

LISER Luxembourg Institute of Socio-Economic Research

uni.lu Université du Luxembourg

References

Berger, Y. G. (2008), 'A note on the asymptotic equivalence of jackknife and linearization variance estimation for the Gini coefficient', *Journal of Official Statistics* **24**(4), 541–555.

Chen, Z. and Roy, K. (2009), 'Calculating concentration index with repetitive values of indicators of economic welfare', *Journal of Health Economics* **28**(1), 169–175.

Davidson, R. (2009), 'Reliable inference for the Gini index', *Journal of Econometrics* **150**(1), 30–40.

Demuynck, T. and Van de gaer, D. (2012), 'Inequality adjusted income growth', *Economica* **79**(316), 747–765.

Donaldson, D. and Weymark, J. A. (1980), 'A single parameter generalization of the Gini indices of inequality', *Journal of Economic Theory* **22**, 67–86.

References ii

Donaldson, D. and Weymark, J. A. (1983), 'Ethically flexible Gini indices for income distributions in the continuum', *Journal of Economic Theory* **29**(2), 353–358.
URL: *http://www.sciencedirect.com/science/article/pii/0022053183900534*

Duclos, J.-Y. and Taptué, A.-M. (2015), Polarization, *in* A. B. Atkinson and F. Bourguignon, eds, 'Handbook of Income Distribution', Vol. 2A, North-Holland, Amsterdam, chapter 5, pp. 301–358.

Fei, J. C. H., Ranis, G. and Kuo, S. W. Y. (1978), 'Growth and the family distribution of income by factor components', *Quarterly Journal of Economics* **92**(1), 17–53.

Fusco, A. and Van Kerm, P. (2020), `bipolar`: Stata module to calculate bi-polarization indices, Statistical Software Components S458775, Boston College Department of Economics.
`http://ideas.repec.org/c/boc/bocode/s458775.html`.

Hampel, F. R. (1974), 'The influence curve and its role in robust estimation', *Journal of the American Statistical Association* **69**(346), 383–393.

References iii

Jenkins, S. P. and Van Kerm, P. (2006), 'Trends in income inequality, pro-poor income growth and income mobility', *Oxford Economic Papers* **58**(3), 531–548.
URL: *http://ideas.repec.org/a/oup/oxecpp/v58y2006i3p531-548.html*

Jenkins, S. P. and Van Kerm, P. (2009), `dsginideco`: Decomposition of inequality change into pro-poor growth and mobility components, Statistical Software Components S457009, Boston College Department of Economics. `http://ideas.repec.org/c/boc/bocode/s457009.html`.

Jenkins, S. P. and Van Kerm, P. (2016), 'Assessing individual income growth', *Economica* **83**(332), 679–703.
URL: *http://dx.doi.org/10.1111/ecca.12205*

Kakwani, N. C. (1977*a*), 'Applications of Lorenz curves in economic analysis', *Econometrica* **45**(3), 719–728.

Kakwani, N. C. (1977*b*), 'Measurement of tax progressivity: an international comparison', *Economic Journal* **87**(345), 71–80.

References iv

Lerman, R. I. and Yitzhaki, S. (1985), 'Income inequality effects by income source: A new approach and applications to the United States', *Review of Economics and Statistics* **67**(1), 151–156.

López-Feldman, A. (2006), 'Decomposing inequality and obtaining marginal effects', *The Stata Journal* **6**(1), 106–111.

O'Neill, D. and Van Kerm, P. (2008), 'An integrated framework for analysing income convergence', *The Manchester School* **76**(1), 1–20.

Peichl, A. and Van Kerm, P. (2007), `progres`: Module to measure distributive effects of an income tax, Statistical Software Components S456867, Boston College Department of Economics. `http://ideas.repec.org/c/boc/bocode/s456867.html`.

Schechtman, E. and Yitzhaki, S. (1987), 'A measure of association based on Gini's mean difference', *Communications in Statistics - Theory and Methods* **16**(1), 207–231.

Schechtman, E. and Yitzhaki, S. (1999), 'On the proper bounds of the Gini correlation', *Economics Letters* **63**(2), 133–138.

References v

Silber, J., Hanoka, M. and Deutsch, J. (2007), 'On the link between the concepts of kurtosis and bipolarization', *Economics Bulletin* **4**(36), 1–6.

Van Kerm, P. (2015), Influence functions at work, United Kingdom Stata Users' Group Meetings 2015 11, Stata Users Group.
URL: *https://ideas.repec.org/p/boc/usug15/11.html*

Van Kerm, P. (2017), Estimation and inference for quantiles and indices of inequality and poverty with survey data: leveraging built-in support for complex survey design and multiply imputed data, United Kingdom Stata Users' Group Meetings 2017 12, Stata Users Group.
URL: *https://ideas.repec.org/p/boc/usug17/12.html*

Wolfson, M. C. (1994), 'When inequalities diverge', *American Economic Review* **84**(2), 353–58.

Yitzhaki, S. (1983), 'On an extension of the Gini inequality index', *International Economic Review* **24**(3), 617–628.

Yitzhaki, S. (1998), More than a dozen alternative ways of spelling Gini, *in* D. J. Slottje, ed., 'Research on Economic Inequality', Vol. 8, JAI Press, Stamford CT, pp. 13–30.

References vi

Yitzhaki, S. and Schechtman, E. (2005), 'The properties of the extended Gini measures of variability and inequality', *METRON International Journal of Statistics* **63**(3), 401–433.

Zhang, X. and Kanbur, R. (2001), 'What difference do polarisation measures make? An application to China', *Journal of Development Studies* **37**(3), 85–98.