Technical Paper Series

PROVIDE PROJECT
The Provincial Decision-making Enabling Project

Technical Paper 2003:4

Measures of Poverty and Inequality: A Reference Paper

Elsenburg
October 2003

PROVIDE PROJECT

The Provincial Decision-making Enabling Project

Overview

The Provincial Decision-Making Enabling (PROVIDE) Project aims to facilitate policy design by supplying policymakers with provincial and national level quantitative policy information. The project entails the development of a series of databases (in the format of Social Accounting Matrices) for use in Computable General Equilibrium models.

The National and Provincial Departments of Agriculture are the stakeholders and funders of the PROVIDE Project. The research team is located at Elsenburg in the Western Cape.

PROVIDE Research Team

Project Leader:	Cecilia Punt
Senior Researchers:	Kalie Pauw
	Esther Mohube
Junior Researchers:	Benedict Gilimani
	Lillian Rantho
	Rosemary Leaver
Technical Expert:	Scott McDonald
Associate Researchers:	Lindsay Chant
	Christine Valente

PROVIDE Contact Details

Private Bag X1
Elsenburg, 7607
South Africa

ceciliap@elsenburg.com
+27-21-8085191
+27-21-8085210

For the original project proposal and a more detailed description of the project, please visit www.elsenburg.com/provide

Measure of Poverty and Inequality: A Reference Paper[1]

Abstract

This paper discusses various measures of poverty and inequality found in the literature. Inequality measures discussed include the range, the variance, the coefficient of variation, the standard deviation of logarithms, the Gini coefficient, Theil's Entropy measure and Atkinson's inequality measure. Of these the mean log deviation, the Theil index and the coefficient of variation have come to be known as the Generalised Entropy class of inequality measures. As far as poverty indicators are concerned the Foster-Greer-Thorbecke measures, a class of generalised decomposable poverty measures, have become very popular in the literature. The paper also discusses some Stata® do-files that were written in order to calculate poverty and inequality measures, with application to the Income and Expenditure Survey data of 1995.

[1] The main author of this paper is Kalie Pauw, Senior Researcher of the PROVIDE Project. This version was revised in December 2004.

Table of contents

List of figures

List of tables

1. Introduction

This note discusses various measures of poverty and inequality. The discussion is fairly technical. The data used is the Stata-format *combined.dta* file, a combined IES/OHS 1995 file developed during an earlier stage of the PROVIDE Project (see *Technical Paper 2003: 2*). The Stata do-files are, as far as possible, generic. The code is easy to understand and can easily be adapted to suit other datasets as well. The paper starts with a discussion of inequality measures (section 2). The structure and content of this section draws mainly on Sen (1997), while articles by McDonald *et al*. (1999), Leibbrandt *et al*. (1999), Woolard (1998) and others are drawn on to illustrate how the theory is applied in practice in South Africa. Section 3 turns to a discussion on poverty measures. This section concentrates on the Foster-Greer-Thorbecke class of poverty measures. Worked examples, results, tables and additional equations are added as an appendix in section 6.

The aim of the paper is not to discuss these results in any detail, nor to draw conclusions. It merely serves as a reference framework for future research. The results are included for those who may wish to verify or replicate results. South African households are divided into urban/rural and racial groups to also illustrate how the within-group inequalities can be calculated. These groups are referred to as 'sub-groups' throughout.

2. Measures of inequality

Inequality can be defined as the dispersion of the distribution of income or some other welfare indicator (Litchfield, 1999). There are various ways to measure inequality. The ones most frequently used in practice usually conform to a certain set of axioms. These axioms or 'desirable properties' are the following (see Litchfield (1999) for more details).

- The Pigou-Dalton Transfer Principle: An income transfer from a poorer to a richer person should register a rise in inequality, or at least not a fall.
- Income scale independence: Inequality measures should be unaffected if there is a uniform proportional change in households' income.
- Decomposability: This requires that overall inequality should be related consistently to constituent parts of the population such as population sub-groups.
- Principle of population: Inequality measures should be invariant to replications of the population. For example, merging two identical datasets should not alter the distribution.

- Anonymity or symmetry: The inequality measure should be independent of any characteristics of individuals (or households) other that their income (or the welfare indicator whose distribution is being measured).

Cowell (1995, cited in Litchfield, 1999) shows that any measure of inequality that satisfies these axioms is a member of the Generalised Entropy (GE) class of inequality measures (see appendix, section 6.3 for more details). Before that, however, some basic inequality measures are discussed. The Stata data file *combined.dta* is adapted in do-file *equality.do* to only include household-level observations (see appendix, section 6.1). This file is saved as *combine3.dta* and used for all calculations and examples in this note. Results are also included in the appendix (section 6.2).

The following notational conventions will be adopted. Inequality measures will be derived by considering the distribution of income over n households, $i = 1,\dots, n$, where y_i is the income level (or some other welfare indicator) of household i. In this paper all inequality measures are based on the adult equivalent per capita income of households.[2] The average income, μ, is defined as (see Table 3)

$$\mu = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad [1]$$

The inequality measures discussed here include the range, the relative mean deviation, the variance, the coefficient of variation, the standard deviation of logarithms, the Gini-coefficient, Theil entropy measure and Atkinson's inequality measures.

2.1. The range

A very crude indicator of inequality is the range (R), described by Sen (1997: 24) as "*perhaps the simplest measure*" of inequality. This measure divides the difference between the highest and lowest income by the mean income.

$$R = {\left(Max_i\, y_i - Min_i\, y_i\right)}\big/{\mu} \qquad [2]$$

If income is evenly distributed all households earn the same, and hence $R = 0$. At the other extreme, if one person earns all the income $R = n$, where n is the size of the population. Clearly the limitation of the range is that it ignores the distribution in between the extremes

[2] The number of adult equivalent household members (E) is given by the transformation $E = \left(A + \alpha K\right)^{\theta}$, where A is the number of adults in each household and K the number of children under the age of 10. Following May (1995, cited in Leibbrandt *et al.*, 2001) we choose $\alpha = 0.5$ and $\theta = 0.9$. These parameters control for the fact that (1) children require a lower level of consumption than adults (α) and (2) larger households benefit from economies of scale (θ).

(Sen, 1997: 25). The range is calculated for all households as well as the sub-groups. Results appear in Table 3 (variable *range*).

2.2. Relative mean deviation

The relative mean deviation improves on the range by not only considering the two extreme income levels. This inequality measure compares the income of each observation with the mean income. It is calculated by taking the sum of the absolute differences between the income of each household and the sample mean income divided by the total income (mean income times number of observations).[3]

$$M = \frac{1}{n\mu}\sum_{i=1}^{n}\left|y_i - \mu\right| \qquad [3]$$

If income is perfectly distributed all households will earn the mean income, and $M = 0$, and $M = 2(n-1)/n$ when one household earns all the income. The main problem with this measure is that it is insensitive to transfers between households who find themselves on the same side of the mean income level (Sen, 1997: 26), thus violating the Pigou-Dalton principle. The relative mean deviation is calculated in do-file *inequality.do* (see Table 3, variable *rmdev*).

2.3. The variance and coefficient of variation

The variance of a stochastic variable is also estimated using the deviation from the mean, but instead of using the absolute differences, these differences are squared. This has the result of accentuating the differences that are further away from the mean (Sen, 1997: 27). The standard deviation is simply defined as the square root of the variance. The latter can be estimated using the following formula.

$$V = \frac{1}{n}\sum_{1=1}^{n}\left(y_i - \mu\right)^2 \qquad [4]$$

From an inequality analysis point of view an attractive feature of the variance (or standard deviation) is that any transfer from a poorer person to a richer person, *ceteris paribus*, will increase the variance and hence the inequality, thus satisfying the Pigou-Dalton principle for inequality measures (Sen, 1997: 27). However, the variance depends on the mean income, and one distribution may show a greater relative variation but have a lower variance if it has a smaller mean. The variance is also not independent of the income scale. If all incomes are doubled, the variance quadruples, thus violating the income scale independence axiom. This is perhaps an "*undesirable property*" (see Litchfield, 1999).

[3] The equation for M in Sen (1997: 25) appears to be faulty. Instead of dividing by $n\mu$ Sen multiplies by $n\mu$.

The coefficient of variation counters this problem by concentrating on the relative variation. It is simply defined as the standard deviation divided by the mean.

$$C = \frac{\sqrt{V}}{\mu} \qquad [5]$$

This inequality measure is a member of the Generalised Entropy measures (see section 6.3). The coefficient of variation has the property that it attaches equal weights to transfers at different levels of income. If a household with income *y* transfers some of its income to another with income (*y* – *d*), the impact is the same whatever the level of *y* (Sen, 1997: 28). The standard deviation (variable *sdy*), variance (variable *vary*) and coefficient of variation (variable *coefvar*) for all households and various sub-groups are reported in Table 3.

2.4. The standard deviation of logarithms

If one wishes to attach greater importance to lower income levels it is necessary to stagger income levels. Since the natural logarithmic function's second order differential is negative, the logarithm of income will be staggered at the top end of the income distribution. Sen (1997: 29) notes that the statistical literature prefers using the logarithm of the geometric mean, but in the income distribution literature the use of the arithmetic means seems more common. The following formula is used to calculate this inequality measure.

$$H = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\log y_i - \log \mu\right)^2} \qquad [6]$$

This measure is also a member of Generalised Entropy class (see appendix, section 6.3). Whereas the coefficient of variation attaches equal weights to transfers at different levels, the standard deviation of logarithms attaches more weight to transfers at the lower end of the income distribution (Litchfield, 1999). Thus, although the logarithmic transformation "*softens the blow in reflecting inequality since it reduces the deviation*", it is useful if one wishes to highlight differences at the lower end of the income scale (Sen, 1997: 29). Table 3 shows the calculated values of the standard deviation of logarithms (variable *sdlog*).

2.5. The Gini coefficient and Gini decomposition

The Gini coefficient is closely related to the concept of income shares of groups or households. It can be defined in terms of a Lorenz curve of a country (see Figure 1). The cumulative percentage of households is plotted against the cumulative share of income, giving rise to a convex Lorenz curve that always lies below the line of perfect equality if income is imperfectly distributed. The line of perfect equality is the 45-degree line in Figure 1. The Gini coefficient can be calculated by dividing the area between the Lorenz curve and the line of perfect equality by the total area underneath the line of perfect equality.

Technically speaking the Gini Coefficient varies between zero and one, although in reality values usually range between 0.20 and 0.30 for countries with a low degree of inequality and between 0.50 and 0.70 for countries with highly unequal income distributions (Todaro, 1997). Table 1 shows the average Gini Coefficient for various groups of countries.

Figure 1: The South African Lorenz curve - 1995

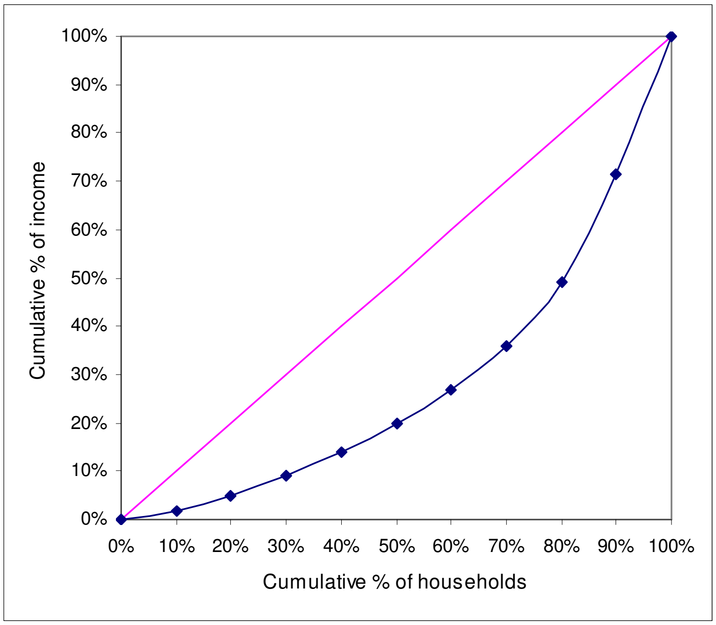


Table 1: Trends in income distribution – 1960 and 1980

Group of Countries	Income Distribution Gini Coefficient 1960	1980
All non-communist developing countries	0.544	0.602
Low-income countries	0.407	0.450
Middle-income, non-oil-exporting countries	0.603	0.569
Oil-exporting countries	0.575	0.612
South Africa (1995)*	0.64	

Source: Adelman (1986) cited in Todaro (1997).
* Own calculation, also see Table 3 and Table 4.

Although the Gini Coefficient and Lorenz curves are useful to give a general idea of the extent of the inequality, there is at least one major weakness. The Lewis two-sector model of economic development (see Todaro, 1997: 76, 142) can be used as an example to illustrate this weakness. In the Lewis model it is assumed that the economy is made up of a traditional sector with low-income workers and a modern sector with high-income workers. The traditional sector is further characterised by zero marginal productivity of labour. Surplus labour can therefore be withdrawn from this sector without a loss in output.

If there is enrichment in the traditional sector, the distribution of income will become more equal, and hence the Lorenz curve will shift closer to the line of perfect equality. Alternatively, in the case of modern sector enrichment, the income distribution becomes more skewed and the Lorenz curve shifts further away from the line of perfect of equality. In both cases it is possible to make an unambiguous statement about the impact of such occurrences on inequality. However, when one considers the Lewis-type modern sector enlargement growth, the outcome is not so clear any longer. In the Lewis-model surplus labour from the traditional sector is absorbed in the modern sector. This leads to both an increase in output and employment in the modern sector. Absolute incomes increase *and* poverty is reduced. The new Lorenz curve will intersect the old one.

This crossing of Lorenz curves can be explained as follows: the poor who remain in the traditional sector still earn the same income as before, but their income is now a smaller fraction of total income due to the growth in total output in the modern sector. This implies that the Lorenz curve is further from the line of perfect equality for low-income groups. At the other end of the income spectrum modern sector workers still earn the same wage as before, but because there are more workers earning this higher wage, the share of the richest income group has declined. The new Lorenz curve is therefore closer to the line of perfect equality for high-income households. The new Lorenz curve therefore crosses the old one. Whether the new Gini Coefficient is higher, lower or the same is irrelevant since no unambiguous welfare judgement can be made. Consequently each country has to be analysed on a case-by-case basis (Todaro, 1997: 145).

There are various formulas that can be used to calculate the Gini coefficient. Sen (1997: 31) shows that the Gini is equal to one half of the relative mean difference, which is defined as the arithmetic mean of the absolute values of differences between all pairs of incomes. With a bit of manipulation it can be shown that

$$G = 1 + \left(\frac{1}{n}\right) + \left(\frac{2}{n^2\mu}\right)\left[y_1 + 2y_2 + \ldots + ny_n\right] \qquad [7]$$

for $y_1 \geq y_2 \geq \ldots \geq y_n$. McDonald *et al.* (1999: 7) follow Stuart (1954) by defining the Gini coefficient in terms of covariances. Formally,

$$G = \frac{2\operatorname{cov}\left(y, F(y)\right)}{\mu} \qquad [8]$$

where y and μ are defined as before, and *F(y)* is the cumulative density function of income, which is uniformly distributed over [0,1].

The Gini coefficient fails the decomposability axiom if the sub-vectors of income overlap. Litchfield (1999) states that although there are "*ways of decomposing the Gini … the component terms of total inequality are not always intuitively or mathematically appealing*". Despite this concern raised, it still remains a popular inequality measure, both to express total inequality and as a decomposable measure. Over the last decade various techniques have been refined for decomposing the Gini coefficient by income sources.

South African households, for example, earn income from a variety of sources. Income can be derived from supplying labour services to firms (variable *inclab*). This includes salaries, bonuses, commission and in-kind transfers from employers. This is the most important source of income for most households. Secondly, income from 'gross operating surplus' (*incgos*) can be earned. This includes income resulting from the ownership of capital, such as interest, rental income or income from a business in which the household has invested. Thirdly, household income can include transfers from other households, i.e. inter-household transfers (*inctrans*). A fourth source of income is income from corporations (*inccorp*), which mainly includes interest, dividends and royalties earned by households due to some association with corporations. It further includes income from annuities, property sales, and insurance claims. A fifth source is transfers from government (*incgov*), which includes unemployment, disability or old-age grants, government pensions or any other type of transfer (in-kind or cash) from government. Finally, a category is created for 'other income' (*incoth*), which includes earnings from being a member of a stokvel, savings drawn and income from other non-specified income sources. The following identity holds for all households:

inctot ≡ inclab + incgos + inctrans + inccorp + incgov + incoth

In general, income (y) can be broken down into $k = 1,\ldots, K$ income source. We define y_k as the income from source k such that $y = \sum_{k=1}^{K} y_k$. An alternative definition for the Gini coefficient follows.

$$G = \frac{2\sum_{k=1}^{K} \operatorname{cov}\left(y_k, F(y)\right)}{\mu} \qquad [9]$$

Equation [9] gives almost exactly the same result as equation [8] (accurate to about 5 decimal places). A Gini coefficient of 0.64 was calculated using both these methods (variables *gini1* and *gini2*). Equation [9] can now be rearranged (see McDonald et al., 1999: 7) to give the following.

$$G = \sum_{k=1}^{K} \left\{ \left[\frac{\operatorname{cov}\left(y_k, F(y)\right)}{\operatorname{cov}\left(y_k, F(y_k)\right)} \right] \left[\frac{2\operatorname{cov}\left(y_k, F(y_k)\right)}{\mu_k} \right] \left[\frac{\mu_k}{\mu} \right] \right\} = \sum_{k=1}^{K} R_k G_k S_k \qquad [10]$$

Here S_k is the share of source k of income in total income, G_k is the Gini coefficient measuring the inequality in the distribution of income component k within the group and R_k is the Gini correlation of income from source k with total income (Leibbrandt *et al.*, 1999: 29). The larger the product of these three components, the greater the contribution of income source k to total inequality as measured by G. S_k and G_k are always positive and less than one, while R_k can fall anywhere in the range [-1,1] since it shows how income from source k is correlated with total income. Some results of a preliminary Gini decomposition exercise are shown in Table 4 (see appendix section 6.4).

2.6. Theil's entropy measure and Theil decomposition

Sen (1997: 34) describes Theil's approach as "*interesting* [and] *rather different from the class of measures we have been looking at*". Theil's approach (1967) is derived from the notion of entropy in information theory. Let x be the probability that a certain event will occur, then the information content $h(x)$ of noticing that the event did in fact occur is a decreasing function of x. In other words, "*the more unlikely an event, the more interesting it is to know that the thing has really happened*" (Sen, 1997: 34). A greater value is therefore attached to the knowledge that the event had occurred. A formula that satisfies this property is the natural logarithm of the reciprocal of x.[4]

$$h(x) = \log \frac{1}{x} \qquad [11]$$

When there are n possible events, 1,…, n, we take the respective probabilities x_1,…, x_n, such that $x_i \geq 0$ and $\sum_{i=1}^{n} x_i = 1$. The entropy – the expected information content – can be viewed as the sum of the information content of each event weighted by the respective probabilities (see footnote 4).

$$H(x) = \sum_{i=1}^{n} x_i h(x_i) = \sum_{i=1}^{n} x_i \log\left(\frac{1}{x_i}\right) \qquad [12]$$

In applying this to income inequality, Theil proposed each x_i be interpreted as the relative share of income accruing to household i. Thus,

[4] Claude Shannon's measure of information (1948, cited in Collier, 1999) can be written as

$$S \equiv \sum_{i=1}^{n} x_i \ln\left(\frac{1}{x_i}\right)$$

where x_i is the probability of an event occurring. Before any event occurs, we know that the probability of it occurring is x_i. If the probability is almost 100%, then you are hardly surprised if it actually occurred. The information value (that it has occurred) is close to zero. Similarly, if a highly unlikely event occurred the information that it had occurred is of high value.

$$x_i = \frac{y_i}{n\mu} \qquad [13]$$

Clearly, the closer each x_i is to $1/n$, the greater $H(x)$ (the entropy), with the maximum value of $H(x)$ equal to log *n* when there is perfect equality. Theil suggested that if the entropy $H(x)$ is subtracted from its maximum value, one would get an index of inequality. Thus, the Theil measure of inequality is defined as follows (the derivation of equation [14] is shown in appendix section 6.5).

$$T = \log n - H(x) = \sum_{i=1}^{n} x_i \log nx_i \qquad [14]$$

The Theil index is a member of the Generalised Entropy class of inequality measures (see appendix, section 6.3). It therefore satisfies the Pigou-Dalton condition, as well as all the other axioms mentioned in section 2. This is illustrated clearly in Collier (1999). In the case of perfect equality $H(x)$ is equal to log *n*, and hence $T = 0$. When there is complete inequality, i.e. one household earns all the income, $T = \log n$. There is no upper limit for inequality, as this depends on the size of the population (Collier, 1999).

Sen (1997: 24) also points out that this index can easily be aggregated over groups. In fact, as stated by Fields (1980, cited in Leibbrandt *et al.*, 2001: 24), the Theil-T statistic is the most commonly cited additively decomposable measure of inequality. It can be used to decompose changes in income inequality into changes of inequality within and between groups, such as gender, race and economic sector (Collier, 1999). If there are *m* groups ($k = 1,\ldots, m$), the following decomposition equation can be used (see section 6.5 in the appendix for an algebraic derivation of this equation).

$$T \equiv \left[\sum_{k=1}^{m} s^k T_k\right] + \left[\sum_{k=1}^{m} s^k \cdot \log\left(\frac{\mu^k}{\mu}\right)\right] \qquad [15]$$

The first term is the weighted average of the Theil indexes for each group, i.e., the weight s_k is the share of group *k*'s income in total income and T_k is the within-group Theil index. The second term is the Theil index using only the group means, hence only inequality between groups is considered (Collier, 1999).

2.7. Atkinson's inequality measures

Atkinson's class of measures has the following general formula (see Litchfield, 1999).

$$A_\varepsilon = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu}\right)^{1-\varepsilon}\right]^{1/1-\varepsilon} \qquad [16]$$

Here ε $(0 < \varepsilon < \infty)$ is an inequality aversion parameter, i.e., the higher the value of ε, "*the more society is concerned about inequality*" (Litchfield, 1999). A_ε ranges from 0 to 1, with zero representing perfect equality. Cowell (1995, cited in Litchfield, 1999) shows that the Generalised Entropy class becomes ordinally equivalent to the Atkinson class when setting $\alpha = 1 - \varepsilon$ (for values of $\alpha < 1$) (see appendix, section 6.3). The formula for Atkinson's class of inequality measures is not included in do-file *inequality.do*.

3. Measuring poverty

3.1. Overview

Poverty measures aim to measure the incidence and depth of poverty. Whereas inequality was defined over the entire distribution, poverty measures only apply to a censored distribution of individuals or households defined as poor (Litchfield, 1999). Although poverty and inequality are usually referred to as if they were related concepts, there is not necessarily a link between the two. A high incidence of poverty does not necessarily mean that a country also has a high degree of inequality and *vice versa*. A desired property of useful inequality measures is that they should be independent of the income scale. Thus, inequality can be high even in countries with no poor persons. However, in most developing countries poverty and inequality are both serious threats.

When trying to identify the poor certain basic steps are followed (see Woolard and Leibbrandt, 2001). Firstly, households or individuals are ranked according to some welfare indicator such as income or expenditure. Next, a poverty line is selected, which separates the rich from the poor. Finally, a poverty profile of the poor households or individuals is constructed using available survey data.

In order to rank households according to a welfare indicator, it is necessary to define poverty first so that a suitable indicator can be selected. The World Bank (as cited in Woolard and Leibbrandt, 2001: 42) loosely defines poverty as the "*inability to attain a minimal standard of living*". Two approaches to measuring 'well-being' or 'standard of living' exist, namely the welfarist approach and the non-welfarist approach. The welfarist approach considers expenditure on all goods and services, including home consumption of home production.[5] The non-welfarist approach is more concerned with specific commodity forms of deprivation, such as inadequate food consumption.

[5] Above it was noted that household (or individual) income could also be used as a measure of welfare. It is often argued that expenditure data is more reliable as households are reluctant to supply information regarding their income. The income data of the IES and OHS 1995 datasets used in this analysis is arguably fairly reliable as households were asked to report figures in such that income equals expenditure (including savings and taxes). Although some inequalities still exist, the income and expenditure totals are very close (Average Total Income: R39272; Average Total Expenditure: R38267, 1995 prices). Alternative measures

The derivation of a poverty line is the most important step in identifying the poor. The literature distinguishes between two types of poverty lines. An absolute poverty line, as the name suggests, is some specific income or expenditure level below which a household is deemed poor. It is often derived from the cost of a bundle of goods required to meet certain basic needs. As the average standard of living of all members of society changes, fewer people will fall below this poverty line. A relative poverty line, on the other hand, is usually defined as the poorest ρ % of the population. Often the median or the second quintile is used as the cut-off point. Thus, as the average standard of living increases, the poorest ρ % of the population will still be regarded relatively poor compared to the remaining $(1 - \rho)$% of the populations. When this approach is used "*the poor are always with us*" (Woolard and Leibbrandt, 2001: 48).

The two most widely used South African poverty lines are the Household Subsistence Level (HSL) (Institute for Planning Research) and the Minimum Living Level (MLL) (Bureau for Market Research). Both are calculated biannually in the major urban areas, and irregularly in rural areas. Woolard and Leibbrandt (2001) compare a number of absolute and relative poverty lines for 1993 and find that the results vary quite substantially.

Table 2: Comparing various poverty lines for South Africa, 1993.

Types of poverty lines	**Rands per month (cut-off)**	**Percentage of individuals below the poverty line**
1. Population cut-off at 40th percentile of households ranked by adult-equivalent expenditure	R301 per adult equivalent	52.8
2. Population cut-off at 50% of national per capita expenditure	R202 per capita	46.9
3. Amount of money needed to achieve a per capita caloric intake of 8500kJ per day (per adult equivalent)	R150 per adult equivalent	40.4
4. Supplemental Living Level (SLL)*	R220 per capita	56.7
5. Minimum Living Level (MLL)*	R164 per capita	44.7

Notes: * Values given are for a family of five, converted to an adult equivalence scale.
Source: Various sources as cited in Woolard and Leibbrandt, 2001.

The welfare indicator in the last column of Table 2 is simply a head-count ratio, i.e., it is a ratio of the number of poor individuals to the total population. However, this tells us nothing about the poverty gap and the depth of poverty, two concepts that will be explained in more detail in section 3.2.

3.2. The Foster-Greer-Thorbecke class of decomposable poverty measures

Foster, Greer and Thorbecke (1984, as cited in Woolard, 1998) proposed a generalised class of decomposable poverty measures. Let y_i be a measure of income, n the size of the

of well-being include per capita income or consumption, per capita food expenditure or caloric intake, the ratio of income (or expenditure) spent on food and educational levels of adult household members (see Woolard and Leibbrandt, 2001).

population, and z the poverty line. If we rank households according to their measure of income and we define households $i = 1,\ldots, q$ as poor and $i = (q+1),\ldots, n$ as non-poor, the Foster-Greer-Thorbecke poverty measure can be expressed as

$$P_\alpha = \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^\alpha \qquad [17]$$

for $\alpha = 1$, 2 or 3.[6] P_α has a different interpretation for each value of α. Note that $(z - y_i)$ is by definition positive for all poor households ($i = 1,\ldots, q$) since all poor households earn less than the poverty cut-off point. When $\alpha = 0$, P_α simply reduces to $P_0 = q/n$, where q is the number of poor. This headcount index (see Table 2) is totally insensitive to the depth of poverty.

When $\alpha = 1$, one is essentially summing the relative poverty gap over all poor households and dividing by the total number of households. This poverty gap index (P_1) measures the depth of poverty because it is a function both the distance of each poor household from the poverty line and the number of poor. Woolard (1998) points out that P_1 has a number of advantages over P_0. Since P_0 is discontinuous at the poverty line, a transfer from a very poor household to a just-poor household that enables the just-poor household to escape poverty will reduce the headcount ratio. This is a violation of the Pigou-Dalton condition. Since P_1 is continuous and concave such a transfer will increase the poverty gap index. However, P_1 nevertheless neglects poverty among the poor. A transfer from one poor household to another will have no impact on P_1, provided the receiving household remains poor after the transfer.

P_2 is also a measure of the depth of poverty. It improves on P_0 and P_1 because it also takes into account the inequality amongst the poor. In fact, it can be shown that P_2 can be decomposed so that it is made up of two components: an amount due to the poverty gap and an amount due to the inequality among the poor, measured in terms of the coefficient of variation. The algebraic derivation of the formula below is provided in the appendix (section 6.7).

$$P_2 = \frac{(P_1)^2}{P_0} + \frac{(P_0 - P_1)^2}{P_0}\left(C_q\right)^2 \qquad [18]$$

C_q denotes the coefficient of variation of income among the poor (see equation [5], section 2.3). Woolard (1998) explains that although this breakdown goes partway in explaining the meaning of P_2, it remains difficult to interpret the measure on its own. One of the advantages

[6] In theory α can take on any value greater than or equal to zero. The larger the value, the more sensitive the measure is to the well-being of the poorest household. Here only values for $\alpha = 0$, 1 or 2 are considered.

of P_2, however, is that an increase in the measured poverty associated with a fall in the living standard will be deemed greater the poorer the household.

4. Concluding comments

Poverty and inequality research has become extremely important in South Africa in recent years, especially given the incumbent government's commitment to the eradication of poverty and the correction of past inequalities. In terms of the calculation of poverty and inequality measures in South Africa some difficulties remain, especially with regards to (1) reliability of South African cross-sectional income data, and (2) the consistency and comparability of such data over time. Poverty measurement is further complicated by the fact that the choice of the poverty line will affect the level of poverty. However, despite these measurement difficulties and data problems, poverty and inequality measures remain useful to gain some understanding of the severity of the social problems in South Africa. This paper provides a useful summary of some of the more frequently used poverty and inequality measures and shows how they can be calculated in Stata, and as such remains a technical reference paper that can be consulted or referenced when doing more applied work in this area.

5. References

Collier, I. (1999). ***Notes on the Theil Index of Inequality***. http://www.wiwiss.fu-berlin.de/w3/w3collie/SOCPOLY/TEACHING/Intro/Intro99/Theil.doc

Cowell, F. (2003). "Theil, Inequality and the Structure of Income Distribution" ***Distributional Analysis Research Programme Discussion Paper***, No. DARP 67, May 2003. Available online: http://sticerd.lse.ac.uk/dps/darp/darp67.pdf

Cowell, F. and Mercader-Prats, M. (1999). "Equivalance Scales and Inequality" ***Distributional Analysis Research Programme Discussion Paper***, No. DARP 27, March 1999. Available online: http://sticerd.lse.ac.uk/dps/darp/darp27.pdf

Leibbrandt, M., Woolard, I. and Bhorat, H. (2001). "Understanding contemporary household inequality in South Africa" in ***Fighting Poverty. Labour Markets and Inequality in South Africa***, UCT Press: Cape Town.

Litchfield, J.A. (1999). Inequality: Methods and Tools. Text for World Bank's Web Site on Inequality, Poverty and Socio-economic Performance: http://www.worldbank.org/poverty/inequal/index.htm

McDonald, S., Piesse, J. and Van Zyl, J. (1999). ***Exploring Income Distribution and Poverty in South Africa***. Paper presented at the Biennial Conference of the Economic Society of South Africa, Pretoria.

Sen, A. (1997). ***On Income Inequality***, Clarendon Press: Oxford.

Woolard, I. (1998). "A Comparison of Urban Poverty". Paper prepared for inclusion into ***Poverty and Inequality in South Africa. Report prepared for the Office of the Executive Deputy President and the Inter-Ministerial Committee for Poverty and Inequality***, May 1998. Available online: http://www.gov.za/reports/1998/poverty/urbanpoverty.pdf

Woolard, I. and Leibbrandt, M., (2001). "Measuring Poverty in South Africa" in ***Fighting Poverty. Labour Markets and Inequality in South Africa***, UCT Press: Cape Town.

6. Appendix

6.1. Do-file calculating inequality measures (inequality.do)

```
#delimit;
set more off;
use combined.dta, clear;

*Household level*;
keep if s2persno==1;

*Optional if inequality measures for sub-groups are needed*;
*keep if s3race == 4;

save combine3.dta, replace;
quietly log using equality.log, replace;

*NB: Select the income measure here*;
gen y = adinc;
label var y "Income measure for inequality analysis";
*Note IM in variable labels stands for 'inequality measure'*;

*Mean income (var: meany)*;
*==================================================*;
egen meany = mean(y);
      label var meany "Mean income";
sum y meany;

*Income share of each household (var: ysh)*;
*==================================================*;
egen toty = sum(y);
      label var toty "National income";
egen n = count(hhid);
      label var n "Total number of obs/hholds";
gen test1 = meany*n;
sum test1 toty;
*Note: test1 and toty should be same*;
gen ysh = y/toty;
      label var ysh "Income share of household";
codebook ysh;
egen test2 = sum(ysh);
sum test2;
*Note: Test must have mean = 1 and 0 std dev.*;
drop test1 test2;

*Range (var: range)*;
*==================================================*;
egen maxy = max(y);
      label var maxy "Maximum income";
egen miny = min(y);
      label var maxy "Minimum income";
gen range = (maxy-miny)/meany;
      label var range "IM: range";
sum miny maxy range n;
*Note: Range should be between 0 and n;

*Relative mean deviation (var: rmdev)*;
*==================================================*;
gen absdev = abs(meany-y)/toty;
egen rmdev = sum(absdev);
      label var rmdev "IM: relative mean deviation";
drop absdev;
sum rmdev;
```

```
*Variance, std dev & coeff of var (var: vary, sdy, coefvar)*;
*==========================================================*;
egen sdy = sd(y);
      label var sdy "IM: Standard dev. of income";
gen vary = sdy^2;
      label var vary "IM: Variance of income";
gen coefvar = sdy/meany;
      label var coefvar "IM: Coefficient of variation";
sum sdy vary coefvar;

*Standard deviation of logarithms (var: sdlog)*;
*===============================================*;

gen lndif = (sqrt(ln(meany)-ln(y)))/n;
egen sumlndif = sum(lndif);
gen sdlog = sumlndif^(0.5);
      label var sdlog "IM: standard dev of logarithms";
drop lndif sumlndif;
sum sdlog;

*Gini coefficient (var: gini)*;
*===============================================*;
*Note: For Gini decomposition see gini.do and giniall.do*;
*      Calculation here just for national gini coefficient*;

sort y;
gen rank = _n;
replace rank = n - rank + 1;
gen ranky = rank*y;
egen sumranky = sum(ranky);
gen gini1 = 1 + (1/n) - ((2/((n^2)*meany))*sumranky);
      label var gini1 "IM: Gini Sen formula";

cumul y, gen(cy);
      label var cy "Cum distribution fn of income";
egen meancy = mean(cy);
      label var meancy "Mean of cum distr";
gen covstep1 = (y-meany)*(cy-meancy);
egen covstep2 = sum(covstep1);
gen covy_cy = covstep2/n;
      label var covy_cy "Covariance of income and F(income)";
corr y cy, cov;
sum covy_cy;
*Note: Check covariance calculation*;
gen gini2 = 2*covy_cy/meany;
      label var gini2 "IM: Gini covariance formula";
sum gini*;

drop rank ranky sumranky meancy covstep*;

*Theil's measure (var: theil)*;
*===============================================*;

gen tstep1 = ysh*ln(n*ysh);
egen theil = sum(tstep1);
sum theil;

quietly log close;

log using imreport.log, replace;

*Final report*;
sum meany range rmdev sdy vary coefvar sdlog gini2 theil;
quietly log close;
```

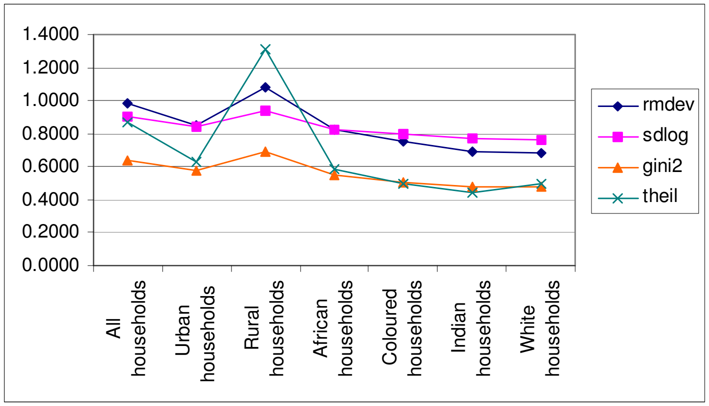
6.2. Inequality results

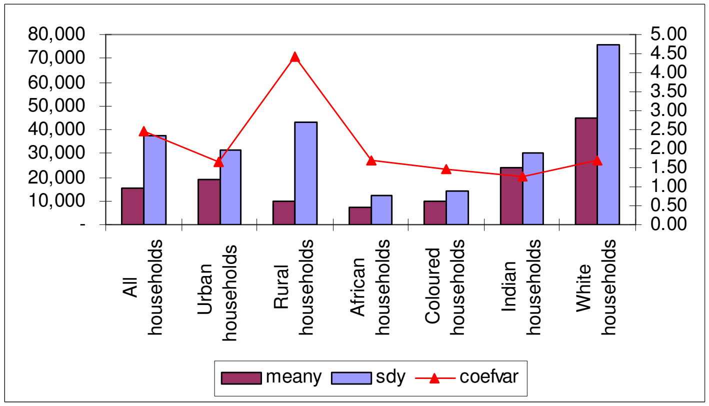
Table 3: Summary of inequality measures for all households and within racial groups and locations - 1995

Inequality measure	**Variable name**	**All households**	**Urban households**	**Rural households**	**African households**	**Coloured households**	**Indian households**	**White households**
Mean income	*meany*	15,101.13	19,216.94	9,759.17	7,413.34	9,848.59	24,169.07	44,702.49
Range	*range*	176.00	66.79	272.34	74.40	48.60	15.20	59.42
Relative mean deviation	*rmdev*	0.9792	0.8507	1.0780	0.8210	0.7553	0.6952	0.6846
Std deviation	*sdy*	37,461.25	31,639.28	43,302.33	12,503.90	14,325.63	30,420.72	75,589.70
Variance	*vary*	1.40E+09	1.00E+09	1.88E+09	1.56E+08	2.05E+08	9.25E+08	5.71E+09
Coefficient of variation	*coefvar*	2.4807	1.6464	4.4371	1.6867	1.4546	1.2587	1.6910
Std dev of logarithms	*sdlog*	0.9027	0.8456	0.9414	0.8270	0.7961	0.7689	0.7664
Gini coefficient	*gini2*	0.6403	0.5718	0.6876	0.5462	0.5065	0.4789	0.4786
Theil's entropy	*theil*	0.8702	0.6265	1.3135	0.5887	0.4947	0.4414	0.4992

Source: Own calculations from IES/OHS 1995 (*combined.dta*)

Figure 2: Comparison of some measures of inequality - 1995





Source: See Table 3

Figure 2 shows a fair degree of correlation between the various measures of inequality. Some further comparisons are drawn between the Theil index and the coefficient of variation (Figure 3) and the Theil index and the range (Figure 4). Litchfield (1999) points out that the various inequality measures may rank the same set of distributions in different ways, simply because of their differing sensitivity to incomes in different parts of the distribution.

Figure 3: Further comparisons – Coefficient of variation and Theil index - 1995

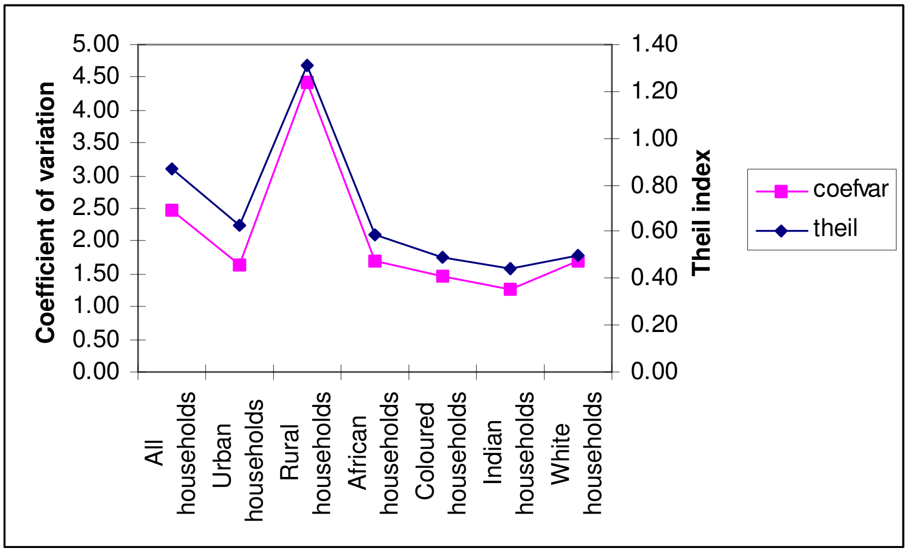
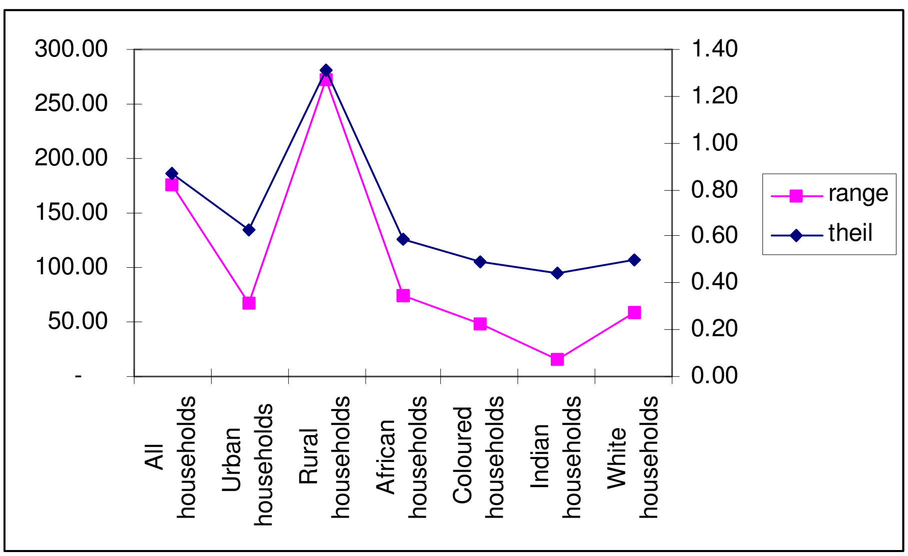


Figure 4: Further comparisons – The range and Theil index - 1995



6.3. Generalised Entropy measures of inequality

The coefficient of variation (section 2.3), the standard deviation of logarithms (section 2.4) and the Theil index (section 2.6) are all members of the Generalised Entropy (*GE*) class of inequality measures. They can all be expressed in terms of the following general formula (see Litchfield, 1999).

$$GE(\alpha) = \frac{1}{\alpha^2 - \alpha}\left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i}{\mu}\right)^{\alpha} - 1\right] \quad [A1]$$

The value of *GE* ranges from 0 to ∞, with zero representing an equal distribution of income. The parameter α ($\alpha \geq 0$) represented the weight given to distances between incomes at different parts of the income distribution. The commonest values of α are 0, 1 and 2. If $\alpha = 0$, more weight is given to distances at the lower end of the income distribution, i.e. *GE* is more sensitive to changes at this end of the distribution. If $\alpha = 1$ equal weights are applied across the distribution, while $\alpha = 2$ gives proportionally more weight to distances between incomes at the higher end of the distribution.

It can be shown that *GE*(0) and *GE*(1) with L'Hôpital's Rule become two of Theil's measures of inequality, namely the mean log deviation[7] and the Theil index.[8]

$$GE(0) = \frac{1}{n}\sum_{i=1}^{n}\log\frac{\mu}{y_i} \quad [A2]$$

$$GE(1) = \frac{1}{n}\sum_{i=1}^{n}\frac{y_i}{\mu}\log\frac{y_i}{\mu} \quad [A3]$$

When $\alpha = 2$ the *GE* measure becomes half the squared coefficient of variation.

6.4. Gini decomposition

Do-files *gini.do* and *giniall.do* was used to decompose the South African Gini coefficient by income source (see section 2.5). The process was repeated for South Africa as a whole, and thereafter for urban households, rural households and the four racial groups, Africans, Coloureds, Indians and Whites. The do-files used appear below, while Table 4 summarises the results. No interpretation is provided here (see Leibbrandt *et al.*, 1999 and McDonald *et al.*, 1999 for more details).

[7] The formula in Litchfield (1999) given here differs slightly from the one given in equation [6].

[8] Compare equation [14].

6.4.1. Master do-file for calculating Gini coeffients (giniall.do)

```
#delimit;
*Start by setting path to the location of combined.dta*;
*Gini decomposition calculations using adult equivalent income data*;

use combined.dta, clear;

set more off;
*Household-level data*;
keep if s2persno==1;

keep  settle inctot inclab incgos
      inctrans inccorp incgov incother
      s3race s3gender s1hhsize hhgroup
      A K E adinc hhgrad pcinc;

*E is the adult-equivalent per capita income of the household*;
*E=(A+0.5K)^0.9 where A (K) is the number of adult (children under 10)*;

gen adinctot = inctot  /E;
gen adinclab = inclab  /E;
gen adincgos = incgos  /E;
gen adinctra = inctrans/E;
gen adinccor = inccorp /E;
gen adincgov = incgov  /E;
gen adincoth = incother/E;

save giniall.dta, replace;

quietly log using gini.log, replace;
*ALL*;
*====================*;
quietly log close;
do gini.do;

quietly log using gini.log, append;
*BY RACE*;
*====================*;

*AFRICAN*;
use giniall.dta, clear;
keep if s3race == 1;
quietly log close;
do gini.do;

quietly log using gini.log, append;
*COLOURED*;
use giniall.dta, clear;
keep if s3race == 2;
quietly log close;
do gini.do;

quietly log using gini.log, append;
*ASIAN*;
use giniall.dta, clear;
keep if s3race == 3;
quietly log close;
do gini.do;

quietly log using gini.log, append;
*WHITE*;
use giniall.dta, clear;
keep if s3race == 4;
quietly log close;
```

```
do gini.do;

quietly log using gini.log, append;
*BY LOCATION*;
*====================*;

*URBAN*;
use giniall.dta, clear;
keep if settle == 1;
quietly log close;
do gini.do;

quietly log using gini.log, append;
*RURAL*;
use giniall.dta, clear;
keep if s3race == 4;
quietly log close;
do gini.do;
```

6.4.2. Sub-do-file for calculating Gini coefficients (gini.do)

```
#delimit;
set more off;

cumul adinctot, gen(cadintot);
cumul adinclab, gen(cadinlab);
cumul adincgos, gen(cadingos);
cumul adinctra, gen(cadintra);
cumul adinccor, gen(cadincor);
cumul adincgov, gen(cadingov);
cumul adincoth, gen(cadinoth);

quietly log using gini.log, append;

corr  adinctot adinclab adincgos adinctra adinccor adincgov adincoth
      cadintot cadinlab cadingos cadintra cadincor cadingov cadinoth,
covariance;

sum   adinctot adinclab adincgos adinctra adinccor adincgov adincoth;

quietly log close;
```

Table 4: Gini decomposition by income source for all households and within racial groups and locations - 1995

		Rk	Gk	Sk	Product			Rk	Gk	Sk	Product
SA	Inclab	0.8611	0.7104	0.5912	0.3616	**African**	inclab	0.8937	0.6898	0.6976	0.4300
	Incgos	0.9043	0.9756	0.1749	0.1543		incgos	0.7557	0.9807	0.0652	0.0483
	inctrans	0.0701	0.9159	0.0234	0.0015		inctrans	0.1151	0.8849	0.0522	0.0053
	inccorp	0.7703	0.9617	0.0924	0.0685		inccorp	0.5446	0.9773	0.0219	0.0117
	Incgov	-0.0197	0.8338	0.0442	-0.0007		incgov	-0.0039	0.7957	0.0836	-0.0003
	Incoth	0.7660	0.9747	0.0738	0.0551		incoth	0.6716	0.9584	0.0795	0.0512
	Gini				**0.6403**		**Gini**				**0.5462**
Urban	Inclab	0.8424	0.6586	0.6455	0.3581	**Coloured**	inclab	0.8871	0.6010	0.7422	0.3957
	Incgos	0.8326	0.9676	0.1314	0.1059		incgos	0.8495	0.9840	0.0741	0.0619
	inctrans	0.0985	0.9430	0.0168	0.0016		inctrans	0.1231	0.9423	0.0202	0.0023
	inccorp	0.6903	0.9437	0.1112	0.0725		inccorp	0.6272	0.9700	0.0461	0.0281
	Incgov	-0.1234	0.8541	0.0368	-0.0039		incgov	-0.1062	0.7957	0.0756	-0.0064
	Incoth	0.6673	0.9678	0.0582	0.0376		incoth	0.6150	0.9707	0.0417	0.0249
	Gini				**0.5718**		**Gini**				**0.5066**
Rural	Inclab	0.6426	0.5948	0.5018	0.1918	**Asian**	inclab	0.7197	0.5393	0.6093	0.2365
	Incgos	0.8259	0.9275	0.2479	0.1899		incgos	0.7790	0.9088	0.2538	0.1797
	inctrans	0.0739	0.9713	0.0077	0.0006		inctrans	0.1896	0.9603	0.0143	0.0026
	inccorp	0.3368	0.8503	0.1462	0.0419		inccorp	0.6667	0.9627	0.0500	0.0321
	Incgov	-0.2337	0.9202	0.0178	-0.0038		incgov	-0.0775	0.8827	0.0254	-0.0017
	Incoth	0.7632	0.9734	0.0785	0.0583		incoth	0.6619	0.9688	0.0463	0.0297
	Gini				**0.4787**		**Gini**				**0.4787**
						White	inclab	0.6426	0.5948	0.5018	0.1918
							incgos	0.8259	0.9275	0.2479	0.1899
							inctrans	0.0740	0.9713	0.0077	0.0006
							inccorp	0.3368	0.8503	0.1462	0.0419
							incgov	-0.2336	0.9202	0.0178	-0.0038
							incoth	0.7632	0.9734	0.0785	0.0583
							Gini				**0.4787**

Source: Own calculations from IES/OHS 1995 (*combined.dta*)

6.5. Theil calculations

In equation [14] the Theil measure was shown to be $T = \sum_{i=1}^{n} x_i \log nx_i$ (section 2.6). This can be derived as follows.

$$
\begin{aligned}
T &= \log n - \sum_{i=1}^{n} x_i \log\left(\frac{1}{x_i}\right) \\
&= \log n - x_1 \log(1/x_1) - x_2 \log(1/x_2) - \ldots - x_n \log(1/x_n) \\
&= \log n + x_1 \log x_1 + x_2 \log x_2 + \ldots + x_n \log x_n \\
&= x_1 \log n + x_2 \log n + \ldots + x_n \log n \\
&\quad + x_1 \log x_1 + x_2 \log x_2 + \ldots + x_n \log x_n \\
&= x_1 \log nx_1 + x_2 \log nx_2 + \ldots + x_n \log nx_n \\
&= \sum_{i=1}^{n} x_i \log nx_i
\end{aligned}
$$

[A4]

Notes:

- $\log(u/v) = \log u - \log v$ and log 1 = 0.
- $\log(uv) = \log u + \log v$
- $\sum_{i=1}^{n} x_i = 1$

The Theil index is also easily decomposable. Following Collier (1999), we start with equation [14]. Since $nx_i = y_i/\mu$ we can rewrite this equation as

$$T = \sum_{i=1}^{n} x_i \log\left(\frac{y_i}{\mu}\right)$$ [A5]

Each individual or household can now be grouped into one of m groups of possibly different size.

$$
\begin{aligned}
T \equiv &\sum_{j=1}^{n^1} x_j^1 \cdot \log\left(\frac{y_j^1}{\mu}\right) \\
&+ \sum_{j=1}^{n^2} x_j^2 \cdot \log\left(\frac{y_j^2}{\mu}\right) \\
&\vdots \\
&+ \sum_{j=1}^{n^m} x_j^m \cdot \log\left(\frac{y_j^m}{\mu}\right)
\end{aligned}
$$ [A6]

Consider the k^{th} term of equation [A6].

$$\sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{y_j^k}{\mu}\right) = \left\{ \underbrace{\sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{y_j^k}{\mu^k}\right)}_{\text{first term}} + \underbrace{\sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{1}{\mu}\right) - \sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{1}{\mu^k}\right)}_{\text{second and third term}} \right. \qquad \text{[A7]}$$

The first term of the kth group can now be rewritten as

$$\begin{aligned}
\sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{y_j^k}{\mu^k}\right) &= \sum_{j=1}^{n^k}\left(\frac{\sum_{i=1}^{n^k} y_i^k}{\sum_{i=1}^{n} y_i}\right)\cdot\left(\frac{y_j^k}{\sum_{i=1}^{n^k} y_i^k}\right)\cdot\log\left(\frac{y_j^k}{\mu^k}\right) \\
&= \left(\frac{\sum_{i=1}^{n^k} y_i^k}{\sum_{i=1}^{n} y_i}\right)\cdot\sum_{j=1}^{n^k}\left(\frac{y_j^k}{\sum_{i=1}^{n^k} y_i^k}\right)\cdot\log\left(\frac{y_j^k}{\mu^k}\right) \\
&= \left(\frac{\sum_{i=1}^{n^k} y_i^k}{\sum_{i=1}^{n} y_i}\right)\cdot\sum_{j=1}^{n^k} s_j^k \cdot \log\left(\frac{y_j^k}{\mu^k}\right) \qquad \text{[A8]} \\
&\equiv s^k \cdot T_k
\end{aligned}$$

while the second and third terms can be rewritten as

$$\begin{aligned}
\sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{1}{\mu}\right) - \sum_{j=1}^{n^k} x_j^k \cdot \log\left(\frac{1}{\mu^k}\right) &= \sum_{j=1}^{n^k} x_j^k \cdot \left[\log(\mu^k) - \ln(\mu)\right] \\
&= \left[\log(\mu^k) - \log(\mu)\right]\cdot\sum_{j=1}^{n^k} x_j^k \\
&= \log\left(\frac{\mu^k}{\mu}\right)\cdot\frac{\sum_{j=1}^{n^k} y_j^k}{\sum_{i=1}^{n} x_i} = s^k \cdot \log\left(\frac{\mu^k}{\mu}\right) \qquad \text{[A9]} \\
&= \frac{n^k \cdot \mu^k}{\sum_{i=1}^{n} y_i}\cdot\log\left(\frac{\mu^k}{\mu}\right)
\end{aligned}$$

Adding up the two previous results over the *m* groups we have the Theil decomposition of inequality (Collier, 1999).

$$T \equiv \left[\sum_{k=1}^{m} s^k T_k\right] + \left[\sum_{k=1}^{m} s^k \cdot \log\left(\frac{\mu^k}{\mu}\right)\right] \qquad \text{[A10]}$$

6.6. Do-files calculating poverty measures (*fgt.do*)

```
#delimit;
set more off;
```

```
*Foster-Greer-Thorbecke Poverty Measures*;

use combined.dta, clear;
keep if s2persno == 1;
save combine3.dta, replace;

egen relpovln = pctile(adinc), p(40);
lab var relpovln "Relative pov line 40th pctile adinc";

*Headcount index: P0*;
*==================*;

gen relpoor = 1 if adinc < relpovln ;
replace relpoor = 2 if adinc >= relpovln;
label define poorlab 1 "poor" 2 "non-poor";
label values relpoor poorlab;
egen countp = sum(relpoor) if relpoor == 1;
egen n = count(hhid);

gen p0 = countp/n;

*Poverty gap index: P1*;
*====================*;

gen incgap = (relpovln - adinc)/relpovln if relpoor == 1;
egen sumincgp = sum(incgap) if relpoor == 1;

gen p1 = sumincgp/n if relpoor == 1;

*Depth of poverty: P2*;
*====================*;

gen gapsq = incgap^2 if relpoor == 1;
egen sumgapsq = sum(gapsq) if relpoor == 1;

gen p2 = sumgapsq/n if relpoor == 1;

egen meanp = mean(adinc) if relpoor == 1;
egen sdp   = sd(adinc) if relpoor == 1;
gen cvpsq    = (sd/mean)^2 if relpoor == 1;

gen p2comp1 = (p1^2)/p0 ;
gen p2comp2 = (((p0-p1)^2)/p0)*cvpsq ;

gen p2check = p2comp1 + p2comp2 ;

sum p0 p1 p2 p2comp1 p2comp2 p2check;

*Checking*;

gen incsq = adinc^2 if relpoor == 1;
egen sumincsq=sum(incsq) if relpoor == 1;
gen cvpsq2=(sumincsq/countp - (meanp^2))/(meanp^2) if relpoor == 1;
gen p2check2 = (((p1^2)/p0) + ((((p0-p1)^2)/p0)*cvpsq2)) if relpoor == 1;
```

6.7. Foster-Greer-Thorbecke calculations

The P_2 measure of inequality can be calculated in two ways, either using equation [17] for $\alpha = 2$ or equation [18] (see section 3.2). The use of equation [18] requires, of course, that equation [17] be used first to calculate P_0 and P_1, i.e. for $\alpha = 0$ and $\alpha = 1$ respectively. The advantage of expressing P_2 in terms of P_0 and P_1 is that it provides a breakdown of the two components of the poverty measure. Suppose there are q poor households. The variance and mean level of income of these poor households can be calculated by substituting q for n in equations [4] and [1] respectively. An expression for $(C_q)^2$ can now be derived.

$$\begin{aligned}(C_q)^2 &= \frac{\frac{1}{q}\sum_{i=1}^{q}(y_i-\mu)^2}{\mu^2} = \frac{\frac{1}{q}\sum_{i=1}^{q}(y_i)^2-\mu^2}{\mu^2} \\ &= \frac{1}{q\mu^2}\sum_{i=1}^{q}(y_i)^2 - 1\end{aligned} \qquad \text{[A11]}$$

Expressions for P_0 and P_1 can easily be derived.

$$P_0 = q/n \text{ and } P_1 = q/n - \frac{q\mu}{nz} \qquad \text{[A12]}$$

From this the two components of P_2 can be calculated using equation [18].

$$\begin{aligned}(P_1)^2 &= \left(\frac{q}{n}\right)^2 - \frac{2q^2\mu}{n^2 z} + \left(\frac{q\mu}{nz}\right)^2 \\ \therefore \frac{(P_1)^2}{P_0} &= \left(\frac{q}{n}\right) - \frac{2q\mu}{nz} + \frac{q\mu^2}{nz^2}\end{aligned} \qquad \text{[A13]}$$

$$\begin{aligned}P_0 - P_1 &= \frac{q\mu}{nz} \\ \therefore \frac{(P_0-P_1)^2}{P_0}(C_q)^2 &= \frac{q\mu^2}{nz^2}(C_q)^2\end{aligned} \qquad \text{[A14]}$$

P_2 is the sum of equation [A13] and [A14]. Substituting $(C_q)^2$ in equation [A14] with equation [A11] and simplifying gives the following.

$$P_2 = \frac{q}{n} - \frac{2q\mu}{nz} + \frac{1}{nz^2}\sum_{i=1}^{q}(y_i)^2 \qquad \text{[A15]}$$

Below P_2 is calculated directly from equation [17] and simplified.

$$\begin{aligned}P_2 &= \frac{1}{n}\sum_{i=1}^{q}\left(\frac{z-y_i}{z}\right)^2 = \frac{1}{n}\sum_{i=1}^{q}\left(1-\frac{2y_i}{z}+\frac{(y_i)^2}{z^2}\right) \\ &= \frac{q}{n} - \frac{2q\mu}{nz} + \frac{1}{nz^2}\sum_{i=1}^{q}(y_i)^2\end{aligned} \qquad \text{[A16]}$$

Equations [A15] and [A16] are the same, which shows the result to hold.

Technical Papers in this Series

Number	Title	Date
TP2003: 1	Creating a 1995 IES Database in STATA	September 2003
TP2003: 2	Creating a 1995 OHS and a combined OHS-IES Database in STATA	September 2003
TP2003: 3	A Standard Computable General Equilibrium Model Version 3: Technical Documentation	September 2003
TP2003: 4	Measures of Poverty and Inequality: A Reference Paper	October 2003
TP2004: 1	SeeResults: A spreadsheet Application for the Analysis of CGE Model Results	November 2004
TP2004: 2	The Organisation of Trade Data for inclusion in Social Accounting Matrix	December 2004
TP2004: 3	Creating a 2000 IES Database in Stata	August 2004
TP2004: 4	Forming Representative Household Groups in a SAM	July 2004

Other PROVIDE Publications

Background Paper Series
Working Papers
Research Reports