Calcutta Statistical Association Bulletin
Vol. 37, March & June, 1988, Nos. 145-146

ON SOME ESTIMATES OF POVERTY MEASURES

P. MAITI and M. PAL

Indian Statistical Institute, Calcutta

ABSTRACT: There are now a number of poverty measures available in the literatures. Some of the measures are alternative to each other and some claimed to be superior in some sense to many others.

While significant work has been done in developing the alternative measures, not much attention has been paid to the problem of estimation of these indices. Estimation does not pose very serious problems in the large sample, but when one deals with a small sample, which may typically be the case in reality, situations become quite different. In fact usual estimators become biased for some of the indices. In this paper, alternative estimators for these cases have been proposed. Other properties of the estimators and some other relevant issues have also been examined.

1. Introduction

There are now a number of poverty measures available in the literature. Some of the measuree are alternative to each other and some are claimed to be superior in some sense to many others.

Though there are controversies on the definition of poor people and on the correct measure of poverty, the fact remains that the poor people exist and the extent of poverty is threatening at least to developing countries like India. Theoretical economists have proposed many measures of poverty index. However, not much attention has been paid to the estimation of such measures. If we really want to ameliorate poverty we must know its extent—a value estimated from a given set of data—regardless of the measure we propose. The estimation of the measures should be as precise as the refinement made on these measures. Unfortunately, this is very much lacking in the literature. The development of estimation of poverty indices is not able to keep pace with development of formulae measuring extent of poverty.

While theoretical economists face problems of identifying a group of people as poor and of aggregating the characteristics of the set of poor people into a 'measure of poverty', the applied economists or statisticians face problems like estimating the number of poor people and the 'value of the measure' based on a sample of observations drawn from a population of interest.

In this paper we investigate the problems of estimation of some measures of poverty starting from head count ratio which is defined as the ratio of number of poor people to the total number of people in the society. We shall first give a set of formulae among many which have been proposed in the literature and then discuss the estimation problems of some of the measures.

2. Indices of poverty

Let N be the number of individuals in the society which is known. Let Y_i be the income of ith individual and Z be the poverty line. If $Y_i \leqslant Z$, we shall call ith individual to be poor. Let us denote S_1 and S_1^c to represent the set of poor and not-so-poor people in the society. Naturally $S_1 \cup S_1^c = S$, which is the set of N people in the society. Hence, $S=\{1, 2, 3, .., N\}$ and $S_1=\{i_1, i_2, ..., i_Q\}$, were Q is the number of poor people which may or may not be known. Thus $i \in S_1$ implis $Y_i \leqslant Z$. Without loss of generality, let us assume that $\{i_1, i_2, ..., i_Q\} \equiv \{1, 2, ..., Q\}$. The corresponding income set can be defined by $I_1=\{Y_1, Y_2, ... Y_Q\}$. Now, we present some of the poverty indices proposed in the literature :

(*i*) *Head Count Ratio* (Booth (1902), Rowntree (1901))

$$X = Q/N.$$

This is one of the oldest measures of poverty mentioned in almost all work on poverty. In fact Bowley (1923), says "There is, perhaps, no better test of the progress of the nation than that which shows what proportion are in poverty" (p. 214).

(*ii*) *Income Gap Ratio* :

$$P_1 = \sum_{i=1}^{Q} \frac{Z - Y_i}{ZQ} = 1 - \frac{\bar{Y}_1}{Z},$$

where $\bar{Y}_1$ denotes the average income of the poor people. This index was first developed by the *US* Social Security Administration [see, Seidl (1987), p. 17].

(*iii*) *Normalized Absolute Deprivation Index* [Seidl (1987), p. 17] :

$$P_2 = XP_1 = \frac{Q}{N}\sum_{i=1}^{Q}\frac{(Z-Y_i)}{QZ} = \frac{Q}{N} - \frac{Q}{N}\frac{\bar{Y}_1}{Z}.$$

(*iv*) *Sen's Poverty Index* [Sen (1976), Sen (1981)] :

$$P_3 = \frac{2}{(Q+1)NZ}\sum_{i=1}^{Q}(Q+1-i)(Z-Y_i)$$

$$= \frac{Q}{N} - \frac{Q}{N}\frac{\bar{Y}_1}{Z} + \frac{Q}{Q+1}\frac{Q}{N}\frac{\bar{Y}_1}{Z}\, LR(Y \mid Y \leqslant Z)$$

$$\simeq \frac{Q}{N} - \frac{Q}{N}\cdot\frac{\bar{Y}_1}{Z} + \frac{Q}{N}\cdot\frac{\bar{Y}_1}{Z}\cdot LR(Y \mid Y \leqslant Z)$$

$$= P_2 + X(1-p_1)\, LR(Y \mid Y \leqslant Z),$$

where $LR(Y \mid Y \leqslant Z)$ is the Lorenz Ratio of incomes of poor which is $\sum\sum_{i,j=1}^{Q} \mid Y_i - Y_j \mid / (2Q^2\, \bar{Y}_1)$

(*v*) *Thon's Poverty Index* [Thon (1979), p 438, Thon (1984), p 61] :

$$P_4 = \frac{2}{N(N+1)Z}\sum_{i=1}^{Q}(N+1-i)(Z-Y_i)$$

$$= \frac{1 - 2\sum_{1}^{N} Y_i^{*}(N+1-i)}{N(N+1)Z},$$

where

$$Y_i^{*} = \begin{cases} Y_i & \text{if } Y_i \leqslant Z \\ Z & \text{if } Y_i > Z. \end{cases}$$

3. Estimation of poverty indices

Here we investigate the problem of estimation of first three indices discussed in the previous section. Most of the poverty measures take account only of the units below the poverty line. Naturally, one can take a sample from the group below the poverty line and estimate the relevant indices. When we have already had a sample from the total population, we need not take another sample from the subpopulation. It will only mean increase of cost and wastage of time and manpower. Instead, we can use the data already available as a part of the bigger survey and give estimates which will be suitable for the subgroup. Here, it should be pointed out that, the usual estimates may lead to erroneous results ; because the sample is not drawn from the subpopulation itself. The estimates are required to be modified suitably. This type of study in sample survey is known as domain of study or study in subpopulation. Another merit in the domain of study is that here we need not have a list of all the units of subpopulation. In the other words, the knowledge of Q, number of people below the poverty line need not be known.

Now, let us discuss how one can estimate the extent of poverty by domain of study approach. To start with, let us assume that a sample of size-n has been drawn from the entire population using SRSWR (Simple Random Sampling with Replacement), the values being $y_1, y_2, \cdots, y_n$, some of which fall below Z and some are above Z.

(I) *Head Count Ratio* $(X = Q/N)$

We have,
$$\hat{X} = \frac{\sum_1^n x_i}{n},$$
where
$$x_i = \begin{cases} 1 \text{ if } y_i \leqslant Z \\ 0 \text{ otherwise.} \end{cases}$$

It simply follows that
$$E(\hat{X}) = X$$
$$V(\hat{X}) = X(1 - X)/n.$$

(II) *Income Gap Ratio* $\left(p_1 = 1 - \frac{\bar{Y}_1}{Z}\right)$:

There are four types of estimates of income gap ratio depending on the extraneous knowledge on Q and/or $\bar{Y}$, the population mean.

(a) Where neither Q, nor $\bar{Y}$ is known,
$$\hat{P}_1^{(1)} = 1 - \frac{1}{Z} \frac{\sum_1^n y_i x_i}{\sum_1^n x_i}.$$

(b) Where only Q is known,

$$\hat{P}_1^{(2)} = 1 - \frac{N}{nZQ} \sum_1^n y_i x_i .$$

(c) Where only $\bar{Y}$ is known,

$$\hat{P}_1^{(3)} = 1 - \frac{1}{Z} \frac{\{n\bar{Y} - \Sigma y_i(1 - x_i)\}}{\Sigma x_i}$$

(d) Where both Q and $\bar{Y}$ are known,

$$\hat{P}_1^{(4)} = 1 - \frac{N}{nZQ} \{n\bar{Y} - \Sigma y_i(1 - x_i)\}.$$

All the estimates are unbiased up to the order $1/n$ except $\hat{P}_1^{(3)}$ and the bias of $\hat{P}_1^{(3)}$ is

$$| B(\hat{P}_{1(3)})| = \left[\frac{(1 - X)}{nXZ} (\bar{Y}_1 - \bar{Y}_2)\right]$$

where $\bar{Y}_2$ is the mean of people above the poverty line. Since contribution of bias is only of the order $1/n$, it will contribute a term of the order $1/n^2$ in the MSE (Mean Square Error). As we are taking MSE upto the order $1/n$, this term is neglected. Comparison of MSEs thus reduces to the comparison of variances only.

So far as the variances are concerned, we have, up to the order $1/n$,

$$V(\hat{P}_1^{(1)}) = \frac{\bar{Y}_1}{nXZ^2} c_1^2$$

$$V(\hat{P}_1^{(2)}) = \frac{\bar{Y}_1^2}{nXZ^2} [c_1^2 + (1 - X)]$$

$$V(\hat{P}_1^{(3)}) = \frac{(1 - X)\bar{Y}_2^2}{nXZ^2}\left[\frac{c_2^2}{X} + \left(\frac{\bar{Y}_1 - \bar{Y}_2}{\bar{Y}_2}\right)^2\right]$$

$$V(\hat{P}_1^{(4)}) = \frac{\bar{Y}_2^2}{nZ^2(1 - X)} [c_2^2 + X],$$

where c_1^2 and c_2^2 are the coefficients of variation of the respective groups (groups of people with income less than or equal to the poverty line has subscript '1' and the other group has subscript '2').

Clearly,

$$V(\hat{P}_1^{(1)}) < V(\hat{P}_1^{(2)}),$$

i.e., $\hat{P}_1^{(1)}$ is a better estimate than $\hat{P}_1^{(2)}$. This implies that the estimate defined by $\hat{P}_1^{(2)}$ does not use the auxilliary informationn very efficiently.

$$V(\hat{P}_1^{(3)}) - V(\hat{P}_1^{(1)}) = \frac{1-X}{nZ^2}\left[\left(\frac{\sigma_2^2}{X} - \frac{\sigma_1^2}{1-X}\right) + (\bar{Y}_1 - \bar{Y}_2)^2\right],$$

where σ_1^2 and σ_2^2 are the variances of incomes of the respective groups. Now, since σ_2^2 is expected to be much larger than σ_1^2, as one usually gets from the behaviour of income distribution that, it has a long upper tail, and since X is not likely to be very close to 1, one would expect that the above quantity is positive. Thus, again, $\hat{P}_1^{(1)}$ has a smaller variance. The last comparison

$$V(\hat{P}_1^{(4)}) - V(\hat{P}_1^{(1)}) = \frac{1}{nZ^2}\left[\left(\frac{\sigma_2^2}{1-X} \quad \frac{\sigma_1^2}{X}\right) + X\right],$$

leads us, by the same argument as given in the previous one, to the same conclusion as in the previous cases that $V(\hat{P}_1^{(1)})$ is less than $V(\hat{P}_1^{(4)})$ as long as X is not very small, which is very unlikely in a country like India.

Thus, we can draw a straightforward conclusion from the above discussion : So far as income gap ratio of India is concerned, the estimate without using any extraneous information is better than the usual estimates using additional information like number of people below the poverty line and mean income of all the people in the population.

(III) *Normalized Absolute Deprivation Index*

$$(P_2 = XP_1 = X - X\bar{Y}_1/Z):$$

Here the estimates proposed are of the form

$$\hat{P}_2 = A - \frac{1}{Z}BC$$

where A is either X or $\hat{X}$, B is either X or $\hat{X}$ and C is one of $(1 - P_1^{(j)})Z$, $j = 1, \ldots, 4$; depending on the extraneous information on Q and/or Y. Thus we should have 16 estimates. Fortunately, we have 8 estimates because some of the terms in B and C cancel out.

Using the same argument as given in the previous section only one estimate turns out to be the best which is

$$\hat{P}_2^* = X - \frac{1}{Z} X \frac{\Sigma y_i x_i}{\Sigma x_i}.$$

Here X needs knowledge of Q. This is in contrast with the best estimate found in the case of income gap ratio where the best estimate corresponds to no knowledge of Q or $\bar{Y}$. This is however quite expected as the first part of normalized absolute deprivation index is nothing but the proportion of people below the poverty line which if known before hand should be incorporated without affecting the second part. When Q is unknown, one can show by a similar argument that $X - \Sigma y_i x_i/(nZ)$ has the smallest variance.

4a. Simultaneous estimation of group means

Notice that, apart from the constant terms we are estimating the mean of incomes below the poverty line. It is also surprising to see that the estimator without using any additional information gives a better fit. Though it is against intuition, the explanation of why the information on Q is unnecessary is not very difficult. Use of the knowledge of Q means making the denominator constant and hence there is no way to compensate the variation of the numerator. Information on $\bar{Y}$, the population mean, becomes unnecessary, because of the peculiarity of income distribution namely that it is highly positively skewed. It is thus envisaged that if we estimate the two means, below and above the poverty lines, simultaneously then the picture might have been lifferent.

The two estimators of means, assuming $0 < \Sigma x_i < n$, are

$$\hat{m}_1 = \frac{\Sigma y_i x_i}{\Sigma x_i} \text{ and } \hat{m} = \frac{\Sigma y_i (1 - x_i)}{\Sigma (1 - x_i)}.$$

without using information on $\bar{Y}$ and

$$\hat{m}_3 = \frac{n\bar{Y} - \Sigma y_i (1 - x_i)}{\Sigma x_i} \text{ and } \hat{m}_4 = \frac{n\bar{Y} - \Sigma y_i x_i}{\Sigma (1 - x_i)}$$

using $\bar{Y}$. Thus we can think of four sets of estimators namely $(\hat{m}_1, \hat{m}_2)$, $(\hat{m}_1, \hat{m}_4)$, $(\hat{m}_3, \hat{m}_2)$ and $(\hat{m}_3, \hat{m}_4)$. The respective generalized variances of the estimators are (to order $O(n^{-2})$)

$$G_1 = \frac{\sigma_1^2 \sigma_2}{n^2 p(1-p)}, \quad G_2 = \frac{\sigma_1^2 (\bar{Y}_1 - \bar{Y}_2)^2}{n^2 (1-p)}$$

$$G_3 = \frac{\sigma_2^2(\bar{Y}_1 - \bar{Y})^2}{n^2 p} \text{ and } G_4 = G_1 + G_2 + G_3,$$

where σ_1^2 and σ_2^2 are variances of incomes below and above the poverty line respectively. Since G_4 is the sum of the other three generalized variances, the estimator (m_3, m_4) is inadmissible. Thus we need only to compare among the other three estimators. Notice that G_1, G_2 and G_3 are inversely proportional to $(\bar{Y}_1 - \bar{Y}_2)^2$, σ_2^2/p and $\sigma_1^2/(1-p)$ respectively.

4b. Illustration

A two parameter lognormal distribution (with parameters μ and σ^2) was fitted by Jain (1977) to consumer expenditure data of 19th NSS round (1964-65) to compare it with three parameter lognormal distribution. He found the estimate of μ and σ^2 for rural India as 3.054 and 0.260 respectively. The estimated poverty line for the year 1964-65 was Rs. 21.30 per capita expenditure per 30 days. G_1, G_2, G_3 can now be calculated, apart from the common factor $e^{4\mu}$, as 0.0491, 0.0582 and 0.5356, the minimum being 0.0491 indicating that again the estimator without using information on $\bar{Y}$ is better.

We have also made some parametric study taking different values of μ, σ^2 and Z. It can be proved that the relative variances depend on Z/e^{μ} and σ^2 only.

Table 1 : Generalized variances of mean estimators for different parametric values of lognormal distribution.

Z/e^{μ}	σ^2	$\sigma_1^2/(1-p)$	σ_2^2/p	$(\bar{Y}_1 - \bar{Y}_2)^2$	G_1	G_2	G_3
0.9	0.20	0.036	0.549	0.536	0.0198	0.0193	0.2943
	0.30	0.048	1.036	0.852	0.0497	0.0409	0.8827
	0.40	0.058	1.713	1.217	0.0994	0.0706	2.0847
	0.50	0,066	2.650	1.626	0.1749	0.1073	4.3089
1.0	0.20	0.065	0.436	0.584	0.0283	0.0374	0.2546
	0.30	0.081	0.870	0.937	0.0705	0.0759	0.8152
	0.40	0.081	1.521	1.329	0.1278	0.1116	2.0214
	0.50	0.101	2.393	1.785	0.2416	0.1803	4.2715
1.1	0.20	0.098	0.365	0.650	0.0358	0.0637	0.2372
	0.30	0.116	0.774	1.036	0.0898	0.1202	0.8019
	0.40	0.129	1.371	1.476	0.1769	0.1904	2.0236
	0.50	0.130	2.226	1.971	0.2894	0.2562	4.3874

From Table 1 it is clear that the case G_3 is ruled out as its value is always greater than the other two values i e., so far as the parametric ranges considered in this table the estimator using information on Y for the group below the poverty line is inefficient compared to he estimator without any using any information or the estimator using information on $\bar{Y}$ in the group mean above the poverty line.

Between G_1 and G_2, it is seen that G_2 is smaller when $Z/e^{\mu} < 1$ In case $Z/e^{\mu} = 1$, G_1 is smaller than G_2 for $\sigma^2 \leqslant \cdot 3$ and it is the reverse is for $\sigma^2 \geqslant \cdot 4$. When $Ze^{\mu} > 1$, G_1 is smaller than G_2 for $\sigma^2 \leqslant \cdot 4$ i.e. $G_1 < G_2$ for small values of σ^2 and large values of Z/e^{μ} (see Fig. 1).

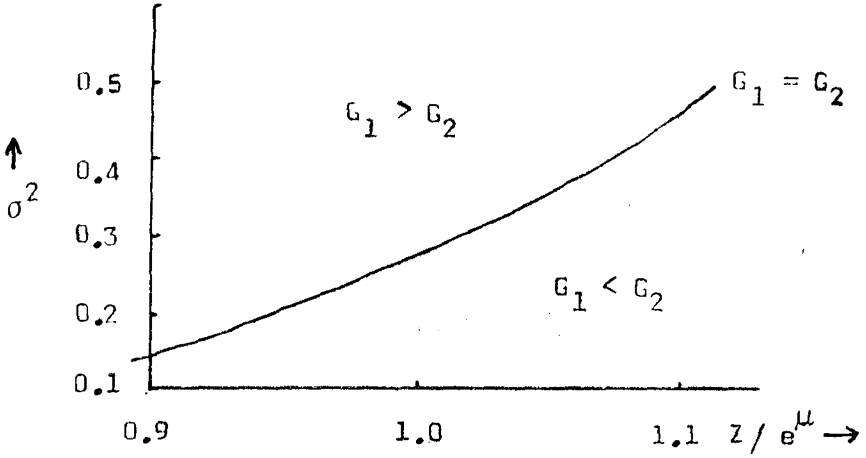

Fig. 1, Comparison between generalized variances G_1 and G_2 corres to the estimates $\hat{m}_1$ and $\hat{m}_2$.

In our example $\sigma^2 = .26$, and $Z/e^{\mu} = 1.0047$ which is approximately equal to 1. That is why we found the estimator without using any extraneous information better than other estimators. Given the same value of $\sigma^2 = .26$, if poverty line is lowered down then the picture becomes just the reverse.

ACKNOWLEDGEMENT

The authors greatly acknowledge Professor Bikas K. Sinha of Indian Statistical Institute, Calcutta, for the valuable discussion they had with him while preparing the manuscript. Comments from a referee were helpful in revising the manuscript.

REFERENCES

Booth, C. (1902) : *Life and Labour of the People of London.* Macmillan, London.

Bowley, A. L. (1923) : *The Nature and the Purpose of the Measurement of Social Phenomena* : P.S. King, London.

Jain, L. R. (1977) : On Fitting a Three-parameter Log-normal Distribution to Consumer Expenditure Data. *Saukhyā C.* 39 61-73.

Maity, P. and Pal, M. (1987) : A Unified Approach on Estimation of Lorenz-Ratio From a Finite Population, Submitted to *Sankhyā B.*

Pal, M. and Maity, P. (1988) : Estimation of Regression Coefficients and other Parameters in a Subpopulation, presented at Indian Econometric Conference at Bangalore ISI, 4-5 January 1988.

Rowntree, B.S. (1901) : *Poverty of A Study of Town Life.* Macmillan, London.

Seidl, Ch. (1987) : *A Survey of Poverty Measures*, Manuscript.

Sen, A. K. (1976) : Poverty : An Ordinal Approach to Management *Econometrica.* **44**, 219-231.

Sen, A. K. (1981) : *Poverty and Famines : An Essay on Entitlement and Deprivation.* Clarendon Press, Oxford.

Thon, D. (1979) : On Measuring Poverty. *Rev. Inc. and Wealth.* **25**, 429-439.

Thon, D. (198 4) : A Poverty Measure, *Indian Econ.* **30**, 55-70.