Are the 2000 Poverty Estimates for India a Myth, Artefact or Real?

The objective of this study is to assess whether the estimates of poverty provided by the government of India for the year 1999-2000 are appropriate, as these estimates have generated a lot of controversy both in India and abroad. We examine this issue using non-parametric methods and provide alternate estimates of poverty for all-India and 16 major states. We compare our poverty estimates with those presented in the literature. Our broad conclusion is that the different methods proposed for correcting poverty estimates in India are unlikely to yield even approximately correct estimates of poverty, or consensus on these estimates, when there are unknown measurement errors due to incomparable surveys.

Gurleen Popli, Ashok Parikh, Richard Palmer-Jones

The issue of measuring poverty using the headcount ratio, in the context of the Indian economy, is widely discussed and disputed among various economists in the government of India, World Bank, and academicians both within and outside India. The discussion has been mainly over the apparent reduction in poverty in the 1990s, as officially noted by the government of India (GoI) using the National Sample Survey (NSS),[1] from 1993-94 (50th round of NSS) to 1999-2000 (55th round of NSS). The estimates of poverty produced on the basis of the 55th round published in February 2001, showed a marked reduction in the headcount poverty measure, which fell from 37.1 (percentage of poor below the poverty line[2]) in 1993-94 to 26.8 per cent in 1999-2000 for rural households, while among the urban households the index fell from 32.9 per cent in 1993-94 to 24.1 per cent in 1999-2000 [Deaton and Dreze 2002].

Debate has emerged between two opposing forces [Deaton and Kozell 2004], one, being the pro-liberalisation[3] group, according to whom liberalisation has reduced poverty as revealed in the official estimates based on the "thick" NSS rounds. Their argument rests largely on the divergence between two sources of information, namely, the national accounts statistics (NAS) that shows increasing economic growth from 1990s, and the NSS, which indicates very stagnant levels of per capita expenditure and poverty up to the 54th round (1993-94). The other view (not in favour of liberalisation) is that the economic growth as promised by the liberalisation process has not trickled down.[4] This lobby, in turn attributes the fall in poverty in the 55th round to the change in survey methodology which took place in that round and incorrectly recorded monthly consumption expenditure.

NSS consumption surveys have used a 30-day recall period for all goods from the 38th round (1983) to the 50th round. Most statistical offices around the world use a shorter recall period for high frequency items, such as food and longer recall period for low frequency goods, such as durables. The NSS experimented in the 51st to 54th rounds with the different recall periods. They compared the traditional 30-day recall questionnaire (Schedule 1) with three reporting periods 7, 30 and 365 days (Schedule 2), applied to different classes of goods. Households were assigned to one or other schedule at random using the same sample design, and it was found that, on average, the experimental 7/30/365 schedule reported higher total expenditures. Shorter reporting periods generated higher rates of consumption so that the 7-day recall period had higher average consumption than the 30-day recall in schedule 1, while 365-day schedule showed lower average consumption.

The schedule used in the 55th round was different from previous quinquennial (so-called thick) rounds and also from both the schedules of the experimental rounds. For the high frequency items, households were asked to report their expenditures for both recall periods (7-day and 30-day). Multiple reporting periods are often used in household surveys, but results cannot be compared with surveys when only the 30-day questions are asked.[5] When the households are asked to report consumption over two periods in the same schedule they are likely to reconcile the reported consumption across two periods. One plausible hypothesis is that reconciliation will probably pull up the rate of consumption at 30-day recall above what it would have been if asked in isolation. If that is the case, 30-day consumption in 55th round is too high compared to the 30-day estimates of consumption from previous rounds. This will overstate the reduction in poverty in the 55th round.

Sundram and Tendulkar (2003a, b), Deaton (2003), Sen (2000) and Sen and Himanshu (2004) are only a few of the studies which analyse in detail how the differences between two schedules affect the measurement of consumption, poverty and inequality. Visaria (2000) pointed out that the estimated poverty was cut by half when the experimental multiple record period (7/30/365 days) schedule was used instead of traditional (30-day) schedule for all items. But such a comparison does not show how trends in poverty have been affected by the change in schedule designs since earlier calculations of poverty were based on the uniform 30-day recall period.

It is important to note that the 30-day recall period was kept in both schedules for some items, namely, fuel and light, miscellaneous goods and services, rents and consumer taxes, and non-institutional medical expenses, which account for a substantial share of the budget.[6] Tarozzi (2004) shows that the distributions of estimated average real per capita total monthly expenditure on 30-day items in the two different schedules used from the 51st to the 53rd rounds of NSS are similar, even though the real per capita total consumption expenditure in Schedule 2 is systematically 15-20 per cent higher than that of the corresponding figures for Schedule 1.

The main objective of this paper is to propose a procedure to estimate per capita expenditure based poverty counts from the 55th NSS round without using data on per capita expenditure from this survey.[7] In Section I of the paper we discuss the data, choice of explanatory variables and summary statistics for two periods 1993-94 and 1999-2000, both being the years of thick rounds of the NSS. Given the belief that the poverty is under recorded in the official counts of GoI in the 55th round, in Section II we provide three different methodologies (parametric, non-parametric and semi-parametric) to obtain "corrected" poverty estimates, taking into account the change in the survey design. In each of these, the dependent variable is a dichotomous poverty dummy where the person in a household below the official poverty line is regarded to be poor. In Section III we present the results and estimation at work. Section IV focuses on comparative advantages of various procedures and summarises the conclusions of the study both for measuring poverty in India and for estimating poverty in the presence of changes in survey schedules.

I
Data and Choice of Explanatory Variables

For our analysis, we use the 50th and 55th rounds of the NSS on consumer expenditure in both rural and urban areas collected in 25 states and seven union territories.[8] The survey periods were from July 1993 to June 1994, for the 50th round and July 1999 to June 2000 for the 55th round. The NSS data are a cross-section of a geographically stratified, clustered random sample of households across India. In addition to information on household consumer expenditure and demographic characteristics, the NSS contains detailed questions on other household characteristics such as the social group, religious affiliation, occupation and educational level of the head of the household. The data on monthly per capita consumption of household is used in conjunction with the official poverty line by the GoI, to classify household in terms of their poverty status. The poverty line is defined for each Indian state and for rural and urban sectors separately. We use the official state and sector poverty lines for the years of the survey, 1993-94 and 1999-2000, to adjust for the inflation from 1993-94 to 1999-2000 for each state and all-India.[9]

The choice of explanatory variables is guided both by economic theory and by the empirical context.[10] The household characteristics play an important role in determining poverty and these are used in one of the approaches to obtain estimates of poverty for 55th round. The standard variables taken to determine poverty both in India and other countries at the household level are educational level and occupational type of the household [Dreze and Srinivasan 1997; Van de Walle and Gunawardena 2001]. To capture the effect of education on the probability of a household being in poverty, we use dummy variables corresponding to the highest educational level completed by the head of the household. Thus, we include dummy variables corresponding to literate, below primary level, literate, at primary level, literate, secondary level and below, literate, at higher secondary level, and literate, graduate level and above (the reference group in our case is households where the head of the household is not literate).

With respect to occupation, we include dummy variables corresponding to four occupational groups in the rural sector – self-employed in agriculture, self-employed in non-agriculture, agricultural labour, and non-agricultural labour; and three occupational groups in the urban sector – self-employed, wage/salary earner, and casual labourer. The reference group is the occupational category termed "others" by the NSS for both rural and urban sectors – these are households which have not been classified as "self-employed in non-agriculture" and earn less than 50 per cent of their total income in any of the categories mentioned. We also use dummy variables for religion[11] and social group affiliations,[12] as there is some evidence that households of certain religious denominations and/or belonging to marginal social groups – the scheduled tribes, the scheduled castes, and other backward castes – may be disadvantaged [Deshpande 2000; Borooah and Iyer 2004; Sundram and Tendulkar 2003c].

In addition to the above explanatory variables, we include in our analysis a number of background and demographic variables. First, we include the generational impact reflected by the age of the household head. We use two variables: age (number of years), and age-squared to reflect the non-linear effects of age on poverty. Second, we incorporate the effect of household size on the probability of the household being in poverty, as previous studies have noted a negative relationship between per capita expenditures and the size of the household [Krishnaji 1984]. Given the possible presence of economies of scale in household consumption, we include household size squared as an additional control variable. Third, we add a dummy for female-headed households because these households are more likely to be in poverty [Dreze and Srinivasan 1997]. Finally, we capture the effect of the demographic composition of the household by adding a set of variables that take into account the proportion of males and females in different age groups [Meenakshi and Ray 2002].

I
Models and Estimation Methods

In this section we present the model we estimate and the methods of estimation used. We start by explaining the notation and defining the problem at hand.

Let x_i be the log of the i^{th} household's monthly per capita consumption expenditure (PCE henceforth) reported when the respondent is given the traditional (30-day recall period for all items) schedule, and z_{js} be the log of the official poverty line for the j^{th} sector (rural or urban) and the s^{th} state of India, to which the i^{th} household belongs. The headcount ratio (HCR) H_{tjs} for any given year t, in the j^{th} sector and s^{th} state is defined as:

$$H_{tjs} = \int_0^{z_{js}} f_t(x)dx = F_t(z_{js}), \qquad \text{...(1)},$$

where $f_t(x)$ is the density function of x_i in year t, and $F_t(z_{js})$ is its cumulative density function (cdf). In subsequent analysis to keep the notation simple we will suppress the subscripts j and s. HCR gives us the proportion of population which lives in households with monthly PCE less than the official poverty line. The empirical counterpart for (1) is,

$$\hat{H}_t = \sum_i \omega_i I(x_i < z), \qquad \text{...(2)}$$

where ω_i's are the individual inflation factors normalised so that they sum to one, and I(.) is the indicator function taking value one whenever the condition in the brackets is satisfied and zero otherwise.

If the survey design had remained the same over time then computation of the HCR for each period is straightforward and can be done using the expression given in (2). However as the recall period was changed in the 55th round, x_i for the 55th round is unobserved and hence HCR for the 55th round cannot be evaluated using (2). Instead what we observe for the 55th round is $\tilde{x}_i$, log of the i^{th} household's monthly PCE reported when the respondent is given the questionnaire with multiple recall periods.

Deaton (2001, 2003) and Deaton and Dreze (2002) propose a method to estimate the HCR for the 55th round, which will be comparable to the poverty estimates from the earlier period. Their approach relies on finding an auxiliary variable using which one can retrieve the distribution of unobserved x_i for the 55th round. One such auxiliary variable is, m_i, which is the log of i^{th} household's reported monthly PCE on selected "30-day items". The specific assumptions they make are: (1) the distribution of the variable m_i is unaffected by the change in the survey design; and (2) the relationship between x_i and m_i is constant over time. The second assumption requires that the probability of being poor remain same over time, once we have the information on m_i. Validity of these assumptions has been tested by both Deaton (2001, 2003) and Tarozzi (2002, 2004).[13]

To use m_i to retrieve information on x_i for the 55th round rewrite (1) as:

$$H_t = \int_0^{\infty}\int_0^{z} f_t(x|m)g_t(m)dxdm = \int_0^{\infty} F_t(z|m)g_t(m)dm \cdot \qquad ...(3)$$

where $F_t(z|m)$ is the conditional distribution of x_i, conditioned on m_i, and $g_t(m)$ is the density function of m_i in year t. The empirical equivalent of (3) will be given by,

$$H_t = \frac{1}{n}\sum_{i=1}^{n} F_t(z|m_i)g_t(m_i). \qquad ...(4)$$

Given the two assumptions made in Deaton's work the "corrected" HCR for the 55th round, H_{55}^{c} can be calculated as:

$$H_{55}^{c} = \frac{1}{n}\sum_{i=1}^{n} F_{50}(z|m_i)g_{55}(m_i). \qquad ...(5)$$

The expression in (5) can now be estimated for the 55th round, given the data available. $g_{55}(m_i)$ is the distribution of m_i, the variable which we have assumed is not impacted by the change in the survey design. This can be estimated using the non-parametric kernel density estimation techniques.

The conditional distribution function $F_t(z|m_i)$, is simply the probability of being poor at time t, conditional on m_i, and can be written as, $F_t(z|m_i) = Pr_t(x_i < z|m_i)$. This probability can be represented as a regression model. The dependent variable of this regression is the poverty dummy, defined as,

$$y_i = \begin{cases} 0 & \text{if} \quad x_i \geq z \\ 1 & \text{if} \quad x_i < z \end{cases} . \qquad ...(6)$$

Deaton in his analysis uses the non-parametric regression model. The model he estimates is: $Pr_t(y_i = 1|m_i) = r_t(m_i) + u_i$, where $r_t(m_i)$ is a non-parametric function. A non-parametric regression differs from an ordinary least square regression, in the fact that it does not force any specific functional form (liner or quadratic for example) on the data. However there is a drawback in his choice of estimation method. In practice the method is equivalent to fitting a linear model in the interval $m_i \pm h$, around every data point, where h is a small window width.[14] As such it carries with it some of the problems associated with estimating a linear regression model for a discrete dependent variable, also called the linear probability model (LPM).

In particular there are two problems associated with LPM – the predicted probabilities from these models do not necessarily lie between 0-1, and the errors are heteroskedastic. While non-parametric estimation takes care of the first problem (the estimated probabilities lie between 0-1), the error terms are still heteroskedastic. This can make the estimators from these regressions inconsistent. Further most of the asymptotic properties for the nonparametric regressions are established for a continuous dependent variable, not for a discrete dependent variable, as is the case here. Keeping in mind that we have a binary dependent variable we propose an alternative estimation method.

The general form of the regression to be estimated is given by,

$$Pr_t(y_i = 1|m_i) = G[\nu(m_i, \beta)] + u_i, \qquad ...(7)$$

where the parameter β reflects the impact of change in m_i on the probability of being poor, and u_i is the random error. The index function $\nu(.,.)$, and the link function G[.] may or may not be known. Model as expressed in (7) is a single index model (SIM), which can be estimated using semi-parametric estimation techniques. We assume the index function $\nu(.,.)$ to have a linear form, $\nu(m_i, \beta) = \beta_0 + \beta_1 m_i = \nu_i$. However no assumption is made on the distribution of the link function $G[\nu_i]$, which is estimated non-parametrically.[15]

There are number of different semi-parametric methods proposed in the literature to estimate the SIM given in (7).[16] The goal is to find efficient estimators for both β and G[.]. The different methods proposed can be classified under two broad headings: (1) weighted average derivative estimation methods (WADE); and (2) pseudo maximum likelihood estimation methods (PMLE). In our analysis we use the first method, particularly we use the density weighted average derivative estimation (DWADE) method.[17]

Once we have estimated the model in (7), we have $\hat{F}_t(z|m_i)$, now we can estimate the "corrected" HCR for the 55th round using (5).

Demographic and Socio-Economic Characteristics of Households

The probability that a household will be poor depends on a number of demographic and socio-economic characteristics of the household. Some of these variables have been identified in Section I above. Deaton's approach does not allow for the use

Table 1: Poverty Estimates

(in Per cent)

	Rural Households		Urban Households	
	1993-94 50th Round	1999-2000 55th Round	1993-94 50th Round	1999-2000 55th Round
Andhra Pradesh	15.8	10.9	38.6	27.3
Assam	45.2	40.4	8.0	7.6
Bihar	57.9	44.1	34.6	33.7
Gujarat	21.6	12.5	28.2	14.8
Haryana	28.3	7.4	16.5	10.2
Himachal Pradesh	30.4	7.6	9.3	4.6
Karnataka	30.1	16.9	39.8	24.7
Kerala	25.3	9.6	24.2	20.0
Madhya Pradesh	40.7	37.5	48.1	38.6
Maharashtra	37.7	23.6	34.9	26.9
Orissa	49.9	48.3	40.7	43.8
Punjab	11.7	6.2	10.9	5.5
Rajasthan	26.3	13.4	31.1	19.5
Tamil Nadu	32.8	20.6	39.9	22.7
Uttar Pradesh	42.2	31.4	35.0	30.6
West Bengal	41.2	31.8	23.0	15.0
All-India	37.6	27.7	33.6	24.7

Notes: Poverty is estimated using the headcount ratio, calculated from the sample of NSS data used in this study.

of more than one auxiliary variable in his analysis. Tarozzi (2004) proposed an alternative method to obtain the corrected estimates for the 55th round. The advantage of his approach over Deaton's is that it allows the use of household characteristics as auxiliary variables in the estimation of the conditional distribution $F_t(z|\varphi_i, m_i)$, where φ_i is a vector of variables describing the household characteristics, and m_i is as defined above. Using the same assumptions as in Deaton's approach he proposes the use of a re-weighting function to obtain the corrected poverty estimates for the 55th round.[18] However there is a caveat in his analysis, as his own findings suggests. The original purpose of estimating the conditional distribution function $F_t(z|\varphi_i, m_i)$, was to recover information about the unobservable x_i for the 55th round. However this is possible only under the two specific assumptions made. Tarozzi's own analysis suggests that while m_i satisfies these two assumptions, the vector φ_i does not, hence limiting the use of his suggested extension of Deaton's approach.

This however does not mean we cannot use the information on household characteristics. For now we take the assumptions underlying the work of Deaton and Tarozzi as given (we will comment on them later), and propose an alternative way to obtain the corrected poverty estimates for the 55th round, taking into account the heteroskedaticity problem. We estimate a probit (one can estimate a logit model as well) model with the poverty dummy as the dependent variable and the vector (φ_i, m_i) as the explanatory variables, for the 50th round. Using the estimated coefficients from the 50th round with the 55th round data we predict the probability of the household being poor. Estimate for poverty, HCR, is then the weighted average of these probabilities, the weights being the household inflation factors ω_i.[19]

In Section IV of the paper we compare results from each of the models and estimation method stated above. We'll also make some comments on application of some of these procedures.

II
Estimation at Work and Results

In Table 1 we report the rural and urban, poverty estimates for India for the selected states, as calculated by us from the data available. The estimates we get are very similar to the GoI poverty estimates. All further estimation is done using our unit record data so the comparisons we make will be to our estimates (reported in Table 1). Using our data we also obtain the corrected poverty estimates for the 55th round using the non-parametric regression, as done by Deaton. We refer to these as the Deaton-Adjusted poverty estimates for the 55th round. These estimates are reported in first column of Table 2a and 2b, for rural and urban India respectively.[20] For all-India and the different states considered here the Deaton-Adjusted poverty estimates are above the GoI poverty estimates, suggesting an underestimation of poverty, by the GoI, for the 55th round.

Single Index Model: Parametric and Semi-parametric Estimates

Next we obtain the corrected headcount ratios for the 55th round by looking at the estimates obtained from the semi-parametric estimation of the conditional distribution $F_t(z|m_i)$. The results are reported in second column of Tables 2a and 2b for rural and urban India respectively. If we use the SIM and estimate it semi-parametrically, then using the same assumptions as made by Deaton we find that the adjusted poverty estimates for the 55th round fall back to the estimates of the 50th round, indicating that there has been no change in poverty over the five years between the two rounds. We get the same result if we use a probit model instead of a semi-parametric SIM. The estimation method (parametric or semi-parametric) does not seem to make a difference; the model specification (non-parametric or SIM) however does make a difference.[21]

The above results are hard to believe, there might be a controversy over how much poverty in India has declined over time, but there is consensus on the fact that it has declined. One explanation for why our adjusted HCR for the 55th round is not different from the HCR for the 50th round could be, that one of the assumptions underlying Deaton's approach is violated, in particular the second assumption. If the probability of being poor does not have a strong relationship with m_i (even if the relationship is constant over time), it would mean that the coefficients β tend to zero, in which case we would get the above result. In fact the estimated coefficients from the semi-parametric method do tend

Table 2a: Corrected Headcount Poverty Ratios for the 55th Round (1999-2000), Rural India
(in Per cent)

	Deaton-Adjusted Estimates	Semi-parametric Estimates	Heteroskedasticity Corrected Probit Estimates	
			No House-hold Covariates	With House-hold Covariates
	(1)	(2)	(3)	(4)
Andhra Pradesh	14.5	15.8	12.0	13.1
Assam	40.4	45.2	44.0	45.5
Bihar	50.4	57.9	53.9	50.6
Gujarat	17.1	21.6	14.4	16.1
Haryana	17.1	28.3	14.1	13.8
Himachal Pradesh	24.7	30.4	21.1	17.8
Karnataka	28.2	30.1	29.2	26.3
Kerala	15.7	25.3	14.8	12.5
Madhya Pradesh	34.5	40.7	39.1	35.4
Maharashtra	30.8	37.7	34.3	30.0
Orissa	50.7	49.9	56.4	50.1
Punjab	6.4	11.7	6.1	7.3
Rajasthan	18.8	26.3	19.0	19.4
Tamil Nadu	23.2	32.8	21.2	18.9
Uttar Pradesh	34.2	42.2	37.1	33.1
West Bengal	32.5	41.2	33.6	32.1
All-India	30.6	37.1	32.8	30.3

Table 2b: Corrected Headcount Poverty Ratios for the 55th Round (1999-2000), Urban India
(in Per cent)

	Deaton-Adjusted Estimates	Semi-parametric Estimates	Heteroskedasticity Corrected Probit Estimates	
			No House-hold Covariates	With House-hold Covariates
	(1)	(2)	(3)	(4)
Andhra Pradesh	30.5	38.6	30.7	28.2
Assam	7.7	8.0	7.3	8.5
Bihar	31	34.6	34.5	32.8
Gujarat	20.2	28.2	21.8	17.5
Haryana	10.4	16.5	10.0	9.6
Himachal Pradesh	7.2	9.3	5.8	4.6
Karnataka	28.5	39.8	31.1	26.8
Kerala	18.3	24.2	17.9	18.1
Madhya Pradesh	39.6	48.1	40.5	38.6
Maharashtra	27.1	34.9	28.2	28.8
Orissa	43.1	40.7	47.6	43.6
Punjab	7.6	10.9	6.4	6.9
Rajasthan	23	31.1	24.7	22.1
Tamil Nadu	29.2	39.9	30.6	25.6
Uttar Pradesh	30.5	35.0	31.8	29.0
West Bengal	18.4	23.0	18.6	17.6
All-India	26.9	33.8	28.4	26.7

to zero.[22] Expenditure on 30-day items, m_i and total expenditure, x_i have a strong correlation (as suggested in Section I of the paper), which means if the model is correctly estimated then the βs should be significantly different from zero (whether or not there is a theoretical reason to believe that there is a relationship).

As they are not significant, another explanation is suggested. Namely, that the results could be due to heteroskedasticity. In the presence of heteroskedasticity the coefficients of the SIM model (whether they are estimated parametrically or semi-parametrically) are inconsistent [Yatchew and Griliches 1985].[23] In particular they tend to zero if the underlying variance is large, giving us results reported above. The source of heteroskedasticity in the SIMs is not due to model misspecification. However we are dealing with a micro level data, which often has problems of heteroskedasticity. This is more likely to be the case, with the most likely source of heteroskedasticity being the household size. We next include the household characteristics in our analysis.

Heteroskedasticity Corrected Poverty Estimates

As we pointed out earlier, the method of estimation – parametric or semi-parametric – does not seem to make a difference in our estimates, so for further analysis we present results only for parametric estimation of SIM, i e, the probit model. We present results from two different specifications, one which includes the household characteristics and the other which does not. The household characteristics that we do include are the variables discussed above.

The regression model is modelled to allow for multiplicative heteroskedasticity [Harvey 1976]. The two variables on which the error variance is assumed to depend on are household size and the square of household size. In each specification we tested for heteroskedasticity, the null hypothesis of no-hetroskedasticity is rejected for both the specifications. We also tested to see if there is any other variable (other than household size and household size square) that might effect the variance of the error term, but we do not find evidence of any other variable impacting the variance. We not only correct for heteroskedasticity by explicitly modelling the error variance, but also calculate Huber-White standard errors.

This model now corrects for both the sources of hetroskedasticity – model misspecification and nature of data. To obtain the adjusted estimates for the 55th round, we use the estimated coefficients from the 50th round and the data of the 55th round to predict the probability of being poor in the 55th round. HCR is then obtained as the weighted average for these predicted probabilities, with the household inflation factors as the weights. The results from this exercise are presented in column three and four of Tables 2a and 2b for the rural and urban India respectively.

For both rural and urban India we find that for most states our estimates of poverty lie above the Deaton-Adjusted estimates for the 55th round.

IV
Summary and Conclusions

In this study, parametric, semi-parametric and non-parametric approaches are used to obtain the headcount measure of poverty for the year 1999-2000 using household expenditure survey data of 1993-94. It is indeed widely known in India and abroad that the surveys for the period 1993-94 and 1999-2000 are non-comparable and the household survey for the latter period was contaminated by the multiple recall periods used. Various investigators have tried to adjust the poverty estimates using different methodologies. The results prove to be sensitive to the underlying model specification.

Deaton (2001) used information on a single auxiliary variable and a non-parametric approach to estimate "adjusted" estimates. His results indicate an underestimation of poverty by the GoI. While GoI suggested a decline of poverty by almost 10 per cent in both rural and urban India, Deaton's estimates suggest a decline of near 7 per cent. His results, however suffer from two caveats – inclusion of only one explanatory variable and the estimation method (using the methodology of continuous variable to a discrete variable). Tarozzi (2004) includes more variables in his non-parametric approach but finds that the GoI estimates are after all not way off the mark (his estimates being closer the GoI estimates).

Our analysis suggests an improvement over Deaton's non-parametric approach, we adjust the methodology of non-parametric estimation to a discrete variable, and find drastically different results. These results are different from those of Deaton and Tarozzi, and indicate no change in poverty over time. As an alternative to the non-parametric method we use a heteroskedasticity incorporated probit model and the estimates of poverty from this model (when no household covariates are used) indicate a decline in poverty by about 5 per cent, which is less than the declines estimated by Deaton's adjustments and the GoI official estimates.

Methods of estimations by themselves cannot be an answer to correction of poverty indices when there exist measurement errors. The underlying model specification and the assumptions made are also very important. Our broad conclusion is that the different methods proposed for correcting poverty estimates in India are unlikely to yield even approximately correct estimates of poverty, or a consensus on these estimates when there are unknown measurement errors due to non-comparable surveys. EPW

Email: g.popli@shef.ac.uk

Notes

[The 50th Round data were made available under the collaborative arrangement between the National Sample Survey Organisation, Government of India, and the Overseas Development Group, School of Development Studies, of the University of East Anglia. The second and third authors work on this paper was funded in part by SSRC Grant No R8256, which also funded access to the 55th Round data. We would also like to acknowledge useful contributions from our colleagues in R8256 Kunal Sen, of the School of Development Studies, University of East Anglia and Amaresh Dubey, Department of Economics, North East Hills University, Shillong, India, and courteous responses from Angus Deaton and Jeff Racine to our requests for clarification. The usual caveats apply.

1 NSS is a household survey done in the country every year, collecting information on the consumption. Every five years a large survey is run with the specific aim of recording household consumption for estimating poverty in the country.

2 The government of India publishes official poverty lines based on an estimated per capita monthly expenditure associated with the consumption of a given minimum caloric intake in 1973-74, updated each year using consumer price indices for each state and sector (rural/urban).

3 Economic liberalisation of India started in 1991, for an overview of the economic reforms and their impact refer to Sachs, Varshney and Bajpai (1999) and references therein.

4 Another argument points to the growing divergence between expenditure poverty and food or calorie poverty [Patnaik 2004; Meenakshi and Vishwanath 2004].

5 Two ways in which the reporting period may effect the expenditure reported are memory lapse and telescoping, these are likely to affect different items in different ways, depending in large part on the salience and frequency of the expenditures.

6 There are six broad categories of goods for which the 30-day recall was used in all schedules, fuel and light, miscellaneous goods, miscellaneous services, non-institutional medical services, rent and consumer cesses and taxes. The first four are important items, and all households report expenditures on the first three. Virtually all households also report

non-institutional medical expenditures. Expenditures on these six categories account for more than 20 per cent of all expenditure (more in urban areas). Total expenditures on these 30-day goods are also highly correlated with total household expenditures and hence these expenditures on comparably surveyed goods might be used to track trends in total expenditures and also in poverty.

7 This is in the spirit of Deaton (2001, 2003).

8 Our work will focus on 16 major Indian states, which together account for 95 per cent of the population.

9 Concern is expressed in the literature over the limitation of the official poverty lines [Subramanian 2005; Deaton and Tarozzi 2000]. The price indices used to update these poverty lines are based on fixed commodity weights that have become outdated over time. Deaton and Tarozzi (2000) have proposed an alternate set of poverty lines based on unit values and quantities consumed obtained from the NSS expenditure surveys themselves. However, a drawback of the Deaton-Tarozzi poverty lines is that they are not available for all states and union territories in India. Clearly using different deflators such as those calculated by Deaton and Tarrozi (2000) and Deaton and Dreze (2002) will result in slightly different results.

10 The tables with the summary statistics of the data, specifically the variables, used in this paper are available from the corresponding author on request.

11 Dummy variables are used for four major religions – Hindu, Muslim, Christian, and Sikh; base category being "other religions".

12 Three dummy variables are used to identify the disadvantaged households– scheduled tribes, scheduled castes, other backward castes; base category being "others" – households not in the disadvantaged social groups.

13 These assumptions are tested using the information available in the thin rounds (51st to 54th), conducted by NSSO every year between 50th and the 55th thick rounds. The validity of using the data from the thin rounds to test the assumptions can be questioned. The second assumption required for the Deaton type adjustments to be valid – that the probability of being poor conditional on the real expenditure on 30-day items is stable across rounds – is much less plausible empirically [Sen and Himanshu 2004].

14 For further details on estimation of non-parametric density and regression refer to Pagan and Ullah (1999, chapters 2 and 3). As a first introduction and to get a non-technical explanation for non-parametric estimation procedures refer to DiNardo and Tobias (2001).

15 For details on the semi-parametric estimation refer to Hardle et al (2004, chapter 6) and Pagan and Ullah (1999, chapter 7).

16 A parametric version of (7) would be a probit (logit) model, where G[.] is assumed to have a cumulative normal distribution (logistic distribution).

17 DWADE is studied by Powell, Stock and Stoker (1998). The PMLE method is detailed by Klein and Spady (1993). Other related studies are Ichimura (1993) and Horowitz (1992).

18 Tarozzi's work is in the spirit of the work done by DiNardo, Fortin and Lemeiux (1996). It requires the additional assumption that the structural relationship between household characteristics and the probability of being poor does not change over time, and that the distributions of the household characteristics variables is not affected by survey design.

19 While our approach still requires the probability of being poor conditional on household characteristics to be stable over time, it makes no assumptions on the distribution of the household covariates – which may/may not be affected by the change in the survey design.

20 We report the "Deaton-Adjusted" estimates to demonstrate that we can replicate his results using our data, any divergence in the subsequent results is only due to different model and or estimation methods used and not due to data discrepancies.

21 Results for probit model are not reported but are available on request from the authors. We also tried the logit model and obtained similar results.

22 Coefficients from the probit model also, though significant, are very small in magnitude.

23 This problem is not addressed by Huber-White robust estimation of standard errors. Deaton (1997, chapter 2) mentions that the inconsistency of estimated parameters in SIM, due to heteroskedsticity, can be ignored if all we are interested in are the estimated probabilities, but should be taken into account if the parameters of the model are of interest, which is the case here.

References

Borooah, V and S Iyer (2004): 'Vidya, Veda and Varna: The Influence of Religion and Caste on Education in Rural India', *Journal of Development Studies*, forthcoming.

Deaton, Angus (1997): *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*, The John Hopkins University Press.

– (2001): 'Adjusted Indian Poverty Estimates for 1999-2000', Research Programme in Development Studies, Princeton University.

– (2003): 'Adjusted Indian Poverty Estimates for 1999-2000', *Economic and Political Weekly*, January 25, pp 322-26.

Deaton, Angus, and Alessandro Tarozzi (2000): 'Prices and Poverty in India', Research Programme in Development Studies, Working Paper No 196, Princeton University.

Deaton, Angus and Jean Dreze (2002): 'Poverty and Inequality in India, A Re-Examination', *Economic and Political Weekly*, September 7, pp 3729-48.

Deaton, A and V Kozel (2004): 'Data and Dogma: The Great Indian Poverty Debate', http://www.wws.princeton.edu/~deaton/papers.html.

Deshpande, A (2000): 'Does Caste Still Define Disparity? A Look at Inequality in Kerala, India', *American Economic Review*, 90, pp 322-25.

DiNardo, John, Nicole M Fortin and Thomas Lemieux (1996): 'Labour Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach', *Econometrica*, Vol 64, No 5, pp 1001-44.

DiNardo, John and Justin L Tobias (2001): 'Nonparametric Density and Regression Estimation', *Journal of Economic Perspective*, Vol 14, No 4, pp 11-28.

Dreze, J and P V Srinivasan (1997): 'Widowhood and Poverty in Rural India: Some Inferences from Household Survey Data', *Journal of Development Economics*, 54, pp 217-34.

Hardle, W, M Muller, S Sperlich and A Werwatz (2004): *Non-parametric and Semi-paramteric Models*, Springer.

Harvey, A (1976): 'Estimating Regression Models with Multiplicative Heteroscedasticity', *Econometrica*, Vol 44, pp 461-65.

Horowitz, J (1992): 'A Smoothed Maximum Score Estimation for the Binary Response Model', *Econometrica*, 60, pp 505-31.

Ichimura, H (1993): 'Semi-parametric Least Squares (SLS) and Weighted SLS Estimation of Single-index Models', *Journal of Econometrics*, Vol 58, pp 71-120.

Klein, R and R Spady (1993): 'An Efficient Semi-parametric Estimator for Binary Response Models', *Econometrica*, Vol 61, pp 387-421.

Krishnaji, N (1984): 'Family-size, Levels of Living and Differential Mortality in Rural India – Some Paradoxes', *Economic and Political Weekly*, 19, pp 248-58.

Meenakshi, J V and R Ray (2002): 'Impact of Household Size and Family Composition on Poverty in Rural India', *Journal of Policy Modelling*, 24, pp 539-59.

Meenakshi, J V and B Vishwanathan (2003): 'Calorie Deprivation in Rural India', *Economic and Political Weekly*, 38(4), pp 369-75.

Pagan, Adrian and Aman Ullah (1999): *Non-parametric Econometrics*, Cambridge University Press.

Patnaik, U (2004): *The Republic of Hunger*, Public Lecture on the occasion of the 50th Birthday of Safdar Hashmi, organised by SAHMAT (Safdar Hashmi Memorial Trust) on April 10, New Delhi.

Powell, J L, J H Stock and T M Stoker (1989): 'Semi-parametric Estimation of Index Coefficients', *Econometrica*, Vol 57, pp 1403-30.

Sachs, Jeffery D, Ashutosh Varshney and Nirupam Bajpai (1999): *India in the Era of Economic Reforms*, Oxford University Press, Oxford.

Sen, Abhijit (2000): 'Estimates of Consumer Expenditure and its Distribution: Statistical Priorities after the NSS 55th Round', *Economic and Political Weekly*, December 16, pp 4499-518.

Sen, Abhijit and Himanshu (2004): 'Poverty and Inequality in India – 1', *Economic and Political Weekly*, September 18, pp 4247-63.

Subramanian, S (2005): 'Unravelling a Conceptual Muddle: India's Poverty Statistics in the Light of Basic Demand Theory', *Economic and Political Weekly*, January 1, pp 57-66.

Sundram, K and S D Tendulkar (2003a): 'Poverty *Has* Declined in the 1990s: A Resolution of Comparability Problems in NSS Consumer Expenditure Data', *Economic and Political Weekly*, January 25, pp 327-37.

– (2003b): 'Poverty in India in the 1990s: Revised Results for All-India and the 15 Major States for 1993-1994', *Economic and Political Weekly*, November 15, pp 4865-72.

– (2003c): 'Poverty among Social and Economic Groups in India in 1990s', *Economic and Political Weekly*, December 13, pp 5263-76.

Tarozzi, Alessandro (2002): 'Estimating Comparable Poverty Counts from Incomparable Surveys: Measuring Poverty in India', Princeton University, processed.

– (2004): 'Calculating Comparable Statistics from Incomparable Surveys, with an Application to Poverty in India', Duke University, processed.

Van de Walle, D and D Gunewardena (2001): 'Sources of Ethnic Inequality in Vietnam', *Journal of Development Economics*, 65, pp 177-207.

Visaria, Pravin (2000): 'Poverty in India during 1994-98: Alternative Estimates', Institute for Economic growth, New Delhi, processed June 9.

Yatchew, Adonis and Zvi Griliches (1985): 'Specification Errors in Probit Models', *Review of Economics and Statistics*, Vol 67, pp 134-39.