WORKING PAPER NO. 55

Estimates of the Number of Very Rich in India

I. Natarajan

NCAER

National Council of Applied Economic Research
Parisila Bhawan, 11-Indraprastha Estate
New Delhi-110002 (India)
Fax: (91-11) 3327164 Tel: (91-11) 3317860-68

September 1995

ISBN 81-85877-26-2

Price:
Rs. 100.00
$ 10.00

Published by

P.K. Krishnaswamy, Registrar & Secretary
National Council of Applied Economic Research
11-Indraprastha Estate, New Delhi-110002

Printed at

Akashdeep Printers, 20, Ansari Road, Daryaganj,
New Delhi-110002.

Preface

NCAER conducted the only national survey of household incomes, savings, and investments, in 1978. Such surveys in a country as big as India, with a part of household transactions taking place by barter, and high incidence of tax evasion, create complications in getting reliable estimates of income and its distribution. This makes the questionnaire and the survey as well as the subsequent analysis, very difficult since incomes and expenditures have to be estimated from the physical assets in possession and the usage of different products and services. The sample required is also very large and becomes even larger when estimates are needed at the state and even lower levels. Perhaps, it is for this reason that NCAER has not been able to conduct another such survey, because of lack of sponsors.

However, NCAER has been able to tackle the problem through a different, though not as efficient, route. Since 1985 NCAER has conducted, till 1994, six surveys of household purchases of selected manufactured consumer goods. In order to capture enough incidents of purchases of different products so as to extrapolate them to the national or sub-national levels, the first four surveys used a sample of 500,000 households (for consumer durables). All these households were asked to answer the question as to which income range their household incomes would fall into the income levels over different years were designed to enable comparison over time by adjusting each year for inflation, using the wholesale price index as a discounting factor. Of course this method ignores variations in household consumption baskets, and regional variations in prices and consumption baskets. But it is the closest that is possible in a survey of this magnitude.

There are five income ranges for which data is collected and extrapolated. The highest level of annual household income in 1992-93 was taken at Rs. 86,000 and above. In the last two years, many companies have become interested in looking at households with far higher levels of income. While some urban sample data is available from other surveys, the samples may not be adequate, and

in any case, ignore the large and fast growing rural markets identified by the same NCAER household survey, of purchases of manufactured consumer goods.

In this paper, Mr. I. Natarajan, Chief Economist, NCAER, who designed all these surveys, has tried to use reported responses of household incomes above Rs. 86,000 per annum, from these surveys, along with the available data on households in the five income ranges, and some other assumptions, to estimate the breakup of urban and rural households in India, nationally, with household incomes above Rs. 86,000 per annum, in different segments upto the highest level of Rs. 5 million and above. This is, as far as we know, the first such scientific attempt to estimate the number of households at high levels of income. We hope in future years to refine our surveys and our estimation procedures, to get even better estimates.

This working paper is intended to provide information to users, and it is also hoped, will stimulate scholars to comment on the estimation procedures so that they can be improved for the future.

Mr. I. Natarajan was assisted by Dr. R.K. Shukla, Statistician, NCAER, in this exercise.

New Delhi
September 1995

S. L. Rao
Director-General

Estimates of the Number of Very Rich in India

The National Council of Applied Economic Research has been conducting a series of market surveys, called Market Information Survey of Households (MISH), during the last 10 years. The surveys, using a multistage, stratified sample design, spread across the length and the breadth of the country, try to estimate the market size and the profile of consumers for a variety of consumer goods, both durable and expendable. Data from MISH have been used extensively by decision-makers in the Government and the corporate sector in the country, and have been widely quoted in the national and international press. The first MISH was carried out for the reference period 1985-86 and the sixth for 1993-94. MISH was not carried out for 1988-89, 1990-91 and 1991-92. For durable consumer goods the sample size was 500,000 households for the first four rounds and 280,000 for 1992-93 and 1993-94. The reduction in the sample size has not affected the quality of the data or the reliability of the estimates. It was made possible by the increasing penetration rates of some of the "luxuries" included in the list of durable goods covered in the survey.

The sample design for these surveys is given below.

Sample Design

A. Coverage

All States and Union Territories of India are covered except Arunachal Pradesh, Manipur, Mizoram, Nagaland, Sikkim, Jammu & Kashmir, Andaman and Nicobar Islands, Dadra & Nagar Haveli, and Lakshadweep.

B. Sample Design for Rural Areas

There are 410 districts in the States/Union Territories covered in the survey. Each of these districts is selected with probability 1, i.e., all are selected. From each district 2 villages are selected with probability proportional to the population of the village in 1991. In all, 820 villages were selected for the 1993-94 round of MISH.

All households in the selected villages were listed through a specially designed proforma. This proforma sought the following information from each household:

- Name of the head of the household
- Location of the house
- Household size
- Main occupation of the head of the household
- Annual household income from all sources

The number of households listed in the selected villages was 99169. Table A provides a profile of the rural sample.

Table A: Profile of the Rural Sample

State/Union Territory	*No. of Districts*	*No. of Villages Selected*	*No. of Households Listed*
Andhra Pradesh	23	46	5330
Assam	23	46	5796
Bihar	42	84	12520
Goa	2	4	428
Gujarat	19	38	4507
Haryana	16	32	3743
Himachal Pradesh	12	24	1731
Karnataka	20	40	4827
Kerala	14	28	3631
Madhya Pradesh	45	90	10576
Maharashtra	30	60	7055
Meghalaya	5	10	957
Orissa	13	26	3041
Punjab	12	24	2479
Rajasthan	27	54	6128
Tamil Nadu	21	42	4713
Uttar Pradesh	63	126	15199
West Bengal	17	34	5565
Chandigarh	1	2	483
Delhi	1	2	193
Pondicherry	4	8	266
TOTAL	**410**	**820**	**99169**

C. Sample Design for Urban Areas

According to the 1991 Census, there are 3,579 cities/towns in the States/Union Territories covered by the survey. The population of cities/towns in India varies from less than 5,000 to over 9 million. There are 53 cities with population exceeding 500,000. All these cities were included in the sample with probability 1. The remaining towns/cities were grouped into 6 strata on the basis of their population size and from each stratum a sample of towns was selected independently.

A progressively increasing sampling fraction with increasing town population class was used for determining the number of sample towns to be selected from each stratum. The sampling fractions were used at the state level. However, when these fractions were actually applied, the number of cities/towns to be allocated were not integers in many cases. In all these cases they were rounded off to the nearest integer. When a rounded off integer was zero, one town was selected from the relevant category. As a result, the actual sampling fraction for each city/town size class turned out to be higher than the one originally planned and is shown in Table B.

Table B: Sampling Fraction for City/Town Groups

City/Town Size Group	*Sampling Fraction*	*No. of Sample Cities/Towns*
Over 500,000	1.00	53
200,000 to 500,000	0.80	65
100,000 to 200,000	0.60	98
50,000 to 100,000	0.30	102
20,000 to 50,000	0.10	97
10,000 to 20,000	0.05	62
Below 10,000	0.04	38
ALL	**0.14**	**515**

In all 515 cities and towns thus selected constituted the first stage sample for urban areas. These accounted for about 14 per cent of the cities/towns in the sample coverage (specified earlier), but, what is more important, covered a major chunk of the urban population.

The Office of Registrar General & Census Commissioner, India had divided each city/town into a number of blocks for building their sampling frame for the 1991 Census. The blocks are compact agglomeration units with well defined boundaries and are of approximately equal population of around 600. A sample of blocks was selected independently from each town/city and these constituted the second stage units for the urban sample.

The number of blocks selected varied between 2 and 30, depending upon the size of the town. The total number of blocks thus selected was 1509.

As in the case of villages, all households in the selected urban blocks were listed. The total number of households listed which are used for estimating the income distribution was 182599. Table C presents the profile of the urban sample.

Distribution of Households by Income

One of the by-products of these surveys is the distribution of households by income. With such a large sample, the surveys could provide income distributions at a much disaggregated level. The distributions are available at the state level for rural and urban areas separately. Further, for urban areas, distributions are available for each state for town groups classified by population. Distributions for individual towns are also available for cities with population of one million and more. Such a unique data set is not available from any other source.

For the year 1993-94, households are classified into five income groups:

Low	Up to Rs. 20,000 per annum
Lower middle	Rs. 20,001 to Rs. 40,000
Middle	Rs. 40,001 to Rs. 62,000
Upper middle	Rs. 62,001 to Rs. 86,000
High	Above Rs. 86,000

Table C: Profile of the Urban Sample

State/Union Territory	No. of Sample Towns/Cities									Total No. of Towns Selected	No. of Sample Blocks	No. of House-holds Listed
	Over 50 Lakhs	10–50 Lakhs	5–10 Lakhs	2–5 Lakhs	1–2 Lakhs	50,000–1 Lakh	20,000–50,000	10,000–20,000	Below 10,000			
Andhra Pradesh	–	2	1	8	11	10	9	2	1	44	133	16074
Assam	–	–	1	–	2	1	2	2	1	9	22	2677
Bihar	–	1	3	5	4	8	8	3	1	33	87	12152
Goa	–	–	–	–	–	1	1	1	1	4	8	903
Gujarat	–	3	1	2	9	8	5	4	2	34	101	11150
Haryana	–	–	1	2	5	3	2	2	1	16	41	5110
Himachal Pradesh	–	–	–	–	1	–	1	1	2	5	11	1274
Karnataka	–	1	2	4	8	5	8	4	3	35	104	12570
Kerala	–	1	2	3	4	3	5	2	1	21	59	7120
Madhya Pradesh	–	2	3	3	8	9	7	9	5	46	121	13256
Maharashtra	1	2	3	10	5	8	10	4	2	45	153	20296
Meghalaya	–	–	–	1	–	–	1	1	1	4	9	751

(Continued)

Table C – *(Contd.)*

State/Union Territory	No. of Sample Towns/Cities									Total No. of Towns Selected	No. of Sample Blocks	No. of House-holds Listed
	Over 50 Lakhs	10–50 Lakhs	5–10 Lakhs	2–5 Lakhs	1–2 Lakhs	50,000–1 Lakh	20,000–50,000	10,000–20,000	Below 10,000			
Orissa	–	–	–	3	2	3	3	3	1	15	35	4094
Punjab	–	1	2	1	4	5	3	2	1	19	53	5390
Rajasthan	–	1	2	3	4	6	7	4	1	28	73	6824
Tamil Nadu	1	2	2	5	8	13	7	4	2	44	145	17579
Uttar Pradesh	–	3	5	10	13	14	13	12	10	80	213	27540
West Bengal	1	–	1	4	10	5	5	2	2	30	104	12628
Chandigarh	–	–	1	–	–	–	–	–	–	1	4	363
Delhi	1	–	–	–	–	–	–	–	–	1	30	483
Pondicherry	–	–	–	1	–	–	–	–	–	1	3	4363
TOTAL	**4**	**19**	**30**	**65**	**98**	**102**	**97**	**62**	**38**	**515**	**1509**	**182599**

However, despite the large sample size, the high income group could not be extended further due to fewer observations. Although a few sample households even reported an annual income of over Rs. 1,000,000, the distribution for income ranges above Rs. 86,000 was erratic due to small sample sizes. This is probably because

— the sample selected is in clusters; and
— the rich households tend to live in clusters.

The distribution of sample observations by income classes, for the year 1993-94, is given in Table D.

Data presented in Table D show that over four-fifths of the households listed reported annual incomes of less than Rs. 50,000, while barely 2.6% of the households had annual incomes in excess of Rs. 86,000. Consequently, the frequencies in the higher ranges of income are both small and exhibit no particular pattern or trend. The number of households reporting more than Rs. 20,000 as monthly income drops sharply, with their combined frequencies being around 200 households only.

By and large, there are greater proportion of the more affluent in urban than in rural areas, with the exception of the highest income group of over Rs. one million per year where the observed frequency is more in the rural.

For the products covered under MISH, disaggregation of income beyond Rs. 86,000 was not considered to be that important as most of these products, with the exceptions of two-wheelers, are priced below Rs. 25,000, a little over three months' income in this group.

However, after the liberalization of the economy, this group has assumed special significance. Multinationals and some Indian manufacturers have started flooding the market with premium products, such as luxury cars, cellular phones, music systems, expensive watches, to name a few, which are likely to be consumed by the very rich. To understand the market for such products and determine its size, it is important to extend this income distribution beyond the range provided in MISH.

Table D: Distribution of Sample Observations by Income Groups

Income Group (Rs. '000)	*Rural*	*Urban*	*Total*
<50	92878	158841	251719
50-62	2156	8922	11078
62-70	1316	3753	5069
70-80	952	4167	5119
80-90	715	2793	3508
90-100	562	2228	2790
100-110	53	310	363
110-120	107	576	683
120-130	43	160	203
130-140	29	80	109
140-160	123	343	466
160-180	20	79	99
180-200	102	161	263
200-250	40	61	101
250-300	37	51	88
300-350	15	12	27
350-400	12	22	34
400-450	3	1	4
450-500	0	14	14
500-600	0	9	9
600-700	0	1	1
700-800	0	3	3
800-900	0	3	3
900-1000	0	5	5
>1000	6	4	10
TOTAL	**99169**	**182599**	**281768**

To derive such a distribution purely through sample surveys would mean increasing the sample several fold. Further, in a cluster sample, where intra-clusteral correlations are very high, the estimates would not be efficient, unless stratification at the block/ village level is introduced. Lack of knowledge of all the blocks/ villages would be a limiting factor for such an exercise.

In view of this, it is proposed to use a theoretical model to estimate the distribution of high income households beyond the range available in MISH. The broad approach adopted is to approximate the existing observed distributions to a theoretical distribution and then extend the theoretical distribution for the desired income ranges.

Theoretical Model

One of the income distributions commonly used to explain the rich is through Pareto's Law, first stated by its author in 1897. According to Pareto's Law, if N(y) is the number of persons or households having an income over y, then

$$N(y) = b.y^{-a}$$

where a and b are parameters.

When this law is turned into a probability distribution, we get

$$F(y) = P(Y <= y) = 1 - \frac{N(y)}{N(y_0)}$$

where y_0 is the lowest income (exogenously fixed) beyond which Pareto's Law applies.

If we set $y^* = \frac{y}{y_0}$ where $y^* > 1$, then

$$F(y^*) = 1 - \frac{1}{(y^*)^a} = 1 - (y^*)^{-a}$$

The corresponding density function is given by

$$f(y^*) = \frac{a}{(y^*)^{(1+a)}} = a.(y^*)^{(-1-a)}$$

and the mean income in the original money units is

$$E(Y) = y_0 \cdot \frac{a}{a-1}$$

Estimated Income Distribution

The parameter '**a**' can be estimated from the observed frequencies generated through MISH.

Since the Pareto Distribution has only a single parameter, equating the expected mean income as given earlier with the observed mean from the MISH sample, it is possible to derive an estimate of '**a**'. This procedure would, however, fail to take into account the dispersion or disparity in income distribution. A better procedure would be to use minimum chi-square method to estimate '**a**'.

The value of y_0 is fixed at Rs. 30,000. The theoretical model is fitted for households with income over Rs. 30,000. The original data were reclassified into finer income groups in order to get a better estimate of '**a**'. In all, the households were distributed into 32 income classes. The upper tail, i.e., above Rs. 86,000 had 20 classes, the ranges for each varying between Rs. 4,000 and Rs. 100,000.

From this distribution the value of '**a**' was estimated through the method of minimum chi-square, which came to 1.94. This value of '**a**' was used to derive the theoretical frequencies of households in the different income groups. On comparing these estimated frequencies with the observed it was found that there were differences between them for the five income groups (low, lower middle, middle, upper middle and high). To make the data consistent, the following procedure was adopted. The number of households for these five income groups was fixed as given in MISH. Within each of these groups, the households were distributed using the relative frequencies as obtained in the model to the different income ranges comprising that group. The observed and estimated distributions of income are presented under Alternative I in Table I and depicted in Chart 1.

Adjusting for Differences with the National Income

The Central Statistical Organisation (CSO), Department of Statistics, Ministry of Planning and Programme Implementation, has estimated the personal disposable income for the year 1993-94 at Rs. 6,314 billion. According to the CSO, personal disposable income closely corresponds to household income. As against this the estimated household income from MISH was Rs. 4,700 billion. Thus, the survey estimate is about 25% less than the CSO estimate. This difference arises out of mainly two reasons:

1. Conceptual
2. Under-reporting of incomes by the households.

The income data collected in the survey do not include certain perks provided by the employers such as reimbursement for travel, medical, etc. They also do not include contributions to provident fund, pension fund by the employer, interest on provident fund, interest accrued but not realised, amount paid towards leave travel and home travel, payments in kind, imputed income of family labour, etc. All these are, however, included in the CSO estimate.

There is also an inherent under-statement of income in surveys, especially at higher levels of income. People tend to suppress their income in surveys particularly when questions relating to its disposition are not asked. There are also no built-in cross checks to detect any discrepancies.

An attempt is made here to work out a distribution of income after adjusting the survey data (relating to disposable income) to match with the CSO estimate.

There are no data available which could form the basis for apportioning the difference between the CSO estimate and the survey estimate between the two causes mentioned earlier. In the absence of this, it was decided to attribute 50 per cent of the difference to conceptual differences and the other half to under-reporting.

It is also felt that conceptual difference exist at all levels of income while under-reporting would pertain only to the relatively

affluent. Accordingly, the original survey distribution is modified as follows.

For the different income categories into which the survey data were classified, the total income was worked out based on the average income of the class and the number of households in that class. Fifty per cent of the difference between the CSO estimate and the survey estimate is added to the total income and then distributed between the different income classes in proportion to the income of that class. The other half of the difference was distributed in the same way but only for households which reported income over Rs. 40,000.

The total income thus attributed to the different classes was divided by the respective number of households to arrive at the new mean income for each class. Obviously this average was much higher than the original mean income reported in the survey and in many cases was beyond the upper limit of the class interval. This necessitated redefining the class intervals so as to make the new average lie in the middle of the redefined class interval. After redefining the class intervals, a new Pareto distribution was fitted to the data. For this alternative the parameter '**a**' was estimated at 1.53. The theoretical distribution was extended to provide estimated number of households at higher levels of income and this data is presented under Alternative II in Table I.

Inclusion of Unaccounted/Black Income

It is widely believed that there exists a black or parallel economy which generates a significant amount of income evading taxes and thereby not getting included in the national income calculations. Several articles have been written on this and numerous estimates have been provided. Some of these estimates have been in the nature of 'informal guesses'. These estimates also vary widely, from less than 10 per cent to over 50 per cent of GDP.

It may be worthwhile to study the impact on the income distribution if this unaccounted income is also taken into account. Since no reliable estimate of this component is available, as a hypothetical case, about 20 per cent of personal disposable income

Chart 1

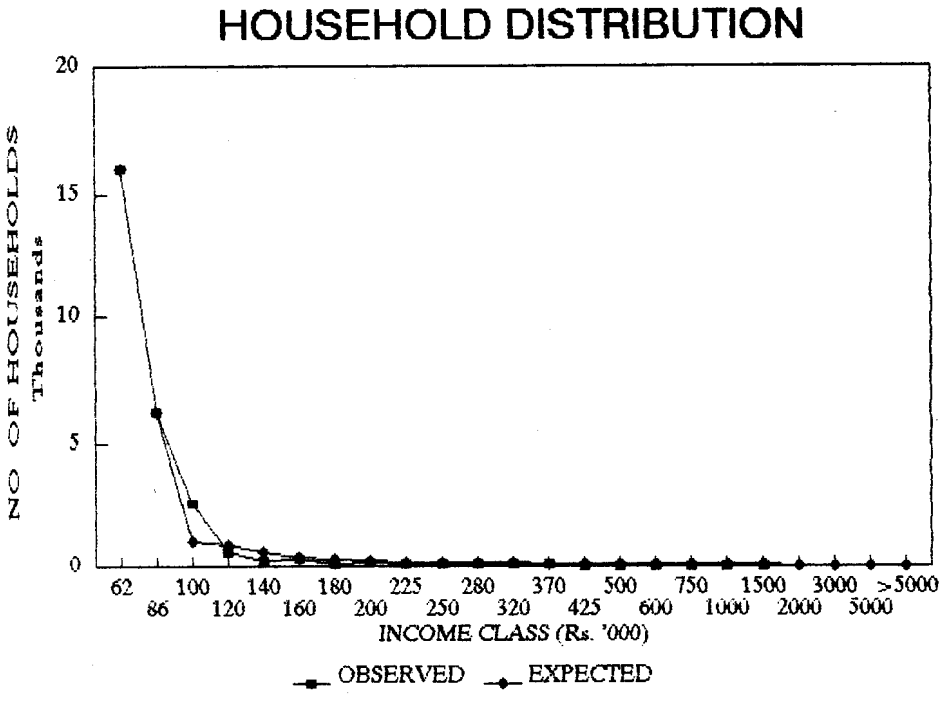
HOUSEHOLD DISTRIBUTION
NO OF HOUSEHOLDS
Thousands
20
15
10
5
0
62
86
100
120
140
160
180
200
225
250
280
320
370
425
500
600
750
1000
1500
2000
3000
5000
>5000
INCOME CLASS (Rs. '000)
OBSERVED
EXPECTED

Table I: Observed and Estimated Distributions of Households by Income Groups

(HHDS in '000)

Income Group (Rs. '000)	*Observed*	*ALT I*	*ALT II*	*ALT III*
<20	90539	90539	81485	81485
20-40	40683	40683	31536	31536
40-62	15962	15962	19865	19865
62-86	6240	6239	8190	6240
86-100	2558	992	2597	2557
100-120	565	868	5187	3785
120-140	163	528	3387	2578
140-160	242	345	1575	1857
160-180	41	238	577	1394
180-200	145	171	435	1081
200-225	28	154	410	1048
225-250	28	111	309	491
250-280	35	96	281	460
280-320	30	89	275	463
320-370	23	74	241	421
370-425	11	54	185	335
425-500	9	47	173	324
500-600	3	38	149	292
600-750	1	31	134	276
750-1000	7	24	117	257
1000-1500	5	18	98	234
1500-2000		6	41	106
2000-3000		5	34	96
3000-5000		2	21	68
>5000		1	18	74
TOTAL	**157318**	**157317**	**157322**	**157321**

Note:

ALT I: Based on Frequencies as Observed in MISH

ALT II: Adjusted for Differences Between Survey and National Income Estimates of Income

ALT III: Adjusted for Unaccounted/Black Income

is assumed to be generated through the parallel economy. It is also assumed that this income accrues only to households at the higher levels of income (above Rs. 62,000 per year). Thus, 20 per cent of personal disposable income is added to these groups of households and the income distribution adjusted for CSO estimates in proportion to the income earned by these groups. As in the earlier case, the new average was worked out for all the classes and the class intervals were redefined to make these averages fall in the middle of the range. As before, a new Pareto distribution was fitted after estimating '**a**' through minimum chi-square. The value of '**a**' was 1.29. The estimated theoretical distribution is presented under Alternative III of Table I.

Distribution for Rural and Urban

MISH data on income are collected for both rural and urban areas. Of the 280,000 households listed for the study about a third are from villages selected from every district in the country. The urban sample also is equally well-spread covering all the states and all classes of towns – from less than 10,000 population to over 5 million.

The income distributions observed from these samples for the two areas are approximated using Pareto's law and following the same procedure as adopted for the country as a whole. The parameter '**a**' was estimated separately for the two areas and the resultant theoretical distributions were estimated to include higher ranges of income.

Since the two distributions were derived independently, the estimated theoretical frequencies for different income classes did not add up to the corresponding totals estimated at the all-India level. The differences, though marginal, need to be reconciled in order to make the data set consistent.

The method used to achieve this is known as "Raking Ratio". This essentially is an iterative process and is used when marginal totals are available and the individual cells do not add up either column-wise, row-wise, or both. In this case the columns are rural

Table II: Estimated Distribution of Households Under Alternative I for Rural and Urban Before Adjusting for All-India Totals

(HHDS in '000)

Income Group (Rs. '000)	*Rural*	*Urban*	*U+R*	*All-India*
<20	74736	15804	90540	90540
20-40	26456	14228	40684	40683
40-62	8620	7344	15964	15962
62-86	2863	3377	6240	6239
86-100	463	506	969	992
100-120	387	463	850	868
120-140	224	296	520	528
140-160	141	201	342	345
160-180	94	144	238	238
180-200	65	107	172	171
200-225	57	99	156	154
225-250	40	74	113	111
250-280	34	66	99	96
280-320	30	63	93	89
320-370	24	54	78	74
370-425	17	41	58	54
425-500	14	37	51	47
500-600	11	32	42	38
600-750	8	28	36	31
750-1000	6	23	29	24
1000-1500	4	19	23	18
1500-2000	1	7	9	6
2000-3000	1	6	7	5
3000-5000	0	3	4	2
>5000	0	3	3	1
TOTAL	**114296**	**43024**	**157320**	**157318**

Table III: Estimated Distribution of Households Under Alternative I for Rural and Urban After Adjusting for All-India Totals

(HHDS in '000)

Income Group (Rs. '000)	*Observed*			*Estimated*		
	Rural	*Urban*	*Total*	*Rural*	*Urban*	*Total*
<20	74736	15804	90540	74736	15804	90540
20-40	26456	14228	40684	26456	14228	40684
40-62	8619	7344	15963	8619	7344	15963
62-86	2863	3377	6240	2863	3377	6240
86-100	1131	1428	2559	464	528	992
100-120	148	416	564	395	473	868
120-140	69	94	163	228	300	528
140-160	108	134	242	142	203	345
160-180	19	22	41	94	144	238
180-200	73	72	145	65	106	171
200-225	14	14	28	56	98	154
225-250	14	14	28	39	72	111
250-280	14	21	35	32	64	96
280-320	12	18	30	29	61	89
320-370	10	13	23	23	51	74
370-425	4	7	11	15	38	54
425-500	2	7	9	13	34	47
500-600	0	3	3	10	28	38
600-750	0	1	1	7	24	31
750-1000	0	7	7	5	19	24
1000-1500	4	1	5	3	15	18
1500-2000				1	5	6
2000-3000				1	4	5
3000-5000				0	2	2
>5000				0	1	1
TOTAL	**114296**	**43025**	**157321**	**114296**	**43024**	**157321**
>500	4	12	16	27	99	125
>1000	4	1	5	5	27	32
>2000	0	0	0	1	7	8
>5000	0	0	0	0	1	1

and urban areas, while the rows are the different income classes. The column totals are the number of households in rural and urban areas which are known, and the row totals are the estimated number of households in the different income classes at the all-India level.

Each row total is distributed to the cells in the row in proportion to the cell frequencies. With this the row totals will match. But the column totals may not match. To make the column totals match, the cell frequencies in each column are revised in proportion to the desired column totals. This would disturb the row totals. To set this right we again distribute the cell frequencies in each row in proportion to the row totals. This process is continued till the system converges and both row and column totals match the desired totals.

The estimated distribution for rural and urban areas before and after these adjustments are presented in Tables II and III.

Adjusting for Differences Between Survey and National Income Estimates

The national income estimate provided by CSO does not give the breakup between urban and rural areas. In the absence of that it is decided to use the proportions between the rural and urban as obtained in the survey to disaggregate the disposable income given by CSO.

It was assumed for the all-India estimates that 50 per cent of the difference between survey estimates & CSO estimates is attributed to conceptual differences while the other half is due to under-reporting.

Conceptual differences between CSO estimates and survey estimates exist both in rural and urban areas. In rural areas, the imputed value of family labour for capital formation is not covered in the survey while it is included in national accounts. Also, the value of home produced goods for self consumption is invariably not reported in the survey.

Under-reporting of income is also likely to be present both in rural and urban areas, as there is an inherent fear/uneasiness in the minds of people when answering questions on income.

However, it is not known how much of this difference at the all-India level is attributable to rural and urban areas individually. In the absence of any indicators, it is decided to apportion this difference between the two areas in proportion to their shares in the all-India income as reported in the survey. Having done so, the differences are distributed in the same way as was done for all-India for each of these two areas. After reclassifying the income ranges, the parameter **'a'** was estimated for both the sets by minimum chi-square method. The theoretical frequencies are then obtained for rural and urban areas separately for the different income classes. The totals for each class were matched with the all-India figure obtained for the corresponding income group. There were minor differences and these were set right using the raking ratio method. The distributions for rural and urban areas under this alternative are presented in Table IV.

Inclusion of Unaccounted/Black Income

The third alternative presented at the all-India level (Table I, Alternative III) was after providing for unaccounted/black income. At the all-India level it was assumed that 20 per cent of personal disposable income is earned additionally by households which went unreported. It is assumed that this unaccounted income, if any, is negligible in rural areas and most of it is generated only in the cities.

Accordingly, no addition to income was made for rural areas and the distribution obtained under Alternative II was retained for Alternative III also for rural areas. Since an exercise was already done at the all-India level after providing for black/unaccounted income, it was felt convenient to derive the urban distribution as a residual after deducting the rural distribution (Alternative II) from all-India Alternative III. The income distribution for rural and urban areas as estimated above is presented in Table V.

Table IV: Estimated Distribution of Households Under Alternative II for Rural and Urban After Adjusting for National Income

(HHDS in '000)

Income Group (Rs. '000)	*Estimated*		
	Rural	*Urban*	*Total*
<20	67262	14223	81485
20-40	21248	10289	31536
40-62	11919	7946	19865
62-86	4711	3479	8190
86-100	1453	1145	2597
100-120	2801	2386	5187
120-140	1957	1430	3387
140-160	1084	491	1575
160-180	329	248	577
180-200	245	190	435
200-225	228	182	410
225-250	170	140	309
250-280	152	129	281
280-320	146	128	275
320-370	127	115	241
370-425	96	90	185
425-500	87	85	173
500-600	74	75	149
600-750	64	69	134
750-1000	55	63	117
1000-1500	43	55	98
1500-2000	17	24	41
2000-3000	14	20	34
3000-5000	8	13	21
>5000	6	12	18
TOTAL	**114294**	**43027**	**157321**
>500	280	332	612
>1000	88	124	212
>2000	27	46	73
>5000	6	12	18

Table V: Estimated Distribution of Households Under Alternative III for Rural and Urban After Adjusting for Unaccounted Income

(HHDS in '000)

Income Group (Rs. '000)	*Estimated*		
	Rural	*Urban*	*Total*
<20	67262	14223	81485
20-40	21248	10288	31536
40-62	11919	7946	19865
62-86	4711	1529	6240
86-100	1453	1104	2557
100-120	2801	983	3785
120-140	1957	621	2578
140-160	1084	773	1857
160-180	329	1066	1394
180-200	245	836	1081
200-225	228	820	1048
225-250	170	322	491
250-280	152	307	460
280-320	146	316	463
320-370	127	294	421
370-425	96	239	335
425-500	87	237	324
500-600	74	218	292
600-750	64	211	276
750-1000	55	202	257
1000-1500	43	190	234
1500-2000	17	89	106
2000-3000	14	83	96
3000-5000	8	60	68
>5000	6	68	74
TOTAL	**114294**	**43027**	**157321**
>500	280	1121	1402
>1000	88	489	577
>2000	27	210	238
>5000	6	68	74

Effect of Change in Parameter 'a'

The parameter **'a'** is also a measure of inequality. A high value of **'a'** corresponds to a steeper decline of frequencies with rising incomes and thus in some sense to a greater degree of equality. This decline is also a function of income. Thus, the change in the frequency as one moves from one income range to another depends both on the value of **'a'** as well as the levels of income at the end point of these two ranges.

It may be recalled, to estimate **'a'** or to fit a Pareto curve to observed data, it is necessary to assume a cut-off point for income y_0, i.e., an income level below which no household exists. This cut-off point also affects the value of **'a'**. It is evident from the form of the curve that **'a'** increases with the cut-off point, as the estimated cumulative frequency for any given income y is given by

$$N(y) = \left[\frac{y_0}{y}\right]^a$$

where y_0 is the income limit and $y_0/y < 1$.

For the three alternatives presented earlier (Table I), an income limit of Rs. 30,000 was assumed. This necessitated leaving out all the households with income below Rs. 30,000 from the sample. To assess the impact of a change in the value of y_0 and therefore that of 'a', it was decided to try a few alternatives by increasing y_0 each time. For each alternative value of y_0, the observed income distribution had to be truncated suitably.

As expected an increase in y_0 led to an increase in the value of 'a'. At the same time, the estimated frequencies in the upper tail of the distribution came closer to the observed. The expected frequency distributions for five values of y_0, namely, Rs. 40,000, Rs. 62,000, Rs. 75,000, Rs. 80,000 and Rs. 86,000, corresponding to the beginning of the middle, upper middle and high income groups are presented in Table VI.

Table VI: Estimated Distribution of Households Under Alternative I Using Different Levels of Cut-off Points for Income

(HHDS in '000)

Income Group (Rs. '000)	*Observed*	*Cut-off Points* (Rs. '000)				
		40 *(2.28)*	*62* *(2.79)*	*75* *(3.67)*	*80* *(3.81)*	*86* *(3.88)*
<20	90539	90539	90539	90539	90539	90539
20-40	40683	40683	40683	40683	40683	40683
40-62	15962	15962	15962	15962	15962	15962
62-86	6240	6240	6240	6240	6240	6240
86-100	2558	1134	1340	1655	1702	1725
100-120	565	939	1021	1092	1098	1100
120-140	163	540	530	495	486	481
140-160	242	337	312	252	243	238
160-180	41	223	193	140	132	128
180-200	145	154	127	83	77	74
200-225	28	134	104	62	57	54
225-250	28	93	68	37	33	31
250-280	35	78	54	26	23	22
280-320	30	69	45	20	17	16
320-370	23	55	33	13	11	10
370-425	11	38	21	7	6	6
425-500	9	31	16	5	4	4
500-600	3	24	11	3	2	2
600-750	1	19	8	2	1	1
750-1000	7	13	5	1	1	1
1000-1500	5	9	3	0	0	0
1500-2000		3	1	0	0	0
2000-3000		2	0	0	0	0
3000-5000		1	0	0	0	0
>5000		0	0	0	0	0
TOTAL	**157318**	**157319**	**157318**	**157318**	**157318**	**157318**

Note:- Figures in parentheses represent value of Parameter **'a'**.

It is seen that for y_0 = Rs. 86,000, the expected frequencies closely match the observed. One would therefore be tempted to use this as the base for working out Alternatives II and III.

But it was already mentioned that the sample was selected from clusters, and since the clusters where the very rich households live are very few, the chances of these clusters getting included in the sample are low. Thus, it is quite likely that the observed frequencies of these households are understated.

Nevertheless, it is thought useful to use these higher income cut-off points to estimate the income distribution for the upper tail. Only Alternative III was tried using the truncated distribution, leaving out households belonging to low and lower middle income households from the sample at the first instance and then using only the high income households in the sample as an extension.

For both these sets, additions to income due to adjustments for consistency with national income as well as provision for unaccounted income were made. As a result of these additions, the cut-off point for income had to be raised to Rs. 75,000 and Rs. 150,000.

Using these cut-off points, the estimated income distributions are presented in Table VII. It is seen that frequencies in the upper tail reduce with the increase in y_0. The estimated number of households with an annual income of over Rs. 1,000,000 reduces from 577,000 at cut-off point of Rs. 30,000 to 495,000 when the cut-off point was Rs. 75,000. It reduced further to about 300,000 when the income ceiling is raised to Rs. 150,000. Similar reductions are found when we consider households with income more than Rs. 2 million and Rs. 5 million.

Thus, these three models provide a range for the number of households in the upper tail.

Table VII: Estimated Distribution of Households Under Alternative III Using Different Levels of Cut-off Points for Income

(HHDS in '000)

Income Group (Rs. '000)	*Cut-off Points* (Rs. '000)		
	30	*75*	*150*
<20	81485	81485	81485
20-40	31536	31536	31536
40-62	19865	19865	19865
62-86	6240	6099	6028
86-100	2557	2563	2470
100-120	3785	3804	3656
120-140	2578	2480	2490
140-160	1857	1689	1389
160-180	1394	1228	1672
180-200	1081	924	1225
200-225	1048	870	1122
225-250	491	655	821
250-280	460	611	650
280-320	463	653	619
320-370	421	573	525
370-425	335	440	388
425-500	324	409	348
500-600	292	352	287
600-750	276	316	245
750-1000	257	276	201
1000-1500	234	230	154
1500-2000	106	95	58
2000-3000	96	79	45
3000-5000	68	50	25
>5000	74	41	17
TOTAL	**157321**	**157321**	**157321**
>500	1402	1438	1032
>1000	577	495	298
>2000	238	170	86
>5000	74	41	17

Effect of Reallocation of Unaccounted/Black Income

Finally, under Alternative III, it was assumed that the unaccounted/black income is generated in the upper middle and high income groups. It would be interesting to see how a change in this assumption would affect the final estimated income distribution of the upper tail.

One more alternative is attempted by distributing this unaccounted income to the middle income group also. Accordingly, the observed distribution is recast and based on this revised distribution the value of parameter **'a'** is estimated. Based on this value of **'a'**, the new theoretical distribution is worked out and presented in Table VIII.

It may be noted that this alternative is in effect equal to taking a smaller amount of unaccounted/black income than the 20 per cent of the National income as assumed under Alternative III. The only difference between Alternative III and the present case is while the 20 per cent of National income is added only to the upper middle and high income groups in the former, it is also distributed to the middle income group in the latter.

But adding an additional amount to the middle income group is not going to affect the upper tail of the income distribution, as these households are not likely to reach a level of more than Rs. one million even after this additional amount. All that the inclusion of the middle income group achieve is to reduce the amount of unaccounted/black income added to upper middle and high income households. For example, if the share of middle income households in the total income earned by these three groups, middle, upper middle and high, is 50 per cent, then only 50 per cent of the unaccounted/black income is added to the upper middle and high income groups. In other words, instead of 20 per cent of the National income added as unaccounted/black money to these two groups under Alternative III, only 10 per cent would be added. Thus, the present case can be viewed as the one which assumes a smaller quantum of unaccounted/black money.

Table VIII: Estimated Distribution of Households Under Alternative III with Unaccounted Income Added to Middle Income Also

(HHDS in '000)

Income Group (Rs. '000)	*Number of Households*
<20	81485
20-40	31536
40-62	20033
62-86	6369
86-100	2891
100-120	2789
120-140	1874
140-160	2752
160-180	2045
180-200	1572
200-225	588
225-250	452
250-280	419
280-320	418
320-370	376
370-425	296
425-500	283
500-600	251
600-750	234
750-1000	213
1000-1500	189
1500-2000	83
2000-3000	74
3000-5000	50
>5000	50
TOTAL	**157320**
>500	1143
>1000	445
>2000	174
>5000	50

It is seen that as a result of this change the estimated number of households with income over Rs. one million has come down from 577,000 to 446,000. Similar reductions are observed at the higher income slabs also.

The estimated number of households in the upper tails under the different alternatives are presented for comparison in Table IX. Alternative I & II are not strictly comparable with the others as they are not adjusted to include unaccounted/black income. If these two are left out, then the estimated households with an income of over Rs. one million varies between 300,000 and about 577,000. A more realistic estimate of this group may perhaps be around 450,000.

Table IX: Distribution of Households in the Upper Tail Under Different Alternatives

(HHDS in '000)

Income Group (Rs. '000)	*ALT I*	*ALT II*	*ALT III*			
			A	*B*	*C*	*D*
>500	125	612	1402	1438	1032	1143
>1000	32	212	577	495	298	445
>2000	8	73	238	170	86	174
>5000	1	18	74	41	17	50

Note:-

"A" Corresponds to Cut-off Point of Rs. 30,000.
"B" Corresponds to Cut-off Point of Rs. 75,000.
"C" Corresponds to Cut-off Point of Rs. 150,000.
"D" With Provision for Unaccounted/Black Income in Income Groups Above Rs. 40,000.