Use of Distributional Assumptions for the Comparison of four Laeken Indicators on EU-SILC Data

Monique Graf

Federal Statistical Office, Statistical Methods Unit

monique.graf@bfs.admin.ch

Swiss Statistics Meeting, Atelier D, November 15, 2007

Outline

Swiss Statistics

Aim

- Gain understanding in the statistical properties of four income-based welfare indices included in the EU-SILC survey (Community statistics on income and living conditions).
- Compare the theoretical relationships between the chosen indices under income model assumptions with the empirical results of the EU-SILC survey...
- in order to investigate whether these relationships are heavily model-dependent or not.
- The comparison will be done on results at the country level.

Literature

A huge literature exist on income distributions and social indicators. Let us mention 3 main references.

- The Handbook on Income distributions by A.B. Atkinson and F. Bourgignon (2000) surveys the economic aspects of income distributions.
- Social Indicators for the monitoring of social inclusion within the EU are discussed by T. Atkinson and coauthors (2002).
- The mathematical aspects of Statistical Size Distributions are collected in the book by C. Kleiber and S. Kotz (2003).

Four Laeken indicators

- The at-risk-of poverty rate (*ARPR*)
- The relative median poverty gap (*RMPG*)
- The income quintile share ratio (*QSR*)
- The Gini index (*GINI*)

These indicators are all scale-free.

Notations

- X' = random variable associated with the (equivalized) income
- $\widetilde{m}$ = the median (equivalized) income
- $0.6\widetilde{m}$ = "poverty line"

The indicators of interest being scale-free, we can scale the income arbitrarily.

Scaling

It is convenient to scale by the median:

- $X = X'/\widetilde{m}$ = scaled (equivalized) income
- 0.6 = "scaled poverty line"

A poor is defined as a person with a scaled equivalized income less than 0.6.

At-risk-of poverty rate (ARPR)

Intensity of poverty

F = cumulative distribution function of the scaled income.

$$ARPR = \Pr(X' < 0.6\widetilde{m}) = \Pr(X < 0.6) = F(0.6)$$

$ARPR$ is the proportion of the population under the poverty line.

Relative median poverty gap (RMPG)

Depth of poverty

$F^{-1}(ARPR/2)$ = scaled median income of the poor (those under the poverty line)

$$RMPG = \frac{0.6 - F^{-1}(ARPR/2)}{0.6}$$

$RMPG$ is the relative gap between the poverty line and the median income of the poor.

Income quintile share ratio (QSR)

Inequality measure

q_{80} (q_{20}): 80th (20th) percentiles of the distribution F.

$$QSR = \frac{\mathrm{E}(X|X > q_{80})}{\mathrm{E}(X|X < q_{20})} \frac{\mathrm{Pr}(X > q_{80})}{\mathrm{Pr}(X < q_{20})}$$

QSR is the ratio of the cumulated incomes of the upper quintile over the cumulated incomes of the lower quintile.

Gini index (GINI)

There are many definitions of the Gini index.
Here is one: In general it is defined for a positive random variable.
X and Y = two independent r.v. with distribution F.

$$GINI(F) = \frac{\mathrm{E}(|X - Y|)}{2\,\mathrm{E}(X)}$$

The index is an inequality indicator measuring the expected absolute difference between two independently selected incomes relative to twice the mean income.

Model assumptions for the scaled income $X', X' > 0$

- Lognormal distribution with median $= 1$:

$$F(x) = \Phi\left(\frac{\ln(x)}{\sigma}\right) \qquad \sigma > 0$$

  where Φ is the standard normal distribution function.

- Fisk (or log-logistic) distribution with median $= 1$:

$$F(x) = \frac{1}{1 + x^{-a}} \qquad a > 0$$

Example: the Gini index

Under one of these models, all scale free indicators are functions of one parameter only, e.g. the Gini index takes the following form:

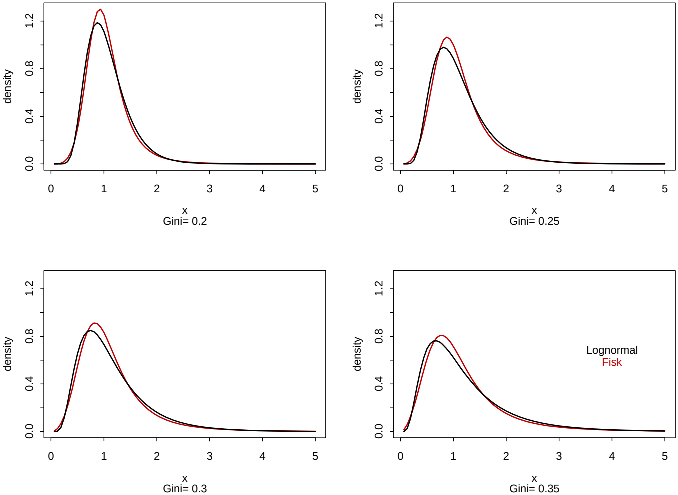
- Lognormal distribution:

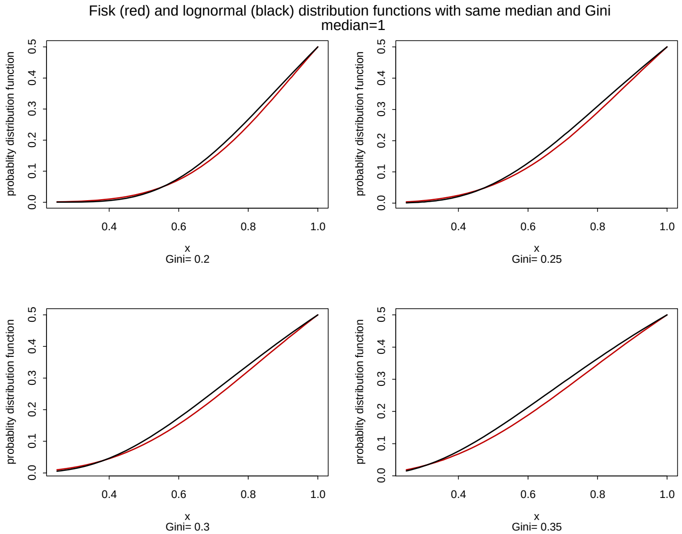
$$GINI = G_L(\sigma) = 2\Phi\left(\frac{\sigma}{\sqrt{2}}\right) - 1$$

- Fisk (or log-logistic) distribution:

$$GINI = G_F(a) = \frac{1}{a}$$

Fisk and lognormal densities with same median and Gini
median=1




Fisk (red) and lognormal (black) distribution functions with same median and Gini
median=1
probablity distribution function
x
Gini= 0.2
Gini= 0.25
Gini= 0.3
Gini= 0.35


Swiss Statistics

Implicit relationships

Under one of these model assumptions, any set of indicators define a parameterized curve.

E.g. the inequality indicators under the lognormal hypothesis, define a curve in the plane:

$$(QSR(\sigma), GINI(\sigma)),\ \sigma > 0$$

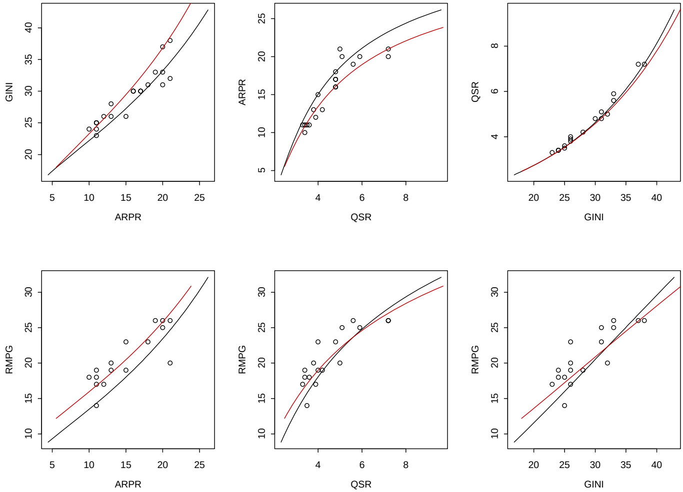
Comparison with EU-SILC 2004 and 2005 estimates

A very large set of Laeken indicators for EU-SILC data can be found on the Eurostat website.

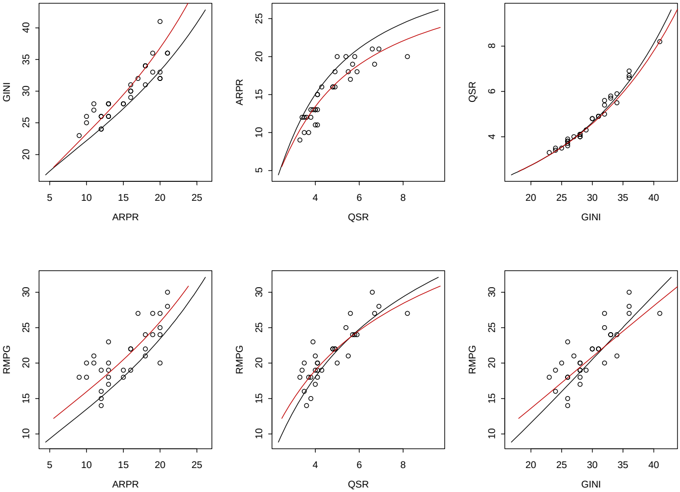
Here we used the May 2007 figures of the four indicators introduced above only for the total population at the country level.
Pairs of indicators are plotted together with the curves under the lognormal and the Fisk assumptions.

no fitting of curves!

Relationship between four Laeken indicators at the country level - Eurostat data 2004 and prediction under the Fisk (red) and lognormal (black) distributions



Relationship between four Laeken indicators at the country level - Eurostat data 2005 and prediction under the Fisk (red) and lognormal (black) distributions



Precision of EU-SILC estimates

The coefficients of variation of the 2004 estimates for 13 European countries have been computed by Osier (2006):

Indicator	min CV (%)	max CV (%)
ARPR	0.5	4.6
RMPG	2.7	7.3
QSR	1.0	4.2
GINI	0.8	4.1

Thus part of the scatter could be attributed to sampling error.

Discussion

- Close concordance between the theoretical lines and the inequality indices *QSR* and *GINI*: in accordance with the fact that these indicators are regarded as essentially measuring the same characteristic.
- The Fisk model seems to outperform the lognormal at least when estimates are computed at the country level.
- This model provides an average relationship and is thus of help for the interpretation.
- An improvement can be expected on using distributions having more than one shape parameter.