Working Paper No. 77
July, 2009

Poverty and Household Size: Facts and Statistical Artefacts

N.Krishnaji

CENTRE FOR ECONOMIC AND SOCIAL STUDIES
Begumpet, Hyderabad-500016

ABSTRACT

This essay argues that while per capita consumption is useful in measuring the incidence of poverty defined in a certain way, it is unsuitable for analysing demographic differentials between the poor and the rich. Indeed, it is shown here that when per capita expenditure (PCE) is used to classify rural households along the poor-to-rich scale, it leads to some paradoxical correlations, the most dramatic of which exhibits increasing infant mortality and general death rates as per capita expenditure rises: that is, better off families, reckoned by the PCE criterion, are shown to experience higher death rates. A somewhat less dramatic but equally puzzling finding shows poor families as big in size by the criterion of per capita expenditure but small in terms of land holdings.

The paper is concerned exclusively with rural India, and it attempts to resolve some of these statistical puzzles through a simple formulation of the interrelationships between household size and variables that describe and govern economic status of families. The analysis shows that the paradoxes arise from ignoring the vital role of income -generating assets, mainly land, and not paying attention to the two-sided nature of the relationship between family size and income. Apart from explaining the paradoxes, the reasoning here suggests an alternate way of looking at poverty, namely, in terms of asset holdings of families. Indeed, tracking the living standards among categories of population, such as that of asset-poor rural labour households, can be an alternative approach to studying trends in poverty - an alternative to the usual pursuit of poverty ratios defined by per capita expenditure. Such an approach, while not necessarily providing the changing numbers and proportions of the poor, would allow us to see how specified groups of population are faring in terms of consumption and deprivation levels, and how in this respect different groups compare with each other.

Acknowledgements

A. Vaidyanathan introduced me decades ago to the household size paradox discussed in this paper. T.Krishna Kumar's comments on an earlier draft have led to many improvements. K.S.James, G.K.Mitra, K.Navaneetham and C.Ravi have all given me much help in putting together the material for the paper. To all of them my sincere thanks.

Poverty and Household Size: Facts and Statistical Artefacts

N.Krishnaji

This essay argues that while per capita consumption is useful in measuring the incidence of poverty defined in a certain way, it is unsuitable for analysing demographic differentials between the poor and the rich. Indeed, it is shown here that when per capita expenditure (PCE) is used to classify rural households along the poor-to-rich scale, it leads to some paradoxical correlations, the most dramatic of which exhibits increasing infant mortality and general death rates as per capita expenditure rises: that is, better off families, reckoned by the PCE criterion, are shown to experience higher death rates. A somewhat less dramatic but equally puzzling finding shows poor families as big in size by the criterion of per capita expenditure but small in terms of land holdings.

The paper is concerned exclusively with rural India, and it attempts to resolve some of these statistical puzzles through a simple formulation of the interrelationships between household size and variables that describe and govern the economic status of families. Our analysis shows that the paradoxes arise from ignoring the vital role of diverse income –generating family assets, mainly land, and not paying attention to the two-sided nature of the relationship between family size and income. Apart from an attempt to explain the paradoxes, the reasoning here suggests an alternate way of looking at poverty, namely, in terms of asset holdings of families. Indeed, tracking the living standards among categories of population, such as that of asset-poor rural labour households, can be an alternative approach to studying trends in poverty – an alternative to the usual pursuit of poverty ratios defined by per capita expenditure. Such an approach, while not necessarily providing the changing numbers and proportions of the poor, would allow us to see how specified groups of population are faring in terms of consumption and deprivation levels, and how in this respect different groups compare with each other.

1. Some Data

A few empirical observations, with some related comments:

Observation 1: *In cross-sections of households, average household size declines as per capita expenditure increases.*

This is a feature of all Indian consumer expenditure data over space and time, as can be readily verified from the reports of the different rounds of the National Sample Survey (NSS) covering all regions of the country. The data for the 59th round for the year 2003 given in Table A1 in the Appendix show, for example, that the estimated mean size among all rural households declines fairly steadily from 6.2 in the lowest monthly per capita expenditure class (below Rs. 225) to 3.8 in the highest one (above Rs. 950). The pattern is similar in respect of farmer households, the mean household size declining from 6.9 to 4.1. The relationship between the two variables is usually exhibited in the form described above, depicting the presumed dependence of household size (S) on per capita expenditure (PCE); besides, some occasional NSS tabulations also demonstrate the reverse relationship of how average PCE systematically decreases as household size increases. For some data in the latter respect see Krishnaji (1984). Some of the theoretical conditions under which such relationships based on grouped observations imply a negative correlation in the population as a whole are stated as proposition (P2) below. However, an analysis of some unit level (that is, household level) data shows, as we shall see later, that the two variables (PCE and S) are indeed inversely correlated in the whole population. Accordingly we restate Observation 1 as follows, for simplicity in exposition:

(O1): *Household size and per capita expenditure are inversely correlated in the rural Indian population.*

The correlation tells us nothing about the direction of causation and not much either about the causes of poverty that could be defined in different ways. Small families (for example, a widow and a child or two) can be very poor; and on the other hand, big-sized, well-endowed families very rich. The larger the number of claimants to a cake the smaller will be the individual share, by simple arithmetic. But the cake can be much bigger for some than for others; so it all depends on how the small and the big cakes are distributed within a given population among families of different size. However, the correlation can - and does – lead to the debatable interpretation that poor families are poor because they are big in size; or to the one that poor families are big because they were poor to begin with for some unfathomable reason, and they produce children as extra hands for income generation, and keep getting bigger and so on, interpretations that basically identify poverty by levels of per capita consumption, relying on the arithmetic of cake-sharing.

To analyse the correlation we must begin by recognising that the total household consumer expenditure is determined by the total household income – which in turn depends on a number of factors that include family size and composition (men, women children, the old-aged, and so on) apart from the all-important one of asset holdings of the family. Thus for example among rural labour households, larger numbers of adults – potential wage earners – lead to higher family incomes. In contrast, larger numbers of children, women and the old-aged (who remain either as non-earning dependents or earn far less than male adults do) lead to lower levels of per capita income and consumption at any level of asset holdings, and so on. Clearly, we can expect the elasticity of income (and hence of consumption) with respect to household size to be less than unity. Among cultivator families, incomes and expenditures depend mainly on landholdings and other assets, but the numbers of men, women, and children within households also influence total incomes through the possible deployment of family members in diverse economic activities, apart from farming itself. More generally, within any homogeneous asset and income group – with an expected relatively low variation in total incomes – lower per capita incomes and expenditures would be associated with larger-sized households by virtue of the cake arithmetic. It is important to see in this respect that even if one assumes rough constancy of birth and death rates within such groups, a variation in family size nevertheless emerges over periods of time - short or long - via the chance mechanism that governs births and deaths from year to year: not all similarly placed families experience a birth or a death in a given period of time. For example, among labour households, four or five years after they get set up, some will have one child, some two children and some none – producing a variation of significance to family size and (the calculated) per capita consumption levels as they emerge. Indeed, young and able-bodied labourer couples are likely to be above the poverty line defined to cover the bare subsistence minimum, but that, in most cases, is sure be a transitory status for them. Moreover, the simple relationship between family size and per capita income breaks down when we compare, say, asset-poor labour and small farmer households with those commanding substantial landholding and other income-earning assets; the latter will surely enjoy higher levels of per capita income than the former irrespective of household size. In other words, the disadvantage of a large family can be offset by the advantage of a higher asset holding. This modified arithmetic has of course to be made more precise, and in that context we present some more empirical observations.

(O2): *When households are classified by land possessed, mean per capita expenditure rises as we move up the land size classes.*

(O3): *By the same classification as in* (O2)*, mean household size also increases as we go up the land size groups*.

These again are universal features of all survey data. Table A2 referring to the year 2003 (NSS 59th round) shows that for rural India as a whole the average monthly per capita expenditure increases from Rs.530 in the lowest landholding group of households with land less than 0.01 hectare (ha) to Rs.728 in the highest one with land over 8 ha. Table A3 shows for the same round that the mean household size increases steadily from 3.34 in the landless group, and 4.01 among tiny holders with land below 0.02 ha, to 11.28 in the highest landholding class with land over 20 ha. Reading observations (O2) and (O3) along with (O1), we get a statistical puzzle that I call the Vaidyanathan paradox:

(O4): *The mean values of per capita expenditure and household size across landholding classes are positively correlated although the variables themselves are inversely correlated in the whole population*.

This is a variant of the well-known Simpson's paradox referring to cases of reversal of conditional orderings and correlations under aggregation; it can be resolved through an analysis of the relevant multivariate distributions. We attempt to do that in the next section through a simple model.

But the conundrum of poverty and family size would remain, Are poor families big or small? By the criterion of per capita expenditure they are big (O1), but looked at in terms of assets, mainly land, they are the smallest (O3). One most striking aspect of Indian data in this respect is the fact that rural labour households (comprising those in agriculture as well as in non-agricultural work) have – and have consistently had over many decades – the smallest household size among all rural classes all over India This is consistent with (O3) since labour families are asset-poor. For data in this respect see Table A4, which shows that rural labour households are about ten percent smaller, on average, than *all* rural households (implying perhaps about 12 to 15 percent lower than in *non-labour* families: the precise numbers depend on the proportion of labour households in the total). These differences have remained so over decades, as the data show.

2. **Household Size**

The heuristic model that follows refers to cultivator households only, and our attempt here is to produce a simple argument that it is consistent with all the different correlations discussed above. No doubt, improvements to this simple model (such as by the inclusion of age-sex composition of households, apart from size, and variables relevant to incomes of non-agricultural households) can lead to more convincing explanations for the observed relationships among *all* rural households as well.

Let *Y, X* and *S* stand respectively for the variables: total household expenditure, landholding (possessed) and household size. (Henceforth, let hh designate 'household' used interchangeably with 'family'.) Further, let lower case letters represent natural logarithms of these variables; that is: $y = ln\ (Y)$, $x = ln\ (X)$ and $s = ln\ (S)$, and so on.

Observations (O1) to (O4) can now be rewritten respectively as displays (1) to (4) below, with *Cov (.,.)* standing for covariance and *E(.)* for the expectation operator for random variables:

(1) $Cov\ (Y/S, S) < 0$;

(2) $E\ (Y/S \mid X = x)$ increases as x increases;

(3) $E\ (S \mid X = x)$ also increases as x increases and, indeed,

(4) $Cov\ [E\ (Y/S \mid X), E\ (S \mid X)] > 0$.

Display (4) says that the mean values of per capita expenditure and household size across landholding classes are positively correlated; while by (1) the underlying variables are inversely related in the whole population. .

Note also that these relationships hold empirically among all rural households, as discussed in the last section. The following model, restricted to cultivator households, can however be extended to all rural families by replacing landholdings with total assets, and allowing for the inclusion, in a suitable form, of variables such as those related to education and professional skills that provide higher incomes to some families, That however is beyond the scope of this paper. Before we present the model we state two statistical propositions, (P1) and (P2), of relevance to our argument:

(P1): *For positive random variables X and Y, Cov (X, Y) and Cov (ln X, ln Y) have the same sign, positive or negative.*

This intuitively obvious result is a direct consequence of the fact that the logarithmic transformation is a strictly increasing one; it follows easily from some theorems in Lehmann (1966).

(P2): *For random variables X and Y, if E (Y| X=x) = m(x) is a non-decreasing (or a non-increasing) function of x, then Cov (Y, X) is non-negative (non-positive).*

This follows from the decomposition of the total covariance as:

$$Cov\ (Y, X) = Cov\ [E\ (Y|X), E\ (X|X)] + E\ [Cov\ (Y, X)|X].$$

The second term equals zero because X becomes a constant. The first term is non-negative (or non-positive) if $m(x)$ is non-decreasing (or non-increasing) – another intuitively obvious result following from Lehmann (1966).

Applying proposition (P2) to statements (2) and (3) we see that the Vaidyanathan paradox is articulated in the simplest of terms by

(4a) $Cov\ [(Y/S), X] > 0$; and $Cov\ (S, X) > 0$, but $Cov\ (Y/S, S) < 0$.

From a purely statistical theoretical angle this is no puzzle because the covariance relation is not a transitive one. Still, all counter- intuitive correlations need probing to discover what lies behind them. So, we now suggest a model in two equations (with variables in logarithms set in lower case as indicated earlier):

$$(5)\quad y = c_1 + b_1 s + b_2 x + u$$

$$(6)\quad s = c_2 + b_3 x + v$$

where c_1 and c_2 are constants; u and v, the residuals, are assumed to be independent with 0 means and finite variances, and also to be independent of s and x. In this model y (total hh expenditure) and s (hh size) – both in natural logarithms – are endogenous and x (land or all assets in the more general case) is the only exogenous variable. While equation (5) – telling us how family income, and hence expenditure, is determined by landholdings and family size – needs no comment, equation (6) needs to be explained. We shall briefly refer, later in the paper, to factors that determine family size – fertility, mortality and the 'age' of the family – and their possible dependence on levels of living and asset holdings. Note further that since the equations are set out in terms of logarithms of all variables, b_1 and b_2 represent the elasticities of total household expenditure with respect to (*w.r.t*) household size and landholdings respectively, and b_3 the elasticity of family size *w.r.t* landholdings.

It follows then that the reduced form equation for per capita expenditure in natural logarithms is:

$$(7)\quad y - s = c_3 + [b_3(b_1 - 1) + b_2]\ x + [u + v(b_1 - 1)],$$

where $c_3 = [c_1 + c_2(b_1 - 1)]$ is a constant. From (6) we get

$$(8)\quad E(s \mid x) = c_2 + b_3 x,$$

which is an increasing function of x if and only if $b_3 > 0$. Similarly equation (7) yields

$$(9) \quad E(y - s \mid x) = c_3 + [b_3(b_1 - 1) + b_2]x \,.$$

From (8) it follows that s and x (and hence S and X, by proposition P1) will be positively correlated if and only if (*iff*) $b_3 > 0$, that is the elasticity of household size *w.r.t.* land is positive.

From (9) we see that $(y - s)$ and x (and hence Y/S and X) will be positively correlated *iff*

$$(10) \quad [b_3(b_1 - 1) + b_2] > 0 \ \text{ or } b_2 > b_3(1 - b_1),$$

a condition that involves all the three elasticities mentioned earlier. Condition (10) also implies that mean PCE increases as land size increases: as can be read from (9).

We have from (8) and (9)

$$(11) \quad Cov\,[E(y - s \mid x), E(s \mid x)] = b_3\,[b_3(b_1 - 1) + b_2]\,Var\,(x)$$

which is positive *iff* (10) holds, along with $b_3 > 0$

Finally, from (6) and (7) we get the total covariance

$$(12) \quad Cov\,(y - s\,,\, s) = b_3\,[b_3(b_1 - 1) + b_2]\,Var\,(x) + (b_1 - 1)\,Var\,(v)$$

in which the second term is negative *iff* $b_1 < 1$; and, further, if it is greater in absolute value than the first term , the overall covariance will be negative.

When we classify households into landholding groups, the first term the right hand side of (12) represents the 'between' group covariance and the second term the 'within' component. Our empirical observations show that the 'between' component is positive and the overall covariance negative (and that is the essence of the paradox); we see from (12) that the conditions required for this to happen involve not only the relevant elasticities but also other characteristics of the relevant multivariate distribution – $Var\,(x)$ and $Var\,(v)$ in our model. As we have suggested earlier, the inverse correlation between family size and per capita expenditure among households within a homogeneous asset group is easy enough to understand; the model provides insights into possible structures that produce the paradox: higher levels of assets surely compensate for the disadvantage of a big family size.

We cannot estimate the model as stated here, however, because equation (5) involving all the variables is not 'identified' in an econometric sense. A modified model would require the addition of exogenous variables relevant to both family size and income generation. However, for illustrative purposes and useful insights, we can look at the decomposition of the covariance as it actually happens in some household level survey data, along with summaries of the conditional means of household size and consumer expenditure across landholding classes.

Table 1 presents the estimated means of household size (S) and monthly per capita expenditure (PCE) among rural households in Andhra Pradesh (AP) classified by land possessed for the years 1983 and 2003.. This summary of household level data shows how both these mean values generally increase as we move up the landholding groups, in accordance with (O2) and O3).

Table 1: Estimated Means of Household Size (S) and Monthly Per Capita Expenditure (PCE in Rs.) by Land Possessed: Rural Andhra Pradesh (AP), 1983 and 2003

Land Possessed (ha)	1983		2003	
	S	PCE	S	PCE
< 0.01	3.62	118	3.46	574
0.01 – 0.2	4.24	124	4.14	646
0.21 – 0.4	4.58	109	4.19	551
0.41 – 1.0	4.80	117	4.08	649
1.01 – 2.0	5.28	122	4.54	654
2.01 – 3.0	5.69	135	4.62	697
3.01 – 4.0	5.93	153	4.92	791
4.01 – 6.0	6.12	156	5.45	768
6.01 – 8.0	6.67	172	5.30	802
> 8.0	6.73	189	5.06	1005

Note: These are estimated from unit level National Sample Survey (NSS) central sample data for Andhra Pradesh relating to the 38th and 61st rounds, using weights appropriate to the sample design, provided by the NSS.

As we have argued, the implied positive component of the covariance 'between' land classes gets swamped, under certain conditions, by the negative component 'within' those land classes, to produce an overall inverse correlation between S and PCE. The unit level data for AP allow us to verify this – for illustration – through an analysis of covariance between and within landholding groups. Tables 2a and 2b show the magnitudes of the statistics that produce the paradox displayed earlier by (4a).

Table 2a: Analysis of Covariance: Monthly Per Capita Expenditure and Household Size by Landholding Groups: Rural Andhra Pradesh 1983

Source	PCE/100	(PCE/100)*S	S
Between land groups (9)	119.0	587.7	4174,6
Within land groups (5775)	4694.5	- 2407.1	25567.3
Total (5784)	4813.5	- 1818.4	27941.9

Note: PCE stands for monthly per capita expenditure (in Rs.) and S for household size. Numbers in parentheses refer to degrees of freedom. The land groups are defined as in Table 1. The computation uses the NSS sample-design based weights

Table 2b: Analysis of Covariance: Monthly Per Capita Expenditure and Household Size by Landholding Groups: Rural Andhra Pradesh 2003

Source	PCE/100	(PCE/100)*S	S
Between land groups (9)	2191.6	1984.3	1458.7
Within land groups (5517)	157386.8	- 8280.4	19028.1
Total (5526)	159578.4	- 6296.1	20486.8

Note: See Note to Table 2a.

3. The Paradox of Death Rates

The resolution of the family size paradox lies thus in a recognition of the endogeneity of both income and household size, and of the effects of other, exogenous factors (such as asset holdings) on both the variables. Among the factors that determine family size is the death rate that we may reasonably expect to be inversely related to income and wealth - and living standards in general. Demographic literature suggests that, apart from mortality rates, fertility rates and the propensity to migrate also depend on the economic status of households. Another factor relevant to household size is that poor families tend to be more 'nuclear', less 'joint' in composition, in comparison to wealthy households, as much historical research on several pre-industrial societies and contemporary Indian evidence shows; for data and reasoning in this respect see Krishnaji (1980: A38 –A43).

If, in this context, we refer to survey data for understanding the nature of dependence of family size on levels of living, we run into another set of paradoxes, the most striking of which is the death rate paradox. In rural areas, with which we are concerned exclusively here, death rates – infant mortality rate (IMR) in particular – have been found in NSS surveys to be positively correlated with per capita expenditure: that is, death rates in survey data exhibit an increase as PCE rises (see Table 3 below). We present this set of outdated data as an illustration of how paradoxes emerge out of misspecification of variables and misclassification of statistics, but it certainly is also a point of departure for further study.

Table 3: Death Rates by Levels of Per Capita Expenditure: Rural India 1964-65

Monthly Per Capita Expenditure (Rs.)	Crude Death Rate	Infant Mor tality Rate
0 - 11	10.03	32.94
11 - 15	10.58	71.71
15 - 21	13.38	122.29
21 - 28	16.06	152.37
28 - 43	17.88	153.13
Above 43	21.81	293.37
All groups	14.75	127.29

Note: These are estimates of annual rates. The crude death rates are per 1000 population; infant mortality rates refer to number of infants dying within a year of birth per 1000 live births.

Source: Report No, 186, National Sample Survey.

These correlations are nonsensical and they obviously arise from using an inappropriate variable (PCE) as a measure of the standard of living. It requires an explanation nevertheless because it was an enduring feature of NSS reports, a decisive source of much empirical information on the Indian economy. (The NSS seems to have discontinued this type of tabulation in later rounds.) Let us note that the paradoxical observation is based on mean values (of death rates) across PCE groups, which conceal much variation within: large families with large land holdings and incomes and small landless families may belong to the same group defined by per capita expenditure. So, as in the case of the family size paradox, we must look at other, exogenous factors that might play a role in producing the heterogeneity and the contradictory correlations 'between' and 'within' homogeneous income and asset classes.

Before we present an explanation, let us note that NSS reports attribute this 'unexpected' result to underreporting of deaths in the poor (that is, low PCE) households and to the possible effects of small sample sizes among the upper PCE groups (that would produce unreliable estimates). The first part of this explanation is false and misleading. For example, if we assume that the IMR in the lowest PCE group is 150 instead of the survey estimate of 32.94 (noting that the NSS estimates for the upper three PCE group are over 150), it would imply that among the poorest roughly only one out five infant deaths are reported, This is absurd for these infant deaths refer to a traumatic family event occurring less than a year before the survey date: we Indians talk freely about births and deaths, and if they happen to be recent ones we talk about nothing else. However it is true that small sample sizes in the upper groups lead to unreliable estimates; but that alone is a sufficient reason to discard this type of tabulation.

If we discard per capita expenditure (or income) as unsuitable for analysing mortality differentials, what other variable can we suggest instead? A clue is provided by the National Family Health Survey (NFHS) data. Table 4 gives the NFHS estimates of infant and Under-5 mortality rates for 2005-06 for rural India, corresponding to households classified by a wealth index (instead of levels of PCE, as in NSS). The steady decline in IMR associated with increasing wealth, as depicted in Table 4, is from about 70.7 in the lowest wealth group to 33.6 in the highest; Under-5 mortality rates exhibit a similar pattern of decline (from 100.9 to 36.2)

Table 4: Infant and Under-5 Mortality Rates by Wealth Index: Rural India 2005-06

Wealth Index Group	Infant Mortality Rate	Under-5 Mortality Rate
Lowest	70.7	100.9
Second lowest	69.2	90.4
Middle	60.6	73.6
Fourth	42.3	49.1
Highest	33.6	36.2
All Households	62.2	82.0

Note: The wealth index has been constructed using 33 different types of household assets. The index groups are quintile-based. For details see the reports of the National Family Health Survey–3.

The plausibility of total resources of a family as a factor relevant to demographic variation across income classes is suggested by equation (6) in our earlier analysis of family size. Empirically, Table 4 supports this, and the reasons are not difficult to fathom. *The prevention of death always assumes priority; and in the event of serious illness in a family, its ability to meet medical and other expenses for saving a life would depend on the total income and asset holdings of the family (which also determine its ability to borrow money) rather than being constrained by family size and income reckoned in per capita terms.*

Regarding death as a household event (independent of family size) suggests also that in some contexts we should compute death rates per household rather than per person as demographers do by convention. The usual rates per person are no doubt useful for many purposes just as per capita income is; so what is suggested here is some rethinking on how the rates (conceptualising them as probabilities of the occurrence of a death) are to redefined when confronted with absurdities like a positive correlation between death rates and per capita expenditure when analysing how incomes influence demographic differentials.

For any group of households defined, say, by per capita expenditure or landholdings, or for the whole population, if the usual crude death rate or death rate per person is CDR and the death rate per household is HDR, we can then write

(13) HDR= CDR * mean household size in that group.

Here we use * for the multiplication sign. Similarly, if we define HIMR as the number of infant deaths per household, then it is easy to verify that

(14) HIMR = IMR * CBR * mean household size in that group

where IMR is the usual ratio of infant deaths to births, and CBR, the crude birth rate, is the ratio of births to total population

The paradoxical empirical observation of death rates increasing as PCE rises can then be rewritten as:

(15) E (CDR | Y/S) and E (IMR | Y/S) both increase as Y/S increases

A formal resolution of these puzzles requires a specification of the probabilities of births and deaths (of infants and persons of all ages as well) as functions of household characteristics, a task not attempted here. Instead, we present a heuristic argument supported by a numerical illustration based on some household level data for Andhra Pradesh.

If we assume that *death rates per household* are roughly the same within a given asset class, then larger households will have smaller *death rates per person* by virtue of equation (13); and these same households will be associated with lower per capita expenditures. Clearly, it is the variations within asset classes that are responsible for the perverse positive correlation between crude death rates and per capita expenditure. The heuristic argument we have suggested here thus neatly ties up the family size paradox with the absurdity of low (high) death rates being associated with low (high) per capita expenditure.

We now use some household level NSS data for Andhra Pradesh (for the 38th round) to provide a numerical illustration of how sensible death rates per household get transformed into unacceptable variations in death rates per person, given the trivariate distribution of landholding, household size and household expenditure, with characteristics as in actually observed data.

Table 5a: Assumed Deaths Rates per Household by Landholdings, Andhra Pradesh, 1983

Land Quartiles	Assumed Deaths per 1000 Households	Mean Household Size (S)	Implied Deaths per 1000 Persons
First	50	4.09	12.2
Second	46	4.86	9.5
Third	43	5.59	7.7
Fourth	40	6.26	6.4

Note: Mean Household size is estimated from the unit level data. See text for further explanations.

Table 5b; Derived Death Rates per Person by Per Capita Expenditure: Andhra Pradesh, 1983

Quartiles by Per Capita Expenditure (PCE)	Derived Deaths Per 1000 Households	Mean Household Size	Derived Deaths Per 1000 Persons
First	44.4	6.02	7.37
Second	44.4	5.62	7.90
Third	44.5	5.13	8.68
Fourth	44.0	4.35	9.77

Notes: The death rates here (by quartiles of PCE) correspond to those by land groups in Table 5a; the procedure is explained in the text.

Table 6a: Assumed Infant Mortality per Household (HIMR) and Crude Birth Rate (CBR) by Landholdings: Andhra Pradesh 1983

Quartiles by Land	Assumed HIMR	Assumed CBR	Mean House -hold Size	Derived IMR
First (lowest)	20	35	4.09	139.7
Second	15	30	4.86	102.8
Third	10	25	5.59	71.6
Fourth	5	20	6.27	39.9

Table 6b: Derived Infant Mortality Rates by Per Capita Expenditure: Andhra Pradesh 1983

Quartiles by per Capita Expenditure	Derived HIMR	Derived CBR	Mean House hold size	Derived IMR
First (lowest)	12.01	27	6.02	73.9
Second	12.04	27	5.62	79.3
Third	12.19	27	5.13	88.0
Fourth	11.32	26	4.35	100.1

We restrict ourselves to the 1713 households in the sample designated as 'self-employed in agriculture' (essentially cultivators, small and big); the restriction arises from the lack of data on asset holdings apart from landholdings, which makes the status of non-cultivator families difficult to determine. These data enable us to estimate, using the appropriate multipliers, the quartiles of the different distributions of our interest here: of land, per capita expenditure, and so on.

Our numerical illustration assigns death rates per 1000 households (HDR) of 50, 46, 43 and 40 to groups defined by land-based quartiles 1 to 4 respectively. We then use (13) to compute the implied death rates per person (CDR) in quartile groups defined by different criteria. Table 5a shows that the calculated CDR values exhibit a sensible pattern across landholding groups: they decline from 12.2 (per mille) in the lowest land quartile to 6.4 in the highest. However, when we reclassify households into quartiles by levels of per capita expenditure, the perverse positive correlation between PCE and CDR emerges; the latter increase from 7.37 per mille in the lowest PCE quartile to 9.77 in the highest. (See Table 5b, which shows, besides, that the reclassification leaves the death rates per household nearly constant, hiding their assumed variation by land classes.) We present similar calculations for IMR in Tables 6a and 6b with assumed plausible values for HIMR and CBR among land classes; a reclassification of households by levels of per capita expenditure is shown here also to lead to the absurd positive correlation between IMR and PCE. In this example we have assumed HIMR to decline from 20 (per thousand) to 5 as land size increases; this corresponds to a decline in the usually computed IMR from 140 to 40 as we move up the land quartiles. The reclassification by PCE quartiles produces an increase in the IMR from 74 to 100.

These illustrative calculations and their implications are consistent with the model we presented earlier as an example of how ignoring the two-sided nature of relationships and the influence of exogenous factors produce paradoxical correlations. In respect of the household paradox, the within land group covariance is negative, the between component positive, and the total negative. In contrast, in the case of death rates, our simulation exercise suggests that the within component (of covariance between death rates per person and PCE) is positive, the between one negative and the total (unacceptably) positive.

We note in passing, however, that in the case of some other demographic variables the two components can be of the same sign. To illustrate this, we now refer to sex ratio (FMR: females per 1000 males) variations in rural India. Table A6 in the Appendix shows that for rural India, 2005-06, the sex ratios are over 1000 in the lowest per capita expenditure groups, while they are far lower (918 and below) in the highest groups; these data are based on the 62nd round of the NSS. Table A7 provides FMR variations classified by wealth status of rural households (to which reference has been made earlier in Table 4). These NFHS-3 data show how sex ratios decline as the wealth status of families improves from the poorest to the richest. *In other words, there is no paradox here: whether by per capita expenditure or by wealth, poorer families have more balanced or distinctly higher sex ratios.* In Tables A8 and A9 we present an analysis of covariance between per capita expenditure and FMR by landholding quartiles for some unit level data relating to Andhra Pradesh (respectively for 1983 and 2003), The analysis shows that both the between and the within components are negative. This happens partly because high female to male ratios are a disadvantage to asset-poor families in terms of income-earning capabilities. On the other hand, well-endowed rural families in many parts of India have a much stronger preference for male progeny in comparison with households having little or no property; the resulting gender biases lead to lower sex ratios. (For a fuller analysis of sex ratios in relation to poverty, see Krishnaji, 1987).

4. Concluding Remarks

There are no easy answers to the question whether poor families are big or small in size. It is even more difficult to discover much substance behind propositions such as that poor families are poor because they are big or that they are big because they are poor, despite the appealing common sense that lies behind these propositions and the statistical evidence usually cited in their support. Above all, it is the narrowness of per capita arithmetic that is responsible for confusion and invalid notions about poverty and the demographic behaviour of the poor.

Although our focus here has been on demographic variation in relation to poverty, the analysis has obvious implications for poverty assessment and income inequalities. Poverty measures based on per capita expenditure are shot through with the difficult problem of how to set the poverty line, what prices to use for the purpose, and so on – a problem that is at the heart of all controversies and the millions of words on the subject. Is there an alternative approach? Yes; and the suggestion here is that we should begin paying attention to relative inequalities.

The poverty line approach leads to absolute measures of poverty and their variations over space and trends in time. While such measures are useful for policy-making of different kinds, for example, in assessing the needs of the poor and determining financial allocations to poverty alleviation programmes etc, there is also a need – obviously less perceived – to assess the trends in relative inequalities, especially under rapid overall growth of the Indian economy in the last couple of decades. By all accounts, the growth processes have exacerbated inequalities: between rural and urban populations, and among different classes. True, the NSS data can be used to estimate aggregative measures of relative inequality, such as the coefficient of variation or the Gini coefficient; bur such measures tend to be far too (numerically) insensitive to trends, particularly because of the sampling errors in their estimation. Relative inequalities surely provide extra information for assessing changing levels of consumption.

For example, the ratio of monthly per capita expenditure (MPCE) among the rural Indian population to that in urban areas declined sharply from 71.6 per cent in 1977-78 to 53.1 per cent in 2004-2005. This is a well-known and well-documented phenomenon. What is perhaps not so well known is that within rural India the ratio of MPCE among labour to that in the entire rural population has increased slightly between 1977-78 and 1999=2000 (from about 80 per cent to 86 per cent, see Table A10). This implies that cultivators are doing much worse than labourers in the present phase of economic growth.

While labourers and self-employed artisans and small cultivators are – and have always been – at the bottom of the socio-economic hierarchy, the scheduled castes and tribes are also among the deprived classes. The NSS does tabulate data relating to such categories of the population. Besides, household level NSS data contain a wealth of information on occupational and other characteristics of families. It is then possible to assess relative inequalities among different classes not only over time but also across space: for example to see how consumption levels among labourers in, say, Punjab and Kerala compare with those in Bihar or Madhya Pradesh.

References

Krishnaji, N (1980): "Agrarian Structure and Family Formation – A Tentative Hypothesis", Review of Agriculture, *Economic and Political Weekly*, Vol 15, pp A38-A43.

Krishnaji, N. (1984): "Family Size, Levels of Living and Differential Mortality in Rural India – Some Paradoxes", *Economic and Political Weekly*, Vol 19, pp 248 –258.

Krishnaji, N. (1987): "Poverty and Sex Ratio – Some Data and Speculations", *Economic and Political Weekly,* Vol 22, pp 892- 897.

Lehmann, E.L. (1966): "Some Concepts of Dependence", *Annals of Mathematical Statistics*, 37, pp 1137 – 1153.

Appendix

Table A1: Estimated Mean Household Size (S) by Monthly Per Capita Expenditure (MPCE) Classes: Rural India 2003.

MPCE (Rs.)	S: All Rural Households	S: Farmer Households
0 - 225	6.2	6.9
226 - 255	6.3	6.7
256 - 300	6.1	6.6
301 - 340	6.0	6.3
341 - 380	5.7	6.0
381 - 420	5.5	5.8
421 - 470	5.2	5.6
471 - 525	5.2	5.3
526 - 615	4.7	5.0
616 - 775	4.6	4.8
776 - 950	4.2	4.4
Above 950	3.8	4.1
All	5.04	5.5

Source: Reports No. 490 & 495, National Sample Survey,.

Tab A2: Average Monthly Per Capita Expenditure (MPCE) by Landholding Classes – Rural India 2003

Land Possessed (Hectares)	MPCE (Rs)
<0.01	530
0.01 – 0.2	545
0.21 – 0.4	535
0.41 – 1.0	525
1.01– 2.0	574
2.01 – 3.0	614
3.01 – 4.0	609
4.01 – 6.0	645
6.01 – 8.0	671
> 8.0	728
All	554

Source: Report No. 490, National Sample Survey.

Table A3: Mean Household Size (S) by Land Owned, Rural India, 2003

Land Owned (Ha)	Mean S
03.34	
< 0.002	4.01
0.002 – 0.005	4.30
0.005 – 0.04	4.70
0.040 – 0.5	4.97
0.5 - 1	5.40
1 - 2	5.70
2 - 3	6.04
3 - 4	6.23
4 - 5	6.78
5 – 7.5	6.95
7.5 - 10	7.13
10 - 20	8.15
Above 20	11.28
All	4.99

Source: Report No. 491, National Sample Survey.

Table A4: Mean Size (S): Labour and All Rural Households, All-India

NSS Round	Year	S: Rural Labour Households	S: All Rural Households
38th	1983	4.6	5.20
43rd	1987-88	4.6	5.07
50th	1993-94	4.4	4.89
55th	1999-2000	4.6	5.04

Source: Reports for different rounds of Rural Labour Enquiry and National Sample Survey.

TableA5: Means of Household Size (S), Total Consumer Expenditure (TCE) and Per Capita Expenditure (PCE) Among Rural Classes, Andhra Pradesh !983 & 2003

Class	1983			2003		
	S	TCE	PCE	S	TCE	PCE
Agricultural labourers, with no Land	3.59	340	103	3.54	1720	508
Agricultural labourers with land	4 .30	413	103	4.04	1891	500
Farmers with land between (0-2) hectares	4.09	541	148	3.92	2439	675
Farmers with land between (2-4) hectares	4.86	586	130	4.15	2439	628
Farmers with land between (4-8) hectares	5.59	704	135	4.53	2970	722
Farmers with land above 8 hectares	6.27	955	167	5.44	3730	765

Note: The entries are estimated from unit level data using procedures suggested in the NSS reports, applying the relevant multipliers.

Table A6: Sex Ratio (FMR) by Monthly Per Capita Expenditure (MPCE) Class: Rural India, 2005-06

MPCE (Rs)	FMR	MPCE (Rs)	FMR
< 235	1112	455 - 510	987
235 - 270	1000	510 - 580	963
270 - 320	977	580 -690	916
320 - 365	1004	690 - 890	907
365 - 410	1043	890- 1155	918
410 - 455	916	> 1155	916

Note: FMR is number of females per 1000 males.
Source: Report No. 523, National Sample Survey, 62nd Round.

Table A7: Sex Ratio by Wealth class: Rural India, 2005-06

Wealth Class	FMR
Poorest	1044
Poorer	1026
Middle	1006
Richer	963
Richest	976

Note: Computed from NFHS–3; FMR is number of females per 1000 males.

Table A8: Analysis of Covariance: PCE and FMR by Land classes: Andhra Pradesh, 1983

Source	PCE/100	(PCE/100)*FMR	FMR
Between land groups (3)	32.15	- 11.35	2.47
Within land groups (1690)	1822.74	- 88.02	1224.17
Total (1693)	1854.89	- 99.37	1226.64

Note: Degrees of freedom are indicated in parentheses.

Table A9: Analysis of Covariance: PCE and FMR by Land classes: Andhra Pradesh 2003

Source	PCE/100	(PCE/100)*FMR	FMR
Between land groups (3)	318.58	- 8.00	3.37
Within land groups (1295)	57036.42	- 429.72	890.23
Total (1298)	57355.00	- 437.72	893.60

Note: Degrees of freedom are indicated in parentheses

Table A10: Monthly Per Capita Expenditure (Rs) by Categories of Population

Year	All Rural (R)	All Urban (U)	Rural Labour (L)	R/U (%)(%)	L/R
1977-78	68.69	96.15	53.60	71.6	78.0
1983	112.31	165.80	91.01	67.7	81.0
1987-88	158.10	249.92	127.81	63.2	80.8
1993-94	284.40	458.00	235.75	61.4	82.9
1999-2000	486.16	854.92	418.42	56.8	86.1
2004-2005	558.78	1052.36		53.1	

Source: Various Rural Labour Enquiry and NSS reports.

Current Working Papers

Working Paper	Paper No.
Fertility Transition in India: Development and Diffusion N. Krishnaji, January 2009	--- 76
Public Health Priorities in India Insights from Changing Patterns of Death by Region, Gender and Age N. Krishnaji, August 2008	--- 75
Girl Children in Domestic Service: The Case of West Bengal Deepita Chakravarty, Ishita Chakravarty, June 2008	--- 74
Population Pressure and Labour Intensification: An Indian Historical Perspective N. Krishnaji, April 2008	--- 73
Turnaround of Indian Railways 2004-06 - A Case Study S. Mahendra Dev, G. Alivelu, March 2008	--- 72
Inclusive Growth in Andhra Pradesh: *Challenges in Agriculture, Poverty, Social Sector and Regional Disparities* S. Mahendra Dev, March 2007	--- 71
The Unqualified Medical Practitioners : *Methods of Practice and Nexus with the Qualified Doctors* K V Narayana, May 2006	--- 70
Millenium Development goals in Asia and Pacipic Region: *Differentiated Progress and Persistent Disparities* S. Mahendra Dev, C. Ravi and S. Galab, March, 2006	--- 69
Banking Sector Reforms and Equity : A Sub-Regional Analysis in Andhra Pradesh K.S. Reddy and P. Prudhvikar Reddy, January, 2006	--- 68
Rural Non-agricultural Employment in India The Residual Sector Hypothesis Revisited C. S. Murty, December, 2005	--- 67
The Differential Impact of User Heterogeneity in Resource Management: A Case Study from Kerala Jeena T Srinivasan, December, 2005	--- 66