Chapter 6. Inequality Measures

Summary

Inequality is a broader concept than poverty in that it is defined over the *entire* population, and does not only focus on the poor.

The simplest measurement of inequality sorts the population from poorest to richest and shows the percentage of expenditure (or income) attributable to each fifth (quintile) or tenth (decile) of the population. The poorest quintile typically accounts for 6-10% of all expenditure, the top quintile for 35-50%.

A popular measure of inequality is the Gini coefficient, which ranges from 0 (perfect equality) to 1 (perfect inequality), but is typically in the range of 0.3-0.5 for per capita expenditures. The Gini is derived from the Lorenz curve, which sorts the population from poorest to richest, and shows the cumulative proportion of the population on the horizontal axis and the cumulative proportion of expenditure (or income) on the vertical axis. While the Gini coefficient has many desirable properties – mean independence, population size independence, symmetry, and Pigou-Dalton Transfer sensitivity – it cannot easily be decomposed to show the sources of inequality.

The best-known entropy measures are Theil's T and Theil's L, both of which allow one to decompose inequality into the part that is due to inequality within areas (e.g. urban, rural) and the part that is due to differences between areas (e.g. the rural-urban income gap). Typically at least three-quarters of inequality in a country is due to within-group inequality, and the remaining quarter to between-group differences.

Atkinson's class of inequality measures is quite general, and is sometimes used. The decile dispersion ratio – defined as the expenditure (or income) of the richest decile divided by that of the poorest decile – is popular but a very crude measure of inequality.

It is often helpful to decompose inequality by occupational group, or by source of income, in order to identify policies that would help moderate inequality.

Learning Objectives

After completing the module on poverty lines, you should be able to:

21. Explain what inequality is, and how it differs from poverty.
22. Compute and display information on expenditure (or income) quintiles.
23. Draw and interpret a Lorenz curve.
24. Compute and explain the Gini coefficient of inquality.
25. Argue that the Gini Coefficient satisfies mean independence, population size independence, symmetry, and Pigou-Dalton Transfer sensitivity, but is not easily decomposable.
26. Compute and interpret generalized entropy measures, including Theil's T and Theil's L.
27. Compute and interpret Atkinson's inequality measure for different values of the weighting parameter ε.
28. Compute and criticize the decile dispersion ratio.
29. Decompose inequality using Theil's T in order to distinguish between-group from within-group components of inequality, for separate geographic areas, occupations, and income sources.

6.1 Definition of inequality

The main focus of this manual is on poverty, which looks at the situation of individuals or households who find themselves at the bottom of the income distribution; typically this requires information both about the mean level of (say) expenditure per capita as well as its distribution at the lower end. But sometimes we are more interested in measuring inequality than poverty per se, and for that reason we have included this relatively brief chapter on measuring inequality.

Inequality is a broader concept than poverty in that it is defined over the *entire* population, and not just for the population below a certain poverty line. Most inequality measures do not depend on the mean of the distribution, and this property of mean independence is considered to be a desirable property of an inequality measure. Of course, inequality measures are often calculated for distributions other than expenditure – for instance, for income, land, assets, tax payments, and many other continuous and cardinal variables.

The simplest way to measure inequality is by dividing the population into fifths (*quintiles*) from poorest to richest, and reporting the levels or proportions of income (or expenditure) that accrue to each level. Table 6.1 shows the level of expenditure per capita, in '000 dong per year, for Vietnam in 1993, based on data from the Vietnam Living Standards Survey. A fifth of the individuals (not households) included in the survey were allocated to each expenditure quintile. The figures show that 8.4% of all expenditures were made by the poorest fifth of households, and 41.4% by the top fifth. Quintile information is easy to understand, although sometimes one wants a summary measure rather than a whole table of figures.

Table 6.1: Breakdown of expenditure per capita by quintile, Vietnam 1993

	Expenditure quintiles					
	Lowest	Low-mid	Middle	Mid-upper	Upper	Overall
Per capita expenditure ('000 dong/year)	518	756	984	1,338	2,540	1,227
% of expenditure	8.4	12.3	16.0	21.8	41.4	100.0
Memo: Cumulative % of expenditure	8.4	20.7	36.7	58.5	100.0*	
Memo: Cumulative % of population	20.0	40.0	60.0	80.0	100.0	

Source: Vietnam Living Standards Survey 1993.
Note: * There is a slight rounding error here.

6.2 Commonly used summary measures of inequality

6.2.1 Gini coefficient of inequality

The most widely used single measure of inequality is the Gini coefficient. It is based on the Lorenz curve, a cumulative frequency curve that compares the distribution of a specific variable (e.g. income) with the uniform distribution that represents equality. To construct the Gini coefficient, graph the *cumulative* percentage of households (from poor to rich) on the horizontal axis and the *cumulative* percentage of expenditure (or income) on the vertical axis. The Lorenz curve shown in figure 1 is based on the Vietnamese data in Table 6.1. The diagonal line represents perfect equality. The Gini coefficient is defined as A/(A+B), where A and B are the areas shown on the graph. If A=0 the Gini coefficient becomes 0 which means perfect equality, whereas if B=0 the Gini coefficient becomes 1 which means complete inequality. In this example the Gini coefficient is about 0.35.

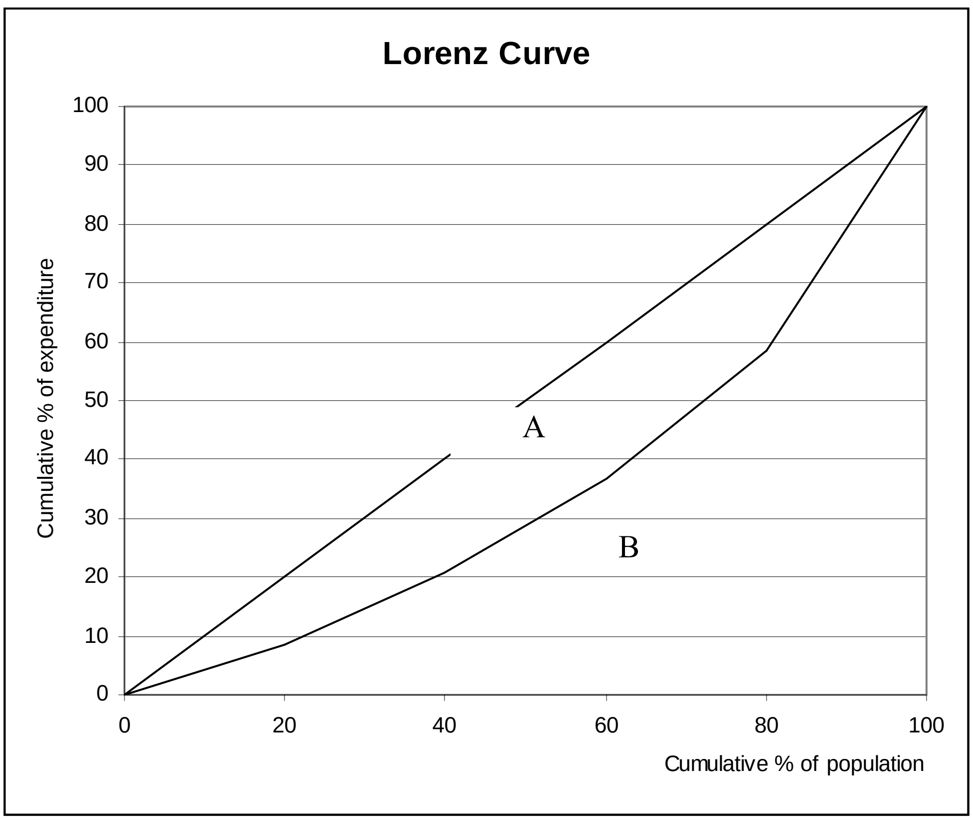

Figure 6.1. ***Lorenz Curve***

Formally, let x_i be a point on the X-axis, and y_i a point on the Y-axis. Then

(6.1) $$Gini = 1 - \sum_{i=1}^{N}(x_i - x_{i-1})(y_i + y_{i-1}).$$

When there are N equal intervals on the X-axis this simplifies to

(6.2) $$Gini = 1 - \frac{1}{N}\sum_{i=1}^{N}(y_i + y_{i-1}).$$

For users of Stata, there is a `gini` command that may be downloaded and used directly (see Appendix 3). This command also has the advantage that it allows one to use weights, which are not incorporated into the two equations shown above.

The Gini coefficient is not entirely satisfactory. To see this, consider the criteria that make a good measure of income inequality, namely:

- *Mean independence.* This means that if all incomes were doubled, the measure would not change. The Gini satisfies this.
- *Population size independence.* If the population were to change, the measure of inequality should not change, ceteris paribus. The Gini satisfies this too.
- *Symmetry.* If you and I swap incomes, there should be no change in the measure of inequality. The Gini satisfies this.
- *Pigou-Dalton Transfer sensitivity.* Under this criterion, the transfer of income from rich to poor reduces measured inequality. The Gini satisfies this too.

It is also desirable to have

- *Decomposability.* This means that inequality may be broken down by population groups or income sources or in other dimensions. The Gini index is not easily decomposable or additive across groups. That is, the total Gini of society is not equal to the sum of the Gini coefficients of its subgroups.
- *Statistical testability.* One should be able to test for the significance of changes in the index over time. This is less of a problem than it used to be because confidence intervals can typically be generated using bootstrap techniques.

6.2.2 Generalized Entropy measures

There are a number of measures of inequality that satisfy all six criteria. Among the most widely used are the Theil indexes and the mean log deviation measure. Both belong to the family of generalized entropy inequality measures. The general formula is given by:

(6.3)
$$GE(\alpha) = \frac{1}{\alpha(\alpha-1)}\left[\frac{1}{N}\sum_{i=1}^{N}\left(\frac{y_i}{\bar{y}}\right)^{\alpha} - 1\right]$$

where $\bar{y}$ is the mean income (or expenditure per capita). The values of GE measures vary between 0 and ∞, with zero representing an equal distribution and higher value representing a higher level of inequality. The parameter α in the GE class represents the weight given to distances between incomes at different parts of the income distribution, and can take any real value. For lower values of α, GE is more sensitive to changes in the lower tail of the distribution, and for higher values GE is more sensitive to changes that affect the upper tail. The commonest values of α used are 0,1 and 2. GE(1) is Theil's T index, which may be written as

(6.4)
$$GE(1) = \frac{1}{N}\sum_{i=1}^{N}\frac{y_i}{\bar{y}}\ln\left(\frac{y_i}{\bar{y}}\right)$$

GE(0), also known as Theil's L, and sometimes referred to as the mean log deviation measure, is given by:

(6.5)
$$GE(0) = \frac{1}{N}\sum_{i=1}^{N}\ln\left(\frac{\bar{y}}{y_i}\right)$$

Once again, users of Stata do not need to program the computation of such measures from scratch; the `GE` command, explained in Appendix 3, allows one to get these measures, even when weights need to be used with the data.

6.2.3 Atkinson's inequality measures

Atkinson has proposed another class of inequality measures that are used from time to time. This class also has a weighting parameter ε (which measures aversion to inequality) and some of its theoretical properties are similar to those of the extended Gini index. The Atkinson class, which may be computed in Stata using the `Atkinson` command, is defined as:

(6.6)

$$A_\varepsilon = 1 - \left[\frac{1}{N} \sum_{i=1}^{N} \left(\frac{y_i}{\bar{y}} \right)^{1-\varepsilon} \right]^{1/(1-\varepsilon)}, \varepsilon \neq 1$$

$$= 1 - \frac{\prod_{i=1}^{N} (y_i^{(1/N)})}{\bar{y}}, \qquad \varepsilon = 1.$$

Table 6.2 sets out in some detail the computations involved in the computation of the Generalized Entropy and Atkinson measures of inequality. The first row of numbers gives the incomes of the ten individuals who live in a country, in regions 1 and 2. The mean income is 33. To compute Theil's T, one first computes y_i/*ybar*, where *ybar* is the mean income level; then compute ln(y_i/*ybar*), take the product, add up the row, and divide by the number of people. Similar procedures yield other generalized entropy measures, and also the Atkinson measures.

Table 6.2: Computing Measures of Inequality

		Region 1					Region 2				
Incomes (=yi)		10	15	20	25	40	20	30	35	45	90
Mean income (ybar)	33.00										
yi/ybar		0.30	0.45	0.61	0.76	1.21	0.61	0.91	1.06	1.36	2.73
ln(yi/ybar)		-0.52	-0.34	-0.22	-0.12	0.08	-0.22	-0.04	0.03	0.13	0.44
Product		-0.16	-0.16	-0.13	-0.09	0.10	-0.13	-0.04	0.03	0.18	1.19
GE(1): Theil's T	**0.080**										
ln(ybar/yi)		0.52	0.34	0.22	0.12	-0.08	0.22	0.04	-0.03	-0.13	-0.44
GE(0): Theil's L	**0.078**										
(yi/ybar)^2		0.09	0.21	0.37	0.57	1.47	0.37	0.83	1.12	1.86	7.44
GE(2)	**0.666**										
(yi/ybar)^.5		0.55	0.67	0.78	0.87	1.10	0.78	0.95	1.03	1.17	1.65
Atkinson, e=0.5	**0.087**										
(yi)^(1/n)		1.26	1.31	1.35	1.38	1.45	1.35	1.41	1.43	1.46	1.57
Atkinson, e=1	**0.164**										
(yi/ybar)^(-1)		3.30	2.20	1.65	1.32	0.83	1.65	1.10	0.94	0.73	0.37
Atkinson, e=2	**0.290**										

Here are some examples of different measures of inequality (Dollar and Glewwe 1999, p.40):

Table 6.3: Expenditure inequality in selected less developed countries			
Country	Gini coefficient	Theil T	Theil L
Côte d'Ivoire, 1985-86	0.435	0.353	0.325
Ghana, 1987-88	0.347	0.214	0.205
Jamaica, 1989	n/a	0.349	0.320
Peru, 1985-86	0.430	0.353	0.319
Vietnam, 1992-93	0.344	0.200	0.169
Source: Reported in Dollar and Glewwe (1999), p.40.			

6.2.4 Decile dispersion ratio

A simple, and widely-used, measure is the decile dispersion ratio, which presents the ratio of the average consumption of income of the richest 10 percent of the population divided by the average income of the bottom 10 percent. This ratio can also be calculated for other percentiles (for instance, dividing the average consumption of the richest 5 percent – the 95th percentile – by that of the poorest 5 percent – the 5th percentile).

The decile ratio is readily interpretable, by expressing the income of the top 10% (the "rich") as a multiple of that of those in the poorest decile (the "poor"). However, it ignores information about incomes in the middle of the income distribution, and does not even use information about the distribution of income within the top and bottom deciles.

6.3 Inequality comparisons

Many of the tools used in the analysis of poverty can be similarly used for the analysis of inequality. In a way analogous to a poverty profile (see chapter 7), one could draw a profile of inequality, which among other things would look at the extent of inequality among certain groups of households. This provides information on the homogeneity of the various groups, an important element to take into account when designing policy interventions.

One may also analyze the nature of changes in inequality over time. One could focus on changes for different groups of the population to show whether inequality changes have been similar for all or have taken place, say, in a particular sector of the economy. In rural Tanzania, although average incomes increased substantially between 1983 and 1991, inequality increased (with the Gini coefficient increasing from 0.52 to 0.72), especially among the poor. This can be linked to important reforms that took place in agricultural price policy, which intensified inequalities, with the poor and less-efficient farmers failing to participate in the growth experienced by wealthier, more efficient farmers (Ferreira, 1996).

It is often instructive to analyze other dimensions of inequality. For instance, in a country where public health provision is well developed and reaches all strata of the population, one could expect to see lower levels of inequality in health outcomes than in income levels, a proposition that could also be tested formally.

6.4 Decomposition of income inequality

The common inequality indicators mentioned above can be used to assess the major contributors to inequality, by different subgroups of the population and regions as well as by income source. For example, average income may vary from region to region, and this alone implies some inequality "between groups." Moreover, incomes vary inside each region, adding a "within group" component to total inequality. For policy purposes it is useful to be able to decompose these sources of inequality: if most inequality is due to disparities across regions, for instance, then the focus of policy may need to be on regional economic development, with special attention to helping the poorer regions.

More generally, in static decompositions, household and personal characteristics, such as education, gender, occupation, urban and rural, and regional location, are determinants of household income. If that is the case, then at least part of the value of any given inequality measure must reflect the fact that people have different educational levels, occupations, genders, and so on. This inequality is the "between-group" component.

But for any such partition of the population, whether by region, occupation, sector or any other attribute, some inequality will also exist among those people within the same subgroup; this is the "within-group" component. The Generalized Entropy class of indicators, including the Theil indexes, can be decomposed across these partitions in an additive way, but the Gini index cannot.

To decompose Theil's T index (i.e. GE(1)), let Y be the total income of the population, Y_j the income of a subgroup, N the total population, and N_j the population in the subgroup. Using T to represent GE(1)

(6.7)

$$T = \sum_{i=1}^{N} \frac{y_i}{N\bar{y}} \ln\left(\frac{y_i N}{\bar{y}N}\right) = \sum_{i=1}^{N} \frac{y_i}{Y} \ln(\frac{y_i N}{Y})$$
$$= \sum_j \left(\frac{Y_j}{Y}\right) T_j + \sum \left(\frac{Y_j}{Y}\right) \ln\left(\frac{Y_j / Y}{N_j / N}\right)$$

This decomposes the inequality measure into two components. The first term represents the within-group inequality and the second term represents the between-group inequality. Similarly, GE(0) can also be decomposed. Using L to represent GE(0):

(6.8)

$$L = \sum_{i-1}^{N} \frac{1}{N} \ln\left(\frac{Y}{Y_i N}\right) = \sum_j \left(\frac{N_j}{N}\right) L_j + \sum_j \frac{N_j}{N} \ln\left(\frac{N_j / N}{Y_j / Y}\right).$$

Exercise: Decompose Theil's T measure of inequality into "within" and "between" components, using the income data provided in Table 6.2. [Hint: "Within" inequality should account for 69.1% of all inequality.]

For a typical decomposition of inequality in expenditure per capita, consider the following simple example, again from Dollar and Glewwe (1999, p.41), which refers to Vietnam in 1993. Using Theil's T, Table 6.4 shows that 22% of the total inequality is attributable to between-group inequality - i.e. to the difference in expenditure levels between urban and rural areas. The remaining 78% of all inequality is due to the inequality in expenditure per capita that occurs within each region.

Table 6.4: Decomposition of expenditure inequality by area, Vietnam, 1993

	Theil T	Between-group inequality	Memo: Population share (%)
All Vietnam	0.200		100
Urban only	0.196	0.044 (22% of total)	20
Rural only	0.136		80

Source: Dollar and Glewwe (1999), p.41.

Similar results were found for Zimbabwe in 1995-96. There a decomposition of Theil's T coefficient showed that the within-area (within rural areas and within urban areas) contribution to inequality was 72 percent, while the between–area (between urban and rural areas) component was 28 percent. In many

Latin American countries, the between-area component of inequality explains a much higher share of total inequality.

Of equal interest is which of the different income sources, or components of a measure of well-being, are primarily responsible for the observed level of inequality. For example, if total income can be divided into self-employment income, wages, transfers, and property income, one can examine the distribution of each income source. If one of the income sources were raised by one percent, what would happen to overall inequality?

Table 6.5 shows the results for the Gini coefficient for income sources in Peru (1997). As the table shows, self-employment income is the most equalizing income source. Thus a 1% increase in self-employment income (for everyone that receives such income) would lower the Gini by 4.9%, which represents a reduction in overall inequality. On the other hand, a rise in property income would be associated with an increase in inequality.

Generally, results such as these depend on two factors:

(1) the importance of the income source in total income (for larger income sources, a given percentage increase will have a larger effect on overall inequality), and

(2) the distribution of that income source (if it is more unequal than overall income, an increase in that source will lead to an increase in overall inequality).

Table 6.5 also shows the effect on the inequality of the distribution of *wealth* of changes in the value of different sources of wealth.

Table 6.5: Peru: Expected change in income inequality resulting from a one percent change in income source, 1997 (as percentage of Gini change)

Income source	Expected change	Wealth sources	Expected change
Self-employment income	-4.9	Housing	1.9
Wages	0.6	Durable goods	-1.5
Transfers	2.2	Urban property	1.3
Property income	2.1	Agricultural property	-1.6
		Enterprises	0

A final example, in the same spirit, comes from Egypt. There it was found that, in 1997, agricultural income represented the most important inequality-increasing source of income, while non-farm income has the greatest inequality-reducing potential. Table 6.6 sets out this decomposition and shows that while agricultural income only represents 25% of total income in rural areas, it accounts for 40% of the inequality.

Table 6.6: Decomposition of income inequality in rural Egypt, 1997				
Income Source	Percentage of households receiving income from this source	Share in total income (%)	Concentration index for the income source	Percentage contribution to overall income inequality
Non-farm	61	42	0.63	30
Agricultural	67	25	1.16	40
Transfer	51	15	0.85	12
Livestock	70	9	0.94	6
Rental	32	8	0.92	12
All sources	100	100		100

References

Dollar, David and Paul Glewwe. 1999.

Ferreira. 1996.

Vietnam: General Statiistical Office. 2000. *Viet Nam Living Standards Survey 1997-1998*, Statistical Publishing House, Hanoi.