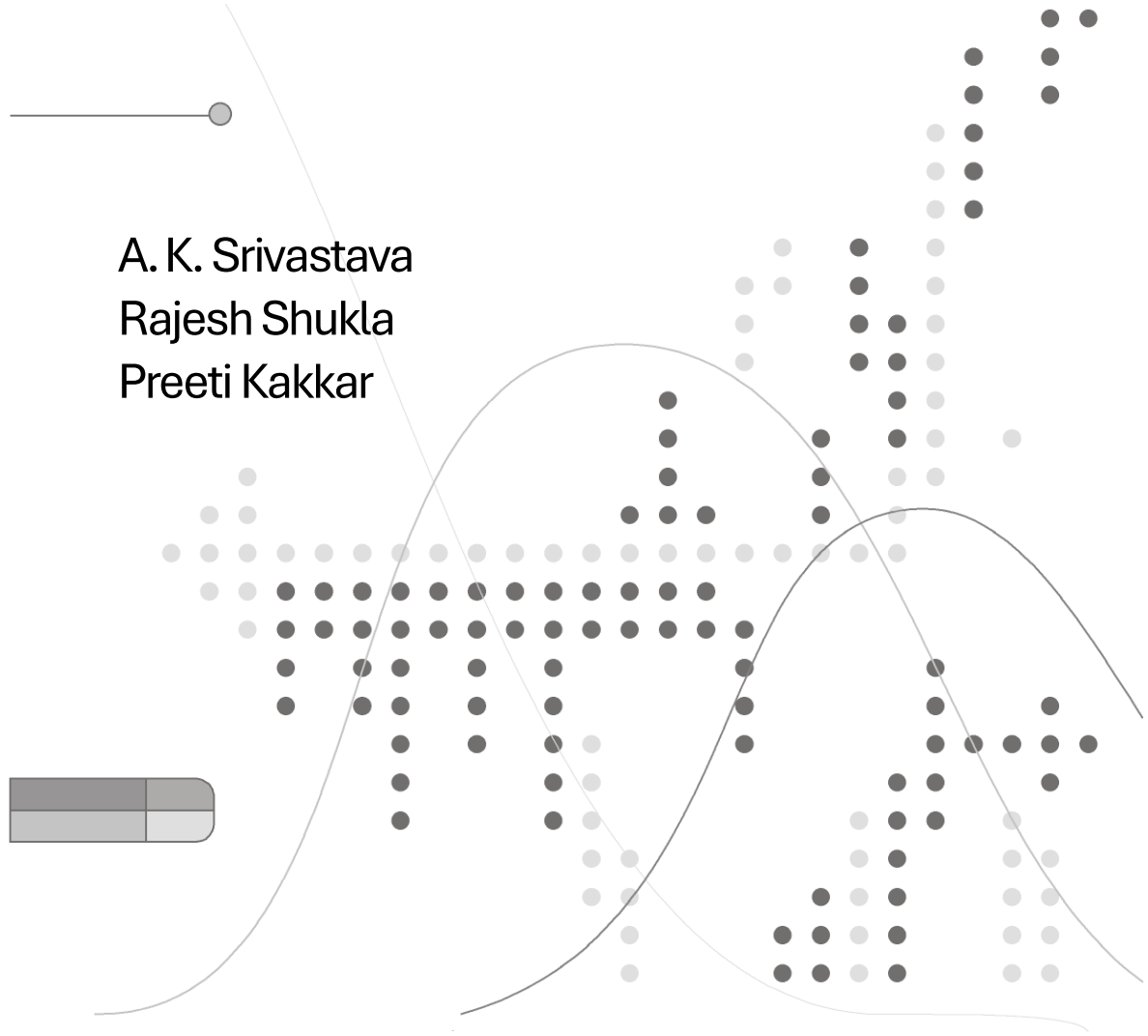
NCAER

NATIONAL COUNCIL OF APPLIED ECONOMIC RESEARCH

Technical Paper

National Survey on Household Income and Expenditure – A Qualitative Assessment

A. K. Srivastava
Rajesh Shukla
Preeti Kakkar



Contents

National Survey on Household Income and Expenditure – A Qualitative Assessment

A.K. Srivastava[1], Rajesh Shukla & Preti Kakkar

Abstract

The National Survey on Household Income and Expenditure (NSHIE) 2004-05 *provides a comprehensive evaluation of household income, expenditure, and savings in India, with a focus on the reliability and quality of data collection. A re-examination of the survey from a qualitative perspective confirms that the national-level estimates are broadly satisfactory, with sampling errors largely within 2 to 3 percent, reflecting the robustness of the sampling design and the adequacy of sample sizes. At the state level, however, higher sampling errors are observed, alongside indications of overestimation in some states and underestimation in others. While non-sampling errors, such as response and non-response biases, are inevitable in large-scale surveys, the adoption of multiple income-related questions has helped improve data quality. The study underscores the need for a structured monitoring mechanism to detect and address potential data inconsistencies during the survey process. Additionally, continuous in-depth analysis of survey data should be an integral part of NCAER's methodology to further enhance the reliability of household economic statistics for policy planning and socioeconomic research.*

Keywords: Household income, expenditure patterns, savings, income distribution, sampling design, estimation methodology, response bias, regional disparities, qualitative assessment, policy implications.

1. Introduction

Large-scale household surveys are important data sources for policy planning and decision making on various issues of national importance. With the changing social and economic scenario of the country, their demand is increasing. The National Council of Applied Economic Research (NCAER) has been conducting various large-scale household surveys on different issues like Human development Index (HDI), Micro Impacts of Macroeconomic and Adjustment Policies (MIMAP), Rural Employment and Demographic Survey (REDS) etc., since its establishment in 1956.

Economic analysts and policy makers identify three main purposes for compiling information on income distribution. The *first* is driven by a desire to understand how the pattern of income distribution can be related to patterns of economic activity and the returns to labour, capital and land, and to the way in which societies are organised. The *second* reflects the concern of policy makers to determine the need for both universal and socially targeted actions on different socio-economic groups and to assess their impact on household income. The *third* is an interest in how different patterns of income distribution influence household wellbeing and people's ability to acquire the goods and services they need to satisfy their requirements.

[1] Dr. A.K. Srivastava, Former Joint Director, IASRI, New Delhi; Dr. Rajesh Shukla, Chief Statistician, NCAER and Ms. Preeti Kakkar, Research Associate, NCAER.

Collecting data on income has always been considered as a difficult proposition due to its propensity for intentional or unintentional respondents' biases. Quite often, data on proxy-variables, which are correlated with the income data are collected and used to infer about income related issues. NCAER has been addressing issues concerned with income distribution through its various surveys. Improvements in the quality of data have been a prime consideration in these surveys.

Market Information Survey of Households (MISH) is an important large-scale household survey repeatedly conducted by NCAER since 1985/86 to estimate market size, penetration for a variety of consumer goods and most importantly to provide a profile of consuming households in terms of income, occupation and location. In India MISH has been the only source providing comparable household income data over time. It used the concept of *'perceived monetary income'* which included all income received by the household as a whole, and by each of its members, during the reference year by asking a single income question *"what is your annual household income from all sources?"* However, the adequacy of the single income question has been a major concern as it has been emphasised repeatedly that there is need for better survey data on household income, improvements in the quality of data and broadening the indicators by which relevant policy issues may be objectively addressed.

The household income coverage was significantly expanded in the 2005 survey renamed as ***National Survey on Household Income and Expenditure (NSHIE)*** to generate more robust and reliable estimates of household income.

In this study, it is proposed to investigate some qualitative aspects of the results from NSHIE, particularly for sampling and non-sampling errors perspective.

2. Objectives

The survey was aimed at capturing household income at the states-level covering both rural and urban India, which is needed to understand the existing regional disparities in terms of changes in the level of living, incidence of poverty, changing consumption patterns, trend in the saving propensities and preferences of different groups of population. Moreover, the information on per capita income of the household sector may also be of use in evaluating the validity of the National Income estimates obtained through conventional methods. All these income statistics can further guide policy makers in framing socio-economic developmental policies and in initiating financial measures for improving economic conditions of the people.

3. Coverage of the Study

Up to primary survey of households was undertaken in 24 major Stats/Union Territories of India covering both rural and urban areas of Andhra Pradesh, Assam, Bihar, Chandigarh, Chhattisgarh, Delhi, Goa, Gujarat, Haryana, Himachal Pradesh, Jharkhand, Karnataka, Kerala, Madhya Pradesh, Maharashtra, Meghalaya, Orissa, Pondicherry, Punjab, Rajasthan, Tamil Nadu, Uttaranchal, Uttar Pradesh, and West Bengal. Territories excluding Jammu & Kashmir, Sikkim,

Arunachal Pradesh, Nagaland, Manipur, Mizoram, Tripura, Andaman & Nicobar Islands, Daman & Diu, Dadra & Nagar Haveli and Lakshadweep. The remaining states were left out due to operational difficulty and accounts for only 3 to 4 per cent of the country's total population.

The target population of the survey was the total population in the country, with states and urban/rural categories as sub-populations or target groups.

4. Sampling Issues in Income Surveys

Review of literature[2] available on issues related to income surveys shows that while collecting income data in the field, problems may arise due to ambiguities in defining income concept, in choice of unit of sampling, sampling frame and reference period of data collection. In addition to these, seasonal effect, lack of availability of accounts for employer households, significant amount of purchases through credit, hidden income generated through wages paid in kind, etc., are other factors affecting reliable income data collection.

Primarily it is necessary to decide about appropriate statistical units and length of the accounting period for data collection otherwise it may lead to inaccurate income estimates. Generally, an accounting period of one year is recommended for income surveys. Defining income appropriately according to the objective is also an important aspect of the survey. The **Canberra City Group Report** has identified a total of 106 components covering the macro and micro concept of income to obtain reliable estimates. Sample design and sampling frames used in the survey are also required to be correctly selected as geographical spread and complete coverage of the statistical unit is important from the point of view of efficiency of estimates. Income distribution statistics will not be robust if some groups are over-represented and others under-represented. Some other issues are also there that discussed in detail in “Report of the Canberra City Group”.

5. Sampling Design

The survey methodology and sampling design adopted is, to some extent, like that used by the National Sample Survey Organization (NSSO) in its Consumption –Expenditure Surveys (CES). This is a household survey, and a notable feature of the survey design is that the sample of households was selected from a wide cross-section of households in the country, covering both rural and urban areas, with the objective of enhancing the precision of the estimates. A three-stage stratified sample design was adopted in which a ready-made frame was used at least for the first two stages, and a sampling frame, i.e., list of households, was developed in the last stage. Sample districts, villages and households formed the first, second and third stage sample units respectively for selection of the rural sample, while cities/towns, urban wards and households were the three stages of selection for the urban sample. Sampling was done independently within each state/UT and estimates were generated at state/UT level. Estimate for all-India was basically the aggregation of estimates for

[2] Report of the Canberra City Group of UN Statistical Commission, called 'Expert Group on Household Income Statistics'

all states/UTs. The rural sample for the survey was selected from a representative number of districts across the country, while the urban sample covered a range from big cities to small towns with populations below 5,000.

Within a state, due to variations in respect of social and economic characteristics, NSS regions are formed considering the homogeneity of crop pattern, vegetation, climate, physical features, rainfall pattern, etc. Thus, a NSS region is a group of districts within a State like each other in respect of agro-climatic features. In the present survey within a state, NSS regions formed the strata for both rural and urban sampling.

5.1. Selection of Rural Sample

In the rural sample design, a sample size of 250 districts was allocated to the 64 NSS regions within the 24 covered States/UTs in proportion to the total number of districts in an NSS region. From each of the NSS regions, the allocated number of districts was selected, as the first-stage sample units, with probability proportional to size and replacement, where rural population of each district as per 2001 Population Census was used as size measure.

Villages formed the second stage of selection procedure. District-wise lists of villages are available from census records (Census 2001) along with population. A total sample of 1,976 villages (second-stage sampling units) was allocated to the selected 250 districts approximately in proportion to rural population of each selected district. The allocated number of sample villages in a selected district was chosen with equal probability sampling approach.

In each of the selected villages, approximately 100 households were selected following equal probability sampling approach for listing purpose and preliminary survey. In preliminary survey, information on land possessed and principal source of income of the listed household was collected for use in stratifying the listed households into 8 strata as follows.

Table 1: Stratification of Rural Households

Classification	Household characteristics	
	Major source of household income	Size of land holding
Stratum 1	Self-employment in agriculture	0 – 2 acres
Stratum 2	Self-employment in agriculture	2 – 10 acres
Stratum 3	Self-employment in agriculture	More than 10 acres
Stratum 4	Labour (agricultural/other casual)	-
Stratum 5	Self-employment in non-agriculture	0 – 2 acres
Stratum 6	Self-employment in non-agriculture	More than 2acres
Stratum 7	Regular salary/wages	0 – 2 acres
Stratum 8	Regular salary/wages	More than 2 acres

This stratification was at second stage of selection and may be called second stage stratification. From each of the above 8 strata, 2 households were selected by following equal probability sampling approach. In case, any of the strata was found to be missing (no household), then households from previous stratum, where additional households were available, were selected so as to get 16 sample households in a selected village.

Sample Size: Following the above sampling design in rural areas, the realized sample of 31,446 households out of preliminary listed sample of 2,11,979 households was spread over 1976 villages in 250 districts and 64 NSS regions covering 24 States/UTs.

5.2. Selection of Urban Sample

According to the 2001 census, there are about 4,850 cities/towns in the states/UTs (excluding Jammu & Kashmir). The population of cities/towns in India varies from less than 5,000 to over a crore. In the urban sample design, within the 24 covered States/UTs, the 64 NSS regions were again treated as the main strata. In each NSS region, towns were categorized into nine groups based on their population, namely big towns and small towns. Of these groups, the first five groups (I to V) consist of 170 cities with a population exceeding 2 lakhs. All these big cities were selected with a probability of one. The remaining four town groups were taken as four separate strata on the basis of their population size and from each stratum a sample of towns was selected independently. These town groups may be called sub-strata which are formed within the main strata of NSS regions in each State.

A progressively increasing sampling fraction with increasing town population class was used for determining the number of towns to be selected from each sub-stratum. From each NSS region, the allocated number of small towns was selected by following an equal probability sampling procedure.

A total sample size of 2,255 urban wards was allocated among the selected small/big towns more or less in proportion to the number of wards in the respective towns, maintaining an equal number of wards allocated to each selected town in a town group. The number of towns and wards in the respective town group is given in Table 2.

The allocated number of wards was selected from each sample town, following equal probability sampling approach. Thus, towns and wards formed the first and second-stage sample units in the urban sample design.

Like in the rural sample design, within a selected ward, a sample of about 100 households was selected for listing and preliminary survey, following equal probability sampling.

Table 2: Sample sizes of towns and wards in the city/town groups

Town class	Town population ('000)	Total towns	Sample towns	Sampling fraction	Wards allocated to selected towns
I	> 10000	3	3	1.00	60
II	5000-10000	3	3	1.00	40
III	1000-5000	29	29	1.00	15
IV	500-1000	37	37	1.00	10
V	200-500	98	98	1.00	6
VI	100-200	219	56	0.26	4
VII	50-100	396	44	0.11	3
VIII	20-50	1,135	28	0.02	3
IX	< 20	2,270	44	0.02	2
Total		**4,190**	**342**	**0.08**	**2,221**

In the preliminary survey, at the time of the listing of the sample households, information on household size, household consumption expenditure for last month (MPCE), and principal source of household income were collected for use in stratifying the listed households into 7 strata as follows:

Table 3: Stratification of Urban Households

Classification	Household characteristics	
	Major source of household income	MPCE
Stratum 1	Regular salary/wages	Less than Rs. 800
Stratum 2	Regular salary/wages	Rs. 801-2500
Stratum 3	Regular salary/wages	More than Rs. 2500
Stratum 4	Self-employment	Less than Rs. 800
Stratum 5	Self-employment	Rs. 801-2500
Stratum 6	Self-employment	More than Rs. 2500
Stratum 7	Labour (agricultural/other casual)	-

From each of the above strata, 2 households were selected at random with equal probability of selection. If there was no household in any of the strata, the shortfall was compensated from the previous stratum, where additional households were available, to get 14 sample households from each selected ward in urban sector for detailed survey.

Sample Size: Following the above sampling design in urban areas, the realized sample of 31,570 households, out of preliminary listed sample of 2,38,813 households, was spread over 2,255 urban wards in 342 towns and 64 NSS regions covering 24 States/UTs.

6. *Estimation*

Based on the sampling design, estimates for Household Income, Expenditure and Savings as well as their distribution along the occupation, age-group, education and land holding size categories etc. were worked out at the national level for Rural and Urban households and were reported in the publication "*How India Earns, Spends and Saves" – A publication by NCAER (2007)*.

To assess the quality of results a deeper analysis of data for the sampling and non-sampling errors is necessary. Some aspects of these errors get smoothened out when one deals with the results at the higher level. In this study we try to work out estimates at State level and develop the estimates for sampling errors for various characteristics. In the process, we also try to look into the non-sampling errors as observed in some of the States.

In the data from NSHIE, the household level information for various characteristics (say income, expenditure, household size etc.) along with the weights to be attached for individual selected households is available. The sampling design is three stage sampling and the final weight for the household is the multiple of weights at the three stages. In case of probability proportional to size sampling (ppswr) followed at the first stage of selection in rural areas, the weights are np_i where n is the sample size of First Stage Units (FSUs) and p_i is the initial probability of selection for the i^{th} unit. At the stages where equal probability is used, the weights are the inverse of sampling fractions at the respective stage.

The broad approach in the estimation of various parameters has been to obtain a weighted estimate of the characteristic at the desired level. One of the limitations observed in the present analysis was that weights for the household level were available, but for working out the estimates of variances, estimates of totals for FSU and even at SSU level were needed. For this purpose, weights at different levels were worked out. Estimation procedure for the sampling design in NSHIE is given in ***Annexure-1***.

7. Results

7.1 Estimates at National Level

Similar to the results as presented in the report - *"How India Earns, Spends and Saves",* we present here some of the national level estimates for Income Expenditure and Savings for rural and urban areas, as also for different categories of households, such as occupation groups; age and education level of chief earner, land holding size classes etc. The estimates of sampling errors measured by percent standard errors (also known as percent CVs) of various estimates are also presented in **Table 4 to 9C**.

Table 4: Estimates of Per Household Income, Expenditure and Savings – All India

Rural/Urban	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Rural	31,445	144.17	51,992 (1.79)	33,054 (1.53)	18,869 (3.01)
Urban	31,570	61.44	95,827 (0.38)	58,417 (0.20)	37,409 (0.29
All India	63,015	205.60	65,041 (1.14)	51,811 (0.96)	41,864 (4.12)

Note: figures in bracket are Per cent Standard Errors

Table 5: Estimates of Per Household Income, Expenditure and Savings *by Size of Land Holdings*

Land holding class	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Landless	9204	55.34	38273 (1.94)	26,891 (1.66)	11,346 (3.75)
Marginal	9743	43.98	46,312 (2.41)	30,818 (2.15)	15,495 (3.79)
Small	5652	19.40	58,484 (2.27)	36,351 (1.74)	22,132 (4.00)
Semi-medium	3178	14.22	73,702 (3.01)	44,410 (2.18)	29,292 (5.06)
Medium	2746	8.09	99,858 (3.79)	51,960 (5.42)	47,898 (4.58)
Large	922	3.14	109,082 (8.80)	52,499 (11.24)	56,583 (8.08)

Note: Figures in bracket are Per cent Standard Errors

Table 6A: Estimates of Per Household Income, Expenditure and Savings by *Major Source of Household Income* - **Rural**

Major source of household income	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Regular Salary/Wages	5,849	14.69	1,03,687 (2.28)	50,838 (1.88)	52,849 (3.32)
Labour	5,733	52.39	28,567 (1.59)	22,883 (1.40)	5,684 (5.32)
Self-Employed Agriculture	11,916	57.76	54,622 (2.22)	35,056 (1.70)	19,566 (4.08)
Self-Employed non-Agriculture	6,710	16.16	67,621 (2.67)	40,827 (2.45)	26,794 (3.93)

Note: Figures in bracket are Per cent Standard Errors

Table 6B: Estimates of Per Household Income, Expenditure and Savings by *Major Source of Household Income* - **Urban**

Major source of household income	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Regular Salary/Wages	11,902	23.24	1,14,551 (1.73)	67,598 (1.57)	46,953 (2.54)
Labour	4,663	14.44	39,626 (1.85)	32,599 (1.69)	7,027 (5.69)
Self-Employed Agriculture	892	1.67	91,218 (7.33)	58,605 (8.52)	32,614 (6.54)
Self-Employed non-Agriculture	12,133	18.92	118,419 (2.15)	66,881 (1.54)	51,539 (3.68)

Note: Figures in bracket are Per cent Standard Errors

Table 6C: Estimates of Per Household Income, Expenditure and Savings by *Major Source of Household Income* – **All India**

Major source of household income	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Regular Salary/Wages	17,751	37.93	1,10,344 (1.38)	61,108 (1.23)	49,235 (2.03)
Labour	10,396	66.83	95,020 (0.58)	54,880 (1.11)	40,141 (4.22)
Self-Employed Agriculture	12,808	59.44	30,957 (2.14)	24,983 (1.67)	5,974 (3.91)
Self-Employed non-Agriculture	18,843	35.08	55,653 (1.69)	35,719 (1.32)	19,934 (2.82)

Note: Figures in bracket are Per cent Standard Errors

Table 7A: Estimates of Per Household Income, Expenditure and Savings *by Age of CWE*– **Rural**

Age of CWE (Chief Wage Earner)	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Less than 25 years	1,708	8.53	39,685 (3.17)	26,334 (2.89)	13,335 (5.70)
25-30 years	7,879	37.80	44,510 (2.43)	29,445 (1.99)	15,055 (4.67)
35-45 years	11,001	52.50	49,247 (1.64)	31,983 (1.27)	17,267 (3.11)
45-55 years	6,952	29.98	60,295 (2.26)	37,189 (1.88)	23,106 (3.68)
More than 55 years	2,995	11.74	70,017 (3.27)	41,179 (2.99)	28,838 (4.87)

Note: Figures in bracket are Per cent Standard Errors

Table 7B: Estimates of Per Household Income, Expenditure and Savings by *Age of CWE*– **Urban**

Age of CWE (Chief Wage Earner)	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Less than 25 years	1,447	2.98	70,794 (5.66)	43,999 (3.96)	26,794 (10.36)
25-30 years	7,762	15.89	82,759 (2.64)	50,693 (1.99)	32,066 (4.69)
35-45 years	11,365	22.47	92,563 (1.74)	57,800 (1.57)	34,762 (2.90)
45-55 years	7,426	13.57	115,676 (2.08)	67,288 (1.77)	48,388 (3.15)
More than 55 years	2,813	5.26	107,524 (3.55)	66,938 (3.21)	40,586 (5.31)

Note: Figures in bracket are Per cent Standard Errors

Table 7C: Estimates of Per Household Income, Expenditure and Savings by *Age of CWE*– **All India**

Age of CWE (Chief Wage Earner)	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Less than 25 years	3,155	11.51	47,745 (2.92)	30,911 (2.34)	16,834 (5.43)
25-30 years	15,641	53.69	55,830 (1.79)	35,741 (1.43)	20,090 (3.31)
35-45 years	22,366	74.97	62,231 (1.20)	39,722 (0.99)	22,509 (2.14)
45-55 years	14,378	43.55	77,553 (1.55)	46,569 (1.31)	30,985 (2.43)
More than 55 years	5,808	17.00	81,616 (2.42)	49,145 (2.20)	32,471 (3.60)

Note: Figures in bracket are Per cent Standard Errors

Table 8A: Estimates of Per Household Income, Expenditure and Savings *by Education of CWE*– **Rural**

Education of CWE (Chief Wage Earner)	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Illiterate	6,481	37.45	35,455 (2.62)	25,543 (1.80)	9,913 (6.76)
Up to Primary	6,012	32.51	41,783 (2.02)	28,576 (1.58)	13,207 (4.42)
Up to Matric	11,985	52.46	52,640 (1.75)	34,046 1.44	18,594 (3.27)
Higher Secondary	3,416	11.61	75,526 (2.95)	44,019 (2.43)	31,508 (4.99)
Graduate and Above	3,498	9.96	114,730 (2.18)	57,528 (2.16)	57,202 (2.96)

Note: Figures in bracket are Per cent Standard Errors

Table 8B: Estimates of Per Household Income, Expenditure and Savings *by Education of CWE*– **Urban**

Education of CWE (Chief Wage Earner)	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Illiterate	757	4.87	49,416 (3.16)	36,684 (2.95)	12,733 (5.43)
Up to Primary	2,982	7.05	58,015 (3.21)	41,343 (2.70)	16,671 (6.49)
Up to Matric	10,917	22.33	76,948 (2.01)	49,689 (1.59)	27,259 (3.49)
Higher Secondary	5,150	9.25	97,648 (2.61)	60,643 (2.50)	37,006 (3.68)
Graduate and Above	10,264	17.78	145,521 (2.17)	80,588 (1.80)	64,933 (3.30)

Note: Figures in bracket are Per cent Standard Errors

Table 8C: Estimates of Per Household Income, Expenditure and Savings *by Education of CWE*– **All India**

Education of CWE (Chief Wage Earner)	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Illiterate	7,238	42.32	37,062 (4.90)	26,825 (4.39)	10,237 (5.85)
Up to Primary	8,994	39.56	44,677 (1.72)	30,852 (1.37)	13,825 (3.74)
Up to Matric	22,902	74.79	59,897 (1.33)	38,716 (1.08)	21,181 (2.42)
Higher Secondary	8,566	20.85	85,334 (1.96)	51,389 (1.75)	33,945 (3.13)
Graduate and Above	13,762	27.75	134,463 (1.65)	72,306 (1.43)	62,157 (2.41)

Note: Figures in bracket are Per cent Standard Errors

Table 9A: Estimates of Per Household Income, Expenditure and Savings *by Income Quintiles* – **Rural**

Per-capita income quintiles	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Q1 (Bottom 20%)	3955	25.76	17,551 (1.48)	19,408 (1.91)	negative
Q2	4602	27.20	26,834 (1.17)	24,197 (1.31)	2,637 (6.53)
Q3	5655	29.65	34,344 (1.05)	28,675 (1.30)	5,669 (3.99)
Q4	7244	30.02	50,385 (1.28)	36,266 (1.52)	14,199 (2.37)
Q5 (Top 20%)	9989	31.54	1,19,633 (1.52)	52,901 (1.82)	66,732 (2.12)

Note: Figures in bracket are Per cent Standard Errors

Table 9B: Estimates of Per Household Income, Expenditure and Savings *by Income Quintiles* – **Urban**

Per-capita income quintiles	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Q1 (Bottom 20%)	4,043	10.98	27,678 (2.02)	32,211 (2.17)	negative
Q2	5,308	11.63	46,750 (1.74)	40,605 (1.92)	6,144 (3.89)
Q3	6,674	12.30	65,736 (1.80)	49,223 (1.91)	16,513 (2.38)
Q4	6,588	12.59	97,165 (1.70)	63,964 (1.84)	33,201 (2.22)
Q5 (Top 20%)	8,957	13.94	215,743 (2.46)	97,005 (2.12)	118,738 (3.11)

Note: Figures in bracket are Per cent Standard Errors

Table 9C: Estimates of Per Household Income, Expenditure and Savings *by Income Quintiles* – **All India**

Per-capita income quintiles	Sample size	Estimated number of households (Million)	Per household income (Rs/annum)	Per household expenditure (Rs/annum)	Per household savings (Rs/annum)
Q1 (Bottom 20%)	7,998	36.74	19,041 (1.00)	21,187 (1.11)	negative
Q2	9,910	38.83	29,353 (0.94)	26,515 (1.07)	2,839 (2.51)
Q3	12,329	41.95	41,220 (1.23)	33,410 (1.34)	7,810 (1.81)
Q4	13,832	42.60	65,235 (1.05)	44,707 (1.30)	20,528 (1.84)
Q5 (Top 20%)	18,946	45.48	153,872 (1.88)	71,088 (1.64)	82,783 (2.46)

Note: Figures in bracket are Per cent Standard Errors

It may be observed that the estimates at national level for various characteristics are obtained quite precisely with standard errors mostly around 2 to 3 percent or less, but in any case, less than 5 percent. It is also observed that the estimates for expenditure are normally more precise than those for income, which is better than the estimates for savings. These results are on the expected lines.

Thus, the reliability of estimates for various characteristics at the national level is quite satisfactory with respect to sampling errors.

7.2 Consistency Checks Applied to the Data

The success of a large-scale survey depends largely upon proper training to the interviewers and supervisors in addition to an efficient sample design and well-designed survey schedules (questionnaires). Necessary precautions were taken at planning as well as the field work stage to maintain the quality of data. Although completed schedules were edited once at the field level, these were again later subjected to manual editing and coding at NCAER.

Five kinds of checks, namely range checks, checks against reference data, skip checks, logical checks and typographic checks were used. Cross-validation of estimates for some key parameters such as household size, sex ratio, distribution of households according to SC, ST, & others, religion, type of dwelling, etc., was done using the results from external sources such as Census 2001 and National Sample Survey of 2003. The survey's estimate of average household size (4.99 members) appeared to be consistent with the estimates obtained from NSS 61st round (4.89 members) and Census 2001 (5.37 members). A similar pattern is also observed in the case of the sex ratio – this study estimates the sex ratio at 927, against 950 by NSS and 933 by Census 2001. All the three data sources are also comparable on some other parameters, such as the distribution of households by socio-religious groups.

Gross income, as estimated by this current study, is found to be around 56 per cent of the personal disposable income provided by the National Accounts Statistics (NAS). Expenditure estimates are also comparable with the NSS 61st round (2004-05) estimates. While the NSS gives an annual monthly per capita expenditure (MPCE) of Rs 725, NSHIE estimates gives this figure at Rs 678. Reliability of estimates has also been checked by evaluation of sampling and non-sampling errors. Sampling errors in terms of standard error and coefficient error have already been discussed in the earlier section.

Although there is no satisfactory procedure for a precise measurement of non-sampling errors, they arise mainly from three sources. *One,* respondents refuse to cooperate and deny information; they supply partial information that may not be usable; or they deliberately provide false information. *Two,* the interviewers are also prone to have some preconceived notions whereby some biases creep into the schedules. *Three,* respondents may not remember all the relevant numbers sought by the interviewers. And this tends to considerably increase the margin

of error in the data collected. In the case of this survey, it was around 3 per cent and largely due to unanticipated reasons such as the psychology of the respondent.

7.3 Estimates at State Level

Estimates at the national level are obtained quite precisely. Besides having enough sample sizes, the national level estimates tend to smooth out many of the aberrations relating to qualitative aspects of the data collected. Some of these aspects are not captured by the sampling errors at the higher levels. If we look at the estimates at somewhat lower levels, some of the qualitative aspects of the data, such as under-estimation or over-estimation, may more evident.

As income, expenditure and savings are related characteristics in which income is the most important characteristic for NSHIE survey, we examine the results for qualitative aspects based on data for income estimates. The following table presents the State-wise household income estimates along with the CVs, as well as estimated MPCEs (from NSSO, 61st round) from respective States. Only nineteen States out of 24 are considered in this analysis as States/UTs of Chandigarh, Delhi, Goa and Pondicherry and Meghalaya are not considered due small sample sizes. MPCEs are not directly comparable to the household income estimates (due to a variety of reasons such as different data sources, errors associated with the estimates as well as varying household sizes in different States), but it provides an indirect face-value assessment of the State level estimates for household income. The CVs provide an idea about the confidence intervals of the income estimates. Abnormal values in the last column (the underlined ones) indicate towards some extent of over or under estimation.

The following points are observed:

i. The sampling errors of the income estimates for most of the States are less than 10 per cent. Only in Uttaranchal the RSE was 30.79%. A closer look at the data indicated that in Uttaranchal, there was a wide amount of variability in the income data.
ii. In some of the States there appeared to be an overestimation of income while in some of the States some underestimation was observed.
iii. The states of Himachal Pradesh and Karnataka are likely to have some over estimation in the income estimates. The income to MPCE ratio was abnormally higher in these two States.
iv. Andhra Pradesh appears to have some amount of underestimation.
v. Kerala also appears to have a smaller value of income to MPCE ratio, but it may be partially explained away due to smaller household sizes in the State. However, this is only an indication. Deeper analysis is needed to further explore whether the differences are due to data quality, or it is due to some assignable causes.

Table 10: State-wise Estimated Household Income -**Rural**

S.no	States	Sample size	Estimated households (Million)	Per household income (Rs/annum)	CVs (%)	MPCE (NSSO, 2004-05)	Ratio Income/ MPCE
1	Himachal Pradesh	512	1.17	96,995	9.89	798	121.5
2	Punjab	767	2.91	67,852	5.99	847	80.1
3	Uttaranchal	480	1.27	58,516	30.79	647	90.4
4	Haryana	752	2.65	76,429	9.09	863	88.6
5	Rajasthan	1,888	7.89	45,786	5.05	590	77.6
6	Uttar Pradesh	4,384	22.44	53,203	4.22	532	100.0
7	Bihar	3,136	14.00	39,378	3.03	417	94.4
8	Assam	1,072	4.15	56,281	11.13	543	103.6
9	West Bengal	1,968	11.89	49,364	4.34	562	87.8
10	Jharkhand	944	4.11	37,126	9.14	425	87.3
11	Orissa	1,376	7.15	35,884	8.00	399	89.9
12	Chhattisgarh	784	3.55	33,907	10.83	425	80
13	Madhya Pradesh	2,112	8.80	40,239	3.70	439	91.7
14	Gujarat	1,440	6.28	62,452	11.53	596	104.8
15	Maharashtra	2,512	11.64	63,766	3.34	567	112.5
16	Andhra Pradesh	2,560	13.35	43,302	8.47	585	74
17	Karnataka	1,648	6.91	82,110	11.23	508	161.6
18	Kerala	848	5.13	67,679	7.59	1013	66.8
19	Tamil Nadu	1,616	8.11	48,850	7.95	602	81.1
All India - Rural		**31,445**	**144.17**	**51,992**	**1.79**		

7.4 Qualitative Improvement due to the Introduction of Multiple Questions on Income

It may be noted that within selected villages/wards, a fixed number (approximately 100) households were selected for listing purposes. Besides questions relating to second stage stratification, a single question regarding annual household income from all sources was asked to each of the listed households. The sample size of listed households in rural areas was 211,979 and in urban areas it was 238,813. After the second stage stratification, sub-samples of households were selected for detailed questioning. Income from different sources including regular salary/wages, self-employment in non-agriculture, wages (agricultural labour), wages (casual labour), self-employment in agriculture and other sources were collected. Pooled over all these sources we get the total annual income of the household from all sources. A total sample size of 31,446 households from rural areas and 31,570 households from urban areas were canvassed for detailed questioning. For these households, we have got responses for total

income from single as well as multiple questions. These responses have been used for the following two purposes:

1) Testing the hypothesis H_0: $\mu_1 = \mu_2$; against H_1: $\mu_1 \neq \mu_2$, where μ_1 and μ_2 are population means for single question and multiple question universes.
2) Estimating the difference of the estimates due to single vs. multiple question responses.

For testing the hypothesis pair wise t-test has been used. For estimating the differences, estimates have been obtained from the same sample based on the sampling design. Taking the multiple question responses as the standard ones the smaller values of responses from single question are taken as underestimation in the household income data.

It is seen that the under estimation due to single question is of the order of 27.5% in rural areas, 39.9% in urban areas and 34.5% at all India level. This figure, however, varies from State to State. In the pair wise test, the numbers of observations for different States are large enough to take the 5% significance level for t- values as 1.96. In most of the States, there is underestimation in single question response for household income. Thus, multiple questions have a definite impact on reducing the response errors towards the household income data.

8. Conclusions

A relook at the analysis of NSHIE 2004-05 data from qualitative perspective suggests that the estimates for various characteristics relating to income, expenditure and savings at all India level are satisfactory. In fact, the sampling errors are broadly around 2 to 3 percent. This reflects the proper choice of sampling design and the ample sizes. The sampling errors have also been worked out for State level estimates. As expected, the errors are slightly higher in the case of state level estimates. In a survey at such a large scale the non-sampling errors in the form of response and non-response errors are inevitable. At best, it can only be reduced to a reasonable level. For income data, precautions have been taken to improve the quality, through multiple questions. While working on the State level estimates, in some of the States, there were indications of over estimation while in some other States some under estimation was observed.

It would be appropriate to follow some monitoring mechanism in the course of the survey to get indications of such likely aberrations in specific areas, so that corrective measures may be taken in time. Also, a deeper analysis of data collected in the surveys conducted by NCAER should be a regular activity in the Organization.

ANNEXTURE 1: ESTIMATION PROCEDURE FOR NSHIE 2004-05

Define the following notations:

s = subscript for s^{th} stratum (NSS region), s=1 to 64
i = subscript for i^{th} FSU (district / city or town)
j = subscript for j^{th} SSU (villages (panchayat ward/ block) in i^{th} district
k = subscript for k^{th} second stage stratum in an SSU, k=1 to 8 for rural and k= 1 to 7 for urban
l = subscript for l^{th} sample household under a particular second stage stratum within a village or block

The following notations are defined for a specific stratum

d = number of sample districts in the stratum (districts are FSUs in the sampling design for rural areas)
N = total number of towns in the stratum (Cities/towns are FSUs in urban area)
n = number of selected towns
U = total size of the stratum (NSS region population)
u = total size of a specific sample district/ city or town
M_i = total number of villages/blocks (SSUs) in the i^{th} FSU
m_i = number of SSUs selected in the i^{th} FSU
K= Number of second stage strata
H_{ij} = total number of households in the $(ij)^{th}$ SSU
H_{ij}' = number of households selected in the $(ij)^{th}$ SSU for listing
H_{ijk} = number of households in $(ijk)^{th}$ second stage stratum after listing
H_{ijk}^{*} = estimated number of households in $(ijk)^{th}$ second stage stratum of households

$$= H_{ijk} \times \frac{H_{ij}}{H_{ij}'}$$

h_{ijk} = number of households selected in the $(ijk)^{th}$ second stage stratum
x, y = observed value of characteristics x, y under estimation
$\hat{X}, \hat{Y}$ = estimate of population total X, Y for the characteristics x, y
Y_{sijkl} = observed value of the characteristic y for the l^{th} household in the k^{th} second stage stratum of j^{th} SSU (village/block) within i^{th} district of s^{th} stratum

Under the above symbols, the estimation procedures are given for a specific stratum for both rural as well as urban areas:

A. Rural - An estimator of population total Y is given by

$$\hat{Y} = \frac{1}{d}\sum_{i=1}^{d}\frac{1}{p_i}\frac{M_i}{m_i}\sum_{j=1}^{m_i}\sum_{k=1}^{K}\hat{Y}_{ijk} \quad \text{...(1)}$$

where, $\hat{Y}_{ijk} = \frac{H^*_{ijk}}{h_{ijk}}\sum_{l=1}^{h_{ijk}} Y_{ijkl}$ ...(2)

and $H^*_{ijk} = H_{ijk} \times \frac{H_{ij}}{H'_{ij}}$

It may be noted that equation (1) may be written as

$$\hat{Y} = \frac{1}{d}\sum_{i=1}^{d}\frac{\hat{Y}_i}{p_i} \quad \text{...(3)}$$

where, $\hat{Y}_i = \frac{M_i}{m_i}\sum_{j=1}^{m_i}\hat{Y}_{ij}$ ...(4)

$$\hat{Y}_{ij} = \sum_{k=1}^{K}\frac{H^*_{ijk}}{h_{ijk}}\sum_{l=1}^{h_{ijk}} Y_{ijkl} \quad \text{...(5)}$$

Thus, the weight attached to (ijkl)th household is

$$W_{ijkl} = \frac{1}{d \times p_i}\frac{M_i}{m_i}\frac{H^*_{ijk}}{h_{ijk}}$$

It may be noted that the three components of this weight are appearing due to three stages of selection. In fact, $W_{ijkl} = W_1 . W_2 . W_3$, where $W_1 = \frac{1}{d \times p_i}$, $W_2 = \frac{M_i}{m_i}$ and $W_3 = \frac{H^*_{ijk}}{h_{ijk}}$

B. Urban- For urban areas, two estimators for population total Y are presented.

i) Estimator-1: First we present the estimator which has been used in the analysis of NSHIE data. Within the main strata of NSS regions, town groups have been considered as sub-strata and allocated numbers of towns are selected within each sub-stratum. In the selected towns, specified numbers of wards are selected within each town. It may be noted that number of wards allocated to each selected town in a town group is same. Let M_i be the number of wards in ith town from which $m_i = m$ wards are selected.

Estimation of total and mean:

The total Y in a substratum is estimated as

$$\hat{Y}_{(1)} = \sum_{i=1}^{n}\sum_{j=1}^{m}\sum_{k=1}^{K}\sum_{l=1}^{h_{ijk}} W_{ijkl}\, y_{ijkl}$$

Here, $W_{ijkl} = \dfrac{H_s}{\sum_{i=1}^{n} m_i}\dfrac{H_{ijk}}{H'_{ij}h_{ijk}}$;

where, $H_s = \sum_{i=1}^{N}\sum_{j=1}^{Mi_i} H_{ij}$ = Total number of households in s[th] town group.

Thus, $$\hat{Y}_{(1)} = \frac{H_s}{\sum_{i=1}^{n} m_i}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\frac{1}{H'_{ij}}\sum_{k=1}^{K}\frac{H_{ijk}}{h_{ijk}}\sum_{l=1}^{h_{ijk}} y_{ijkl}$$

$$= \frac{H_s}{\sum_{i=1}^{n} m_i}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\hat{\bar{Y}}_{ij}$$

Where, $\hat{\bar{Y}}_{ij} = \dfrac{1}{H'_{ij}}\sum_{k=1}^{K}\dfrac{H_{ijk}}{h_{ijk}}\sum_{l=1}^{h_{ijk}} y_{ijkl}$ = Estimate of average y (income, say) in (ij)th Ward.

$$= \sum_{k=1}^{K}\sum_{l=1}^{h_{ijk}} W^{*}_{ijkl}\, y_{ijkl}$$

Here, $$W^{*}_{ijkl} = \frac{H_{ijk}}{H'_{ij}h_{ijk}} = W_{ijkl}\frac{\sum_{i=1}^{n} m_i}{H_s}$$

From the data, W_{ijkl} are available. For estimation of variances, $\hat{\bar{Y}}_{ij}$ will be required and W^{*}_{ijkl} may be worked out as above.

Since m_i = m (say) in a sub-stratum (city group),

$$\hat{Y}_{(1)} = \left[\frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m}\hat{\bar{Y}}_{ij}\right]H_s$$

The first component is a simple mean of the estimated averages type estimator, while H_s is the total number of households in the entire sub-stratum (Town group).

$$\hat{\bar{Y}}_{(1)} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{m}\sum_{j=1}^{m}\hat{\bar{Y}}_{ij} = \frac{1}{n}\sum_{i=1}^{n}\hat{\bar{Y}}_{i}$$

We observe that $\hat{Y}_s = \hat{\bar{Y}}_s H_s$. This estimator is biased and the bias vanishes if average income (if income is the characteristic) in a district is uncorrelated with the population in the district.

ii) Estimator-2: An unbiased estimator of the population total Y on the line of estimation procedure as developed for rural areas is given as

$$\hat{Y}_{(2)} = \sum_i \sum_j \sum_k \sum_l W'_{ijkl} Y_{ijkl}$$

where, $W'_{ijkl} = W'_1 \times W'_2 \times W'_3$

$W'_1 = 1$, for large towns

= $\frac{N}{n}$, for small towns

$W'_2 = \frac{M_i}{m_i}$, and

$W'_3 = \frac{H_{ijk}^*}{h_{ijk}}$

here also $H^*_{ijk} = H_{ijk} \times \frac{H_{ij}}{H_{ij}'}$

For working out H^*_{ijk}, the number of households in ijth ward i.e. H_{ij} should be known. In NSHIE data, this information in selected wards was not available. This was probably one of the reasons why this estimator was not considered and the first estimator i.e. $\hat{Y}_{(1)}$ was worked out. However in case H_{ij} is known, this estimator which is unbiased may be considered.

In fact,

$$\hat{Y}_{(2)} = W'_1 \sum_{i=1}^{n} \hat{Y}_i$$

$$= W'_1 \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} \hat{Y}_{ij}$$

$$= W'_1 \sum_{i=1}^{n} \frac{M_i}{m_i} \sum_{j=1}^{m_i} \sum_{k=1}^{K} \frac{H_{ijk}}{h_{ijk}}^* \sum_{l=1}^{h_{ijk}} Y_{ijkl}$$

C. Overall Estimate for Aggregates

The estimates have been developed for a specific stratum (say, sth). Overall estimates for a state will be obtained by aggregating the strata estimates. Thus, for the state level, $\hat{Y} = \sum_s \hat{Y}_s$.

Further, national level estimates will be obtained by aggregating the state level estimates over all the states.

The Standard Errors are obtained by

$$SE(\hat{Y}) = \sqrt{\hat{V}(\hat{Y})}$$

Percent Relative SE (RSE) is given by

$$RSE(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} x100$$

D. Estimates of Ratios

Let $\hat{X}$ and $\hat{Y}$ be the estimates of the aggregates X and Y for two characteristics x and y at the stratum level. Then the combined ratio estimates $\hat{R}$ of the ratio $R = \frac{Y}{X}$ will be obtained as $\hat{R} = \frac{\hat{Y}}{\hat{X}}$. Again the overall estimates for a state will be obtained by aggregating the strata estimates.

E. Estimates of Standard Errors

The estimated variances of the above estimates will be as follows:

i) Rural

As described in (3), for a stratum in rural areas, where PPSWR has been used for selection at first stage (districts) in rural areas,

$$\hat{Y} = \frac{1}{d}\sum_{i=1}^{d}\frac{\hat{Y}_i}{p_i}$$

Defining $Z_i = \frac{\hat{Y}_i}{p_i}$, $\hat{Y}$ is given as

$$\hat{Y} = \frac{1}{d}\sum_{i=1}^{d} Z_i = \overline{Z}$$

An estimator of variance for $\hat{Y}$ is given by

$$\hat{V}\left(\hat{Y}\right) = \frac{1}{d(d-1)}\sum_{i=1}^{d}(Z_i - \overline{Z})^2$$

ii) For aggregate at state level

$$\hat{Y} = \sum_s \hat{Y}_s$$

$$\hat{V}(\hat{Y}) = \sum_s V(\hat{Y}_s)$$

The Standard Errors are obtained by

$$SE(\hat{Y}) = \sqrt{\hat{V}(\hat{Y})}$$

Percent Relative SE (RSE) is given by

$$RSE(\hat{Y}) = \frac{SE(\hat{Y})}{\hat{Y}} x100$$

iii) Urban

(i) Estimator $\hat{Y}_{(1)}$

An estimator of variance for $\hat{Y}_{(1)}$ is approximately given by ignoring the contributions due to third stage of selection towards the sampling variance, an estimator for the variance of $\hat{\bar{Y}}$ is given as,

$$\hat{V}(\hat{\bar{Y}}_{(1)}) = \left(\frac{1}{n} - \frac{1}{N}\right) s_b^2 + \frac{1}{nN} \sum_{i=1}^{n} \left(\frac{1}{m} - \frac{1}{M_i}\right) s_i^2$$

Where,

$$s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\hat{\bar{Y}}_i - \hat{\bar{Y}})^2$$

$$s_i^2 = \frac{1}{m-1} \sum_{j=1}^{m} (\hat{\bar{Y}}_{ij} - \hat{\bar{Y}}_i)^2$$

(ii) Estimator $\hat{Y}_{(2)}$

An estimator of variance for $\hat{Y}_{(2)}$ is approximately given by

$$\hat{V}\left(\hat{Y}_{(2)}\right) = N^2 \left(\frac{1}{n} - \frac{1}{N}\right) \times {s_b}^2$$

$$where \qquad {s_b}^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(\hat{Y}_i - \frac{1}{n} \sum_{1=1}^{n} \hat{Y}_i\right)^2$$

Here the components for second and third stages have been ignored and to that extent, variances are slightly overestimation.

iv) For ratios $\hat{R}$

For strata with PPSWR selection at first stage:

$$\hat{MSE}(\hat{R}) = \frac{1}{\hat{X}^2}\left[\sum_s \frac{1}{d(d-1)}\sum_{i=1}^{d}\left\{\left(Z_i - \hat{R}Z_i^*\right)^2 - d\left(\bar{Z} - \hat{R}\bar{Z}^*\right)^2\right\}\right]$$

where $\hat{R}$ is the combined ratio estimate for the ratio $R = \frac{Y}{X}$

$$Z_i = \frac{\hat{Y}_i}{p_i}; \qquad Z^*{}_i = \frac{\hat{X}_i}{p_i};$$

and $$\bar{Z} = \frac{1}{d}\sum_{i=1}^{d}\hat{Z}_i; \qquad \bar{Z}^* = \frac{1}{d}\sum_{i=1}^{d}\hat{Z}_i^*$$

For strata with SRSWOR selection at first stage:

$$\hat{MSE}(\hat{R}) = \frac{1}{\hat{X}^2}N^2\left(\frac{1}{n} - \frac{1}{N}\right)\left(\frac{1}{n-1}\right)\left[\sum_{i=1}^{n}\left(\hat{Y}_i - \hat{R}\hat{X}_i\right)^2 - n\left(\sum_{i=1}^{n}\frac{\hat{Y}_i}{n} - \hat{R}\sum\frac{\hat{X}_i}{n}\right)^2\right]$$

v) Estimates of RSE

$$\hat{RSE}\left(\hat{Y}\right) = \frac{\sqrt{\hat{Var}\left(\hat{Y}\right)}}{\hat{Y}} \times 100$$

$$\hat{RSE}\left(\hat{R}\right) = \frac{\sqrt{\hat{MSE}\left(\hat{R}\right)}}{\hat{R}} \times 100$$