The econometrics of inequality and poverty
Lecture 8: Confidence intervals and testing

Michel Lubrano

October 2016

Contents

1 Introduction **3**

2 The simple case of FGT indices **3**
2.1 The asymptotic distribution of FGT indices 4
2.2 Testing for equality 4
2.3 Empirical application using the FES 5

3 Variances of indices in stratified samples **5**
3.1 How the FES is sampled 6
3.2 Indices in stratified samples 6
3.3 Variances of indices in stratified samples 7
3.4 Variances of indices in IID samples 7
3.5 Empirical illustration using the BHPS 8

4 Standard deviation for the Gini **9**
4.1 The regression method 9
4.2 Davidson's method 10
4.3 A first Bayesian approach 11
4.4 Empirical application 11
4.5 The R programs 12
4.6 Inequality measurements and the lognormal distribution 13
4.7 Properties of mixture models 14

5 Estimation of the Gini index with grouped data **15**

6 Decomposable inequality indices **17**
6.1 Definition 17
6.2 Inequality indices for mixtures of distributions 18
6.3 Index decomposition for the log normal 19
6.4 A Bayesian approach for computing the variance of the GE index 20

7 Testing for stochastic dominance **21**
7.1 Hypotheses 21
7.2 Asymptotic distribution of empirical dominance curves 21
7.3 An example 22
7.4 Inference 23
7.5 Three dominance tests 24
7.6 A simplified test 25
7.7 Illustration 25

8 Exercices **28**

1 Introduction

This chapter details various procedures to compute standard errors for different classes of inequality and poverty indices. These computations are sometime easy as for the FGT indices, sometime slightly more complicated as for the Atkinson index or the generalised entropy indices. The Gini is a special case that has to be treated separately. The chapter will end with the special topic devoted to testing for stochastic dominance.

For long, it was supposed that it was no use to associate a confidence interval to income inequality and poverty measures. It was thought that anyway, they were very small due to the important size of the survey samples. However, we have seen that sometimes the data are presented in groups for confidential reasons. An some other times, one is interested in subgroup decomposition, so that at the end samples are not so large. Take for instance the case of isolated parents. So the size argument does not hold. What is true on the contrary is that these standard deviations might be complicated to evaluate.

The poverty measures and poverty indices that we want to estimate are complex functions of the observed sample, as for instance quantiles. We must use special techniques in order to evaluate their standard deviations. The paper by Berger and Skinner (2003) is a useful reference. It is possible to obtain in some cases analytical formulae, most of the using linearisation techniques; or we have to use resampling techniques like the bootstrap in order to provide a sampling standard deviation. Davidson (2009) compares several methods to compute the variance of the Gini index and provides a good approximation method. His results are extended for the poverty index of Sen and to that of Sen-Shorrocks-Thon.

The Bayesian approach can provide an interesting solution to this question while avoiding some of the bias problems which are attached to the bootstrap approach when the latter is not correctly designed. Suppose that we have adjusted a parametric distribution to the income distribution using a Monte Carlo method. We have then a collection of draws for this distribution. If the chosen distribution is simple enough, an analytical expression for many indices is available. For each value of the parameters, we can then compute the corresponding value of the indices, for instance the Gini. Standard deviations and small sample distribution are then trivially obtained. If we want to consider a richer family of distributions, we can turn to mixtures of densities, considering mixtures of log-normals for instance as we have analytical results. In general, the overall index will be obtained as a weighted sum of individual indices, provided the considered index is decomposable.

2 The simple case of FGT indices

The case considered by Kakwani (1993) is relatively simple. This simplicity is explained by the considered indices, the Foster, Greer, and Thorbecke (1984) poverty indices which are linear and decomposable. These indices are computed as sums of independent identically distributed random variables. In this case, the central limit theorem can be applied directly. This would not be the case with the Sen (1976) index, as it involves the rank of the variables (Gini index over the poor).

2.1 The asymptotic distribution of FGT indices

The Foster, Greer, and Thorbecke (1984) index is defined conditionally on a given value of α

$$P_\alpha = \int_0^z \left(\frac{z-x}{z}\right)^\alpha f(x)dx \qquad \alpha \geq 0. \tag{1}$$

This index can be estimated in a relatively simple way. Let us consider a sample of n households where adult equivalent income is $x_1, \cdots, x_n$. Let us suppose that the observations x_i are ordered by increasing order and that q is the rank of the last poor ($q = \text{Max}_i \; i \; \mathbb{1}(x_{(i)} \leq z)$. A consistent estimate for P_α is given by

$$\widehat{P}_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-x_{(i)}}{z}\right)^\alpha \tag{2}$$

in the general case and by $\widehat{P}_0 = q/n$ for $\alpha = 0$. Applying the central limit theorem provides first the asymptotic normality of this estimator

$$\sqrt{n}(\widehat{P}_\alpha - P_\alpha) \sim N(0, \sigma^2). \tag{3}$$

The variance σ^2 is defined as

$$\sigma^2 = \text{E}(\widehat{P}_\alpha - P_\alpha)^2 = \int_0^z \left(\frac{x-z}{z}\right)^{2\alpha} f(x)dx - P_\alpha^2. \tag{4}$$

A natural estimator for this variance is given by

$$\widehat{\sigma}^2 = \widehat{P}_{2\alpha} - \widehat{P}_\alpha^2. \tag{5}$$

The standard deviation of $\widehat{P}_\alpha$ will be estimated as $\widehat{\sigma}/\sqrt{n}$ and noted $\hat{\sigma}_P$. Then the random variable

$$t = \frac{\widehat{P}_\alpha - P_\alpha}{\hat{\sigma}_P} \tag{6}$$

is asymptotically normal with zero mean and unit variance. Let us call $t_{0.05}$ the critical value for the normal at the 5% level, we can built the following confidence interval

$$\widehat{P}_\alpha - t_{0.05}\hat{\sigma} \leq P_\alpha \leq \widehat{P}_\alpha + t_{0.05}\hat{\sigma}. \tag{7}$$

2.2 Testing for equality

The test proposed by Kakwani (1993) is relatively simple as it is equivalent to testing the equality of two means of two independent samples. This is a well designed problem in the statistical literature. Let us consider two independent samples of respective size n_1 and n_2. We consider the asymptotic distributions of $\sqrt{n_i}\widehat{P}_i$ with variance σ_i^2. We have omitted the index α in P_α just for the clarity of notations. The the standard deviation of the estimated difference $\widehat{P}_1 - \widehat{P}_2$ is equal to

$$SE(\widehat{P}_1 - \widehat{P}_2) = \sqrt{\frac{\hat{\sigma}_1^2}{n_1} + \frac{\hat{\sigma}_2^2}{n_2}} \tag{8}$$

because the samples are independent and the statistics

$$\eta_y = \frac{\widehat{P}_1 - \widehat{P}_2}{SE(\widehat{P}_1 - \widehat{P}_2)} \tag{9}$$

is asymptotically normal with zero mean and unit variance. This statistics allows us to test that two poverty indices are equal against the hypothesis that they are different.

2.3 Empirical application using the FES

We have four different samples for the FES and we can suppose that they are roughly independent. Poverty has changed a lot during the years covered by these samples as we have already documented. We can ask the question to know if poverty is statistically different over the years or if these differences result simply from sampling errors. Due to the large sample sizes and also

Table 1: FGT indices and their sampling errors

	1979	1988	1992	1996
P_α	0.0186	0.0313	0.0369	0.0272
S.D.	0.0383	0.0531	0.0681	0.0540

Kakwani test

	1979	1988	1992	1996
1979	0.0			
1988	-15.52	0.0		
1992	-21.63	-7.71	0.0	
1996	-10.09	4.36	11.40	0.0

The FGT indices were computed for $\alpha = 2$ and z being 60% of the median.

the fact that the sample are distant of several years, all the values of the FGT poverty indices are statistically different. As this is a bilateral test, the 5% critical value is 1.96. We must also note that the variance of the indices could not be computed for some years when $\alpha = 1$.

3 Variances of indices in stratified samples

In this section, we shall study the influence of the sample design on the computation of standard errors for inequality indices. Many results can be found in the literature concerning IID samples, leading many times to complicated formulae. An account of this literature is given in Biewen and Jenkins (2006). The major contribution of this paper is to provide generic formulae for the

generalised entropy index and for the Atkinson index when the sample design is complex. As a by-product, it also provides expressions for the standard deviation of these indices for the iid case which are more simple than the previous ones found in the literature.

3.1 How the FES is sampled

The previous results were in a way simple because we considered IID samples. But as soon as there are weights of clusters, the calculation of variances for most indices become complicated. We begin by a short illustrative presentation of the FES to situate the question. From section 2 of Berger and Skinner (2003), we have a description of the sampling design for the FES.

The FES is a multistage stratified random sample of $n = 6630$ private household in 1999 drawn from a list of postal addresses. Postal sectors are the primary sample units and are selected by probability proportional to a measure of size, after being arranged in strata defined by standard regions, socioeconomic group and ownership of cars. The sample for Northern Ireland is drawn as a random sample of addresses with a larger sampling fraction than for Great Britain. Under the FES sampling design, all households in Great Britain are selected with equal first-order inclusion probabilities. All households in Northern Ireland are likewise selected with a fixed inclusion probability, greater than that in Great Britain. Out of the about 10000 households selected for the target sample, about 66% are contacted and co-operate fully in the survey. Response probabilities have been estimated in a study linking the target sample to the 1991 census (Elliot, 1997; Foster, 1998). These response probabilities multiplied by the sampling inclusion probabilities generate basic survey weights d_k for each household k. These weights will be referred to as prior weights and will be treated as fixed, independent of the sample. The prior weights d_k are adjusted to agree with control totals by using the raking procedure proposed by Deville et al. (1993) and are fully described in Section 5. The resulting weights are denoted w_k and termed the raking weights. Unlike the prior weights, these weights are sample dependent.

With this example, we understand that the design of a survey is not a simple matter.

3.2 Indices in stratified samples

In a stratified sampling design, we have $h = 1, ..., L$ strata. In each strata there are $i = 1, ..., N_h$ clusters. In each cluster, there are $j = 1, ..., M_i$ individuals. Of course, N_h is not constant over the strata and the number of individuals is not constant over the clusters. Biewen and Jenkins (2006) show that if y_{hij} is the individual income, it suffices to compute the two quantities, using the weight w_{hij} if necessary

$$U_\alpha = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} w_{hij} y_{hij}^\alpha$$

and

$$T_\alpha = \sum_{h=1}^{L} \sum_{i=1}^{N_h} \sum_{j=1}^{M_i} w_{hij} y_{hij}^\alpha \log y_{hij}$$

Given U_α and T_α, Biewen and Jenkins (2006) show that the Atkinson index is

$$I_A^\epsilon = 1 - U_0^{-\epsilon/(1-\epsilon)} U_1^{-1} U_{1-\epsilon}^{1/(1-\epsilon)}$$

when $\epsilon \neq 1$. For the GE family, we have

$$I_{GE}^\alpha = \frac{1}{\alpha^2 - \alpha}(U_0^{\alpha-1} U_1^{-\alpha} U_\alpha - 1)$$

except when $\alpha = 0$ or $\alpha = 1$. T_α serves for these cases. The complete formulae are given in Biewen and Jenkins (2006).

3.3 Variances of indices in stratified samples

The two indices can be seen as a function f of T_k population statistics. Biewen and Jenkins (2006) then consider the first order Taylor expansion of $f(T)$

$$f(\hat{T}) = f(T) + \sum_{k=1}^{K} \frac{\partial f(T)}{\partial T_k}(\hat{T}_k - T_k).$$

The variance of the inequality index is approximated by the variance of

$$\sum_{k=1}^{K} \frac{\partial f(T)}{\partial T_k} \hat{T}_k.$$

After some computations detailed in Biewen and Jenkins (2006), we have

$$\text{Var}(\hat{I}) = \sum_{h=1}^{L} \frac{n_h}{n_h - 1} \sum_{i=1}^{n_h} \left(\sum_{j=1}^{m_i} w_{hij} s_{hij} - \frac{1}{n_h} \sum_{i=1}^{n_h} \sum_{j=1}^{m_i} w_{hij} s_{hij} \right)^2.$$

It just remains to compute the s_{hij} for the two decomposable indices. For the GE, we have

$$s_{hij} = \frac{1}{\alpha} \hat{U}_\alpha \hat{U}_1^{-\alpha} \hat{U}_0^{\alpha-2} - \frac{1}{\alpha - 1} \hat{U}_\alpha \hat{U}_1^{-\alpha-1} \hat{U}_0 \alpha - 1 y_{hij} + \frac{1}{\alpha^2 - \alpha} \hat{U}_0^{\alpha-1} \hat{U}_1^{-\alpha} y_{hij}^\alpha.$$

For the Atkinson index, the formula is slightly longer and given in Biewen and Jenkins (2006).

3.4 Variances of indices in IID samples

In the case of IID samples, the above formulae still apply, but with a much simplified expression, and in particular lead to simpler computations than those indicated in Cowell (1989). We first have to compute

$$\hat{\mu}_\alpha = \frac{1}{n} \sum w_i x_i^\alpha.$$

The two indices are estimated using

$$I_{GE} = \frac{1}{\alpha^2 - \alpha}(\mu_0^{\alpha-1} \mu_1^{-\alpha} \mu_\alpha - 1),$$

and

$$I_A = 1 - \mu_0^{-\epsilon/(1-\epsilon)} \mu_1^{-1} \mu_{1-\epsilon}^{1/(1-\epsilon)}.$$

The variance of these indices is obtained as

$$\text{Var}(\hat{I}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} \left(w_i z_i - \frac{1}{n} \sum w_i z_i \right)^2 .$$

We just have to compute the value of z_i for each index. For the Atkinson index, we have:

$$z_i^A = \frac{\epsilon}{1-\epsilon} \mu_1^{-1} \mu_{1-\epsilon}^{1/(1-\epsilon)} \mu_0^{-1/(1-\epsilon)} + \mu_0^{-\epsilon/(1-\epsilon)} \mu_{1-\epsilon}^{1/(1-\epsilon)} \mu_1^{-2} x_i - \frac{1}{1-\epsilon} \mu_0^{-\epsilon/(1-\epsilon)} \mu_1^{-1} \mu_{1-\epsilon}^{\epsilon/(1-\epsilon)} x_i^{1-\epsilon}.$$

For the Generalised Entropy index:

$$z_i^{GE} = \frac{1}{\alpha} \mu_\alpha \mu_1^{-\alpha} \mu_0^{\alpha-2} - \frac{1}{\alpha-1} \mu_\alpha \mu_1^{-\alpha-1} \mu_0^{\alpha-1} x_i + \frac{1}{\alpha^2-\alpha} \mu_0^{\alpha-1} \mu_1^{-\alpha} x_i^\alpha.$$

Thus for both indices, we have a formula which is easy to program.

3.5 Empirical illustration using the BHPS

We report here some of the empirical results of Biewen and Jenkins (2006) concerning the BHPS. This survey, devoted to the UK, is particularly relevant because variables identifying clusters and strata are made available. So we can compare standard deviations of indices when the sample design is taken into account and when it is not.

Table 2: Income inequality in Britain: BHPS 2001
Complex survey estimators

	Estimator	SE 1	SE 2	SE 3
MLD	0.1702	0.0058	0.0058	0.0040
Theil	0.1653	0.0074	0.0074	0.0056
A(1)	0.1565	0.0049	0.0049	0.0034
A(2)	0.3770	0.0289	0.0288	0.0263

The BHPS is composed of 250 primary sampling units or clusters and of 75 strata. There are 9 782 individuals who come from 4 060 households. Inequality indices are computed over the individuals who receive a fraction of their household income according to a square root equivalence scale. The estimators reported in Table 2 use individual data and individual sample weights. Standard errors are computed with three different options. In column 2 (SE 1), the reported standard error accounts for stratification and clustering. Column 3 accounts for clustering at the household level only. Column 4 ignores replication of observations at the household level, clustering and stratification. There is not much of a difference between SE 1 and SE 2. However,

ignoring all clustering and stratification leads to under estimated standard errors. We conclude that, at least for the BHPS, it is essential to take into account the fact that several individuals come from the same household. These results are in a way paradoxical as stratification should produce a smaller variance.

4 Standard deviation for the Gini

We have detailed many different expressions for the Gini and derived the associated estimation procedures when no assumption is done concerning the data generating process. In particular, we show that the Gini could be seen as the covariance between the cumulative density of an observation and its rank. More precisely, we had

$$\begin{aligned} G &= 1 - 2\int_0^1 L(p)dp \\ &= \frac{2}{\mu}\int_0^\infty yF(y)f(y)dy - 1 \\ &= \frac{2}{\mu}\left[\int_0^\infty yF(y)f(y)dy - \frac{\mu}{2}\right] \end{aligned}$$

This formula opens the way to an interpretation of the Gini coefficient in term of covariance as

$$\text{Cov}(y, F(y)) = \text{E}(yF(y)) - \text{E}(y)\text{E}(F(y)).$$

Using this definition, we have immediately that

$$G = \frac{2}{\mu}\text{Cov}(y, F(y)).$$

Using this result, we showed in Lecture 4 that an estimator for the Gini could be

$$\hat{G} = \frac{2}{n^2\bar{y}}\sum i\, y_{(i)} - \frac{n+1}{n}$$

4.1 The regression method

Using some of the results of Ogwang (2000), Giles (2004) showed that the Gini can be estimated using a weighted regression

$$i\sqrt{y_i} = \theta\sqrt{y_i} + \epsilon_i.$$

As a matter of fact, this regression give an estimate for θ which is equal to

$$\hat{\theta} = \frac{\sum i\, y_{(i)}}{\sum y_{(i)}}.$$

This estimator of θ can be plugged into the definition of the previous estimator of the Gini as we could perfectly say that

$$\widehat{G} = \frac{2}{n \sum y_i} \sum i \, y_{(i)} - \frac{n+1}{n}.$$

The variance of the Gini is then a simple linear function of the variance of θ and

$$\text{Var}(\hat{G}) = \frac{4\text{Var}(\hat{\theta})}{n^2}.$$

And the variance of $\hat{\theta}$ come directly from the OLS estimation of the regression. We must note three points

- The regression come from the initial regression $i = \theta + \nu_i$ supposing a form of heteroskedasticity as $\epsilon_i = \sqrt{y_{(i)}}\nu_i$. This assumption has to be tested as underlined by Giles (2004) himself.
- More importantly, the order statistics are correlated and that correlation is ignored in this regression. This entails a bias in the estimation of the variance of the Gini.
- Why not use a more robust estimator for the variance of $\hat{\theta}$ to solve this point?

This is certainly the simplest way of computing a standard deviation for the Gini coefficient. Giles does not give the asymptotic distribution of his estimator. However, we can suppose that, if we ignore the correlation problem, the estimator unbiased and the asymptotic distribution is Gaussian. So that an asymptotic confidence interval can be computed. If we want to solve properly these questions, we have to turn to Davidson (2009) which is slightly more complex paper.

4.2 Davidson's method

Davidson (2009) gives an alternative expression for the variance of the Gini which is not based on a regression, but simply on the properties of the empirical estimate of $F(x)$. Recall that for instance the Gini can be evaluated as

$$\frac{1}{\mu} \int_0^\infty F(y)(1 - F(y))dy.$$

So an estimator of the Gini can be based on the natural estimator of $F(x)$. This property is used here. We do not describe the derivation of the method, but simply the final result. If we note $\hat{I}_G$ the numerical evaluation of the sample Gini, we have:

$$\hat{Var}(\hat{I}_G) = \frac{1}{(n\hat{\mu})^2} \sum (\hat{Z}_i - \bar{Z})^2, \tag{10}$$

where $\bar{Z} = (1/n) \sum_{i=1}^n \hat{Z}_i$ is an estimate of $E(Z_i)$ and

$$\hat{Z}_i = -(\hat{I}_G + 1)x_{[i]} + \frac{2i-1}{n} x_{[i]} - \frac{2}{n} \sum_{j=1}^{i} x_{[j]}.$$

This is however an asymptotic result which is general gives lower values than those obtained with the regression method of Giles. Small sample results can be obtained if we adjust a parametric density for y and use a Bayesian approach.

4.3 A first Bayesian approach

The Gini index in a log normal density is equal to $2\Phi(\sigma/\sqrt{2}) - 1$. In a Bayesian framework, the posterior density of σ^2 under a non-informative prior is an inverted gamma density with n degrees of freedom and scale parameter equal to $s^2 = \sum(\log(x_i) - \bar{x})^2$ where $\bar{x} = \frac{1}{n}\sum \log(x_i)$.

We can simulate draws of σ^2 from this posterior density, for each draw evaluate the corresponding Gini coefficient. Once we have got m different draws of the Gini, we can compute its mean and standard deviation. The standard deviations computed in this way are the comparable to the bootstrap ones.

4.4 Empirical application

Using the FES data set, we compute the Gini index for the four years and provide a bootstrap estimation of its standard deviation as programmed previously in `R`.

Table 3: Standard errors for the Gini index using FES data

year	1979	1988	1992	1996
Gini	0.25634	0.30735	0.32140	0.29758
Bootstrap	0.00233	0.00341	0.00367	0.00323
Davidson asymptotic	0.00233	0.00337	0.00368	0.00325
Bayesian lognormal	0.00221	0.00263	0.00281	0.00263
Bayesian mixture of LN	0.00210	0.00320	0.00385	0.00330
Giles	0.00585	0.00527	0.00552	0.00595

Three methods, the bootstrap, Davidson and the Bayesian approach based on a fitted lognormal distribution using a diffuse prior give comparable results. On average, the standard deviations given by the method of Giles are twice as large as those of the other methods.

Using the fit of a simple lognormal produces a nice approximation, but has a tendency to underestimate the standard deviation. The line *Bayesian mixture of LN* corresponds to a method explained below and corrects for this underestimation, because it is based on a mixture of lognormal which has a better fit than the simple lognormal. values are very close to those produced by the asymptotic method.

It is important to have a correct method available for computing standard errors. Using a t test, we can show that inequality in 1996 reached a level which is significantly lower than that of

1992 ($t = 2.12$) when using the bootstrap standard deviations. This difference is not significant when using the Giles standard deviations ($t = 1.01$).

4.5 The R programs

```
# Compute Gini standard deviations for FES data

rm(list=ls())
library(ineq)
library(boot)
library(xtable)

rinvgamma2 = function(n,df,scale){scale/rchisq(n,df)}

BGini = function(x,ndraw){
  # Bayesian SD based on the lognormal fit
  lx = log(x)
  n = length(x)
  s2 = sum( (lx-mean(lx))^2 )
  sig2 = rinvgamma2(ndraw,n,s2)
  G = 2*pnorm(sqrt(sig2/2),mean=0,sd=1) - 1
  return(sd(G))
}

stdavidson = function(y){
  # Asymptotic SD Davidson
  n=length(y)
  G=Gini(y)
  mu=mean(y)
  zhat=rep(0,n)
  for(i in 1:n){
    zhat[i] = -(G+1)*y[i]+ ((2*i-1)/n)*y[i]-2*sum(y[1:i])/n
  }
  zbar = mean(zhat)
  varG = sum((zhat-zbar)^2)/(n*mu)^2
  return (sqrt(varG))
}

stgiles = function(y){
  # Gile's regression method
  n=length(y)
  s=0
  g=rep(0,n)
```

```
  for( i in 1:n){
    g[i]=i*sqrt(y[i])
  }
  ols=lm(g~sqrt(y))
  return(2*sqrt(diag(vcov(ols))[2])/n)
}

bt = function(y){
  # Bootstrapping
  r = boot(y,function(d,i){a <- Gini(d[i])},R=999)
  return( sd(r$t) )
}

data1=read.table("fes79.csv",header=F,sep=";")
data2=read.table("fes88.csv",header=F,sep=";")
data3=read.table("fes92.csv",header=F,sep=";")
data4=read.table("fes96.csv",header=F,sep=";")

y79 = sort(data1[,1])/223.5*223.5
y88 = sort(data2[,1])/421.7*223.5
y92 = sort(data3[,1])/546.4*223.5
y96 = sort(data4[,1])/602.4*223.5

TA = rbind(c(Gini(y79),Gini(y88),Gini(y92),Gini(y96)),
 c(bt(y79),bt(y88),bt(y92),bt(y96)),
 c(BGini(y79,1000),BGini(y88,1000),BGini(y92,1000),BGini(y96,1000)),
 c(stgiles(y79),stgiles(y88),stgiles(y92),stgiles(y96)),
 c(stdavidson(y79),stdavidson(y88),stdavidson(y92),stdavidson(y96))
  )

xtable(TA,digits=5)
```

4.6 Inequality measurements and the lognormal distribution

This section builds heavily on Lubrano and Ndoye (2011).

Cowell (1995) offers the different analytical expressions provided by the lognormal distribution on commonly used inequality indices that we reproduce in Table 4. Each measurement depends on the single shape parameter σ. $\Phi(.)$ is the cumulative distribution of the standard normal distribution and $\mu(F)$ is the income average of the considered population having distribution F. The Generalised Entropy (GE) index is sensitive to the behaviour of the upper tail for large positive values of α; for α negative, the index is sensitive to changes in distribution that affect the lower tail. The parameter $\epsilon \geq 0$ characterises (relative) inequality aversion for the Atkinson

Table 4: Inequality measurements and lognormal distribution

Inequality index	General expression	Lognormal expression
Gini index		
I_G	$\frac{1}{\mu}\int_0^\infty F(y)(1-F(y))dy$	$I_G(\sigma) = 2\Phi(\sigma^2/2) - 1$
Generalized Entropy		
I_{GE}^α	$\frac{1}{\alpha^2-\alpha}\int\left[\left(\frac{y}{\mu(F)}\right)^\alpha - 1\right] f(y)dy$	$I_{GE}^\alpha(\sigma) = \frac{\exp\left((\alpha^2-\alpha)\sigma^2/2\right)-1}{\alpha^2-\alpha}$
Atkinson index		
I_A^ϵ	$1 - \frac{1}{\mu(F)}\left(\int y^{1-\epsilon} f(y)dy\right)^{\frac{1}{1-\epsilon}}$	$I_A^\epsilon(\sigma) = 1 - \exp\left(-\frac{1}{2\epsilon\sigma^2}\right)$

index, inequality aversion being an increasing function of ϵ. The Atkinson index may be viewed as a particular case of the GE index with $\alpha \leq 1$ and $\epsilon = 1 - \alpha$.

The GE class include a number of other inequality indices such as the mean logarithmic deviation index ($I_{MLD} = \lim_{\alpha\to 0} I_{GE}^\alpha$), Theil's index ($I_{Theil} = \lim_{\alpha\to 1} I_{GE}^\alpha$) and the coefficient of variation ($1/2I_{CV}^2 = \lim_{\alpha\to 2} I_{GE}^\alpha$). For the lognormal distribution, the MLD and the Theil index become the same and are both equal to $\sigma^2/2$.

4.7 Properties of mixture models

Mixture models have nice properties that will be of direct interest for our purpose. Those properties are directly related to the linearity of the model. In any finite mixture, the overall cumulative distribution is obtained as the weighted sum of the individual cumulative distributions so that in our case:

$$F(x) = \sum_{j=1}^{k} p_j F_j(x|\mu_j, \sigma_j^2).$$

The first moment $\mu(F)$ of X is obtained as a linear combination of the first moment of each member of the mixture

$$\mu(F) = \sum_{j=1}^{k} p_j \mu(F)_j.$$

That property extends to the un-centred higher moments.

We can use directly these properties in order to derive the expression of the Gini index for a mixture of lognormals. A Gini index can be written as a function of the overall cumulative distribution, using the integral expression given in Table 4:

$$I_G(\mu, \sigma^2, p) = \frac{1}{\mu(F)}\int_0^\infty F(x)(1-F(x))\,dx,$$

where $\mu(F)$ is the overall mean of the mixture. Let us develop this expression for a mixture of k

elements.

$$I_G(\mu, \sigma^2, p) = \frac{1}{\sum_{j=1}^{k} p_j \mu(F)_j} \int_0^{\infty} \sum_{j=1}^{k} p_j F_j(x)(1 - \sum_{j=1}^{k} p_j F_j(x))\, dx,$$

As the cumulative of the lognormal is $F_j(x) = \Phi(\frac{\log x - \mu_j}{\sigma_j})$, the Gini index can be obtained as the result of a simple numerical integral ($\Phi(.)$ being directly available in any numerical package).

But this integral has to be evaluated for every draw of the MCMC experiment. We thus get m evaluations of the Gini index. Summing over all the draws, we get an estimate for the mean index:

$$\hat{I}_G = \frac{1}{m} \sum_{t=1}^{m} I_G(\mu^{(t)}, \sigma^{(t)}, p^{(t)}),$$

The standard deviation can be obtain in a similar way by summing the squares

$$\hat{I}_G^2 = \frac{1}{m} \sum_{t=1}^{m} I_G(\mu^{(t)}, \sigma^{(t)}, p^{(t)})^2,$$

so that the small sample variance is obtained as $\hat{I}_G^2 - (\hat{I}_G)^2$.

For decomposable indices, it is possible to go a step further on as decomposability implies that the overall index can be expressed as a weighted sum of individual indices (plus a remainder) as we shall now see.

5 Estimation of the Gini index with grouped data

The Gini index is quite easy to estimate on individual data as we have seen above, but when we have grouped data, the task is more difficult. We could adjust a parametric density on the grouped data, and deduce the corresponding value of the Gini coefficient from the estimated parameters. However, this might be restrictive, because it depends on the fit of the particular parametric density. Another avenue was proposed in the literature, notably with Gastwirth (1972), which is based on the property that the Gini coefficient is equal to twice the surface between the Lorenz curve and the first diagonal. So we must first estimate a Lorenz curve, using a semi-parametric method. With grouped data, the Lorenz curve is represented as a sequence of straight lines when it should be a curve. From the sequence of straight lines, we can deduce a lower bound G_L while an upper bound G_U has to be found, which would correspond to an unknown curve. Which value should be chosen between these two bounds? Schader and Schmid (1994) recommends to use as a point estimate a particular linear combination of these two bounds which is:

$$\hat{G} = \frac{1}{3} G_L + \frac{2}{3} G_U.$$

We have now to explain how the Lorenz curve is estimated with grouped data. Let us suppose that we have k income classes. The lower bound x_1 is for instance zero with $x_1 = 0$ and we suppose that the last class is open so that the upper bound is $x_{k+1} = +\infty$. So the last class is unbounded, which is a rather frequent case. The class frequencies are nc_i with $n = \sum_{i=1}^{k} nc_i$.

Let (p_i, y_i) be the cumulative population share and income share with starting point $(p_0, y_0) = (0, 0)$ and terminal point $(p_k, y_k) = (1, 1)$. The intermediate points of the Lorenz curve are defined as

$$p_i = \sum_{j=1}^{i} nc_j/n, \tag{11}$$

$$y_i = L(p_i) = \sum_{j=1}^{i} \mu_j nc_j/n. \tag{12}$$

This method is just the generalisation of the natural estimator of the Lorenz curve when raw data are available, except that here we compute the cumulative partial sum of income using the mean of each class weighted by its relative frequency. But as raw data are not available, we must provide an estimate for the mean value of each cell, μ_j. This is quite easy when the cells are bounded, because in this case

$$\mu_i = \frac{x_{i-1} + x_i}{2}.$$

However, when the last cell is unbounded, we have to make a Pareto assumption and estimate the corresponding Pareto parameter α. We have already seen how α can be estimated in Chapter 6, but let us recall the simple method of Quandt (1966). The method assumes that the Pareto density can be adjusted using the two last classes, so:

$$\hat{\alpha} = \frac{\log(nc_{k-1} + nc_k) - \log(nc_k)}{\log(x_k) - \log(x_{k-1})}. \tag{13}$$

Then the mean of the last open class is found as being:

$$\mu_k = x_k \frac{\hat{\alpha}}{\hat{\alpha} - 1},$$

provided that $\hat{\alpha} > 1$. Once the means are defined, we are in a position to explain how the lower bound of the Gini is estimated as a simple surface:

$$G_L = 1 - \sum_{i=1}^{k} (y_i + y_{i-1})(p_i - p_{i-1}).$$

This is a lower bound because the underlying Lorenz curve corresponds to a series of linear interpolation segments. The upper bound is obtained as the sum of the lower bound and a factor Δ which is more complicated and required the evaluation of the overall mean $\mu = \sum_{i=1}^{k} \mu_i \, nc_i/n$. We have:

$$\Delta = \frac{1}{\mu} \left[\sum_{i=1}^{k-1} \left(\frac{nc_i}{n} \right)^2 \frac{(\mu_i - x_{i-1})(x_i - \mu_i)}{x_i - x_{i-1}} + \left(\frac{nc_k}{n} \right)^2 (\mu_k - x_{k-1}) \right].$$

We thus have an estimate of the Gini coefficient. But for the while no standard deviation is associated to this estimate. We could think about an adaptation of Davidson (2009) method.

6 Decomposable inequality indices

Decomposability is a very convenient property because it is essentially a linearity property. It means that sums can be decomposed in series of partial sums, which is very convenient to compute variances. In this section, we characterise decomposable indices which cover essentially the generalised entropy indices and the Atkinson index. We then explore how this decomposability property can be used to compute variances when the income distribution is modelled as a mixture of log-normal densities. This section draws heavily on Lubrano and Ndoye (2011).

6.1 Definition

A decomposable inequality index can be expressed as a weighted average of inequality within subgroups, plus inequality between those subgroups.

Let $I(x, n)$ be an inequality index for a population of n individuals with income distribution x. $I(x, n)$ is assumed to be continuous and symmetric in x, $I(x, n) \geq 0$ with perfect equality holding if and only if $x_i = \mu$ for all i, and $I(x, n)$ is supposed to have a continuous first order partial derivative. Under these assumptions, Shorrocks (1980) defines additive decomposition condition as follows :

Definition 1 *Given a population of of any size $n \geq 2$ and a partition into k non-empty subgroups, the inequality index $I(x, n)$ is decomposable if there exists a set coefficients $\tau_j^k(\mu, n)$ such that*

$$I(x, n) = \sum_{j=1}^{k} \tau_j^k I(x^j; n_j) + B,$$

where $x = (x^1, \ldots, x^k)$, $\mu = (\mu_1, \ldots, \mu_k)$ is the vector of subgroup means, $\tau_j(\mu, n)$ is the weight attached to subgroup j in a decomposition into k subgroups, and B is the between-group term, assumed to be independent of inequality within the individual subgroups. Making within-group transfers until x_i^j in each subgroup and letting u_n represent the unit vector with n components, we obtain $B = I(\mu_1 u_{n_1}, \ldots, \mu_k u_{n_k})$.

The already defined family of Generalised Entropy indices is decomposable. For a given parameter α for inequality aversion, it is defined by:

$$I_{GE}(\alpha) = \frac{1}{\alpha^2 - \alpha} \int \left[\left(\frac{x}{\mu} \right)^{\alpha} - 1 \right] f(x) dx$$

where $\alpha \in (-\infty, +\infty)$. For α large and positive values, the index is sensitive to changes in the distribution that affect the upper tail, typically for $\alpha > 2$. For $\alpha < 0$, the index is sensitive to changes in distribution that affect the lower tail. In empirical works, the range of values for α is typically restricted to $[-1, 2]$ (see Shorrocks (1980)) because, otherwise, estimates may be unduly influenced by a small number of very small incomes or very high incomes.

This family includes half the coefficient of variation squared for $\alpha = 2$, the Theil coefficient for $\alpha = 1$ and the mean logarithmic deviation (MLD) for $\alpha = 0$:

$$\begin{aligned} I_{Theil}(F) &= \int \frac{x}{\mu} log\left(\frac{x}{\mu}\right) f(x)dx, \\ I_{MLD}(F) &= -\int log\left(\frac{x}{\mu}\right) f(x)dx. \end{aligned}$$

The Atkinson index is expressed as

$$I_A^\epsilon = 1 - \frac{1}{\mu}\left[\int x^{1-\epsilon} dF(x)\right]^{\frac{1}{1-\epsilon}},$$

where $\epsilon \geq 0$ is a parameter defining (relative) inequality aversion. This time, high values of ϵ correspond to a high aversion for income inequality among the poor.

There is a close relation between the Generalised Entropy index and the Atkinson index. They are ordinally equivalent for cases $\alpha \leq 1$ and $\epsilon = 1 - \alpha$ so that we have the relation:

$$I_A^\epsilon = 1 - \frac{1}{\mu}\left[(\alpha^2 - \alpha) I_{GE}^\alpha + 1\right]^{\frac{1}{\alpha}}.$$

Consequently, the Atkinson index is decomposable, thanks to the properties of the GE, but this decomposition is an indirect one. See Cowell (1995) for more details.

6.2 Inequality indices for mixtures of distributions

The decomposability property is a very powerful one as we can use it to find the analytical expression of a GE index for a mixture of densities. For a mixture model with k components in $f(.)$ and weights given by p_j, the expression of the GE index is

$$\begin{aligned} I_{GE}^\alpha &= \frac{1}{\alpha^2 - \alpha}\int\left[\left(\frac{x}{\sum_{j=1}^k p_j\mu_j}\right)^\alpha - 1\right]\sum_{j=1}^k p_j f_j(x)dx \\ &= \sum_{j=1}^k p_j \frac{1}{\alpha^2 - \alpha}\int\left[\left(\frac{x\mu_j}{\mu_j\sum_{j=1}^k p_j\mu_j}\right)^\alpha - 1\right] f_j(x)dx, \\ &= \sum_{j=1}^k p_j \left(\frac{\mu_j}{\sum_{j=1}^k p_j\mu_j}\right)^\alpha \frac{1}{\alpha^2 - \alpha}\int\left[\left(\frac{x}{\mu_j}\right)^\alpha - 1\right] f_j(x)dx \\ &\quad + \frac{1}{\alpha^2 - \alpha}\left[\sum_{j=1}^k p_j \left(\frac{\mu_j}{\sum_{j=1}^k p_j\mu_j}\right)^\alpha - 1\right], \end{aligned}$$

Let us define

$$\tau_j = p_j\mu_j / \sum_{j=1}^k p_j\mu_j,$$

and let I^j_{GE} denote the generalised entropy family index with parameter α for the group j then

$$I^{\alpha}_{GE} = \underbrace{\sum_{j=1}^{k} p_j^{1-\alpha}\tau_j^{\alpha} I^j_{GE}}_{withinGE} + \underbrace{\frac{1}{\alpha^2-\alpha}\left(\sum_{j=1}^{k} p_j^{1-\alpha}\tau_j^{\alpha} - 1\right)}_{betweenGE}. \tag{14}$$

The most popular variants of this specific class of GE family are the Theil and the MLD since they are the only zero homogeneous decomposable measures such that the weights of the within-group-inequalities in the total inequality sum to a constant (see Bourguignon (1979)). In a mixture, these two indices are

$$I_{Theil} = \underbrace{\sum_{j=1}^{k} \tau_j I^j_{Theil}}_{withinTheil} + \underbrace{\sum_{j=1}^{k} \tau_j log\left(\frac{\tau_j}{p_j}\right)}_{betweenTheil}. \tag{15}$$

$$I_{MLD} = \underbrace{\sum_{j=1}^{k} p_j I^j_{MLD}}_{withinMLD} - \underbrace{\sum_{j=1}^{k} p_j log\left(\frac{\tau_j}{p_j}\right)}_{betweenMLD}. \tag{16}$$

6.3 Index decomposition for the log normal

The log-normal distribution is widely used to model the income distribution. We have seen that even if it fits the data reasonably well, it is not sufficient to give a correct account of all the details of the income distribution. A mixture of at least two log-normal distributions does a far better job as already seen using the FES data.

The log-normal density has the particular property that first the generalised Entropy index has an analytical expression given by

$$GE = \frac{\exp((\alpha^2-\alpha)\sigma^2/2) - 1}{\alpha^2-\alpha}$$

The second property which can be easily verified is that the Theil index and the MLD are the same with

$$I_{Theil} = I_{MLD} = \frac{\sigma^2}{2}$$

For a mixture of k log-normal densities, the value τ_j is simplified into

$$\tau_j = \frac{p_j \exp(\mu_j + \sigma_j^2/2)}{\sum p_j \exp(\mu_j + \sigma_j^2/2)}.$$

But note that even if $I^j_{Theil} = I^j_{MLD}$, as we have a different decomposition for the two indices in a mixture, when estimated within a mixture of log-normals, the two indices will not be the same. This is due to the weights which are the p_j for the I_{MLD} and the τ_j for the I_{Theil}.

6.4 A Bayesian approach for computing the variance of the GE index

Suppose now that we have adjusted a mixture of log-normal densities on a given income series. We have to use a MCMC approach and thus obtain random draws for the parameters. As we have an analytical expression for the GE index, we can transform each draw of the parameter into a draw of the GE index. It is then very easy to estimate the posterior density of the GE index, compute its mean and standard deviation. Using the Family Expenditure survey and ignoring the sample design, we get estimation results reported in Table 5.

Table 5: Estimates and standard errors of GE index and Gini using the FES

	GE ($\alpha = 0.5$)	Theil	MLD	Gini
1979	0.108	0.104	0.107	0.255
	(0.0041)	(0.0027)	(0.0027)	(0.0021)
1988	0.168	0.148	0.160	0.307
	(0.0093)	(0.0043)	(0.0057)	(0.0032)
1992	0.182	0.165	0.176	0.321
	(0.0086)	(0.0042)	(0.0057)	(0.0035)
1996	0.168	0.150	0.146	0.295
	(0.0087)	(0.0097)	(0.0095)	(0.0033)

7 Testing for stochastic dominance

This problem is much more complex than the previous one. We no longer have to consider a simple index number, but a complete curve, the dominance curve. We thus have to compare curves, which means comparing two sets of points instead of two points.

When we are interested in poverty, the meaningful concept is restricted dominance. Whenever we speak about poverty, we have to define a poverty line, using for instance half the mean or half the median. If we want to make robust comparisons, it is better to select a rather wide interval instead of just a point. We thus consider the interval $[z_*, z^*]$ which corresponds to two extreme value for the poverty line. We have two samples A and B for which we have computed two dominance curves at the order s that we note $F_s^A(x)$ and $F_s^B(x)$ for the two samples. We recall that dominance curves are given by (as shown in a previous chapter using integration by parts):

$$F_s(x) = \int_0^x F_{s-1}(t)dt = \frac{1}{(s-1)!}\int_0^x (x-t)^{s-1} f(t)\, dt.$$

They are functions of x for a given s. For analysing stochastic dominance at the order 1, we consider $s = 1$, and so on.

7.1 Hypotheses

We can distinguish three different type of hypothesis that can be in turn the null and the alternative:

1. $H_0 : \delta_s(x) = F_s^A(x) - F_s^B(x) = 0 \qquad \forall x \in [z_*, z^*]$. The two distributions corresponding to samples A and B cannot be distinguished.
2. $H_1 : \delta_s(x) = F_s^A(x) - F_s^B(x) \geq 0 \qquad \forall x \in [z_*, z^*]$. The two distributions are ranked, distribution B clearly dominates distribution A.
3. H_2 : no restriction on $\delta_s(x)$. There is no possibility to rank the two distributions. They can be anything.

If we were in an uni-dimensional framework, H_0 would correspond to a point hypothesis, H_1 would lead to a unilateral test and H_2 to a bilateral test. But here these hypotheses have to be verified either for the sample points contained in the interval $[z_*, z^*]$, or over a fixed grid of equidistant covering the same interval $[z_*, z^*]$. We shall focus our attention on the latter solution.

7.2 Asymptotic distribution of empirical dominance curves

Let us consider the following grid of K equidistant points

$$z = [z_k] = z_*, z_2, \cdots, z_{K-1}, z^*. \tag{17}$$

Davidson and Duclos (2000) derive a fundamental result for what follows:

$$\sqrt{n}(\hat{F}_s(z) - F_s(z)) \sim \mathrm{N}(0, \Sigma). \tag{18}$$

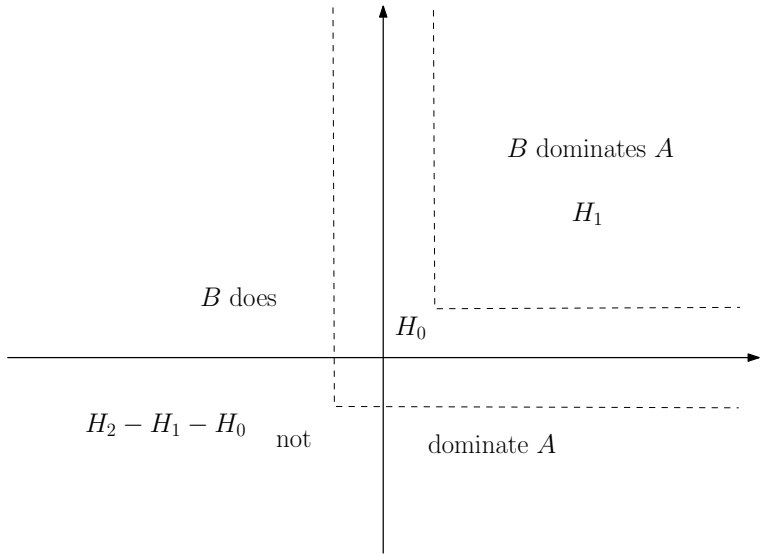

Figure 1: Tests of dominance and non-dominance
Graph of $\delta(x)$ for x_1 and x_2

They derive expressions for Σ for a whole range of cases and the reader should refer to their paper for details. This matrix expresses the correlation structure that exists between the different values of $\hat{F}_s(.)$ when the latter is computed over a grid. The proposed dominance tests take into account this correlation structure. We call $\hat{\delta}_s(z)$ the estimated difference between two dominance curves computed of the grid $[z_k]$. If the two populations A and B are independent, we have

$$\sqrt{n}(\hat{\delta}_s(z) - \delta_s(z)) \sim \mathrm{N}(0, \Omega = \Sigma_A + \Sigma_B). \tag{19}$$

7.3 An example

The best way to understand the type of tests that can be built using (19) is to follow the example given in Davidson (2010). Distribution B dominates distribution A means $\delta(x) = F_A(x) - F_B(x) \geq 0$ and its sample equivalence is $\hat{F}_A(x) - \hat{F}_B(x) \geq 0$. Dominance in the sample cannot lead to rejection of dominance in the population. Non dominance in the sample cannot lead to rejection of non dominance in the population. This relation between sample and population lead to the presence of zones of uncertainty when building a test. To illustrate this point, we can consider two distributions defined on the same support and consisting of three points each, x_1, x_2, and x_3 by increasing order. Then by definition, $F_A(x_3) = F_B(x_3) = 1$. Inference on stochastic dominance can be led considering the first two points for which we compute $\hat{\delta}(x_i)$, for $i = 1, 2$. Distribution B dominates distribution A if $\hat{\delta}(x_i) \geq 0$. We can visualise in Figure 1 the plot of $\hat{\delta}_1$ and $\hat{\delta}_2$. Dominance and H_1 corresponds to the upper right quadrant. Equality and H_0 corresponds to the zone between the two doted lines.

7.4 Inference

The first task is to estimate the dominance curve. Suppose that we have n independent draws y_i from the population. The theoretical dominance curve is

$$F_s(x) = \frac{1}{(s-1)!}\int_0^x (x-y)^{s-1} dF(y),$$

for a given value s, usually $s = 1$ or $s = 2$. A natural estimator can easily be found from this notation, using the Monte Carlo interpretation of $dF(y)$ with

$$\hat{F}_s(x) = \frac{1}{(s-1)!}\frac{1}{n}\sum_{i=1}^{n}(x-y_i)^{s-1}\mathbb{1}(y_i < x).$$

From Davidson and Duclos (2000), we deduce that the variance-covariance matrix of this estimator can be estimated by

$$\hat{\Sigma} = \frac{1}{((s-1)!)^2}\frac{1}{n}\sum_i (x-y_i)^{s-1}(x'-y_i)^{s-1}\mathbb{1}(y_i < x)\mathbb{1}(y_i < x') - \hat{F}_s(x)\hat{F}_s(x').$$

The dimension of this matrix corresponds to the dimension of the grid. It matrix gives the covariances existing between two estimated points of the dominance curve.

Let us now consider two independent samples from two independent populations A and B. We want to compare these two populations. We have to consider the distribution of the estimated difference between the two dominance curves:

$$\hat{\delta}(x) = \hat{F}_s^A(x) - \hat{F}_s^B(x).$$

We have stated that the asymptotic distribution of $\hat{\delta}(x)$ is normal with zero mean and variance the sum of the two variances when A and B are independent. So we have just to use the two estimated variance-covariance matrices $\hat{\Sigma}^A$ and $\hat{\Sigma}^B$.

Remark:

Davidson and Duclos (2000) consider the other case when we have independent draws of paired income y_i^A, y_i^B of the same population, for instance before and after tax income. We have now to estimate the variance-covariance Ω in one shot, using a formula which is not so that different as we have:

$$\Omega = \frac{1}{((s-1)!)^2}E(x-y_A)^{s-1}(x'-y_B)^{s-1}\mathbb{1}(y_A < x)\mathbb{1}(y_B < x') - F_A^s(x)F_B^s(x').$$

This covariance matrix can be consistently estimated using the natural estimator for $F_A^s(x)$ and $F_B^s(x')$. While the expectation can be estimated by

$$\frac{1}{n}\sum_i (x-y_i^A)^{s-1}(x'-y_i^B)^{s-1}\mathbb{1}(y_A < x)\mathbb{1}(y_B < x').$$

Of course in this case, we must have the same number of observations in A and B.

7.5 Three dominance tests

The more simple test consists in testing equality H_0 against the alternative H_2. It corresponds to a bilateral test and a mere generalisation of the test of Kakwani (1993) that we detailed in section 2. This test was first proposed in Beach and Davidson (1983) and later in Dardanoni and Forcina (1999) for testing the equality of two Lorenz curves. This test is implemented by means of a Wald statistics:

$$T_{02} = \hat{\delta}(z)'\Omega^{-1}\hat{\delta}(z), \tag{20}$$

which is distributed as a χ^2 with K degrees of freedom, K being the size of the grid.

Dardanoni and Forcina (1999) propose two other tests that have a much more complex implementation, even if they use the asymptotic normality result (19). The first test compares H_0 (equality of the two dominance curves) with H_1 that represents dominance of B over A. The null hypothesis is thus an hypothesis of equivalence against the alternative of dominance. The second test compares H_1 with H_2. In this case, the null hypothesis becomes dominance and the alternative represents the most general form of non-dominance. In order to describe these tests, we must be able to characterise an hypothesis in a multidimensional space. For this, Dardanoni and Forcina (1999) define a distance function of a vector with respect to a space. Thus, for instance the distance of vector δ with respect to the space described by the null hypothesis H_0 is defined as:

$$d(\delta, H_0, \Omega) = \min_{y \in H_0} (\delta - y)'\Omega^{-1}(\delta - y).$$

In order to compute this distance, we have to solve a quadratic programming problem. The required statistics are

$$T_{01} = n\hat{\delta}(z)'\Omega^{-1}\hat{\delta}(z) - \min_{y \geq 0} n(\hat{\delta}(z) - y)'\Omega^{-1}(\hat{\delta}(z) - y)$$

for the test de of equality against dominance and

$$T_{12} = \min_{y \geq 0} n(\hat{\delta}(z) - y)'\Omega^{-1}(\hat{\delta}(z) - y).$$

for the test of dominance against non-dominance.

These tests are problematic for two reasons. First, they are difficult to compute as they require the use of quadratic programming for each case. Second, they have a complicated asymptotic distribution, based on a mixture of χ^2 distributions.

However, as underlined by Dardanoni and Forcina (1999) and also by Davidson and Duclos (2006), *the test of the null hypothesis of non-dominance (the space described by H_2 in our notations where we have subtracted the space described by H_1) against the alternative hypothesis of dominance (H_1 in our notations) leads to a much simplified result.* Simplification comes from the fact that no information is lost if we neglect the correlation structure of Ω. We are then back to the test proposed by Howes (1993) for a grid and by Kaur, Prakasa Rao, and Singh (1994) for all the sample points. This test consists in computing separately the K values of the Student statistics and then to take their minimum. Consequently, our test statistics is:

$$T_{21} = \min_{z_i} \hat{\delta}(z_i)/\omega_{ii}.$$

This statistics has two advantages, compared to the previous one, which are their exact counter parts. It is easy to compute. Its asymptotic distribution is simple as it is a $\mathrm{N}(0,1)$ under the null. Davidson and Duclos (2006) recommend to use this test. They however show that it produces coherent results only if we truncate the tails of the distributions, regions where we have not enough observations. This trimming operation becomes natural when we test for restricted dominance provided the bounds z_* et z^* are adequately chosen.

7.6 A simplified test

Davidson and Duclos (2013) propose a simplified test which can be applied for testing restricted first order stochastic dominance. This test is based on the minimum t statistics and has non-dominance as the hull hypothesis. If we succeed in rejecting this null, we may legitimately infer the only other possibility, namely dominance. Considering F_A and F_B two cumulative distributions of samples of population A and B with respective sample sizes N_A and N_B. The t statistics associated to a restricted grid on z is given by

$$\frac{\sqrt{N_A N_B}(\hat{F}_A(z) - \hat{F}_B(z))}{\sqrt{N_B \hat{F}_A(z)(1-\hat{F}_A(z)) + N_A \hat{F}_B(z)(1-\hat{F}_B(z)}}.$$

For higher order dominance, it is not possible to find such simple analytical expressions and one has to rely on the approach developed in the previous subsections.

7.7 Illustration

We have already explored the FES data sets and shown that many changes occurred over the period 1979-1996 concerning inequality and poverty. The main debate was that the increase in inequality was accompanied by a large increase in the mean income. So the level of the overall welfare became a hot topic. Using those data, we can first compute means, poverty lines and various indices. We are then going to investigate the evolution of the first order dominance

Table 6: Poverty and inequality indices for FES data

	1979	1988	1992	1996
Mean	83.08	102.89	109.63	111.46
Gini	0.256	0.307	0.321	0.298
Poverty line	44.54	52.42	55.15	56.52
P_0	0.135	0.180	0.196	0.151

curves over a large interval containing the poverty line. Comparing these curves will shed some light on the evolution of the income distribution, focussing our attention on the least favoured

households. Considering the large interval $[20, 100]$, we are going to use the simplified test of Davidson and Duclos (2013) for testing restricted stochastic dominance at the order one.

The results given in Table 7 are particularly interesting. We have stochastic dominance all the time provided we restrain our attention to a region over the poverty line. The situation is improved for individuals below the poverty line only between 1992 and 1996. That result does not extends when comparing 1996 to 1978.

Table 7: First order stochastic dominance
Rejection of the null of non-dominance

z	1979-1988	1988-1992	1992-1996	1979-1996
20.00	-0.964	-2.225	2.861	-0.272
28.89	-1.668	-2.790	4.369	-0.006
37.78	5.088	-1.363	4.952	8.479
46.67	9.100	1.332	7.826	17.834
55.56	8.781	3.061	8.970	20.523
64.44	8.314	3.270	7.615	18.987
73.33	11.115	2.916	4.658	18.485
82.22	13.264	3.153	3.259	19.452
91.11	15.214	3.611	2.258	20.825
100.00	15.975	4.651	1.159	21.479

We give in Figure 2 the graph of the dominance curves corresponding to the same interval as that used to compute the tests reported in Table 7. The shape of these curves is well in accordance with the reported values of the t tests. But the test give a clear answer to know the points where the curves are statistically different. In particular, it is difficult to compare the curves 1992-1996 for low incomes. Only the tests gives a clear answer.

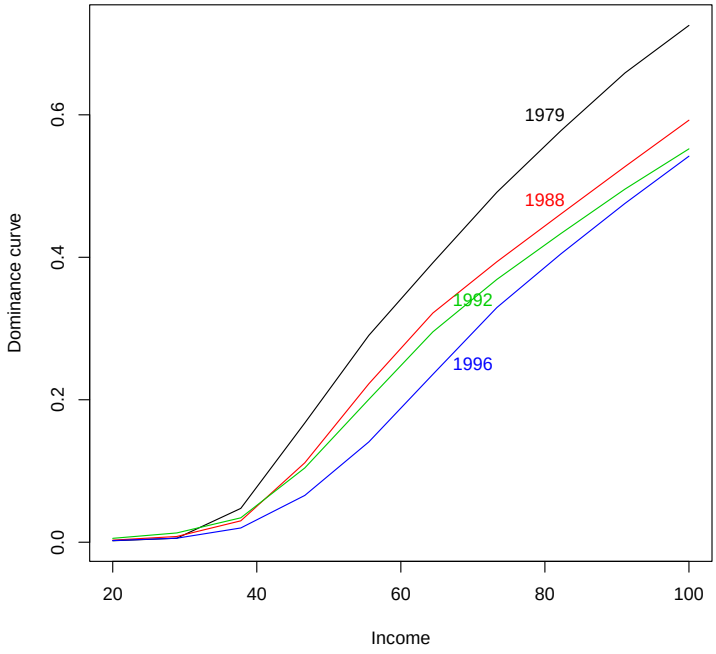

Figure 2: Dominance curves using the FES survey

8 Exercices

- Imagine a method of estimation for the poverty deficit curves and the dominance curves using order statistics.
- Using the grouped data provided in Table 1 of chapter 5, draw an histogram for each year of the US income distribution. Draw the corresponding dominance curves at the order 1.

References

BEACH, C. M., AND R. DAVIDSON (1983): "Distribution-Free Statistical Inference with Lorenz Curves and Income Shares," *The Review of Economic Studies*, 50, 723–735.

BERGER, Y. G., AND C. J. SKINNER (2003): "Variance Estimation of a Low-Income Proportion," *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 52, 457–468.

BIEWEN, M., AND S. P. JENKINS (2006): "Variance Estimation for Generalized Entropy and Atkinson Inequality Indices: the Complex Survey Data Case," *Oxford Bulletin of Economics and Statistics*, 68(3), 371–383.

BOURGUIGNON, F. (1979): "Decomposable income inequality measures," *Econometrica*, 47, 901–920.

COWELL, F. (1995): *Measuring Inequality*, LSE Handbooks on Economics Series. Prentice Hall, London.

COWELL, F. A. (1989): "Sampling variance and decomposable inequality measures," *Journal of Econometrics*, 42, 27–41.

DARDANONI, V., AND A. FORCINA (1999): "Inference for Lorenz curve orderings," *The Econometrics Journal*, 2(1), 49–75.

DAVIDSON, R. (2009): "Reliable Inference for the Gini Index," *Journal of Econometrics*, 150, 30–40.

——— (2010): "Inference on income distributions," Discussion Paper No 2010-48, GREQAM, Marseille.

DAVIDSON, R., AND J.-Y. DUCLOS (2000): "Statistical inference for stochastic dominance and for the measurement of poverty and inequality," *Econometrica*, 68, 1435–1464.

——— (2006): "Testing for Restricted Stochastic Dominance," working paper 2006-36, ECINEQ.

——— (2013): "Testing for Restricted Stochastic Dominance," *Econometric Reviews*, 32(1), 84–125.

FOSTER, J., J. GREER, AND E. THORBECKE (1984): "A class of decomposable poverty measures," *Econometrica*, 52, 761–765.

GASTWIRTH, J. L. (1972): "The Estimation of the Lorenz Curve and Gini Index," *The Review of Economics and Statistics*, 54(3), 306–316.

HOWES, S. (1993): "Asymptotic Properties of Four Fundamental Curves of Distributional Analysis," Unpublished paper, STICERD, London School of Economics.

KAKWANI, N. (1993): "Statistical inference in the measurement of poverty," *Review of Economics and Statistics*, 75(4), 632–639.

KAUR, A., B. PRAKASA RAO, AND H. SINGH (1994): "Testing for Second-Order Stochastic Dominance of Two Distributions," *Econometric Theory*, 10, 849–866.

LUBRANO, M., AND A. J. NDOYE (2011): "Inequality decomposition using the Gibbs output of a Mixture of lognormal distributions," Discussion Paper 2011-19, GREQAM.

QUANDT, R. E. (1966): "Old and new methods of estimation and the Pareto distribution," *Metrika*, 10(1), 55–82.

SCHADER, M., AND F. SCHMID (1994): "Fitting parametric Lorenz curves to grouped income distribution: a critical note," *Empirical Economics*, 19(3), 361–370.

SEN, A. (1976): "Poverty: an Ordinal Approach to Measurement," *Econometrica*, 44(2), 219–231.

SHORROCKS, A. F. (1980): "The Class of Additively Decomposable Inequality Measures," *Econometrica*, 48(3), 613–625.