Technical clarification regarding high multiplier value in PLFS 2022-23

Subsequent to the release of PLFS unit level data for the survey period July 2022 – June 2023 some data users have indicated about a high multiplier value present in the unit level data for the State Assam.

The clarification regarding the mentioned high multiplier value is given below

1. **Outline of the PLFS Sampling Design in place upto 2023-24**

A stratified multi-stage sampling design considering NSS regions (comprising of a number of districts) in a State/UT as the basic stratum or primary sampling units or first stage units (FSUs) for rural and urban areas has been used in the Periodic Labour Force Survey (PLFS) 2022-23 and 2023-24.

For urban areas, the FSUs are Urban Frame Survey (UFS) blocks, while for rural areas, they are villages based on the 2011 Population Census (or Panchayat wards in Kerala). The ultimate sampling units (USUs) are households. The selection of FSUs and USUs have been done using Probability Proportion to Size (PPS) and Simple Random Sampling Without Replacement (SRSWOR) method, respectively.

With a view to ensure completeness of the coverage in the survey, a provision of selection of some sample villages from the universe of uninhabited villages (according to 2011 population census) has been kept in the sampling design of PLFS. As a convention, the population of each of these uninhabited villages is considered as '*one'* for the purpose of selection in the sample by PPS method.

2. **Procedure for Selection of Survey Units:**

In the PPS method, units from a population or universe are selected with probability proportional to a size measure and consequently, units with larger sizes have a higher probability of being selected in the sample. The selection probability, P (selection), for each unit is given by the formula:

$$\text{P (selection)} = \frac{z_i}{\sum_i z_i}$$

where z_i is the size of the i-th FSU and $\sum_i z_i$ is the total size (Census Population in rural sector) of the NSS region.

3. **Sampling Weight or Multiplier:**

Sampling weight is a multiplicative factor assigned to each selected unit in the sample. It represents the number of similar, unselected units (such as villages or UFS blocks) within the same stratum. The sampling weight, an inverse of inclusion probability, ensures that the sample accurately reflects the entire population, and the sample results can be generalized to represent the entire population. This weight depends inversely on selection probability.

$$\text{Sampling wight} \propto \frac{1}{\text{P (Selection)}} = \frac{\sum_i z_i}{z_i}$$

It is evident from the expressions of the selection probability and sampling weight that for selected units with very small size (i.e., z_i) in comparison to the overall stratum size (i.e., $\sum_i z_i$), the selection probability will be close to zero and the corresponding sampling weights, being the inverse of the selection probability, will be very high.

4. **Implications of smaller size on Multiplier:**

Since the sampling weight is dependent on the size of the selected unit, in rare cases, when a smaller-sized unit is selected, it will have a high sampling weight, indicating that the small unit represents a larger portion of the population. Therefore, the observed characteristics of the selected unit are superimposed on the units which were not selected.

5. High Multiplier value in PLFS 2022-23:

(i) In the rural sector of Assam, a total of 327 FSUs were surveyed, and among them, one uninhabited (small-sized) village was selected. The selection of such a small unit by the PPS method is a rare occurrence.

(ii) Since in PLFS:2022-23 and 23-24, NSS regions within a state/UT have been considered as the basic stratum or primary sampling unit and population figures of NSS regions are considered as the size measure for the purpose of selection by PPS method, the numerator of the sampling weight, i.e., $\sum_i z_i$ are usually large numbers.

(iii) Further, consequent upon selection of a village with negligible population i.e., z_i (as the village was uninhabited according to 2011 population census) assigned a very high sampling weight to the selected village.

(iv) Such an exceptionally high sampling weight disproportionately influenced the state-level estimates for Assam.

(v) As explained above, the situation involving the high multiplier has arisen due to the adoption of a specific sampling technique and occurrences of such incidents are extremely rare. NSS has been using PPS sampling in PLFS since its inception in 2017 for selection of about 12,000 FSUs in each year and such a rare incidence as consequence of selection of a very small-sized unit has happened only once in 2022-23 survey for rural Assam.

The users of the PLFS data for 2022-23 is hereby sensitised about the occurrence of these high multiplier value in the unit level data and is requested to exercise caution while using the mentioned unit level data in their analysis by appropriately accounting for the implications of using the specific high multiplier value in the estimates produced.

6. Remedial action taken:

The following measures have been introduced in the revised version of PLFS that has begun from January 2025 for generation of monthly estimates of key labour force indicators:

(i) Instead of PPS sampling, Simple Random Sampling Without Replacement (SRSWOR) has been adopted for selection of the primary sampling units, i.e., villages or urban blocks and also the households. This will eliminate the occurrences of incidences of disproportionately high values of sampling weights as in this method all units will have equal probability of getting selected in the sample.

(ii) A special stratum at all-India level, consisting of all uninhabited villages have been constituted and for the sake of completeness of coverage, provision for selection of few villages from this stratum has been made. Estimates generated from the samples of this stratum will have contribution only to the estimates at all-India level.

****