

The State of Indian Economic Statistics: Data Quantity and Quality Issues

By

Pranab Bardhan

University of California at Berkeley

International reputation at one time.

Now that reputation is in shambles, to put it bluntly.

Our data-collecting statistical outfits are now weaker than many other developing countries

This is hurting the analysis of some of the most important basic elements of the Indian economy

For example: Growth, Poverty, Inequality

Growth

Service sector now constitutes nearly 58 per cent of national income.

With high growth (not just in software, business processing and finance)

Yet severe data problems

There is no separate price index for services yet. (The Rangarajan Commission report of 2001 pointed to the special problems of the price deflator used for the service sector).

- ✓ If the service sector price deflator were to grow at the same rate as the GDP deflator (as it did over the 40 years 1951-91), the service sector growth rate in recent years would have been significantly lower than the official rate reported.
- ✓ Base year value added is adjusted to current prices by using consumer price indices. But consumer prices are not appropriate for value added; and consumer prices for the employed wage workers are not appropriate for the self-employed in service sector.
- ✓ Problems arising from variations in quality of existing products that are difficult to observe and the advent of new products.
- ✓ Problems even in estimating the output of services from the private corporate sector.
 - Many companies do not submit their annual audited accounts to the registrar of companies.
 - Companies provide data after a long lag.
 - Estimates are made on the basis of growth in paid-up capital for a (small and non-random) RBI sample of companies. Highly unreliable.

- ✓ **A large fraction (about half) of service sector output is in the unorganised sector. (In Trade it is 75%, in Community, Social and Personal Services it is 70%). Benchmark estimates are prepared using the estimates of the value added per worker from the Economic Census, multiplied by the no. of workers. Then this is extrapolated for a particular year, using, for example, consumer price index data.**
- **Large errors of undercounting in Economic Census (sometime because it misses enterprises without fixed premises)**
- **Economic Census should be every 5 years, but the frequency has been erratic**
- **Economic Census is carried out by a large no. of ad hoc enumerators with little training**
- **Workforce estimates are made through hybrid methods: applying worker-to-population ratios of NSS to the population projections of the Census. Such hybrid methods are obviously unsatisfactory, but NSS underestimates of total population, compared to the Census, are growing over time.**

(A committee recently reported that NSS underestimates rural population by 6% and urban population by as large as 21%).

- Inappropriateness of consumer price indices for value added, already mentioned before.**
- Ad hoc assumptions like: for nearly 3 decades CSO has been assuming that the net value added for unorganized non-banking financial sector (money lenders, indigenous bankers, etc.) is just one-third of that of the organized financial sector**

Many such outdated ratios abound in different spheres of Indian official statistics

Major part of the manufacturing is also in the unorganized sector.

Periodic surveys of NSS of unorganized manufacturing enterprises: unincorporated proprietary and partnership enterprises.

The standard presumption, not always followed, is that the cut-off point between NSS and ASI is firms employing 10 workers. To this day NSS and ASI do not use a common sampling frame. The sample size for ASI has seriously declined in recent years, resulting in large sample errors. Economic Census used for sampling frame for NSS has the same problems as mentioned before.

Our informal sector itself is dualistic.

We do not have much information of the linkage between the organized and unorganized sector.

We know very little about the quality of employment in the informal sector—work conditions, work duration, safety, etc.

Issues of Green GDP (9% depreciation of natural capital)

Poverty

Poverty line is more a conceptual issue, not a data issue

Data problems relate to the consumption expenditure figures and appropriate price indices for the poor

- **Large and increasing discrepancy between NSS estimate and that from National Accounts Statistics (hhd- non-hhd, smaller set of products in the NSS data, recall lapse in case of non-food items or for items which are less frequently purchased)**
- **Non-response errors large and increasing. A recent analysis of field investigator reports suggests that for NSS 61st Round data about a quarter of the sample households, all-India, were not 'capable' of answering or were too 'busy or reluctant' (this percentage is above 40% in Bihar, Jharkhand and Uttarkhand). This non-response is in general higher for upper caste households than for SC/ST households**

Price Index

- **Until recently, problems of not covering the self-employed poor**
- **Unit value from NSS data, for bottom groups**
- **A new consumer price index**

**Non-income indicators of poverty can be estimated from National Family Health Survey. (Discontinued?)
Large sample, but not panel data.**

Inequality

- No income data (except NCAER) so the usual Gini coefficients cited are misleading in comparing with other countries. Doubts about NCAER data.
- Non-response error may be larger for better-off households
- As housing expenditure becomes more important over time, any bias in not fully capturing the consumption stream accruing to owner-occupiers in NSS data may make inequality estimates more problematic
- For the urban sector, sometimes NSS consumption data had been combined with income tax data to get some idea of the income distribution. But the Govt. of India has stopped publishing data on Income Tax Revenue Statistics after 1999-2000

More important than inequality of outcome is **inequality of opportunity**—depends on inequality of land and other assets, education, locality and social status.

Land and Asset data from NSS every 10 years.

I understand that the sample sizes of the All-India Debt and Investment Surveys have declined significantly.

As financial assets become more important, their disclosure and valuation problems more serious.

Very little data on educational inequality by households. Some attempts to estimate Gini of years of schooling for the adult population

A lot of estimates available now for Latin American countries on inequality of opportunity—what proportion of variations in an individual's income is explained by circumstances beyond his or her control (like ethnicity, gender, parental occupation or education, etc.).

The only measure of such opportunity inequality for India I have seen is by A. Singh (2011), on the basis of data from IHDS jointly carried out by NCAER and University of Maryland

Dynamic Inequality

Intergenerational Mobility crucial for understanding the changing nature of inequality in our highly stratified society.

But none of the usual surveys give us panel data, following the same household, which one would have liked for such mobility analysis.

Cross-section correlations between parent's education and son's or daughter's education—and also siblings correlations-- have been computed on the basis of National Family Health Survey data by Emran and Shilpi (2012)

Analysis of changing caste barriers to mobility on the basis of NSS consumption survey data by Hnatkowska, Lahiri, and Paul (2011).

But these surveys have two major disadvantages, compared to IHDS, for this purpose:

- Sample selection bias on account of co-residency**
- Father's current occupation, not job history**

On the basis of IHDS data analysis of occupational mobility and educational mobility has been done for father-son matched data by Motiram and Singh (2012) and Azam and Bhatt (2012) respectively

Since NCAER sample may have problems, it is high time NSS introduces a module in its regular surveys to capture at least some aspects of inter-generational mobility. Such a module may, for example, include questions about job history of parents and children (including those who are no longer resident in the household), years of schooling, natal family data for married women, etc.

Very little intra-household data: for example, no way of finding out gender inequality in ownership of land holdings in India

This will be increasingly important to know as agriculture is passing more into the hands of women

So our data are highly deficient both in quantity and quality, even in answering the basic questions of growth, poverty and inequality in our economy.

I have heard stories about administrative problems in our statistical bureaucracy

- **Paucity of staff and dependence on ad hoc part-time workers. On account of inadequacy of staff in NSS and CSO sample sizes for many surveys are shrinking**
- **Reduction in the number of sanctioned posts of field investigators and their supervisors at NSS.**
- **Not enough resources spent—as the economy becomes more complex, we need more not less resources for the data collection machinery**
- **Staff positions have remained vacant; more lucrative jobs in the private sector, so the average quality of staff declining**
- **Generally low priority is given to the need for statistics in government circles**

The National Statistical Commission should be made a permanent and independent statutory body; the various statistical units of the Ministries should all autonomously report to NSC

The general situation seems so dire that the whole statistical community needs to get involved in focusing attention on this job, and in helping restore the former prestige of Indian data quality. ISI is the obvious place to start this, as the great tradition of Professor Mahalanobis in this respect is in some danger now.