

Reliability of Our Statistical System

A Rejoinder to G Raveendran

SANJAY KUMAR, N K SHARMA

In this brief rejoinder to G Raveendran's article "Good and Bad Statistics" (*EPW*, 16 September 2023), the authors argue that there is no empirical or theoretical ground for treating rates and ratios from the National Sample Survey as more reliable, under the assumption of similar underestimation in the numerator as well as the denominator.

Concerns about the deterioration of the official statistical system have frequently been expressed, more so after the release of the 2011–12 gross domestic product (GDP) series and thereafter on the non-release of the Consumption Expenditure Survey (CES) 2017–18 data, delay in releasing the Periodic Labour Force Survey (PLFS) data in 2019, etc. The current round of debate started with Prime Minister's Economic Advisory Council (PM-EAC) member Shamika Ravi's article questioning the sample selection in the National Sample Survey (NSS). This was followed by a series of articles by Bibek Debroy, chairperson PM-EAC, Pronab Sen, former chairperson, National Statistical Commission (NSC) as well as former Chief Statistician of India (CSI), P C Mohanan and Amitabh Kundu (the former having been the acting chairperson, NSC in 2019), Raveendran, etc.

The objective of this article is to highlight the gravity of the problem of population underestimation in NSS, which, though rightly pointed out by Ravi (2023), got distracted in the articles by Sen (2023) and Raveendran (2023), from its real import by looking at the underestimation issue in isolation without considering the methodology actually followed by the NSS. Although Sen (2023) rightly referred to the definitional differences in the NSS

and the census in classifying areas as rural or urban, he clarified the issue only partially.¹ More importantly, whether this underestimation in the NSS estimates could be attributed entirely (or even largely) to the said definitional differences, appears to have been missed by Sen (2023) altogether.

Raveendran (2023), while seeking to support NSS estimates, in fact, ended up trivialising the issue of population underestimation by putting forward the argument that instead of aggregates, the NSS gives rates and ratios as its surveys are not designed to obtain absolute estimates of rural or urban population and that the rates and ratios remain invariant to changes in population totals since, in his view, any correction in the denominator (that is, the population total) is equally applicable to its numerator (that is, the population satisfying the given attribute). It is hence claimed that NSS rates/ratios are considered reliable.

On this aspect, Raveendran's support is rather *ipse dixit*, devoid of any theoretical or empirical underpinnings, for many reasons. For instance, first, NSS population estimates for total (that is, rural+urban) which is free from any classification issue, have been consistently lower than those from census or census-based projections (Table 1). Second, the extent of observed underestimation in NSS can be considered as a lower bound only, because the census population itself is undercounted by nearly 2.3% as per post-enumeration survey results for Census 2001 as well as Census 2011 (RGI 2014). Third, from a simple analysis, it is seen that the estimated number of households in various NSS rounds is generally closer to those emanating from the nearest census. Fourth, the household

Table 1: Population Underestimation in NSS vs Census Projections

Round/Segment	Rural			Urban			All India
	Male	Female	Total	Male	Female	Total	
50th (1993–94)	11.3	10.7	11.0	19.0	18.4	18.7	13.1
55th (1999–2000)	6.2	4.9	5.6	16.5	15.1	15.8	8.4
61st (2004–05)	8.1	6.3	7.2	21.9	19.8	20.9	11.1
66th (2009–10)	10.1	10.0	10.1	20.6	19.8	20.2	13.1
68th (2011–12)	8.1	7.4	7.7	18.1	18.8	18.4	11.1

Source: Derived by using interpolated RGI population projections as on 1 March of the relevant two years to the intervening 1 January (for 1 January 2000, 2001 figures extrapolated, 1 January 1994 population taken from Visaria [2002]) and the NSS population from the respective survey reports.

The views expressed are personal.

Sanjay Kumar (sanjay.kumar61@nic.in) retired as additional director general and N K Sharma (nareshkr.sharma@nic.in) retired as director general from the Ministry of Statistics and Programme Implementation, Government of India.

size in NSS, vis-à-vis census, is lower (Table 2) and the said definitional differences cannot account for the observed gap. It is germane to recognise that lower household size in NSS may arise from the selection of smaller households (a design issue) and/or omission of member(s) within households (a non-sampling error). Fifth, it may be easily verified from NSS microdata that the levels of population underestimation vary spatially across age–sex classes besides fluctuating temporally. These variations would naturally have repercussions on the reliability of the resulting rates/ratios.

Table 2: Underestimation in Household Size, NSS vs the Closest Census

Round/Segment	Rural	Urban	All India (%)
50th (1993–94) vs Census 1991	12.2	16.7	13.4
55th (1999–2000) vs Census 2001	6.3	12.5	7.7
66th (2009–10) vs Census 2011	8.1	14.5	9.8
68th (2011–12) vs Census 2011	9.3	15.9	11.2

Source: Derived by using the total rural/urban populations and the relevant numbers of households from the respective NSS round reports and the census results.

The issue of population underestimation in NSS is organically tethered to the issue of reliability of the rates/ratios,

each of which is an outcome of two internally generated figures. A cursory appreciation of issues in the previous paragraph leads to the inexorable inference that the NSS rates/ratios generated from such disparately varying components render them unusable for the same reasons as those for its aggregates. The assumption of uniform underestimation in the numerator as well as the denominator is not based on any empirical evidence. Rather, evidence to the contrary abounds. For instance, there can be no indicator broader than average household size but the denominator (that is, number of households) of this ratio is not underestimated to the extent of the numerator (that is, population) in NSS.

It seems that almost everyone is uncritically presuming the sampling errors to be the root cause of the problem. Statistically speaking, repeated samples from a universe should result as much in overestimation as in underestimation vis-à-vis the true value. Unless it is argued that census projections are on the higher side by quite

a large margin (which, as stressed above, may rather be underestimated), the absence of even a single occurrence of population overestimation in NSS, prima facie, defies the statistical odds. It is, therefore, likely that non-sampling errors are very large and unidirectional.

That non-sampling errors, perhaps growing over time, may indeed have impacted the NSS results may be hypothesised from the fact that in 1972–73, the consumption expenditure from the NSS was 0.18% higher in food and 16.7% lower in non-food (overall 5.47% lower) in comparison to the private final consumption coming from the national accounts data. In 1993–94, NSS consumption estimates were lower by 39.76% (food), 47.77% (non-food), and 43.66% (overall) and stood, in 2011, at 28.29% (food), 54.77% (non-food), and 45.24% (overall) lower than the national accounts data (MOSPI 2015). NSS consumption results for 2017–18 were not even released in view of “the data quality issues” (MOSPI 2019). Also, the employment–unemployment data from the NSS has been questioned at various fora.

EPWRF India Time Series (www.epwrfits.in)

Rural India Statistics

The following 16 subject modules of EPWRF India Time Series (EPWRF ITS) online database contain data series relating to the rural sector in India:

- Price Indices
- Consumption Expenditure Statistics
- Employment Statistics
- Labour Statistics
- Economic Census
- Agricultural Statistics
- Livestock Statistics
- Agricultural Wages in India
- Wage Rates in Rural India
- Annual Survey of Industries
- Banking Statistics
- Microfinance Statistics
- Health Statistics
- Educational Statistics
- Environment Statistics
- Telecom Statistics

The EPWRF ITS has 32 modules covering a range of macroeconomic, financial and social sector indicators on the Indian economy.

EPWRF India Time Series is an e-ShodhSindhu consortium approved online database.

For further details, visit www.epwrfits.in | For subscription details, write to us at its@epwrf.in

In fact, in view of the gravity of this issue, the government had constituted a committee, as noted in NSSO (2008) as follows:

Some of the observations made by the Bimal Roy Committee on “Underestimation of Population in NSS” may be referred to. It appears from the report of the Committee that the reasons for underestimation are mainly attributable to the sources of non-sampling errors, and are more likely to be different for different segments of the population and may vary from round to round. This makes it difficult to construct correction factor to adjust for variable degrees of underestimation for different segments of the population ... Moreover, the above observations on the magnitude and direction of divergence suggest strongly calculating separate adjustment factor for each class/group/category/domain.

However, despite a lapse of nearly 17 years of the said committee report, no such separate adjustment factors appear to have evolved so far. Hence, Raveendran’s (2023) simplistic defence of NSS rates/ratios is altogether misplaced. This aspect requires a more nuanced understanding and careful handling.

In truth, these issues require a holistic resolution. For minimising non-sampling errors and improving survey data quality in NSS, real-time validation checks may go a long way. Also, presently there is no system of independent post-survey checks in NSS, on the lines of the post-enumeration surveys adopted in the population census, where the simple exercise of mere “counting of heads” and involving no sampling is known to undercount the population. Obviously, implications of non-sampling errors on the estimates coming from canvassing a lengthy questionnaire, as in NSS, would be much more complex and multidimensional.

NSS is, no doubt, an old organisation in the field of survey execution. Age, however, in itself, may not necessarily be an indicator of continued relevance in the dynamic socio-econo-techno environment we are living in as it requires continual institutional adaptation and reinvention. It nonetheless needs to be appreciated that NSS has been transparent in sharing its microdata, regardless of the fact that it is this sharing which has resulted in more questions being raised by the users with regard to the NSS data. However, such a level of transparency

is not visible in the case of other macro indicators like the index of industrial production, consumer price index, national accounts aggregates, etc, though the relevant microdata may be easily shared even while following the government’s extant Guidelines for Statistical Data Dissemination (GSDD) 2019. It is hoped that this openness shall generate a well-informed debate on the quality issues with respect to these macro-indicators.

Most of the commentators, for the present, appear to be in agreement that the official statistical system is collapsing. Some others may even aver that it has already collapsed. However, what is not in dispute is that to remain relevant and useful, the system needs a thorough overhaul and significant upgradation. The solution for this, however, does not lie in setting up more expert committees and/or according statutory status to NSC since the malaise is much more deep-rooted than perhaps being fully realised. With the authority, albeit limited, at its disposal, what improvements the setting up of NSC has brought about during the last 18 years, remain to be assessed.

NOTE

- 1 The population projections incorporate, by way of urban–rural growth differential, the likely urbanisation since the last census which, among others, includes the emergence of new

census towns during the inter-central period. The same, however, would not be so classified in NSS until these are notified just before the next census. In fact, if an area is notified as urban but no urban frame survey has since been conducted there, it would continue to be classified as rural in NSS.

REFERENCES

- Debroy, Bibek (2023): “The Janus of India’s Official Statistics,” *Indian Express*, 10 July.
- Mohanani, P C and A Kundu (2023): “Narratives in Search of Data,” *Indian Express*, 12 July.
- MoSPI (2015): “Report of the Committee on Private Final Consumption Expenditure,” Ministry of Statistics and Programme Implementation, Government of India, Table 1, p 11.
- (2019): Press Note on “Household Consumption Expenditure Survey,” Ministry of Statistics and Programme Implementation, Government of India, viewed on 25 September 2023, <https://pib.gov.in/PressReleasePage.aspx?PRID=1591792>.
- NSSO (2008): “Review of Concepts and Measurement Techniques in Employment and Unemployment Surveys of NSSO,” NSSO (SDRD) occasional paper/1/2008, viewed on 25 September, https://www.mospi.gov.in/sites/default/files/workshop/nsc_paper_SDRD_12jan09.pdf.
- Raveendran, G (2023): “Good and Bad Statistics,” *Economic & Political Weekly*, Vol 58, No 37, pp 68–71.
- Ravi, Shamika (2023): “The Sample Is Wrong,” *Indian Express*, 7 July.
- RGI (2014): “Census of India 2011: Report on Post Enumeration Survey,” Registrar General and Census Commissioner, Government of India, p iii.
- Sen, Pronab (2023): “Statisticians Are Not Stupid,” *Indian Express*, 10 July.
- Visaria, P (2002): “Workforce and Employment in India, 1961–94,” *National Income Accounts and Data Systems*, B S Minhas (ed), New Delhi: Oxford University Press, Table 1.4, p 13.

EPWRF India Time Series (www.epwrfits.in)

Agricultural Wages in India

The **EPW Research Foundation** has added a module on Agricultural Wages in India to its online database, EPWRF India Time Series (EPWRFITS).

This module provides month-wise data on Agricultural wages for:

- Agriculture Operations
 - Ploughing
 - Sowing
 - Weeding
 - Reaping
 - Harvesting
- Rural Skilled Labour:
 - Carpenter
- ❖ Contains data for 21 major states for men and women
- ❖ Presents quarterly and annual series (calendar year, financial year and agricultural year), derived as averages of the monthly data

With this addition, the EPWRFITS has 32 modules covering a wide range of macroeconomic, financial and social sector indicators of the Indian economy.

For subscription details, visit www.epwrfits.in or write to us at its@epwrf.in

EPWRF India Time Series is an e-ShodhSindhu consortium approved online database.