Measuring Inequality

Carsten Schröder

July 12, 2017

Aims and scopes of the lecture

The lecture will make you familiar with the

- measurement (graphs and indices) of inequality,
- empirical application of the concepts using STATA,
- techniques for statistical inference and modeling of distributions.

Aims and scopes of the lecture

Bottom line. Get acquainted with the techniques required for conducting empirical research on inequality.

Aims and scopes of the lecture

After the semester, you should be ready to give data-based answers to questions such as:

- How unequal is the German income distribution?
- Is the German distribution more unequal than the French distribution?
- Has inequality increased in the last years?
- Do different inequality measures tell the same story?
- Are the findings statistically significant?
- How large is the inequality reduction due to public interventions (taxes and transfers)?

Examination

- There will be a written exam at the end of the semester.
- Question set covers topics from lecture and tutorial, theory and programming.

Data and data access

We use the public sue file of the German *Socio-Economic Panel* (SOEP): http://www.diw.de/de/soep

1. Characteristics of SOEP
   - Household panel for 1984-20XX
   - Numerous socio-economic characteristics at the household and individual level
2. SOEP public use file
   - Small subsample of the SOEP (non-representative)
   - Limited set of variables for only one wave (2011)
   - Further infos will be provided in the tutorial

Basic readings and materials

For the measurement of inequality, consult:

1. Cowell, F. (2010): Measuring Inequality, 3rd ed.[1]
2. Further readings are provided in each section.

[1] Free version: http://darp.lse.ac.uk/papersDB/Cowell_measuringinequality3.pdf.

Basic readings and materials

Further materials provided in blackboard:

1. Lecture notes
2. Exercises
3. STATA codes

Table of Contents

Definition. Economic inequality is the differences in the distribution of material resources within or between populations.

Introduction

Empirical research on inequality has a long tradition.

- **Pareto** (1897): constant income distribution over time and space
- **Kuznets** (1955): inverted-u shape relationship between inequality and development
- **Harrison and Bluestone** (1988): "great u-turn"; (recent) rise of inequality is pervasive (globalization; new technologies)
- **Deininger and Squire** (1996): inter-temporal variation of inequality less important than cross-country variation

Why should we care about inequality?

Inequality rose recently in many societies and is for the following reasons (one of) the most important problem we are facing today:

- Economically: credit rationing prevents the poor from undertaking investments in education and weakens innovations.
- Socially: social cohesion is weakened because people are less likely to trust each other and lower community involvement.
- Politically: economic elites may be generated, that use their power to (re)design the political system and gain special treatments- creating a vicious circle.
- Fairness and justice considerations.

How to compress the information content of an income distribution?

- Empirical studies on inequality, usually, rely on micro data (at person or household level).
- Micro-data bases easily contain 10,000s or even 100,000s of observations.
- Somehow we need to handle this flood of information
  - We need to feed the data in a computer.
  - We need to find ways to compress the flood of information.

Introduction

Measurement concepts

- Diagrams: Graphical representation of the distribution.
  - Examples: Frequency distribution, Lorenz curve, Parade of dwarfs
- Indices: Scalar representation of the distribution.
  - Examples: Gini, Theil, or Atkinson index
- Rankings: Ordering of distributions.
  - Comparing distributions by means of indices and/or diagrams.

Dimensions to keep under control.

- **Focal variable**
- **Reference unit**
- **Adjustment for differences in needs**
- **Weighting**
- **Comprehensiveness of measure**
- **Data source**
- **Data preparation**

Dimensions to keep under control.

- **Focal variable**: income, expenditure, wealth, ...
  "The relative advantages and disadvantages that people have, compared with each other, can be judged in terms of many different variables, e.g. their respective incomes, wealths, utilities, resources, liberties, rights, quality of life, and so on. The plurality of variables on which we can possibly focus (the focal variables) to evaluate interpersonal inequality makes it necessary to face, at a very elementary level, a hard decision regarding the perspective to be adopted" Sen, 1992

Introduction

Dimensions to keep under control.

- **Reference unit**: household, person, tax unit, ...

Dimensions to keep under control.

- **Reference unit**: household, person, tax unit, ...

$$\begin{bmatrix} \text{Household unit} & \text{Person} & \text{Indiv. income} \\ 1 & 1 & 100 \\ 2 & 1 & 0 \\ 2 & 2 & 100 \\ 2 & 3 & 200 \\ 2 & 4 & 100 \end{bmatrix}$$

$\rightarrow$ Distribution of household incomes: $(100, 400)$

$\rightarrow$ Distribution of individual incomes: $(100, 0, 100, 200, 100)$

$\rightarrow$ Distribution of per capita income, household level: $(100, 100)$

Dimensions to keep under control.

- **Adjustment for differences in needs**
  - If reference units differ in composition, members of different units with the same unit material resources may attain different living standards.
  - To obtain a measure that reflects differences in living standards across reference units, resources must then be adjusted for differences in needs.
  - Equivalence scale. Convert material resources of different reference units in welfare-equivalent units.

Dimensions to keep under control.

- **Adjustment for differences in needs**
  - Instrument of conversion: Equivalence scale (ES).
    - OECD: $ES^{OECD} = 1 + 0.5 \times (n_{adults} - 1) + 0.3 \times n_{children}$
    - Square-root: $ES^{SR} = (n_{members})^{0.5}$
  - Outcome of conversion: Equivalent income ($EI = \frac{income}{ES}$)

Dimensions to keep under control.

- **Adjustment for differences in needs**

$$\begin{bmatrix} \text{Household} & & \text{Household} & & \\ \text{unit} & \text{Persons} & \text{income} & ES^{SR} & EI \\ 1 & 1 & 100 & 1 & 100 \\ 2 & 4 & 400 & 2 & 200 \end{bmatrix}$$

Dimensions to keep under control.

- **Weighting**
- Alternative approaches suggested in the literature
  - Weight all reference units equally
  - Weight reference units by size
  - Weight reference units by needs ($= ES$)

Dimensions to keep under control.

- **Welfare weighting**

$$\begin{bmatrix} \text{Household} & & \text{Household} & & \\ \text{unit} & \text{Persons} & \text{income} & ES^{SR} & EI \\ 1 & 1 & 100 & 1 & 100 \\ 2 & 4 & 400 & 2 & 200 \end{bmatrix}$$

$\rightarrow$ EI distr.; equal weighting: $(100, 200)$

$\rightarrow$ EI distr.; size weighting: $(100, 200, 200, 200, 200)$

$\rightarrow$ EI distr.; needs weighting: $(100, 200, 200)$

Dimensions to keep under control.

- **Comprehensiveness of measure**
  - Should we consider the value of home production?
  - Should we consider the value of non-cash transfers (e.g., free education opportunities)?
  - Say, the focal variable is ...
    - ... income, should it reflect inter-household transfers, or imputed rents?
    - ... expenditures, how should we deal with expenditures for durable goods (depreciation)?

Dimensions to keep under control.

- **Data source**
  - Survey data
    - Pros: availability; detailed background information (household size, education, ...)
    - Cons: item non-response; false responses; representativeness
    - Example for Germany: Socioeconomic Panel (SOEP)
  - Administrative data
    - Pros: validity of the data
    - Cons: non-availability; limited background information; difficult to interpret
    - Example for Germany: Employment histories from German Pension Insurance

Dimensions to keep under control.

- **Data preparation**
  - What to do if ...
    - ... data are top/bottom coded (imputation methods)?
    - ... data are censored?
    - ... data are implausible (negative disposable income)?
    - ... we want to conduct comparative analysis but data come from different sources?
    - ...

References to this section

- Harrison, B, and B. Bluestone (1988). The Great U-Turn. New York: Basic Books.
- Deininger, K., and L. Squire (1996). A new data set measuring income inequality. World Bank Economic Review, 10, 565591.
- Kuznets, S. (1955). Economic Growth and Income Inequality. American Economic Review, 45, 1-28.
- Pareto, V. (1897). Cours d'économie politique. Lausanne and Paris: Rouge and Pichon. Reprinted in Oeuvres Complètes, ed. by G.-H. Bousquet and G. Busino. Genève: Librairie Droz, 1964.
- Sen, A. (1992). Inequality Reexamined, Clarendon Press, Oxford.

Table of Contents

Charting inequality

Methods to compress information content of micro data

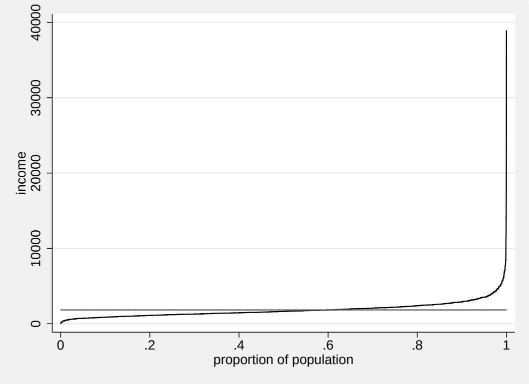
- Diagrams
- Indices
- Rankings

Charting inequality: diagrams

Parade of dwarfs. Charts income against the cumulated sorted population share.

Charting inequality: diagrams

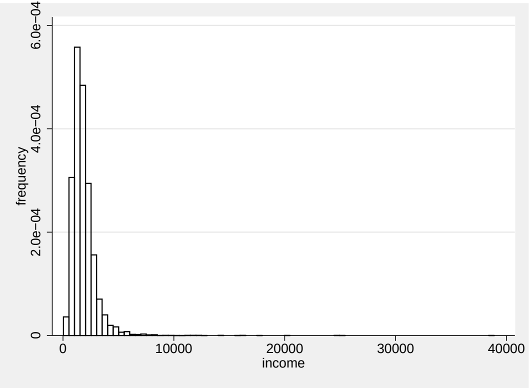
Figure. Parade of dwarfs, Germany 2011



Note. Equivalent disposable income (square root scale); horizontal line is mean eq. inc; SOEP 2011

Frequency distribution of income. Charts the relative frequency (density) of income y, $f(y)$.

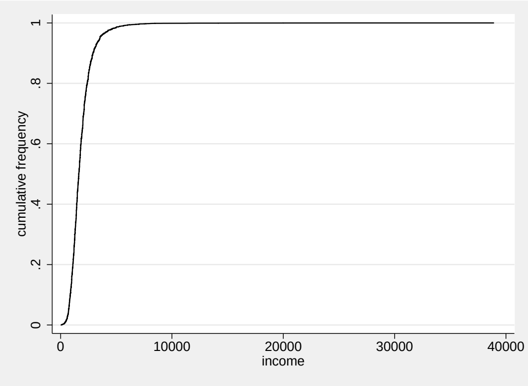
Figure. Frequency distribution of income, Germany 2011



Note. Equivalent disposable income (square root scale); SOEP 2011

Cumulative frequency distribution of income. Charts the cumulative frequency of income y, $F(y)$.

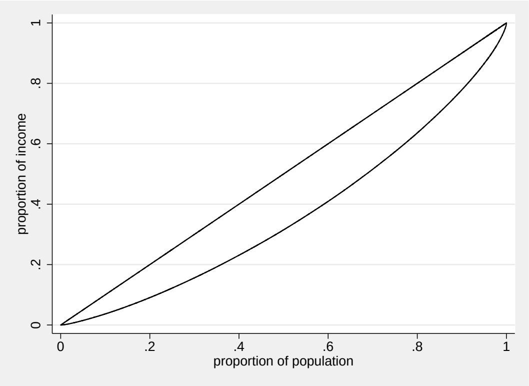
Figure. Cum. frequency distribution of income, Germany 2011



Note. Equivalent disposable income (square root scale); SOEP 2011

Lorenz curve. Charts the cumulative income share, $\phi(y) = L(F)$, against the cumulative population share (of the ordered population), $F(y)$.

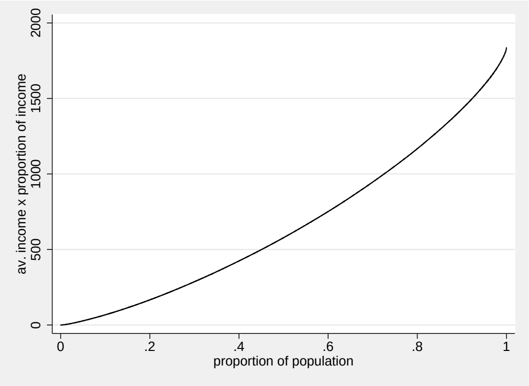
Figure. Lorenz curve, Germany 2011



Note. Equivalent disposable income (square root scale); SOEP 2011

Generalized Lorenz curve. Charts the cumulative income share multiplied with average income, $\bar{y}L(F)$, against the cumulative population share (of the ordered population), $F(y)$.

Figure. Generalized Lorenz curve, Germany 2011



Note. Equivalent disposable income (square root scale); SOEP 2011

Charting inequality: inequality indices

1. Description of the income distribution by a single number.
2. Sacrifice information but allow immediate comparisons of distributions
3. Different inequality measures have different properties (e.g., how they respond to changes in distribution). A systematic analysis of the properties of different measures is provided in Section 4.

Charting inequality: inequality indices

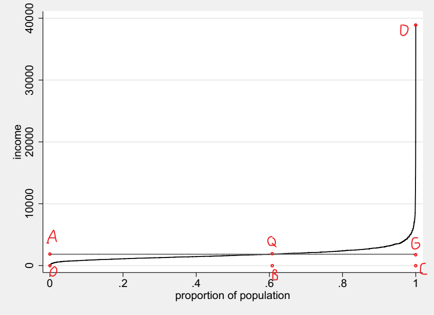
Basic notation for discrete distributions

1. n: population size
2. i: observation unit (person, household, etc.)
3. y_i: income of i
4. $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$: arithmetic mean
5. $y^{\star} = exp(\frac{1}{n}\sum_{i=1}^{n} log(y_i)) = [y_1 * y_2 * y_3 * ... * y_n]^{\frac{1}{n}}$: geometric mean

Range. Difference between minimum and maximum value of income.

$$R = y_{max} - y_{min}$$

Figure. Parade of dwarfs, Germany 2011



Note. Equivalent disposable income (square root scale); horizontal line is mean eq. inc; SOEP 2011

$$R = \overline{CD}$$

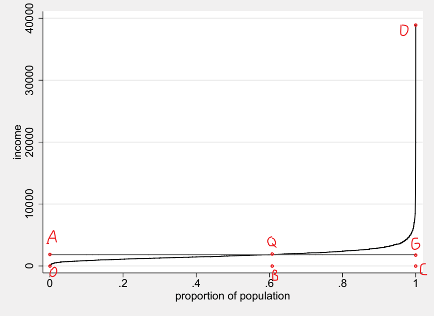
Range.

- Strength
  - Describes extreme incomes
- Weakness
  - Ignores much of the information about the distribution

Relative mean deviation.

$$M = \frac{1}{n}\sum_{i=1}^{n}|\frac{y_i}{\bar{y}} - 1|$$

Figure. Parade of dwarfs, Germany 2011



Note. Equivalent disposable income (square root scale); horizontal line is mean eq. inc; SOEP 2011

$$M = \frac{\text{area OAQ} + \text{area QGD}}{\text{area OCGA}}$$

Charting inequality: inequality indices

Relative mean deviation.

- Strength
  - Describes average deviation from mean income
- Weakness
  - Insensitivity to income reallocation among the group of people below (above) the mean

Variance.

$$V = \frac{1}{n}\sum_{i=1}^{n}[y_i - \overline{y}]^2 \tag{2.1}$$

Charting inequality: inequality indices

Variance.

- Strength
  - Describes the dispersion of the frequency distribution
- Weakness
  - Doubling all incomes (and thus leaving the shape of the distribution unchanged) quadruples V

Coefficient of variation.

$$c = \frac{\sqrt{V}}{\bar{y}} \tag{2.2}$$

Coefficient of variation.

- Strength
  - Describes the dispersion of the frequency distribution
  - Scale invariant (multiplying all incomes with the same factor leaves c unchanged)
- Weakness
  - Hard to interpret from economic / welfare perspective
  - Particular transfer effect: Transferring € 1 from a unit with an income of €500 to a unit with €400 has the same effect on c as transferring the same €from a unit with € 50,100 to a unit with € 50,000.

Logarithmic variance.

$$v = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{y_i}{\bar{y}}\right]^2 \tag{2.3}$$

Logarithmic variance.

- Strength
  - Describes the dispersion of the frequency distribution
  - Scale invariant
- Weakness
  - Hard to interpret from economic / welfare perspective
  - Particular transfer effect: An income transfer among two units at the distribution's top can have ambiguous effects on v.

Variance of the logarithms.

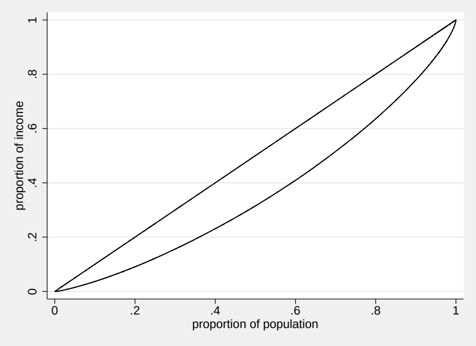
$$v_1 = \frac{1}{n}\sum_{i=1}^{n}\left[\log\frac{y_i}{y^*}\right]^2 \tag{2.4}$$

Variance of logarithms. Weaknesses and strengths are similar to logarithmic variance.

Gini.

$$G = \frac{1}{2n^2\bar{y}} \sum_{i=1}^{n} \sum_{j=1}^{n} |y_i - y_j| \tag{2.5}$$

Figure. Lorenz curve, Germany 2011



Note. Equivalent disposable income (square root scale); SOEP 2011

$G = 2(\text{area L \& 45 degree line})$

Charting inequality: inequality indices

Gini coefficient.

- Strength
  - Intuitive, easy to interpret, and scale invariant
- Weakness
  - Particular transfer effect: Transferring € 1 from a unit with an income slightly above to a unit with an income slightly below the average has a larger effect on G than the same transfer among two households at the distribution's bottom or top.

Some indices for Germany:

year	$\bar{y}$	M	R	V	c	G	v
1991	1262.46	.353	27740.92	540156.2	.582	.251	.267
1996	1374.92	.341	14708.96	521189.7	.525	.243	.222
2001	1512.96	.341	16268.11	559701.9	.494	.244	.220
2006	1619.96	.382	67232.84	1178343	.670	.274	.271
2011	1835.60	.381	38850.88	1232336	.605	.272	.274

Note. Data from SOEP, 1991-2011.

Charting inequality: rankings

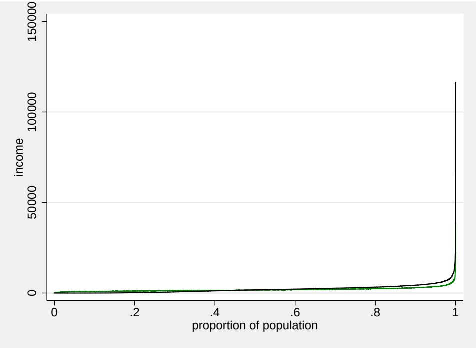
Types of rankings:

1. Quantiles
   - Example: How large is the difference between gross and disposable income in the second decile of the distributions of gross and disposable income?
2. Shares
   - Example: How large is the income share of the second decile of the distributions of gross and disposable income?

Charting inequality: rankings

Quantiles.

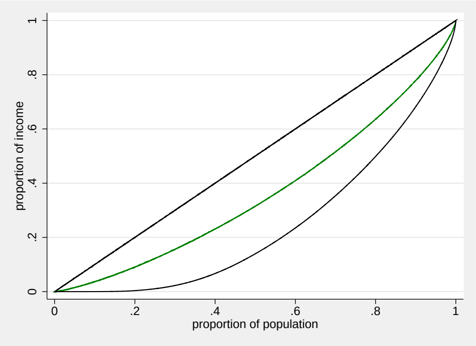
Figure. Parade of dwarfs, Germany 2011



Note. Data from SOEP 2011. Green: disposable inc. Black: gross inc.

Shares.

Figure. Lorenz curves for gross and disposable income, Germany 2011



Note. Data from SOEP 2011. Green: disposable inc. Black: gross inc.

Table of Contents

Analysing inequality

So far, we have discussed some measures of inequality that came about "by accident." Now we will construct the measures from "first principles."

Methods to derive inequality indices.

1. Measures based on **social welfare functions**
2. Measures based on **information theory**
3. Measures based on **sets of principles**

Analysing inequality: social welfare functions

Basic idea.

- Use a social welfare function *SWF* to rank all possible states of society in the order of societal preferences.
- **Example.** If SWF for distribution A is larger than for distribution B, than A is preferred to B.
- Hence, once agreed on the SWF, the approach enables the researcher to say, in effect "you tell me how the SWF looks like, and I will tell you the value of the inequality statistic."

Characteristics of SWF

1. Individualistic and nondecreasing
2. Symmetric
3. Additive
4. Strictly concave
5. Constant relative inequality aversion

The SWF is **individualistic** and **nondecreasing**, if the welfare level in state A, W_A, can be written:

$$W_A = W(y_{1A}, y_{2A}, ..., y_{nA})$$

and, if $y_{iB} \geq y_{iA}\ \forall i$ implies, ceteris paribus, that $W_B \geq W_A$, which in turn implies that state B is at least as good as state A.

If SWF is **individualistic** and **nondecreasing**, W depends on the individual incomes and does not go down if y_i goes up.

The SWF is **symmetric** if it is true that, for any state,

$$W(y_1, y_2, ..., y_n) = W(y_2, y_1, ..., y_n) = W(y_n, y_2, ..., y_1).$$

If SWF is **symmetric**, units are treated anonymously: $dW = 0$ if any two units simply swapped incomes.

The SWF is **additive** if it can be written:

$$W(y_1, y_2, ..., y_n) = \sum_{i=1}^{n} U_i(y_i) = U_1(y_1) + U_2(y_2) + ... + U_n(y_n) \tag{3.1}$$

where U_i is a function of y_i alone.

If SWF is **additive**, the effect of a change in unit i's income on W is independent of the material situation of the other units.

An **individualistic, nondecreasing, symmetric, and additive** SWF can written as:

$$W(y_1, y_2, ..., y_n) = \sum_{i=1}^{n} U(y_i) = U(y_1) + U_2(y_2) + ... + U(y_n) \quad (3.2)$$

and a change in W as:

$$dW = U'(y_1)\Delta y_1 + U'(y_2)\Delta y_2 + ... + U'(y_n)\Delta y_n$$

Let us call $U(y_i)$ the welfare index for i. The rate at which the welfare index increases, the **social marginal utility** or **welfare weight**, is

$$U'(y_i) = \frac{dU(y_i)}{dy_i} > 0$$

The SWF is **strictly concave** if

$$\frac{dU'(y_i)}{dy_i} < 0$$

The SWF has **constant relative inequality aversion** if $U(y_i)$ can be written:

$$U(y_i) = \frac{(y_i)^{1-\epsilon} - 1}{1-\epsilon}, \text{if } \epsilon \neq 1 \tag{3.3a}$$

$$U(y_i) = ln(y_i), \text{if } \epsilon = 1 \tag{3.3b}$$

(or in cardinally equivalent form), where $\epsilon > 0$ is the **inequality aversion parameter.**

The relative inequality aversion parameter, ϵ, is crucial for:

1. Maximum amount of sacrifice
2. Social utility
3. Welfare weights

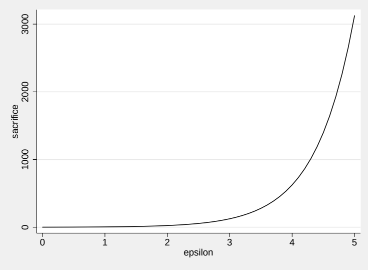
Maximum amount of sacrifice and ϵ

- ϵ determines the maximum amount of sacrifice if transfers from the top to the bottom cause a reduction in total income (efficiency losses).
- To see this, suppose that
  - $n = 2$ with $U(y_i) = \frac{(y_i)^{1-\epsilon}-1}{1-\epsilon}$, and $\frac{y_2}{5} = y_1$
  - Unit 1 receives $\Delta y_1 = 1$ from unit 2
  - How much can the transfer cost, expressed in income units of unit 2?

Problem to solve.

$$\begin{aligned}
dW = 0 = U'(y_1)\Delta y_1 + U'(y_2)\Delta y_2 \\
&\rightarrow 0 = y_1^{-\epsilon} * 1 + y_2^{-\epsilon}\Delta y_2 \\
&\rightarrow -y_1^{-\epsilon} = y_2^{-\epsilon}\Delta y_2 \\
&\rightarrow -\left(\frac{y_1}{y_2}\right)^{-\epsilon} = \Delta y_2 \\
&\rightarrow -\left(\frac{y_2}{y_1}\right)^{\epsilon} = \Delta y_2 \\
&\rightarrow -5^{\epsilon} = \Delta y_2
\end{aligned}$$

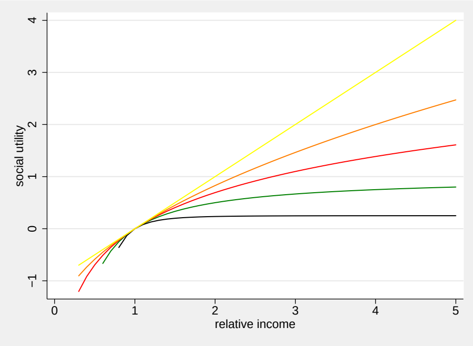
Maximum amount of sacrifice and ϵ



Note. See example.

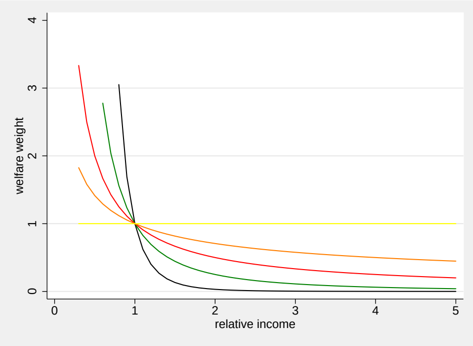
Interpretation. Suppose "poor" unit 1 receives €1. For $\epsilon = 0/0.5/1/2/5$, $dW = 0$ if "rich" unit 2 can sacrifice up to €1/2.24/5/25/3,125.

Social utility and ϵ



Note. $\epsilon = 0$: yellow; 0.5: orange; 1: red; 2: green; 5: black

Welfare weights and ϵ



Note. $\epsilon = 0$: yellow; 0.5: orange; 1: red; 2: green; 5: black

Putting SWF to work.

Theorem 1. If social state A **dominates** the state B according to their quantile ranking, then $W_A > W_B$ for any individualistic increasing, additive, and symmetric SWF.

Example. Assume 3 persons in four states with distributions $A = (3, 5, 6)$, $B = (4, 5, 6)$, $C = (5, 3, 7)$, $D = (5, 4, 5)$

- W_B vs. W_A: Every person is as least as well-off in B compared with A. Hence B dominates A, $W_B > W_A$.
- W_C vs. W_A: Because of symmetry assumption, C is equivalent to $C' = (3, 5, 7)$. As C' dominates A, $W_C > W_A$.
- W_D vs. W_A: Due to lack of dominance, A and D cannot be ordered. This is a general message: **sometimes, distributions cannot be ranked unambiguously**.

Theorem 2. Let the state A have an associated distribution $(y_{1A}, y_{2A}, ...y_{nA})$ and state B have $(y_{1B}, y_{2B}, ...y_{nB})$, where total income in A and B is identical, $\sum_{i=1}^{N} y_{iA} = \sum_{i=1}^{N} y_{iB}$. Then the Lorenz curve for A **lies wholly inside** the Lorenz curve for B if and only if $W_A > W_B$ for any individualistic, increasing, symmetric, and strictly concave SWF.

Problem. Ranking of distributions A and B requires $\sum_{i=1}^{N} y_{iA} = \sum_{i=1}^{N} y_{iB}$. However, this is a rare case.

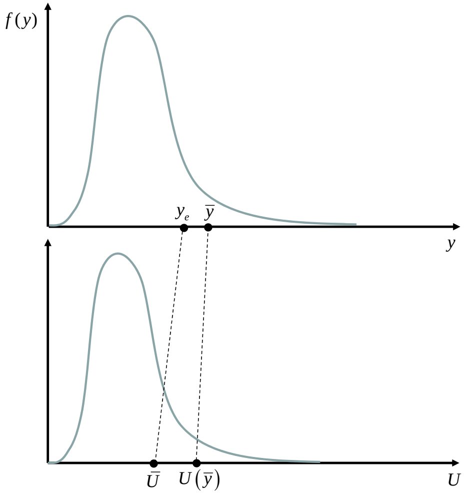
Theorem 3. The generalized Lorenz curve for state A **lies wholly above** the generalized Lorenz curve for B if and only if $W_A > W_B$ for any individualistic, additive increasing, symmetric, and strictly concave SWF.

From SWF to SWF-based indices.

Given the income distribution and the SWF (with the underlying social utility functions), we can derive the following information:

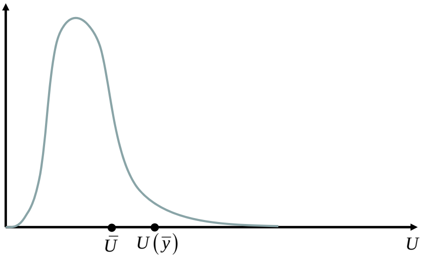
- The level of social welfare, $W = W(y_1, ..., y_n)$
- The distribution of income and of social utilities.
- The social utility, $U(\bar{y})$, for each unit were total income be distributed perfectly equally ($y_1 = y_2 = ... = y_n = \bar{y}$).
- The income, y_e which, if received by all units, would result in the same social welfare as the existing distribution yields.

Figure. Distributions of income and social utilities



Dalton index.

Figure. Distributions of social utilities



$$D_\epsilon = 1 - \frac{\bar{U}}{U(\bar{y})} = 1 - \frac{\frac{1}{n}\sum_{i=1}^{n}[y_i^{1-\epsilon} - 1]}{\bar{y}^{1-\epsilon} - 1}$$

Atkinson critique. D_ϵ is sensitive to the level from which social utility is measured.[2]

Atkinson, therefore, suggested an index that is based on income:[3]

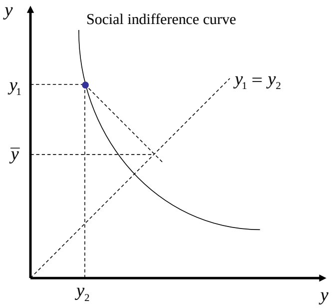
$$A_\epsilon = 1 - \frac{y_e}{\bar{y}} = 1 - \frac{U^{-1}(\bar{U})}{\bar{y}} = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\bar{y}}\right]^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}}$$

[2] As an example, adding a non-zero amount to the social utilities changes D_ϵ.

[3] For $\epsilon = 1$, $A_1 = 1 - \frac{1}{\bar{y}}(\prod_{i=1}^{n} y_i)^{1/n}$

Atkinson index.

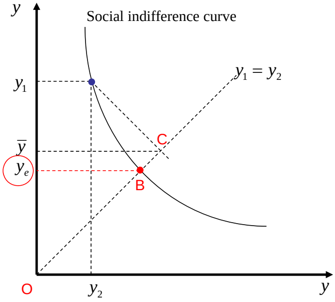
$$A_{\varepsilon} = 1 - \frac{y_e}{\bar{y}}$$



Atkinson index.

$$A_\varepsilon = 1 - \frac{y_e}{\overline{y}} = \frac{\overline{BC}}{\overline{OC}}$$

"equally distributed equivalent income": indicates the level of income per head which, if equally shared, would generate the same level of social welfare as the actual distribution



Explicit formulation of y_e using $U(y_i) = \frac{y_i^{1-\epsilon}-1}{1-\epsilon}$:

$$\bar{U} = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i^{1-\epsilon}-1}{1-\epsilon} \doteq \frac{y_e^{1-\epsilon}-1}{1-\epsilon}$$

$$\rightarrow \frac{1}{n}(\sum_{i=1}^{n} y_i^{1-\epsilon}) - \frac{n}{n} \doteq y_e^{1-\epsilon} - 1$$

$$\rightarrow \frac{1}{n}(\sum_{i=1}^{n} y_i^{1-\epsilon}) \doteq y_e^{1-\epsilon}$$

$$\rightarrow \left[\frac{1}{n}(\sum_{i=1}^{n} y_i^{1-\epsilon})\right]^{\frac{1}{1-\epsilon}} \doteq y_e$$

Ordinal equivalence of Dalton and Atkinson index.

Step 1. Express D_ϵ using A_ϵ

$$D_\epsilon = 1 - \frac{\bar{U}}{U(\bar{y})} \quad \rightarrow \quad 1 - D_\epsilon = \frac{\bar{U}}{U(\bar{y})} \text{ (O1)}$$
$$A_\epsilon = 1 - \frac{y_e}{\bar{y}} \quad \rightarrow \quad 1 - A_\epsilon = \frac{y_e}{\bar{y}} \quad \rightarrow \quad y_e = \bar{y}(1 - A_\epsilon) \text{ (O2)}$$

$$\bar{U} = U(y_e) \text{ (O3)}$$

(O3) in (O1): $1 - D_\epsilon = \frac{U(y_e)}{U(\bar{y})}$ (O4)

(O2) in (O4): $1 - D_\epsilon = \frac{U(\bar{y}(1-A_\epsilon))}{U(\bar{y})} \rightarrow D_\epsilon = 1 - \frac{U(\bar{y}(1-A_\epsilon))}{U(\bar{y})}$

Ordinal equivalence of Dalton and Atkinson index.

Step 2. Compute first derivative of $D_\epsilon = 1 - \frac{U(\bar{y}(1-A_\epsilon))}{U(\bar{y})}$ w.r.t. A_ϵ

$$\frac{\partial D_\epsilon}{\partial A_\epsilon} = -\frac{U(\bar{y})U'_{A_\epsilon}(\bar{y}(1-A_\epsilon))(-\bar{y}) - 0U(\bar{y}(1-A_\epsilon))}{[U(\bar{y})]^2}$$

$$\frac{\partial D_\epsilon}{\partial A_\epsilon} = \bar{y}\frac{U'_{A_\epsilon}(\bar{y}(1-A_\epsilon))}{[U(\bar{y})]} > 0$$

Q.e.d.

Analysing inequality: information theory

Basic idea.

- Information theory is concerned with the problem of 'valuing' the information that a certain event out of a number of possible events has occurred.
- Suppose events are $1, 2, 3, \ldots$ with probabilities $p_1, p_2, p_3, \ldots$
- The lower the probability that a certain event has occurred, the more exciting is its occurrence.

Notation and concepts.

- $h(p_i)$: information content of event i with probability p_i
  1. $h(p_i = 1) = 0$
  2. $p_i > p_j \rightarrow h(p_i) < h(p_j)$
  3. $h(p_i p_j) = h(p_i) + h(p_j)$
- The single functional form that complies with (1-3) is $h(p) = -log(p)$[4].
- $H(p) = -\sum_{i=1}^{n} p_i ln(p_i)$: information content (entropy) of the distribution
- $H(p)$ is the largest if $p_1 = p_2 = p_3 = \ldots$

[4] We need the minus as $0 \leq p \leq 1$.

Theil's idea.

- Interpret events as units (persons, households)
- Interpret probabilities p_i as income shares, $s_i = \frac{y_i}{\sum_{i=1}^{n} y_i}$
- **Theil index: maximum entropy minus entropy of the distribution**

Maximum entropy from an income distribution (requires that $s_i = \frac{1}{n} \ \forall i$):

$$H^{max}(s) = -\sum_{i=1}^{n} p_i ln(p_i)$$
$$= -\sum_{i=1}^{n} \frac{y_i}{\sum_{i=1}^{n} y_i} ln(\frac{y_i}{\sum_{i=1}^{n} y_i})$$
$$= -n\frac{1}{n} ln\left(\frac{1}{n}\right) = -ln\left(\frac{1}{n}\right) = ln\left(n\right)$$

Deriving the Theil index.

$$T = H^{max}(s) - \overbrace{\left(-\sum_{i=1}^{n} s_i \ln(s_i)\right)}^{H(s)}$$

$$T = ln(n) + \sum_{i=1}^{n} s_i \ln(s_i)$$

Deriving the Theil index continued.

$$T = ln\left(n\right) + \sum_{i=1}^{n} s_i \ln\left(s_i\right) \quad \Big| \text{ Multiply } ln(n) \text{ with } \sum_i s_i = 1$$

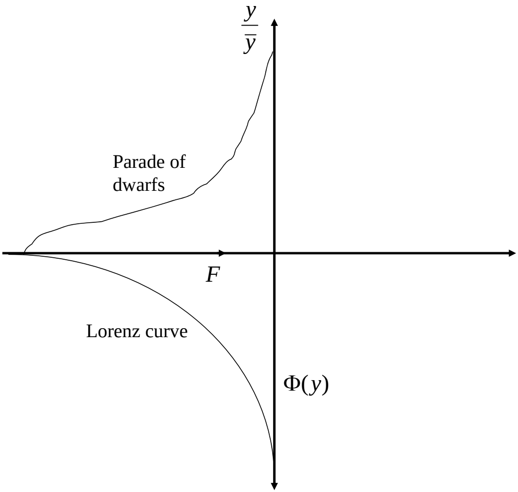
$$T = \sum_{i=1}^{n} s_i \ln n + \sum_{i=1}^{n} s_i \ln s_i$$

Deriving the Theil index continued.

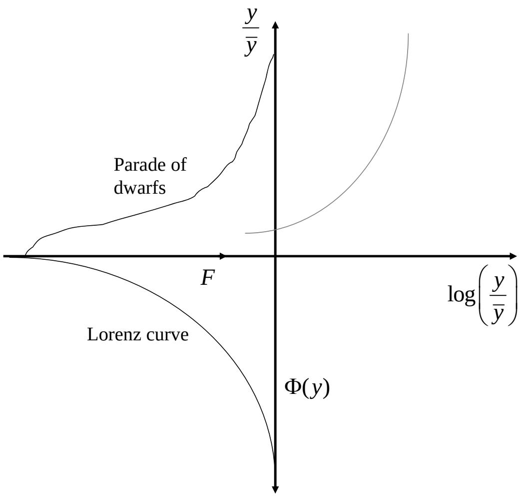
$$T = \sum_{i=1}^{n} s_i \left(\ln(n) + \ln(s_i)\right)$$

$$T = \sum_{i=1}^{n} s_i \ln(ns_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{y_i}{\bar{y}} ln\left(\frac{y_i}{\bar{y}}\right)$$

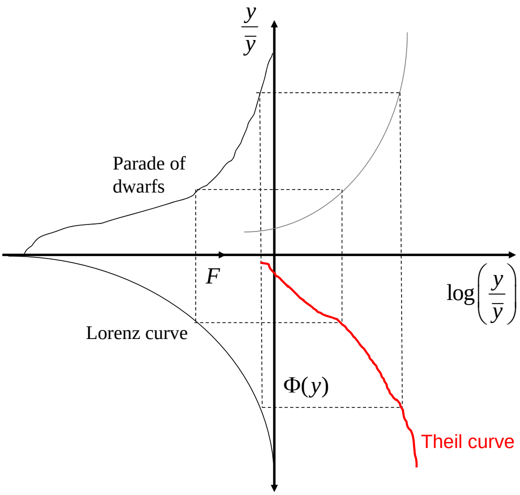
Theil curve and Theil index.



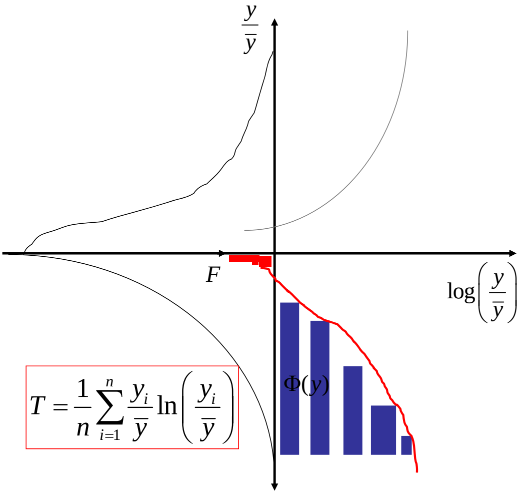
Theil curve and Theil index.



Theil curve and Theil index.

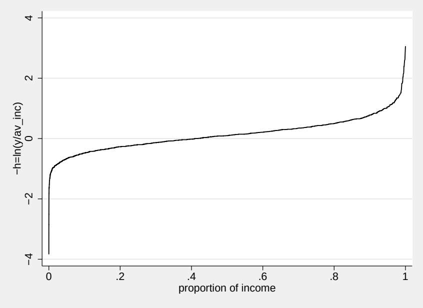


Theil curve and Theil index.



$$T = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{\bar{y}} \ln\left(\frac{y_i}{\bar{y}}\right)$$

Theil curve Germany 2011.



Note. Equivalent disposable income (square root scale); SOEP 2011

What happens to the Theil index if income is transferred from "rich" unit (2) to "poor" unit (1)?

Assume s_1 increases to $s_1 + \Delta$ ($\Delta > 0$), and s_2 decreases to $s_2 - \Delta$

$$\begin{aligned} T &= \sum_{i=1}^{n} \frac{1}{n} \ln\left(\frac{1}{n}\right) + \sum_{i=1}^{n} s_i \ln\left(s_i\right) \\ \Delta T &= \Delta s \left[ln(s_1) + s_1 \frac{1}{s_1} \right] - \Delta s \left[ln(s_2) + s_2 \frac{1}{s_2} \right] \\ &= \Delta s ln(s_1) - \Delta s ln(s_2) \\ &= \Delta s [ln(s_1) - ln(s_2)] \\ &= -\Delta s \ln\left(\frac{s_2}{s_1}\right) < 0 \text{ since } s_2 > s_1 \end{aligned}$$

Interpretation

- Theil index decreases when income is transferred from rich to poor
- The decrease depends on the **ratio** (not on the absolute difference) of the transfer recipient's and the donator's income.

Generalized entropy index.

The Theil index is a special case of the generalized entropy index,

$$E(\alpha) = \frac{1}{n(\alpha^2 - \alpha)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1 \right]$$

Proof $E(\alpha = 1) = \frac{1}{n(\alpha^2 - \alpha)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1 \right] = T$.

- For $\alpha = 1$ the denominator $n(\alpha^2 - \alpha)$, is zero.
- To derive an expression for $E(1)$ (and $E(0)$), apply the rule of *de l'Hôpital*:
  - The limit of an undefined ratio between two functions of the same variables is equal to the limit of the ratio of their first derivatives
  - Function 1: $n(\alpha^2 - \alpha)$
  - Function 2: $\left(\frac{y_i}{\bar{y}} \right)^{\alpha} - 1$

Proof continued.

- Derivatives:
  - Function 1: $\frac{\partial\left(n(\alpha^2-\alpha)\right)}{\partial\alpha} = n(2\alpha - 1)$
  - Function 2:[5] $\frac{\partial\left(\left(\frac{y_i}{\bar{y}}\right)^{\alpha}-1\right)}{\partial\alpha} = \left(\frac{y_i}{\bar{y}}\right)^{\alpha} ln\left(\frac{y_i}{\bar{y}}\right)$
- Hence it follows, $E(1) = \frac{1}{n(2-1)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}}\right)^{1} ln\left(\frac{y_i}{\bar{y}}\right)\right]$

  $\Rightarrow E(1) = \frac{1}{n} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}}\right) ln\left(\frac{y_i}{\bar{y}}\right)\right] = T$ **Q.e.d.**

[5] Remember: first derivative of $y(x) = a^x$ is $y' = a^x ln(a)$.

Two other standard cases.

- **Mean logarithmic deviation** ($\alpha = 0$)
  - Denominator $n(\alpha^2 - \alpha)$ becomes zero. Using *de l'Hôpital*,

$$E(0) = \frac{1}{n(2 \cdot 0 - 1)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^0 ln \left(\frac{y_i}{\bar{y}} \right) \right] = -\frac{1}{n} \sum_{i=1}^{n} \left[ln \left(\frac{y_i}{\bar{y}} \right) \right]$$

- **Half squared coefficient of variation** ($\alpha = 2$)
  - Denominator is different from zero. Using standard expression for $E(\alpha = 2)$,

$$E(2) = \frac{1}{n(2^2 - 2)} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^2 - 1 \right] = \frac{1}{2n} \sum_{i=1}^{n} \left[\left(\frac{y_i}{\bar{y}} \right)^2 - 1 \right]$$

Entropy indices for Germany:

year	$E(0)$ Mean log dev.	$E(1)$ Theil	$E(2)$ Half squared CV
2000	.010	.099	.117
2001	.102	.103	.122
2002	.106	.109	.135
2003	.120	.131	.210
2004	.120	.136	.246
2005	.118	.125	.165
2006	.129	.141	.225
2007	.128	.137	.197
2008	.126	.136	.213
2009	.121	.125	.161
2010	.127	.134	.177
2011	.128	.133	.183

Note. Data from SOEP, 2000-2011.

Basic idea. Structural approach to inequality.

- Formulate elementary properties that we think an inequality index ought to have.
- Identify the class of mathematical functions that have these properties.

Principle 1. Weak principle of transfers (WPT).
Under the weak principle of transfers the following is always true:

- Consider any two individuals, one with income y, the other with income $y + \delta$ where $\delta > 0$
- Then transfer a positive amount of income $\Delta > 0$ from the richer to the poorer person, where $\Delta < \delta/2$ (no re-ranking).
- Inequality should then decrease.
- The WPT, however, does not indicate by how much inequality should decrease.

Theorem 4.
Suppose the distribution of income in social state A could be achieved by a simple redistribution of income in social state B (so that total income is the same in each case) and the Lorenz curve for A lies wholly inside that of B. Then, as long as an inequality index satisfies the weak principle of transfers, that inequality index will always indicate a strictly lower level of inequality for state A than for state B.

Principle 2. Income scale independence (ISI).
If everyones income changes by the same proportion, $\lambda > 0$, then the value of the inequality index should remain the same.

Under income scale independence the following is always true:

- Consider a distribution of incomes $(y_1, y_2, ..., y_n)$.
- Then multiply all incomes with the same factor $\lambda > 0$. For example, we do not measure income in EUR but in US dollar.
- Inequality should then remain the same.

Principle 3. Principle of population (PP).
If the distribution y is the k-fold replication of x, that is, $y = (y^1, y^2, ..., y^k)$, where $y^j = x$ for each $j = 1, 2, ..., k$, then the inequality index should remain the same.

Under principle of population the following is always true:

- Given $u = (10, 20, 30)$, let $v = (10, 10, 20, 20, 30, 30)$.
- The principle of population demands that $I(u) = I(v)$.
- The principle of population helps us to view inequality in average terms.
- Hence, principle of population is helpful if we want to compare inequality levels of differently-sized populations, say US and LX.

Principle 4. Decomposability (D) (by subgroups).
Suppose a population can be decomposed in different mutually-exclusive subgroups, $j = 1, ..., k$, of size n_j. Let inequality for subgroup j be denoted I_j, and inequality for the total population I_{total}. An inequality index is decomposable, if there can be found some aggregation function Ξ such that

$$I_{total} = \Xi(I_1, I_2 \ldots, I_k; \overline{y_1}, \overline{y_2} \ldots, \overline{y_k}; n_1, n_2 \ldots, n_k)$$

Under decomposability the following is always true: Total inequality should be a function Ξ of

- inequality in each subgroup, $I_1, ..., I_k$;
- mean income in each subgroup, $\overline{y_1}, \overline{y_2} \ldots, \overline{y_k}$;
- population size of each subgroup;
- **and nothing else**!

Theorem 5.
Any inequality index that simultaneously satisfies the properties of the weak principle of transfers, decomposability, scale independence and the population principle must be expressible either in the form,

$$E_\theta = \frac{1}{\theta^2 - \theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\overline{y}}\right]^\theta - 1\right],$$

or as $J(E_\theta)$, some ordinally-equivalent transformation of E_θ, where θ is a real parameter.

Limitation of Theorem 5.

- Generalized entropy indices *and* ordinally-equivalent indices meet WPT, ISI, PP, and D.
- Accordingly, distributions cannot be ranked in an unambiguous order.
- E.g., for $\theta = 2$ the generalized entropy formula is ordinally equivalent to the variance V and the coefficient of variation c.
- E.g., for $\theta < 1$ the generalized entropy formula is ordinally equivalent to the Atkinson and the Dalton index for distributions with a given mean income.

Principle 5. Strong principle of transfers (SPT).

- Consider two individuals, one with income share s_1, the other with s_2, with $s_2 > s_1$.
- Then transfer a positive amount of income from the richer to the poorer person (no re-ranking).
- Inequality should then decrease.
- The decrease only depends on

$$d = h(s_1) - h(s_2)$$

Corollary of Theorem 5.

Any inequality index that simultaneously satisfies the properties of the strong principle of transfers, decomposability, scale independence and the population principle must be expressible in the form,

$$E_{\theta} = \frac{1}{\theta^2 - \theta}\left[\frac{1}{n}\sum_{i=1}^{n}\left[\frac{y_i}{\overline{y}}\right]^{\theta} - 1\right].$$

Implications of corollary of Theorem 5:

- Theorem 4 shows that WPT amounts to comparing Lorenz curves: Distribution A is less unequal than B if the Lorenz curve of A lies wholly inside that of B.
- The distance concept, $d = h(s_i) - h(s_j)$, gives structure to the valuation of income differences: Defining a particular distance concept results in a class of indices.
- E.g., suppose $h(s_i) = s_i$. Then the absolute differences $d = s_i - s_j$ matter.
- E.g., suppose $h(s_i) = ln(s_i)$. Then the relative log-differences $ln\left(\frac{s_i}{s_j}\right)$ matter.

Principles and indices:

- As explained above, the entropy class meets all the five criteria.
- Others fail in at least one dimension:
  - The Gini index, for example, violates SPT and D.
  - The logarithmic variance and the variance of logarithm violate WPT and thus SPT.
  - The variance and the Dalton index violate ISI.

Index	Principle of Transfers	Distance concept	Decomposable?	Independent of income scale and interval population size?	Range in interval $[0, 1]$?
Variance, V	strong	Absolute differences	Yes	No: increases with income	No
Coeff. of variation, c	weak	As for variance	Yes	Yes	No
Relative mean deviation, M	just fails	0, if incomes on same side of $\bar{y}$, or 1 otherwise	No	Yes	Yes
Logarithmic variance, v	fails	Differences in (log-income)	No	Yes	No
Variance of logarithms, v_1	fails	As for logarithmic variance	No	Yes	No
Gini, G	weak	Depends on rank ordering	No	Yes	Yes
Atkinson index, A_ϵ	weak	Difference in marginal social utilities	Yes	Yes	Yes
Dalton index, D_ϵ	weak	As for Atkinson index	Yes	No	No
Theil index, T	strong	Proportional	Yes	Yes	No
Generalised entropy, E_θ	strong	Power function	Yes	Yes	No

Note: 'just fails' means a rich-to-poor transfer may leave inequality unchanged rather than reducing it.

References to this section

- Atkinson, A. B. (1970). On the measurement of inequality. Journal of Economic Theory, 2, 244263.
- Bosmans, K. (2007). Comparing degrees of inequality aversion. Social Choice and Welfare, 29, 405-428.
- Bourguignon, F. (1979). Decomposable income inequality measures. Econometrica, 47, 901920.
- Ebert, U. (2007). Ethical inequality measures and the redistribution of income when needs differ. Journal of Economic Inequality, 5, 263-278.
- Shorrocks, A. F. and J. E. Foster (1987). Transfer-sensitive inequality measures. Review of Economic Studies 54, 485498.
- Theil, H. (1967). Economics and Information Theory. Amsterdam: North Holland.

Table of Contents

Review and outlook

- Review
  - Up till now we have compressed the information content of an income distribution into a single number or a graph.
  - No preconception of the general pattern which the distribution ought to take.
- Outlook
  - Now we use mathematical formulas that approximate the income distribution.
  - These formulas make comparisons of distributions easier.

Applications.

- Useful for representing particular parts of the income distribution where a distinctive regularity of form is observed.[6]
- Useful for filling in gaps of information if database is incomplete.[7]
- Useful for characterizing the solutions to economic models of the income distribution process.[8]

[6] E.g., Pareto distribution can be used for description of top percentiles.
[7] E.g., imputation of item-non response in survey data.
[8] E.g., linkages between (normal) distributed abilities and labor market outcomes.

Functional form. A functional form of a distribution is a mathematical representation which defines the distribution function or the density function of not just a single distribution, but of a whole family of such distributions. Each family member has common characteristics, identified by certain parameters.

The idea of a model

Example.

- Suppose a particular formula for the density function fits a distribution nicely, i.e.,

$$f = \varphi(y; a, b) \tag{1}$$

- Parameters: a and b (e.g., the mean and the variance).
- The function gives a smooth representation of the density function for any y.
- If the functional form $\varphi(.)$ nicely fits the distributions of interest, then distributional comparisons can build on comparisons of parameters.

The idea of a model

Quantile. Given a functional form, we can compute its quantiles. Quantiles slice a distribution in intervals of equal probability. More generally, a quantile is a function $x = q(p)$ that gives the value of x such that $F(x) = p$:

$$x = q(p) = F^{-1}(p), \tag{2}$$

with $F(.)$ denoting the cumulative distribution function.
Thus, $q(p)$ is the income level below which we find a proportion p of the population.
E.g., the median is $x = q(0.50)$.

Aims of the analysis

- We will study two most important functional forms in distribution analysis:
  1. Lognormal
  2. Pareto
- We will (a) characterize the forms' properties; (b) provide graphical representations; (c) relate the forms to inequality indices; (d) estimate the parameters.

The lognormal distribution

- Lognormal originates from the normal distribution.
- Normal distribution is useful in many applications.

The lognormal distribution

Normal distribution $N(x, \mu, \sigma^2)$. The normal (or Gaussian) distribution is a continuous probability distribution, defined by,

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \text{and} \tag{3}$$

$$F(x, \mu, \sigma) = 0.5 \left(1 + erf\left(\frac{x-\mu}{\sigma 2^{0.5}}\right)\right) \tag{4}$$

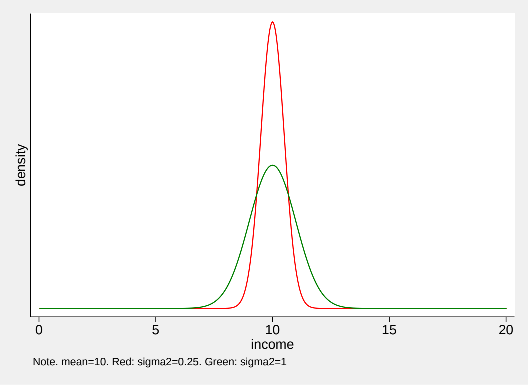
The parameter μ is the mean of the distribution (and also its median and mode); σ the standard deviation; σ^2 its variance.[9]

- $N(x, \mu, \sigma^2)$ peaks at $x = \mu$
- The higher is σ^2, the more 'spread out' will be the distribution.

[9] $erf(x)$ is the Gaussian error function. If $\mu = 0$ and $\sigma = 1$, the distribution is called the standard normal distribution with density $\frac{e^{-0.5x^2}}{(2\pi)^{0.5}}$.

The lognormal distribution

Figure. Normal distribution



The lognormal distribution

- Normal distribution is symmetric around the mean.
- Income distributions, however, usually have a heavy right tail (are positively skewed).
- Hence, the normal distribution is not a good description of income distributions.
- Lognormal distributions do a better job.

The lognormal distribution

Basic idea of the lognormal.

- Suppose we deal with the distribution of income (y).
- Suppose we find that $x = log(y)$ has the normal distribution.
- Then y is said to be lognormally distributed.

The lognormal distribution

Lognormal distribution $\bigwedge(x, \mu, \sigma^2)$. The lognormal distribution is a continuous probability distribution, defined by,

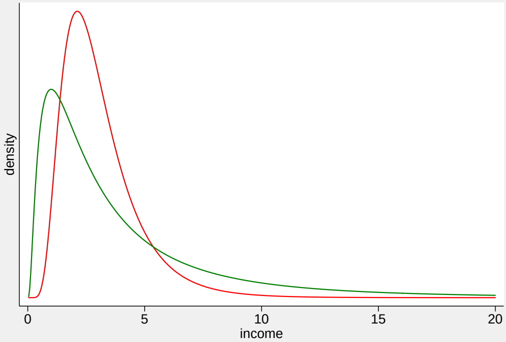
$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}, x > 0, \text{and} \tag{5}$$

$$F(x, \mu, \sigma) = \Phi\left(\frac{ln(x) - \mu}{\sigma}\right), \tag{6}$$

with mean μ, standard deviation σ of the variable's natural logarithm, and Φ denoting the normal.

The lognormal distribution

Figure. Lognormal distribution



Note. mean=1. Red: sigma2=0.25. Green: sigma2=1

The lognormal distribution

Attractive features of lognormal

1. Immediate relationship to standard normal distribution.[10]
2. Symmetrical Lorenz curves
3. Non-intersecting Lorenz curves
4. Immediate interpretation of parameters
5. Preservation under loglinear transformations

[10] The standard normal has $\mu = 0$, and $\sigma = 1$.

Feature 1. Immediate relationship to standard normal distribution

Deriving the lognormal requires only three steps:

1. Compute $x = log(y)$.
2. Standardize x, $z = \frac{x-\mu}{\sigma}$
3. Obtain $f(z)$ $(F(z))$ from the standard normal distribution.

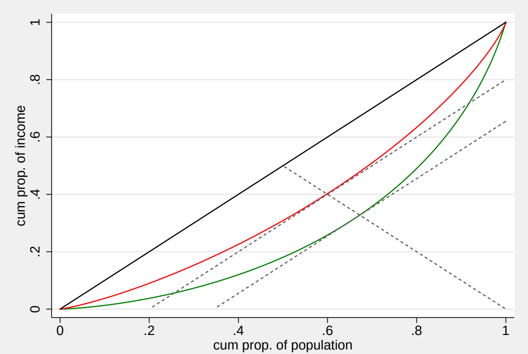
Feature 2. Symmetrical Lorenz curves

- The Lorenz curve from a lognormal distribution is symmetric (around a point that is defined by the mean).[11]
- Hence, if the Lorenz curve is not symmetric, the lognormal will not be a good description of the data.

[11] This is because the lognormal distribution is symmetric on a log-scale of y.

The lognormal distribution

Figure. Lorenz curves from lognormal distributions



Note. mean=1. Red: sigma2=0.25. Green: sigma2=1

Feature 3. Non-intersecting Lorenz curves

- Given, any two members of the lognormal family of distributions, the Lorenz curves do not intersect.
- Hence, one distribution will unambiguously exhibit greater inequality than the other.[12]

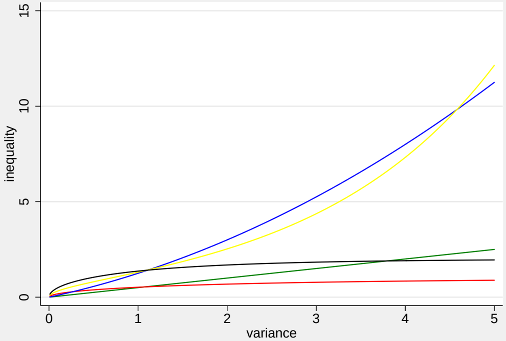
[12] Follows from Theorem 2, which basically states that inequality in a distribution A is lower compared to B if the Lorenz curve of A lies wholly inside the Lorenz curve of B.

Feature 4. Immediate interpretation of parameters

- The lognormal is defined by two parameters, μ and σ^2.
- Max. of Lorenz curve is normalized to one. Hence, μ is irrelevant for an inequality comparison based on Lorenz curves.
- Thus each lognormal-based Lorenz curve ($L(F) = \Phi(\Phi^{-1}(F) - \sigma)$) is uniquely determined by σ^2.[13]
- So σ^2 (or σ) is a satisfactory inequality measure if the distribution is lognormal.

[13] For proof, see Sarabia, J. M. (2008). Parametric Lorenz Curves: Models and Applications. In D. Chotikapanich (Ed.), Modeling Income Distributions and Lorenz Curves (Economic Studies in Inequality), pp. 167-190. Springer.

Figure. Lognormal parameter σ^2 and inequality indices



Note. Green: Theil; Red: Gini; Log var: blue; coef. of var: yellow; rel. mean dev.: black;

Feature 5. Preservation under loglinear transformations

- Suppose x is distributed $N(x; \mu; \sigma^2)$ and $z = a + bx$, then z is normally distributed with $N(z; a + b\mu; b^2\sigma^2)$.[14]
- Suppose y is lognormally distributed $\Lambda(y; \mu; \sigma^2)$ and $w = Ay^b$, then w is lognormally distributed with $\Lambda(w; ln(A) + b\mu; b^2\sigma^2)$.

[14] Remember the basic properties of the variance: (1) $Var(X + a) = Var(X)$; (2) $Var(bX) = b^2 Var(X)$.

Feature 5. Preservation under loglinear transformations (contd.)

- A useful application:
  - In many countries, $t = y - Ay^b$, constant residual progression $(A > 0; 1 > b > 0)$[15], is a reasonable approximation of the income tax.
  - Then disposable income is $w = y - t = y - y + Ay^b = Ay^b$.
  - Suppose y is lognormally distributed, the distribution of after-tax income is also approximately lognormal.

[15] The "residual" is, $res = Ay^b$, and the elasticity, $\frac{\partial res}{\partial y}\frac{\partial y}{\partial res} = b$. The smaller is b, the more progressive is the tax.

The lognormal and inequality indices

Index	Definition	$\Lambda(y;\mu;\sigma^2)$
Variance	$V = \int [y-\bar{y}]^2\, dF$	$e^{2\mu+\sigma^2}\left[e^{\sigma^2}-1\right]$
Coeff. of. var.	$c = \frac{V^{0.5}}{\bar{y}}$	$\left(e^{\sigma^2}-1\right)^{0.5}$
Rel. mean dev.	$M = 0.5\int \lvert\frac{y}{\bar{y}}-1\rvert dF$	$\left[2N\left(\frac{\sigma}{2}\right)-1\right]$
Log. variance	$\upsilon = \int\left[log\left(\frac{y}{\bar{y}}\right)\right]^2 dF$	$\sigma^2+0.25\sigma^4$
Var. of logs	$\upsilon_1 = \int\left[log\left(\frac{y}{y^\star}\right)\right]^2 dF$	σ^2
Gini	$G = 1-2\int \Phi dF$	$2N\left(\frac{\sigma}{2^{0.5}}\right)-1$
Atkinson	$A_\epsilon = 1-\left[\int\left(\frac{y}{\bar{y}}\right)^{1-\epsilon} dF\right]^{\frac{1}{1-\epsilon}}$	$1-e^{-0.5\epsilon\sigma^2}$
Gen. entropy[a]	$E_\theta = \frac{1}{\theta^2-\theta}\left[\int\left(\frac{y}{\bar{y}}\right)^\theta dF-1\right]$	$\frac{e^{0.5\left[\theta^2-\theta\right]\sigma^2}-1}{\theta^2-\theta}$
Mean log dev.	$L = \int \log\left(\frac{y}{\bar{y}}\right) dF = E_0$	$\frac{\sigma^2}{2}$
Theil	$T = \int \frac{y}{\bar{y}} log\left(\frac{y}{\bar{y}}\right) dF = E_1$	$\frac{\sigma^2}{2}$

Note: [a] $\theta \neq 0,1$; Λ is the lognormal distribution.

The lognormal distribution: Estimation with SOEP

Basics

1. Database: SOEP 2011
2. Income concept: Disposable household income per capita
3. Weighting: by size and SOEP frequency weights

Estimation with maximum likelihood (ML)

1. $x_1, ..., x_n$: Independent and identically distributed incomes from a distribution with an unknown probability density function $f_0(.)$.
2. We surmise that f_0 belongs to the lognormal with the parameter vector $\theta = (\mu, \sigma^2)$: the parametric model.
3. θ_0: unknown true parameter vector.
4. With ML we find an estimator vector $\hat{\theta}$ for θ_0.

Estimation with maximum likelihood (ML) (cont.)

Step 1. Specify the joint density function for all observations for the IID sample:

$$f(x_1, x_2, \ldots, x_n \mid \theta) = f(x_1|\theta) \times f(x_2|\theta) \times \cdots \times f(x_n|\theta). \qquad (7)$$

Step 2. An alternative view is to consider the $x_1, x_2, ..., x_n$ to be "parameters" of $f(.)$, whereas θ will be the function's variable. This function is the likelihood,

$$\mathcal{L}(\theta \,|\, x_1, \ldots, x_n) = f(x_1, x_2, \ldots, x_n \mid \theta) = \prod_{i=1}^{n} f(x_i|\theta), \qquad (8)$$

The lognormal distribution: Estimation with SOEP

Estimation with maximum likelihood (ML) (cont.)

Equivalently, the log of the likelihood is,

$$\ln \mathcal{L}(\theta \mid x_1, \ldots, x_n) = \sum_{i=1}^{n} \ln f(x_i|\theta), \tag{9}$$

and the average log-likelihood is,

$$\hat{\ell} = \frac{1}{n} \ln \mathcal{L}. \tag{10}$$

ML estimates θ_0 by finding a value of θ that maximizes $\hat{\ell}(\theta|x)$.

The lognormal distribution: Estimation with SOEP

Required ado-packages: lognfit; lognpred; plogn; qlogn

Central command lines:

Estimate parameters; store density and cumulative density as variables

lognfit $inc[w = W]$, pdf(pdffit) cdf(cdffit)

$\star$ Save parameters μ and σ^2:

matrix list $e(b)$

scalar $sc\_sigma = \_b[v : \_cons]$

scalar $sc\_mu = \_b[m : \_cons]$

scalar $sc\_sigma2 = sc\_sigma^2$

$\star$ Visual inspection of fit

local $mu = sc\_mu$

local $sigma = sc\_sigma$

qlogn inc, param(‘mu’ ‘sigma’)

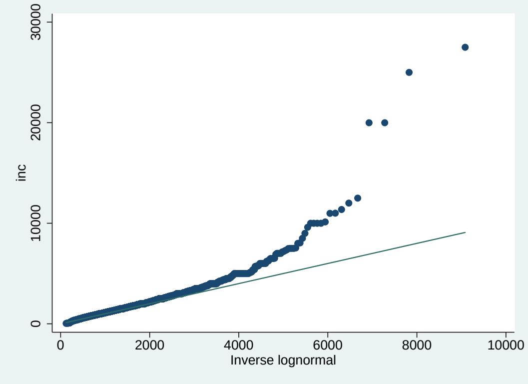
Figure. SOEP estimates of lognormal parameters

```
ML fit of lognormal distribution                 Number of obs   =      11698
                                                 Wald chi2(0)    =          .
Log likelihood = -91040.063                      Prob > chi2     =          .

------------------------------------------------------------------------------
         inc |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
m            |
       _cons |   6.968331   .0050502  1379.81   0.000     6.958433     6.97823
-------------+----------------------------------------------------------------
v            |
       _cons |   .5462178    .003571   152.96   0.000     .5392187    .5532169
------------------------------------------------------------------------------
```

The lognormal distribution: Estimation with SOEP

Figure. Symmetry plot



The lognormal distribution: Inequality indices from parameters

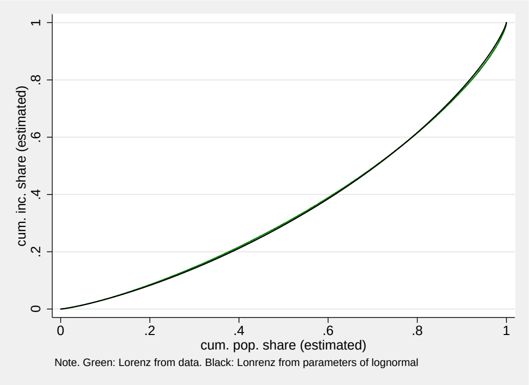
Table. Observed and estimated inequality indices

Index	From parameter est.	From plain data
Gini	0.301	0.299
Theil	0.149	0.160

Note. Data is SOEP 2011.

The lognormal distribution: Estimation with SOEP

Figure. Lorenz curve from lognormal



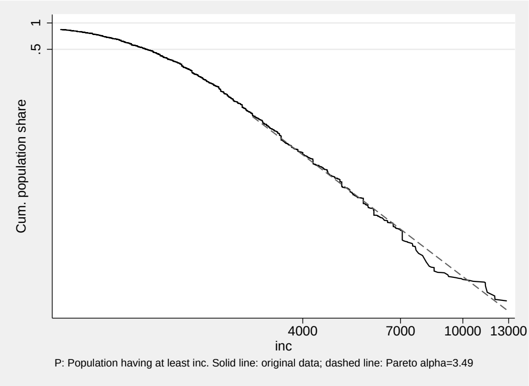
The Pareto distribution

Pattern observed for many countries and points in time.

- For the upper income range, a log-scale plot of $P = 1 - F(y)$ (ordinate) against y looks close to a straight line.
- The absolute value of the slope of the line is known as Pareto's α.

The Pareto diagram

Figure. Pareto distribution



Note. SOEP 2011. Equivalent income (square-root scale); size weighted.

The Pareto distribution

Pareto distribution $\Pi(y, \underline{y}, \alpha)$. The Pareto distribution is a continuous probability distribution, defined by,

$$f(y, \underline{y}, \alpha) = \frac{\alpha \underline{y}^{\alpha}}{y^{\alpha+1}}, \tag{11}$$

and

$$F(y, \underline{y}, \alpha) = 1 - \left[\frac{\underline{y}}{y}\right]^{\alpha}. \tag{12}$$

Parameter α is Pareto's α; $\underline{y}$ is a lower-income bound.

The Pareto distribution

Attractive features of Pareto distribution

1. Linearity of Pareto diagram
2. Van der Wijk's law
3. Non-intersecting Lorenz curves
4. Immediate interpretation of parameters
5. Preservation under loglinear transformations

The Pareto distribution

Feature 1. Linearity of Pareto diagram

- From $F(y) = \Pi(y, \underline{y}, \alpha) = 1 - \left(\frac{\underline{y}}{y}\right)^{\alpha}$ it is easy to compute
  - the proportion of the population having y or less;
  - the Lorenz curve $\Phi = 1 - [1 - F(y)]^{\frac{\alpha-1}{\alpha}}$; and
  - average income $\bar{y} = \frac{\alpha}{\alpha-1}\underline{y}$.

The Pareto distribution

Feature 2. Van der Wijk's law

- Take an arbitrary base income, y_b.
- Then the average income of the population with an income of y_b or larger is $By_b = \frac{\alpha}{\alpha - 1} y_b$.

Feature 2. Van der Wijk's law

- The proportionality $B = \frac{\alpha}{\alpha - 1}$ between average income equal or above y_b and y_b can be interpreted as an inequality index: **average to base index**.
- The larger is α, the smaller is B. So, the relative gap between your income (y_b) and the average income of all richer units gets smaller.

Proof of Van der Wijk's law

- Average income of everyone equal or above y_b is,

$$z(y_b) = \frac{\int_{y_b}^{\infty} uf(u)du}{\int_{y_b}^{\infty} f(u)du} = \bar{y}\frac{(1-\Phi(y_b))}{1-F(y_b)} \tag{13}$$

  - Nominator $((1-\Phi(y_b)))$: income share of population with y_b or larger
  - Denominator $(1-F(y_b))$: fraction of the population with y_b or larger
- Plug in the definitions for $\bar{y}$, Φ and F:[16]

[16] Substituting $\bar{y}$ and Φ gives
$z(y_b) = \frac{\alpha}{\alpha-1}\underline{y}\frac{(1-F(y_b))^{\frac{\alpha-1}{\alpha}}}{1-F(y_b)} = \frac{\alpha}{\alpha-1}\underline{y}\left(1-F(y_b)\right)^{-1/\alpha}$. Substituting F gives
$z(y_b) = \frac{\alpha}{\alpha-1}\underline{y}\left(\left(\underline{y}/y_b\right)^{\alpha}\right)^{-1/\alpha} = \frac{\alpha}{\alpha-1}y_b$.

Proof of Van der Wijk's law

- $$z(y_b) = \frac{\alpha}{\alpha - 1}\underline{y}\left[\frac{\underline{y}}{y_b}\right]^{\alpha-1}\left[\frac{\underline{y}}{y_b}\right]^{-\alpha} \tag{14}$$
  $$z(y_b) = \frac{\alpha}{\alpha - 1}y_b \tag{15}$$
- So the average income above y_b is proportional to y_b.

Feature 3. Non-intersecting Lorenz curves

- Given any two members of the Pareto family of distributions, the Lorenz curves, $\Phi = 1 - [1 - F(y)]^{\frac{\alpha-1}{\alpha}}$, do not intersect.
  - **Proof.** Suppose we chose an arbitrary value of F for a given α_0. We need to check how $[1 - F(y)]^{\frac{\alpha-1}{\alpha}}$ varies with α. The derivative is:[17] $[1 - F(y)]^{\frac{\alpha-1}{\alpha}} \, ln \, [1 - F(y)] \, \frac{1}{\alpha^2} < 0$.
  - Increasing α means that $[1 - F(y)]^{\frac{\alpha-1}{\alpha}}$ gets smaller and Φ larger: For $\alpha_1 > \alpha_0$, $L(\alpha_1) > L(\alpha_0)$. Q.e.d.

[17] The derivative of $y = a^{f(x)}$ with respect to x is $y' = f'(x) a^{f(x)} ln(a)$.

Feature 4. Immediate interpretation of parameters

- Since Pareto distributions do not cross and $L(\alpha_1) > L(\alpha_0)$ for $\alpha_1 > \alpha_0$, it follows:
- Any scale invariant measure indicates less inequality in the distribution with the higher α.[18]

[18] See Theorem 4.

Feature 5. Preservation under loglinear transformations

- Suppose y is Pareto distributed with parameters $\underline{y}$ and α, so that $F(y) = 1 - \left[\frac{y}{\underline{y}}\right]^{-\alpha}$.
- Suppose we have a constant residual progression tax, $t = y - Ay^b$, so that disposable income is $w = Ay^b$.
- Minimum disposable income is, $\underline{w} = A\underline{y}^b$.
- Since $\frac{w}{\underline{w}} = \frac{Ay^b}{A\underline{y}^b} = \left(\frac{y}{\underline{y}}\right)^b \Rightarrow \frac{y}{\underline{y}} = \left(\frac{w}{\underline{w}}\right)^{\frac{1}{b}}$
- Substituting F, it follows $F(w) = 1 - \left[\frac{w}{\underline{w}}\right]^{-\frac{\alpha}{b}}$. [19] So, w is again Pareto distributed.

[19] From F it follows that $\frac{y}{\underline{y}} = (1-F)^{\frac{-1}{\alpha}}$. Now, in the equation replace $\frac{y}{\underline{y}}$ by $(1-F)^{\frac{-1}{\alpha}}$.

Inequality indices and functional forms

Index	Definition	$\Lambda(y;\mu;\sigma^2)$	$\Pi(y;\underline{y};\alpha)$
Variance	$V = \int [y-\bar{y}]^2\, dF$	$e^{2\mu+\sigma^2}\left[e^{\sigma^2}-1\right]$	$\frac{\alpha \underline{y}^2}{(\alpha-1)^2(\alpha-2)}$
Coeff. of. var.	$c = \frac{V^{0.5}}{\bar{y}}$	$\left(e^{\sigma^2}-1\right)^{0.5}$	$\left(\frac{1}{\alpha(\alpha-2)}\right)^{0.5}$
Rel. mean dev.	$M = 0.5\int \lvert\frac{y}{\bar{y}}-1\rvert dF$	$\left[2N\left(\frac{\sigma}{2}\right)-1\right]$	$\frac{(\alpha-1)^{\alpha-1}}{\alpha^{\alpha}}$
Log. variance	$\upsilon = \int \left[log\left(\frac{y}{\bar{y}}\right)\right]^2 dF$	$\sigma^2+0.25\sigma^4$	$log\left(\frac{\alpha-1}{\alpha}+\frac{1}{\alpha}+\frac{1}{\alpha^2}\right)$
Var. of logs	$\upsilon_1 = \int \left[log\left(\frac{y}{y^\star}\right)\right]^2 dF$	σ^2	$\frac{1}{\alpha^2}$
Gini	$G = 1-2\int \Phi dF$	$2N\left(\frac{\sigma}{2^{0.5}}\right)-1$	$\frac{1}{2\alpha-1}$
Atkinson	$A_\epsilon = 1-\left[\int\left(\frac{y}{\bar{y}}\right)^{1-\epsilon} dF\right]^{\frac{1}{1-\epsilon}}$	$1-e^{-0.5\epsilon\sigma^2}$	$1-\left[\frac{\alpha-1}{\alpha}\right]\left[\frac{\alpha}{\alpha+\epsilon-1}\right]^{\frac{1}{1-\epsilon}}$
Gen. entropy[a]	$E_\theta = \frac{1}{\theta^2-\theta}\left[\int\left(\frac{y}{\bar{y}}\right)^{\theta} dF-1\right]$	$\frac{e^{0.5\left[\theta^2-\theta\right]\sigma^2}-1}{\theta^2-\theta}$	$\frac{1}{\theta^2-\theta}\left[\left[\frac{\alpha-1}{\alpha}\right]^{\theta}\frac{\alpha}{\alpha-\theta}-1\right]$
Mean log dev.	$L = \int \log\left(\frac{y}{\bar{y}}\right) dF = E_0$	$\frac{\sigma^2}{2}$	$log\left(\frac{\alpha}{\alpha-1}\right)-\frac{1}{\alpha}$
Theil	$T = \int \frac{y}{\bar{y}} log\left(\frac{y}{\bar{y}}\right) dF = E_1$	$\frac{\sigma^2}{2}$	$log\left(\frac{\alpha-1}{\alpha}\right)+\frac{1}{\alpha-1}$

Note: [a] $\theta \neq 0, 1$; Λ is the lognormal distribution; Π is the Pareto distribution.

The Pareto distribution: Estimation with SOEP

Required ado-packages: none

Central command lines:

Estimate parameters; store density and cumulative density as variables

```
sort inc
sum W
scalar sc_N = r(sum)
generate cum_pop_share= sum(W)/sc_N
generate P = 1−cum_pop_share
gen lnP = ln(P)
gen lninc = ln(inc)
reg lnP lninc if inc > 3500 [w = W]
matrix list e(b)
scalar sc_alpha=_b[lninc]
```

The Pareto distribution: Estimation with SOEP

Figure. SOEP estimation of Pareto parameter

```
      Source |       SS       df       MS              Number of obs =     888
-------------+------------------------------           F(  1,   886) =88683.29
       Model |  850.485743     1  850.485743           Prob > F      =  0.0000
    Residual |  8.49687001   886  .009590147           R-squared     =  0.9901
-------------+------------------------------           Adj R-squared =  0.9901
       Total |  858.982613   887  .968413318           Root MSE      =  .09793

------------------------------------------------------------------------------
         lnP |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lninc |  -3.488937   .0117158  -297.80   0.000    -3.511931   -3.465943
       _cons |   25.47338   .0988367   257.73   0.000      25.2794    25.66736
------------------------------------------------------------------------------
```

The Pareto distribution: Inequality indices

Table. Estimates of inequality indices

Index	$\hat{\alpha}$	Plain data	Plain data ($y \geq 3500$)
Gini	0.167	0.272	0.170
Theil	0.064	0.133	0.074

Note. Data is SOEP 2011. Equivalent income using square-root scale.

Three flexible functional forms

- Generalized Gamma distribution

$$f(y; a, \beta, p) = \frac{ay^{ap-1}e^{-\left(\frac{y}{\beta}\right)^a}}{\beta^{ap}\Gamma(p)} \quad 0 \leq y \tag{16}$$

- Generalized Beta distribution of the first kind

$$g(y; a, b, p, q) = \frac{ay^{ap-1}\left(1 - \left(\frac{y}{b}\right)^a\right)^{q-1}}{b^{ap}B(p, q)} \quad 0 \leq y \leq b \tag{17}$$

- Generalized Beta distribution of the second kind

$$h(y; a, b, p, q) = \frac{ay^{ap-1}}{b^{ap}B(p, q)\left(1 + \left(\frac{y}{b}\right)^a\right)^{p+q}} \quad 0 \leq y \tag{18}$$

Three flexible functional forms

- GB1 and GB2 are four-parameter distributions and contain the three-parameter GG (and a number of other distributions) as a special case.
- Included special cases: lognormal, exponential, Singh-Maddala (having the Pareto type II distribution as limiting case ($a = 1$)
- Strategy:
  - Estimate GB1 and GB2.
  - Perform test on parameters to see if limiting case applies.

Three flexible functional forms

Figure. Distribution tree

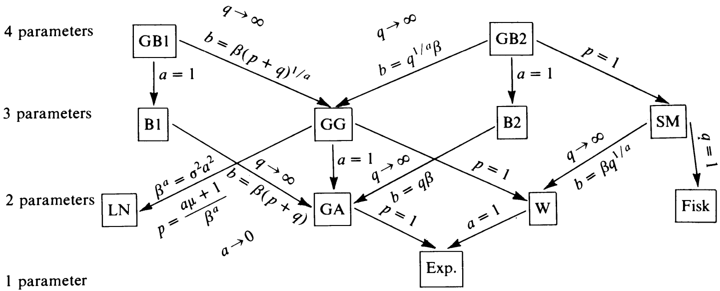


FIGURE 1—Distribution trees.

B1(2): beta of the first (second) kind; LN: lognormal; GA: gamma; W: Weibul; SM: Singh-Maddala.

Source: McDonald (1984), p. 648.

A concluding remark

Warning.

- Functional forms are convenient, but they do not do miracles.
- Usually, a functional form does not fit equally well over the whole income range. For example, sometimes it fits nicely over a wide income range but not at the distribution's bottom or top.
- Sensitivity to specification
  - Example: lower income bound for Pareto distribution is set by the researcher.
  - With the SOEP 2011 data (eq. disp. income per quarter), we obtain $\alpha(\underline{y} = 2000) = 3.460$ $\alpha(\underline{y} = 3000) = 3.555$; $\alpha(\underline{y} = 9000) = 3.478$.

References to this section

- Freund, J.E., and Perles, B.M. (2007): Modern Elementary Statistics, New Jersey: Pearson Prentice Hall.
- Jenkins, S.P. (2004): Fitting Functional Forms to Distributions using ml. Presentation at Second German Stata Users Group Meeting, Berlin. http://www.stata.com/meeting/2german/Jenkins.pdf
- Kleiber, C., and Kotz, S. (2003): Statistical Size Distributions in Economics and Actuarial Sciences. Hoboken, NJ: John Wiley.
- McDonald, J.B. (1984): Some Generalized Functions for the Distribution of Income, Econometrica, 52, 647-664.
- Pareto, V. (2001): On the Distribution of Wealth and Income. In: Baldassarri, M., and Ciocca, P. (Eds.), Roots of the Italian School of Economics and Finance: From Ferrara (1857) to Einaudi (1944), Vol. 2, 231-276, Houndmills: Palgrave.
- Sarabia, J.M. (2008): Modeling Income Distributions and Lorenz Curves, in: Volume 5 of the series Economic Studies in Equality, Social Exclusion and Well-Being, 167-190.

Table of Contents

The basic idea of resampling methods.

- Estimate the precision of a statistic (say, the Gini index),
    1. by using a subset of the original database (**jackknife**), or
    2. by drawing random samples from the original database (**bootstrap**)

Bootstrap

The basic idea of the bootstrap.

- The original population is unknown as our sample data is only an excerpt of it.
- Inference about a population from the sample data, however, can be modeled by resampling the sample data and perform inference thereon.
- In bootstrap resamples, the sample is treated as the population, and hence the quality from the resample data is measurable.

Procedure.

- Suppose our sample data is of size N with population distribution $\mathfrak{F}$.
- We want to understand the variability of a statistic $\hat{S} = S(\mathfrak{F})$ (say, the Gini index).
- The procedure is as follows:

Bootstrap

1. Generate a bootstrap sample, b, from the sample data by resampling with replacement[20], so that the bootstrap sample size is again N.
2. Compute the bootstrap statistic, S_b, from the bootstrap sample, b.
3. Repeat the procedure a large number of times (usually $B = 1,000$ repetitions)
4. From the boostrap statistics, $\{S_b\}_{b=1}^{B}$ we can assess the variability of the statistic $\hat{S}$ based on the sample data.

[20] Replacement allows that an observation can be included in the bootstrap sample several times.

Inspection of variability.

1. Graphical: Histogram of $\{S_b\}_{b=1}^{B}$
2. Confidence intervals (CI)

CI type 1: normal approximation CI.

- The symmetric $1-\alpha$ confidence interval is,

$$\left(\hat{S} - SD(\{S_b\}_{b=1}^{B}) z_{(1-\alpha)/2}; \hat{S} + SD(\{S_b\}_{b=1}^{B}) z_{(1-\alpha)/2}\right), \quad (19)$$

  where $z_{(1-\alpha)/2}$ is defined such that $N(z_{(1-\alpha)/2}) = 1 - \frac{\alpha}{2}$ and $N(.)$ denotes the standard normal distribution, and SD is the standard deviation from the bootstrap statistics.[21]
- Justifiable if the estimator to obtain $\hat{S}$ has an asymptotically normal distribution (as most estimators do).

[21] For example, for computing the 95 percent normal CI, it is $\hat{S} \pm 1.96 SD(\{S_b\}_{b=1}^{B})$.

CI type 2: percentile CI.

- The asymmetric $1-\alpha$ confidence interval is,

$$\left(c_{\alpha/2}(\{S_b\}_{b=1}^{B}); c_{(1-\alpha)/2}(\{S_b\}_{b=1}^{B})\right), \tag{20}$$

  where $c_{\alpha/2}$ and $c_{(1-\alpha)/2}$ are the centiles of $(\{S_b\}_{b=1}^{B})$
- Typically $\hat{S}$ will not lie at the midpoint of $c_{\alpha/2}$ and $c_{(1-\alpha)/2}$.
- Therefore, the asymmetric percentile CI will provide a better coverage than the normal approximation CI when $\{S_b\}_{b=1}^{B}$ is skewed.

The Bootstrap: Estimation with SOEP

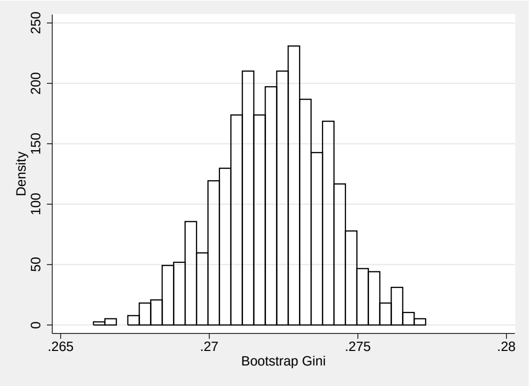
Required ado-packages: none

Central command lines:

```
global B = 1000
gen giniboot = .
gen boot=_n
forvalues b = 1(1)$B{
        qui gen w`b'= runiform()
        qui sum w`b'
        qui replace w`b'=w`b'/r(mean)
        qui inequal7 inc [w=hweight*w`b']
        qui replace giniboot=`r(gini)' if boot==`b'
        qui drop w`b'
}
```

Boostrap

Figure. Histogram of Bootstrap Gini coefficients



Note. SOEP 2011. Equivalent income (square-root scale); size weighted.

Table. Bootstrap confidence intervals for Gini

	CI_{lo}	Gini (obs.ed)	CI_{up}
Normal CI	.269	.272	.276
Percentile CI	.268	.272	.276

Note. Data is SOEP 2011; equivalent disposable income (square-root scale); size weighted.

Jackknife

The basic idea of the Jackknife.

- Similar to Bootstrap: resample and derive statistics from the Jacknife samples.
- Different is the resampling: Jackknife samples are derived by leaving out one observation at a time from the sample set.

Example.

- Suppose we have an income distribution (see next page).
- We compute the Theil index from the sample and the Theil indices from the Jackknife samples.

Table. Estimates of inequality indices

y	W	N	$\bar{y}$	T	N_{JK}	$\bar{y}_{JK}$	T_{JK}
50	1	12	291.667	.139	11	313.636	.094
100	1	12	291.667	.139	11	309.091	.116
150	1	12	291.667	.139	11	304.545	.131
200	1	12	291.667	.139	11	300	.142
250	2	12	291.667	.139	10	300	.159
300	1	12	291.667	.139	11	290.909	.152
350	1	12	291.667	.139	11	286.364	.152
400	1	12	291.667	.139	11	281.818	.150
450	1	12	291.667	.139	11	277.273	.146
500	2	12	291.667	.139	10	250	.133

Note. W is the frequency weight.

The average over T_{JK} is .137, its standard deviation is .0199.

Confidence interval.

- Following Wolter (1985), the Jackknife variance estimator is,

$$V_{jk} = \frac{N-1}{N} * \sum_{JK=1}^{N} (S_{JK} - \hat{S})^2, \tag{21}$$

  with N denoting the number of Jackknife replications, and S_{JK} the Jacknife statistic leaving out observation JK.[22]

- The Jackknife standard deviation and standard error is,

$$SD_{jk} = (V_{jk})^{0.5} \text{ and } SE_{jk} = \frac{SD_{jk}}{N^{0.5}} \tag{22}$$

[22] In case of weighted data, it $V_{jk} = \frac{N^w - 1}{N^w} * \sum_{JK=1}^{N} w_k (S_{JK} - \hat{S})^2$ with $N^w = \sum_k w_k$.

Confidence interval (cont.).

- Then the 95 percent CI for large N is:

$$CI_{lo} = \hat{S} - 1.96SD_{jk}; CI_{up} = \hat{S} + 1.96SD_{jk} \tag{23}$$

The Jackknife: Estimation with SOEP

Table. Confidence intervals for Gini

	CI_{lo}	Gini	CI_{up}
Boot Normal CI	.269	.272	.276
Boot Percentile CI	.268	.272	.276
Jackknife CI	.261	.272	.283

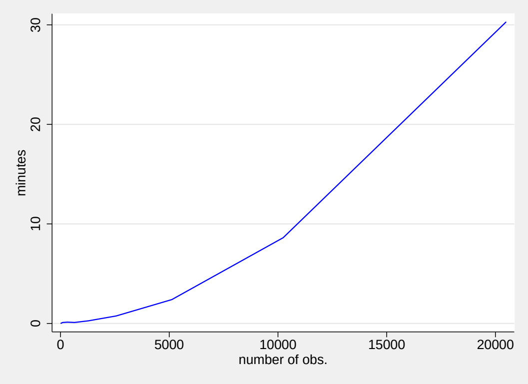
Note. Data is SOEP 2011; equivalent disposable income sq-scale.

Comparing the Bootstrap and the Jackknife

- Both methods estimate the variability of a statistic from the variability of that statistic between subsamples (rather than from parametric assumptions).
- Usually, both give rather similar results.
- Advantages of the Jackknife over the Bootstrap:
  1. Results are replicable as it does not rely on (random) bootstrap weights.
  2. Bootstrap is time intensive for large samples (see next pages).
  3. Shortcuts for jackknife help to reduce the computational burden.

Computational burden for bootstrap

Figure. Computing time for bootstrap



Note. Entropy, Gini, Atkinson index are computed for artificial dataset.

Shortcuts for the Jackknife

- At first glance, computing Jacknife statistics for large sample seems rather time consuming:
  - For a sample of size N (N denotes the un-weighted number of obs.), N statistics need to be computed.
- Clever programming, however, can heavily reduce the computational burden.
- Next page gives an example for the mean.

Shortcuts for the Jackknife

Suppose we have a distribution of incomes, y, and W is the sampling weight. We want to obtain mean incomes from all Jackknife samples.

This can be achieved with only 2 (not N) runs through the data:

- Run 1:
  - sum y $[w = W]$
  - scalar $scsum = r(sum)$
  - scalar $scN = r(sum\_W)$
- Run 2:
  - gen $jkmean = (scsum - y * W)/(scN - W)$

Table. Jackknife of the mean

y	W	$\bar{y}_{JK}$
50	1	$(500 - 1 * 50)/(5 - 1) = 112.5$
100	1	$(500 - 1 * 100)/(5 - 1) = 100.5$
150	1	$(500 - 1 * 150)/(5 - 1) = 87.5$
100	2	$(500 - 2 * 100)/(5 - 2) = 100$
500	5	

Shortcuts for the Jackknife

Basic strategy (see Karoly and Schroeder (2015)):

1. Compute statistic of interest, θ, and some auxiliary statistics from the full sample.
2. Write statistic of interest without obs. i, $\theta_{(i)}$, as a function of
   - θ,
   - auxiliary statistics from full sample, and
   - characteristics of i.

Shortcuts for the Jackknife

Computing $\theta_{(i)}$ requires information on a single observation (i) and not $N-1$ observations.

Shortcuts for the Jackknife

Theil index from overall sample

$$\theta_T = \frac{1}{N\bar{y}} \left(\sum_{i=1}^{N} w_i y_i ln(y_i) \right) - ln(\bar{y}) \tag{24}$$

with

- normalized weight: $w_i = \frac{\omega_i}{\frac{1}{N}\sum_{i=1}^{N} \omega_i}$,
- number of observations: $N = \sum_{i=1}^{N} \omega_i$
- arithmetic mean of income: $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} w_i y_i$

Theil index from subset with *i*th observation left out

$$\theta_{T(i)} = \frac{1}{(N - w_i)\bar{y}_{(i)}} \left(\sum_{j \neq i} w_j y_j ln(y_j) \right) - ln(\bar{y}_{(i)}) \qquad (25)$$

with

$$\bar{y}_{(i)} = \frac{N\bar{y} - w_i y_i}{N - w_i} \qquad (26)$$

Initially, from (25)

$$\theta_{T(i)} = \frac{1}{(N-w_i)\bar{y}_{(i)}}\left(\sum_{j\neq i} w_j y_j ln(y_j)\right) - ln(\bar{y}_{(i)})$$

$$\theta_{T(i)} = \frac{1}{(N-w_i)\bar{y}_{(i)}}\left(\sum_{i=1}^{N} w_i y_i ln(y_i)\right) - \frac{w_i y_i ln(y_i)}{(N-w_i)\bar{y}_{(i)}} - ln(\bar{y}_{(i)}) \quad (27)$$

Shortcuts for the Jackknife

Substituting (26), $\bar{y}_{(i)} = \frac{N\bar{y}-w_i y_i}{N-w_i}$, into (27) gives

$$\theta_{T(i)} = \frac{1}{(N-w_i)\bar{y}_{(i)}}\left(\sum_{i=1}^{N} w_i y_i ln(y_i)\right) - \frac{w_i y_i ln(y_i)}{(N-w_i)\bar{y}_{(i)}} - ln(\bar{y}_{(i)})$$

$$\theta_{T(i)} = \frac{1}{(N-w_i)\frac{N\bar{y}-w_i y_i}{N-w_i}}\left(\sum_{i=1}^{N} w_i y_i ln(y_i)\right) - \frac{w_i y_i ln(y_i)}{(N-w_i)\frac{N\bar{y}-w_i y_i}{N-w_i}} - ln\left(\frac{N\bar{y}-w_i y_i}{N-w_i}\right)$$

$$\theta_{T(i)} = \frac{1}{N\bar{y}-w_i y_i}\left(\sum_{i=1}^{N} w_i y_i ln(y_i)\right) - \frac{w_i y_i ln(y_i)}{N\bar{y}-w_i y_i} - ln\left(\frac{N\bar{y}-w_i y_i}{N-w_i}\right) \quad (28)$$

Rewriting (24) as

$$\sum_{i=1}^{N} w_i y_i ln(y_i) = [\theta_T + ln(\bar{y})]\, N\bar{y} \tag{29}$$

and substituting into (28) gives

$$\theta_{T(i)} = \frac{1}{N\bar{y} - w_i y_i}\left(\sum_{i=1}^{N} w_i y_i ln(y_i)\right) - \frac{w_i y_i ln(y_i)}{N\bar{y} - w_i y_i} - ln\left(\frac{N\bar{y} - w_i y_i}{N - w_i}\right)$$

$$\theta_{T(i)} = \frac{N\bar{y}}{N\bar{y} - w_i y_i}(\theta_T + ln(\bar{y})) - \frac{w_i y_i ln(y_i)}{N\bar{y} - w_i y_i} - ln\left(\frac{N\bar{y} - w_i y_i}{N - w_i}\right) \tag{30}$$

Shortcuts for the Jackknife

Eq. (30), reveals: $\theta_{T(i)}$ can be expressed as a function of

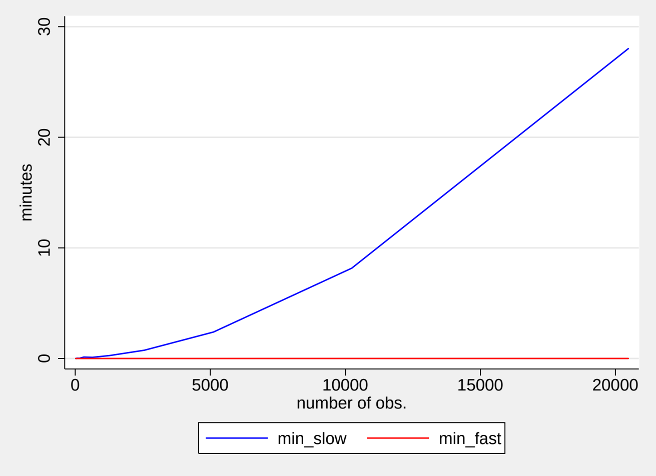
1. θ_T,
2. auxiliary statistics from full sample, N, $\bar{y}$, and
3. characteristics of i: w_i, y_i.

$$\theta_{T(i)} = \frac{N\bar{y}}{N\bar{y} - w_i y_i}(\theta_T + ln(\bar{y})) - \frac{w_i y_i ln(y_i)}{N\bar{y} - w_i y_i} - ln(\frac{N\bar{y} - w_i y_i}{N - w_i})$$

Shortcuts for the Jackknife

Hence, to compute $\theta_{T(i)}, \ldots, \theta_{T(N)}$ requires one additional pass through the data.

Quantitative assessment of the computer time reduction



Shortcuts for the Jackknife

Karoly and Schroeder (2015) provide shortcuts for:

1. Variance
2. Coefficient of variation
3. Variance of logs
4. Mean log deviation
5. Atkinson index[23]

Shortcuts for Gini are provided in Karagiannis and Kovacevic (2000) and Yitzhaki (1991).

[23] Yitzhaki (1991) provides an efficient procedure for the Gini index.

Shortcuts for the Jackknife

$$\theta_V = \frac{1}{N-1} \sum_{i=1}^{N} w_i (y_i - \bar{y})^2 \tag{31}$$

$$\theta_{V(i)} = \theta_V \frac{(N-1)^2}{(N-2)(N-w_i)} - \frac{N w_i (N-1)}{(N-w_i)^2 (N-2)} (\bar{y} - y_i)^2 \tag{32}$$

Shortcuts for the Jackknife

$$\theta_{CV} = \frac{(\theta_V)^{0.5}}{\bar{y}} \tag{33}$$

$$\theta_{CV(i)} = \frac{(\theta_V \frac{(N-1)^2}{(N-2)(N-w_i)} - \frac{Nw_i(N-1)}{(N-w_i)^2(N-2)}(\bar{y} - y_i)^2)^{0.5}}{\frac{1}{N-w_i}(N\bar{y} - y_i w_i)} \tag{34}$$

Shortcuts for the Jackknife

$$\theta_{VL} = \frac{1}{N-1}\sum_{i=1}^{N} w_i(x_i - \bar{x})^2 \tag{35}$$

$$\theta_{VL(i)} = \theta_{VL}\frac{(N-1)^2}{(N-2)(N-w_i)} - \frac{Nw_i(N-1)}{(N-w_i)^2(N-2)}(\bar{x} - x_i)^2 \tag{36}$$

with: $\bar{x} = ln(y^*)$; $y^* = exp(\frac{1}{N}\sum_{i=1}^{N} w_i ln(y_i))$; $x_i = ln(y_i)$

Shortcuts for the Jackknife

$$\theta_{MLD} = \frac{1}{N} \sum_{i=1}^{N} w_i ln(\frac{\bar{y}}{y_i}) \tag{37}$$

$$\theta_{MLD(i)} = \frac{N}{N - w_i}(\theta_{MLD} - ln(\bar{y})) + \frac{w_i ln(y_i)}{N - w_i} + ln(\frac{N\bar{y} - y_i w_i}{N - w_i}) \tag{38}$$

Shortcuts for the Jackknife

$$\theta_{A_1} = 1 - \frac{y^*}{\bar{y}} = 1 - \frac{exp\left[\frac{1}{N}\sum_{i=1}^{N} w_i ln(y_i)\right]}{\bar{y}} \tag{39}$$

$$\theta_{A_1(i)} = 1 - \frac{exp(\frac{N}{N-w_i} ln(y^*) - \frac{ln(y_i)w_i}{N-w_i})}{\frac{N\bar{y}-w_i y_i}{N-w_i}} \tag{40}$$

Shortcuts for the Jackknife

$$\theta_{A_2} = 1 - \frac{N}{\sum_{i=1}^{N} w_i \frac{\bar{y}}{y_i}} \tag{41}$$

$$\theta_{A_2(i)} = 1 - \frac{N - w_i}{\frac{N\bar{y} - w_i y_i}{\bar{y}(N - w_i)} \frac{N}{1 - \theta_{A_2}} - \frac{w_i(N\bar{y} - w_i y_i)}{y_i(N - w_i)}} \tag{42}$$

References to this section

- Karagiannis, E., and Kovacevic, M. (2000): A Method to Calculate the Jackknife Variance Estimator for the Gini Coefficient, Oxford Bulletin of Economics and Statistics 62, 119-122.
- Karoly, L.A., and Schroeder, C. (2015): Fast Methods for Jackknifing Inequality Indices, Applied Econometrics, 37(1), 125-138.
- Poi, B.P. (2004): From the help desk: Some bootstrapping techniques, The Stata Journal, 4, 312-328.
- Wolter, K. (1985): Introduction to Variance Estimation, Springer, New York.
- Yitzhaki, S. (1991): Calculating Jackknife Variance Estimators for Parameters of the Gini Method, Journal of Business and Economic Statistics, 9, 235-239.

End

Now you are ready to start your on empirical research on inequality. Enjoy!