

# The Indian Official Statistical System Revisited

P. Maiti

*Indian Statistical Institute, Kolkata and Member Working Group, Kolkata,  
India*

T. J. Rao

*Indian Statistical Institute CRRao AIMSCS, Hyderabad, India*

J. K. Ghosh

*Purdue University, West Lafayette, USA*

---

## Abstract

Certain discrepancies in datasources for estimating the household consumption expenditure to derive Gross Domestic Product(GDP) of India are discussed in this paper. Simple implementable strategies are suggested to improve the estimation of GDP.

*AMS (2000) subject classification.* Primary 62D05, 62P20, Secondary 97K80.  
*Keywords and phrases.* Statistical system in India, estimates of household consumption expenditure from National Accounts Statistics (NAS) and from National Sample Survey (NSS), informal sector, gross domestic product, statistical inference, Central Statistical Office, National Sample Survey Office, National Statistical Commission.

---

## Abbreviations used in this paper

**ASNMI** : Annual Survey of Non Manufacturing Industries

**BSLLD** : Basic Statistics for Local Level Development.

**CSO** : Central Statistical Office

**EC** : Economic Census

**EPM** : Enquete Permanente aupres des Menages

**FISIM** : Financial Intermediation Services Indirectly Measured

---

The views expressed in this paper are those of the authors and do not reflect the views of the Institutions to which they belong

**GDP** : Gross Domestic Product

**GVA** : Gross Value Added

**GVAPW** : Gross Value Added Per Worker

**HCE** : Household Consumption Expenditure

**HCES** : Household Consumer Expenditure Survey

**ICLS** : International Conference of Labour Statisticians

**ILO** : International Labour Organization

**MoSPI** : Ministry of Statistics and Programme Implementation

**MRP** : Mixed Reference Period

**NAS** : National Accounts Statistics

**NCEUS** : National Commission for Enterprises in the Unorganized Sector

**NIPFP** : National Institute of Public Finance and Policy

**NSC** : National Statistical Commission

**NSS** : National Sample Survey

**NSSO** : National Sample Survey Office

**PFCE** : Private Final Consumption Expenditure

**RGI** : Registrar General of India

**SAM** : Social Accounting Matrix

**WG** : Working Group

## 1 Introduction

Ghosh et al. (1998) reviewed the evolution of Statistics in India starting from about 300 B.C. to the late nineties. The Indian Statistical System has been praised by many statistical experts in the past for its coverage and accuracy, but vast changes were taking place in the country even when the above review was written. The Indian Economy had moved towards liberalization and the so called Hindu growth rates, slow and unchanging, had become a thing of the past. The growth raised new questions about official

statistics some of which have been discussed in two papers (Ghosh (1998) and Ghosh and Maiti (2004)).

It is in the spirit of the Fundamental Principles of Official Statistics of the UN Statistical Commission and the functions of the National Statistical Commission of the Government of India that we take a fresh look at the present official statistical system of the country.

The effect of the growth that we have mentioned above on the poor has been a subject of debate and even some reduction of the head count ratio for the poor has been questioned by a section of the economists among others.

This debate has been further fuelled by the coverage and accuracy of Indian statistical data in recent times. For example, calculation of the Gross Domestic Product (GDP) for India has been based on National Accounts Statistics (NAS) data. The Central Statistical Office (CSO), earlier known as the Central Statistical Organisation of the Ministry of Statistics and Programme Implementation (MoSPI) releases estimates of expenditure on GDP by item break-up, namely Private Final Consumption Expenditure (PFCE), Govt. Final Consumption Expenditure, Gross Fixed Capital Formation, Change in Stock, Exports less Imports, and residuals with total adding up to GDP. Estimates of Consumption Expenditure based on Household Consumer Expenditure Surveys (HCES) of National Sample Survey Office (NSSO) do differ from the PFCE obtained by NAS and difference between the two figures has been steadily increasing, currently reaching a staggering magnitude of about 50 %. The then Prime Minister of India, Dr. Manmohan Singh has been deploring this publicly, as can be seen from his speech at the inauguration of the Platinum Jubilee Celebrations of the Indian Statistical Institute (ISI), exhorting statisticians to examine what has gone wrong with statistical data. This is the major problem we mentioned earlier.

Supporting Dr. Manmohan Singh's concern on the quality of Indian data on official statistics, recently Professor Pranab Bardhan during the Eighth International Calcutta Triennial Symposium held in 2012 referred to the diminishing quality of current Indian Official Statistics, especially data relating to GDP of the country. We have studied this problem and believe it is not only of importance to India but to the other developing countries as well.

It appears that there are both theoretical and practical issues of fundamental importance to all countries having a large unorganized informal sector. Moreover, the issues may be of importance even to countries without such an unorganized/informal sector.

The role of this sector in the Indian Economy is discussed in Section 5. We shall first move on to the controversy regarding the estimates of the

Private Final Consumption Expenditure(PFCE) of the National Accounts Statistics(NAS) versus the estimates of the Consumption Expenditure based on the Household Consumer Expenditure Surveys(HCES) of the National Sample Survey(NSS). Next, having noticed that the informal sector is one of the main contributors to the GDP, we shall briefly discuss an earlier use of adaptive sampling to get a better coverage of this sector. However, for a large scale survey like that of NSS, it is not always feasible to employ adaptive (network) sampling and hence we shall propose an easily implementable strategy within the NSS set up. This strategy was prompted by the importance given to the panchayats in rural and municipalities in urban sector in view of the 73rd and 74th amendments of the Constitution of India which makes collection of decentralized data possible. The success of this technique was earlier noted in Ghosh (1999) and Maiti (2007). There has been a remark that NSS has been adopting the same old sampling design for a long time. Our suggestion of a new and simple design could be an answer to that in this particular case. The paper ends with a brief note on the current status and some concluding remarks.

## **2 Controversy Relating to NAS vs. NSS Estimates of Household Consumption Expenditure**

Central Statistical Office(CSO) publishes National Accounts Statistics (NAS) for the country and releases estimates of expenditure on Gross Domestic Product(GDP) by item break-up of which Private Final Consumption Expenditure(PFCE) is an important component. CSO's estimate of PFCE is derived following the 'commodity flow approach'. This consists of obtaining quantum and value of different commodities flowing finally into the consumption process of the households and the private Non-Profit Institutions Serving Households(NPISH) from the quantum and value of the commodities produced and available. In other words, the inputs for PFCE of NAS consist of expenditure by households (including Non-Profit Institutions Serving Households (NPISH)) on non-durable consumer goods and services and all durable goods except land and buildings. Thus household consumption expenditure plays an important role in the calculation of GDP.

The National Sample Survey (NSS), a 64 year old successful phenomenon, initiated by Prof. Mahalanobis, is a nation-wide, large-scale continuous survey operation organized in the form of successive Rounds employing sampling techniques. Among other things, surveys of this organization include household consumption expenditure ascertained directly by canvassing a well structured schedule of enquiry covering broadly every item of consumption. Besides the PFCE of NAS mentioned above, these Consumer Expenditure

surveys by the NSS Office also provide estimates of household consumption expenditure, with the exception of NPISH. Some of the inputs for obtaining these estimates are food and non-food items, fuel and light, clothing, footwear, durable goods for domestic use. Even though one expects that the NAS estimates and the NSS estimates should be close, it is seen from Table 1 that there is a widening divergence between these two estimates over a period of time.

It is noted that the gap between the two sets of estimates has been widening progressively. For example, a 5.5 percent difference increased to 10.4 in 1977-78 which further went up to 25.6 percent and 38.1 percent in 1983-84 and 1993-94 respectively. Thus within a period of two decades, the percentage difference increased almost seven times. It is also observed that the NAS estimates were found to be moving at a faster rate than NSS estimates.

For releasing poverty estimates, the erstwhile Planning Commission needed the above data and it has been suggested by the task force of 1979 to adjust the NSS-based size distribution by a correction obtained by shifting the NSS distribution uniformly to the right by the ratio of per capita PFCE to per capita HCE.

However, the Expert Group on Estimation of Proportion and Number of Poor, constituted by the Planning Commission in 1993 regarded this

Table 1: Divergence between NAS and NSS Estimates (Rs. Crores) of household consumption expenditure

Year	Source	Total Expenditure
(1)	(2)	(3)
1972-73	NAS	35160
	NSS	33210
	%difference	5.5
1977-78	NAS	63084
	NSS	56530
	% difference	10.4
1983-84	NAS	146084
	NSS	108668
	% difference	25.6
1993-94	NAS	574772
	NSS	355770
	% difference	38.1

Source : NSSO-CSO (2005), 1 Crore = 10 million.

adjustment as inappropriate and even inadequate and recommended to abandon this procedure and suggested that the NSS remains the best available source of assessing poverty incidence.

Thus, the controversy relating to the NAS and NSS estimates referred to above had an important role in the Indian Official Statistical System and received much attention.

Reverting to the Indian scenario, as mentioned in the introduction, the widening gap between the two sets of estimates displayed in Table 1 is disturbing. Earlier studies by Minhas (1988), Minhas and Kansal (1990) and Minhas et al. (1986) and the later paper by Sundaram and Tendulkar (2001), among several others, led to a detailed cross-validation study of estimates of private final consumption expenditure (PFCE) available from national accounts and consumption expenditure from household survey by an Expert Group, Government of India (2005). This Report as well as the Report of CSO, Government of India (2008) attribute the deviations to coverage, reference time-frame, unmatched classification schemes, treatment of cooked meals, notional components in NAS estimate of PFCE such as imputed rents on owner-occupied dwellings and financial intermediation services indirectly measured (FISIM), differential implicit prices. They also suggested improved methods of deriving NAS estimates and data collection in the consumer expenditure surveys and studies on validity of data base and examined the extent of deficiency and ascertained reasons. Since the reports are thorough and widely discussed, we shall not comment on these in detail.

We will only observe that the three papers, of which Minhas is a co-author, seem to show an interesting trend. The first paper (Minhas, 1988) provides a detailed analytical scrutiny of both the estimates. In each case, he looks at the major components in the two estimates and compares them with corresponding estimates in the previous years, or other analogous estimate for the corresponding components. The study of this kind for major food grains reduces the discrepancy to 12 %, much less than the original figure and after providing additional ways of further reduction of the discrepancy, goes on to hope that the two estimates of NAS and NSS could be reconciled. The detailed comparison shows that sometimes NSS figures are larger than they should be as well as smaller due to bad response. No systematic bias or other defect can be identified. One can summarize Minhas's final conclusion as follows: While both methods can improve their estimates by paying heed to lapses pointed out in the paper, the NSS estimate is based on real data and sound methodology as well as trained investigators.

The NAS, on the other hand, seems rather weak because at various stages, subjective, arbitrary adjustments are made without clear justification.

This is evident, for example, from the use of partial data on 'stocks', quality of data on 'marketable surplus' based on old reports and data on 'intermediate uses' with outdated ratio adjustments. Further data on production of consumer goods in the unorganized sector and on trade and transport margins were found to be weak (Minhas et al., 1986).

The third paper by Minhas and Kansal (1990) in the series which is in the same spirit, made some observations on the 1983 data. Here, the authors provide the margin of uncertainty in the estimates of per capita expenditure using the different sets of estimates provided by the NAS in its revisions. They also offer critical comments on the rationality of the 'judgemental and accommodative' adjustments. Minhas and Kansal (1990) seem to feel that the NAS data is fluid, probably in the sense of being unstable, compared with NSS estimate and make no attempt on further reduction in the divergence between the estimates.

As suggested by Minhas later (Comments on Expert Group Report, 2005), it is hoped that the newly constituted National Statistical Commission gives some thought to this problem.

### 3 Similar Scenario in some other Developing Countries

We shall first refer to Bhalla (2002) who deals with this aspect with several interesting new points for not only India, but many other developing countries as well, though he has a partial analysis of actual figures.

We have mentioned towards the end of Introduction that this is not an unfamiliar issue. A note from the European Commission states that national accounts rely on several sources to estimate consumption expenditure, both from the demand and supply sides and the member states are legally obliged to transmit the data to Eurostat. However with respect to the household surveys conducted in the member countries, the note questions their standards.

While discussing the Final Consumption Expenditure of Households in South Korea, Han (2011) attributes the differences between National Accounts Statistics and Household Income & Expenditure Survey to coverage (by subjects or by items), bias due to the limitation of sampling method and classification system among others. While China, Kazakhstan, Laos and Vietnam use the mean of survey data with a deflation factor of 3 % to 12 % as the mean for national accounts, this divergence is gauged with Engel curve adjustments for Sri Lanka (see Bhalla (2002)).

Robilliard and Robinson (2003) discussed the problem of reconciling household surveys and national accounts data using a Cross Entropy estimation method. The household data is from the 'Living Standards

Measurement Survey' for Madagascar called EPM 93 (Enquete Permanente aupres des Menages) and the macro aggregates are from the Social Accounting Matrix (SAM 95). Here again, the weighted sum of household incomes falls short by 15.2 % compared to SAM figure. The method successfully uses the information provided by the national accounts data for Madagascar to re-estimate the household weights used in the survey so that the survey results are consistent with the aggregate data.

The aim of this statistical exercise (Robilliard and Robinson, 2003) is to estimate a set of household survey weights consistent with extraneous information from National Accounts data. The sampling design of a survey of  $N$  households provides prior survey probabilities  $p_i^*$  using which data on a vector  $x_i$  of observed characteristics for each household (hh) such as hh size, hh expenditure etc. is collected  $i = 1, 2, \dots, N$ . Further, we have aggregates or weighted averages of some of the hh information from alternative sources, say  $y = (y_1, y_2, \dots, y_R)$ . It is required that  $y$  be consistent with the distribution of sampling probabilities (sample weights)  $(p_1, p_2, \dots, p_N)$ . Usually  $N$  is very large and  $R$  is small corresponding to a few macro economic and demographic adding-up constraints. To solve this problem, the estimation procedure minimizes the Cross Entropy measure of distance between the new estimated probabilities and the prior, namely

$$\text{Min.} \quad \sum p_i \ln(p_i/p_i^*)$$

subject to the moment constraints

$$\sum_1^N p_i = 1$$

We have briefly introduced the problem here and for full details we refer to Robilliard and Robinson (2003). Here the application of Cross Entropy criterion does not require specification of an explicit likelihood function (see, Golan (1998) and Zellner (1990)). Thus Robilliard and Robinson's procedure adjusts the prior probabilities  $p_i^*$  using the new information to generate posterior estimated weights as solution to the above problem, which (Zellner, 1988) terms as an 'efficient processing rule'. The possibility of such a study in the Indian context is suggested in Rao (2010b).

#### 4 Certain Recommendations from Various Quarters

The recommendations of the Government of India (2005) and the Report of CSO, Government of India (2008) perhaps need to be implemented for quality data from both sides-NAS as well as NSS.

It has been observed that NSS provides a valid estimate based on direct observations relating to the survey period. Furthermore, unlike NAS, NSS avoids recourse to adjustments based on arbitrary assumptions. Thus NSS remains to be the best available data source (see for example, Government of India (1993) and Sundaram and Tendulkar (2001)).

As per recommendations of the Rangarajan Commission, the subcommittee of the National Commission for Enterprises in the Unorganized Sector (NCEUS), reiterates among other things, the need of an Annual Survey of Non- Manufacturing Industries (ASNMI) and also hopes that the MCA21 Programme of the Ministry of Company Affairs eventually provides details of current data on the missing private corporate sector.

Objections to NAS data have also been raised by Rajaraman (2012). It was pointed out that CSO estimates are based on estimates submitted by states, but the states do not follow any uniform procedure. The CSO tries to correct this at a later stage. Rajaraman recommends that a common acceptable procedure be evolved for all states to follow, with correction at a later stage by CSO.

Reverting back to the question of accuracy of NSS estimates in relation to the choice of the Reference Period, we point out that the NSSO Expert Group on Non Sampling Errors, Government of India (2003) reported the suitability of different reference periods for measuring household consumption. This, though pilot but rigorous study, looked at the aspects discussed by Mahalanobis and Sen (1954) as well. As a start, a Mixed Reference Period (MRP) has been used in the Surveys and the Commission may try to implement the Group's recommendations which are expected to improve the quality of data.

One of the main criticisms against NSS consumption data is that non-response has increased over time and is expected to increase further as consumer preferences increase - the better off a person is, the more likely is non-response. With the advent of information technology and the computer-savvy nature of urban respondents, India also could take a bold step of conducting some of their surveys using CAPI/CATI in urban areas. While discussing the strengths and weaknesses of the consumer expenditure survey from a Bureau of Labour Statistics Perspective, (Garner et al., 2009) noted that there is an evidence that some consumer expenditure estimates are biased due to item non-response and measurement error. In this context, we also refer to Ravallion (2000) and Deaton (2001) among others and note that NSS is not alone.

Assuming that a Working Group (WG) precedes a particular Round of NSS, constitution of which should be encouraged, it is suggested that the

sources of non sampling errors faced in a similar previous Round of the NSS cycle should be thoroughly discussed and the participants of the Training Conferences are made aware of these so that non sampling errors do not creep in during the course of the current survey Round. The WG can plan the questionnaires in such a way as to minimize the non sampling errors. Care should be taken to include questions on new technological innovations which gain faster acceptance by the public. For example, while shortening the schedules, it may be remembered that items such as mobile phones/internet should be included which have become very popular in the country. (The schedule of 63rd round briefly mentions mobile handsets and the schedule of 67th round had a block on internet usage as well).

In the context of poverty studies, estimated distribution of consumer expenditure provided by NSS is used which is based on a general purpose sampling design with emphasis on point parameters. However, here the interest is in the estimation of tail end frequencies which are subject to large sampling errors (Murthy, 1977). Referring to this, Minhas (1988) suggests a cautious use of NSS data. This is reiterated by the Expert Group on Estimation of Proportion and Number of Poor constituted by the Planning Commission which cautions about the hazardous nature of carrying out pro-rata adjustments in the observed size distribution of consumer expenditure in a particular NSS round. Furthermore, Srivastava et al. (2003) also refer this point again quoting Murthy (1977) that for a study of economic inequality with respect to highly skewed distributions based on expenditure or land holdings, estimation of the corresponding frequency distributions is itself more meaningful.

This problem is aggravated in India (and other developing countries) by the presence of a huge black market, which would typically evade direct methods based on questions and answers, but may be obtained as a latent variable in public consumption. For example, one of the data sources could be adding up airfares and including it in NAS which could take care of expensive foreign travels by the rich. Other possible sources are bullion and jewellery transactions, expenses at grand weddings and parties as well as holiday homes and resorts, real estate and immovable property transactions, foreign remittances and gifts etc., just to quote a few. A 2012 Report of the National Institute of Public Finance and Policy (NIPFP) on the quantification of black money in the Country seems to have estimated that unlawful wealth would exceed 10 percent or more of GDP. A similar earlier study seems to have estimated that black money generated in India was 19 percent to 21 percent of GDP.

## 5 Informal Sector

Economists and policy makers do not distinguish between the terms ‘unorganized sector’ and ‘informal sector’ and often use them interchangeably. As per the concepts of the 15th International Conference (ILO, 1993) of Labour Statisticians (ICLS), informal sector is defined as consisting of those units that are engaged in production of goods or services with the primary objective of generating employment and incomes to the persons concerned. These are small scale units operating at a low level of organization with ‘little or no division between labour and capital as factors of production’. The labour is mostly casual and drawn from personal and social contacts without any contracts or guarantees.

According to the National Statistical Commission (NSC), more than 90 percent of work force and about 50 percent of the national product are accounted for by the informal sector. More and more informal activities are being successfully undertaken by the weaker sections of the society. These high levels of informality may have led to the high levels of growth of the economy during the past two decades.

The NSC may soon implement the recommendations of the Committee of NSC on Unorganised Sector Statistics, Government of India (2012) to improve the data bases from existing NSS Employment and Unemployment and Enterprise Surveys, Labour Bureau Annual Surveys of Employment and Unemployment, comprehensive scheme for study of cost of cultivation of Ministry of Agriculture, Censuses on Micro, Small and Medium Enterprises, Economic Censuses in the unorganized sector. Having recognized the data requirement for monitoring the long run transition from a predominantly rural and agricultural economy and society to predominantly urban and non-agricultural one, an NSC expert committee has identified certain data gaps in the existing system. By its very nature, the informal sector is characterized by fragmentation and random clusters that are difficult to identify. Also, they hardly maintain any accounts or keep good records. In India, as one of the city groups of the United Nations Statistical Commission to address methodological issues related to informal sector, an Expert Group on Informal Sector Statistics, popularly known as the ‘Delhi Group’, was set up in 1997. Furthermore in 2004, the government has constituted the National Commission for Enterprises in the Unorganized Sector (NCEUS), one of the terms of reference of which was to review the status of unorganized/informal sector in India including the nature of enterprises, their size, spread and scope, and magnitude of employment. To achieve this objective, the Commission constituted a Task Force in 2008 which re-

viewed the existing methodologies for estimating the contribution of unorganized/informal sector to GDP. This subcommittee observed that the 'labour input method' (O.E.C.D., 2002) is not feasible in the Indian context and the 'method of apportioning' proposed by Kolli and Hazra (2005) is confined to GDP estimates of unorganized sector and subsequently developed a modified approach which suggests apportioning by using estimates of labour inputs and productivity differentials of both the organized and unorganized sectors.

This modified apportioning method was found to be an improvement over other existing methods. According to the estimates worked out by the subcommittee by using the modified apportioning method, as high as 50 percent of GDP was contributed by the unorganized/informal sector in 2004-05.

## **6 Alternative Sampling Designs Suggested**

It is believed that Unorganised Manufacturing and Trade contributes a big share of India's national income and capturing and surveying this sector fairly accurately is a demanding task. The surveys conducted by NSSO are facilitated as a follow up of the Economic Censuses. One of the recent ones, namely the 5th Economic Census(EC), conducted in 2005 is defective in the sense that it could not cover all enterprises without fixed premises and the enterprises carried out in residential premises due to field difficulties. Thus in general, ECs suffer from the undercount of enterprises, especially own account enterprises and hence its use as a sampling frame for the follow-up surveys is questionable. In the 62 nd Round of NSS, it was decided for the first time to use a list frame of enterprises and survey them in addition to the usual area frame. This was done to net sufficient number of larger enterprises so as to get a reliable estimate of Gross Value Added (GVA) per worker, number of workers, their earnings and other parameters. However, only around a quarter of these units from the list frame could be surveyed since the list was inaccurate. The recently concluded 6th EC is expected to provide a good sampling frame free of errors. It also made some changes relating to coverage of non-agricultural activities and introduced schedules so as to help create a Business Register as recommended by the Rangarajan Commission.

It is difficult to enumerate the people in India who are engaged in various sparsely scattered small scale industries in unorganized/informal sector in order to estimate the parameters needed for national accounting. A traditional stratified two-stage sampling design scheme is ineffective for some of the enterprises because of failures to capture their sparse locations and scatter.

Alternatively, an adaptive sampling scheme was used by Chaudhury et al. (2004) to capture the sparse and scattered enterprises. The suggested adaptive sampling scheme involves:

- (i). Construction of networks based on a well defined neighborhood; i.e., defining certain neighborhoods and formation of networks for every population unit.
- (ii). Extending the initial sample by appropriate network formations as in (i) above i.e., if one starts with any sample (ultimate stage unit being a village in rural areas), a corresponding adaptive sampling just adds to it the units in the networks of all the sampling units.

As an illustration, let us consider the problem of estimating the total number of earners through specific small scale industries in villages in the unorganized sector in various districts of West Bengal. Even though, the earner's individual earnings are meagre, their collective contribution to the nation's GDP is believed to be quite substantial and it is thus of interest for national accounting. As explained above, even a follow up survey of unorganized sector using the EC frame fails to capture such scattered rural earners. Chaudhury et al. (2004) estimate the total number of such earners by using a conventional stratified two stage sampling scheme and an adaptive sampling scheme netting villages that contained the specified categories of earners. They find that adaptive sampling technique is more efficient and suitable in such a case.

It may be noted that adaptive sampling procedure makes the final sample size a random variable. Operationally, the adaptive sampling scheme will be more complicated and would be difficult for a nationwide large scale survey like NSS for the following reasons:

- (a) There will be additional work and the associated manpower needed would be of considerable size in forming the neighborhoods and networks on nationwide large scale surveys.
- (b) Handling a sampling design with varying sample size would be very complicated both operationally and technically for purposes of making inference, e.g. getting unbiased estimates.

However, we have presented adaptive sampling technique here briefly, because for certain small scale surveys in a scenario like this, this technique is very useful as seen in the work of Thompson(2002) and Chaudhuri et al. (2004). In view of the difficulties in the large scale application of adaptive sampling, we propose next another simple implementable method.

## 7 An Easily Implementable Strategy

The alternative strategy proposed is based on the earlier work (Ghosh and Maiti, 2004) on panchayats (local bodies in villages like municipalities in towns) wherein the authors refer to the recommendation of the NSC regarding the strengthening of data from panchayats. (See also Ghosh (1999) referring to the State of Madhya Pradesh and Adhikari et al. (1998) referring to West Bengal). One of the comments of the Roy- Iyer Review Committee of erstwhile NSS Organisation is that, since its inception, for most of its sample surveys, NSS has been using a standard sampling design without any innovative methodological studies. We feel that our suggestion below is an answer in that direction. We suggest that panchayats be guided to participate in the survey and be rewarded monetarily for such work - initially in pilot areas under the supervision of an investigator from NSS.

As against the adaptive and other sampling schemes discussed above, we shall now propose a simple sampling scheme as follows: The sampling design for the survey of enterprises in the informal sector should be cluster sampling with one 'gram panchayat' being regarded as a cluster of villages. The list of all the enterprises for the villages combined, under the gram panchayat can be made available at the panchayat level.

Taking one gram panchayat as the unit of analysis, appropriate estimation procedure can be developed without disturbing much of the existing NSS methodology. One could also use a village as the ultimate sampling unit. This would provide two estimates of total number of units in the Informal Sectors and the two estimates for the same quantity will help validate each.

Similarly in urban areas, a municipality can be considered as a cluster of wards and the list of enterprises for all wards combined under the municipality can be made available at the municipality level, and the appropriate estimation procedure can be developed without much difficulty. The advantages of this scheme are:

- (i) It is easy to operate in a large scale survey compared to the adaptive sampling scheme;
- (ii) Taking panchayat as a unit, means reduction of sparsity and scatteredness in the enterprises of the informal sector in a localized village, as a panchayat is a natural cluster of villages;
- (iii) Information on all types of economic activity (including rickshaw pulling, thelawalla(hand cart pulling), small grocery shops, small jewellery shops etc.) in the panchayat area could be regularly recorded in the

panchayat office and recording of such enterprises could be made mandatory;

- (iv) Thus, information on relevant items can be easily obtained. One of the reasons for suggesting this alternative is that since decentralized planning is already under way, one can reap the benefits with a proper strengthening of the decentralized statistical system which ensures that at least one major chunk of the unorganized sector could be netted through this design. This is in line with the recommendations of Rangarajan Commission and the Report on Basic Statistics for Local Level Development (BSLLD) of CSO, Government of India (2014).

Taking advantage of the local government agencies, perhaps a stronger statistical system could be evolved as recommended by Rangarajan Commission. As an illustration, we mention that recently, a statistical information system for decentralized planning for the district of Howrah, West Bengal has been developed successfully (Maiti (2007)). As a member of the Working Group on preparation of Annual and Perspective Plan for the district of Howrah in the state of West Bengal, under the 11th Five Year Plan(2007-12), one of the authors (Maiti) was involved in the : (i) Creation of a computerised data base for decentralized planning in the rural sector of the district of Howrah, i.e. collection and use of statistics at the grass-root levels for planning and decision making at the level of panchayats, (ii) Development of a useful software for a computerized data base, (iii) Preparation of fourteen different Block wise Human Development Reports as well as Report for the rural sector of the District as a whole and (iv) Preparation of one data based Gram Panchayat plan under decentralized planning.

## 8 Current Status and Concluding Remarks

It may be noted that the ‘Industrial Statistical Act’ of 1942 is very old and the ‘Collection of Statistics Act, 1953’, was found to be weak. Faced with this, after several discussions, the second NSC has rewritten these. A law to enforce the collection of statistics which was recently enacted by the Government of India under the “Collection of statistics Act 2008” is to be implemented by the new Commission.

The recommendation of the Rangarajan Committee regarding the creation of a permanent and *statutory* apex body, to be named as - National Commission of Statistics— *through an Act of Parliament*, independent of the Government in respect of policy making, coordination and maintaining quality standards of core statistics has been implemented by a Government Order. This body is called National Statistical Commission (NSC) which

we have referred to earlier. One feels that with the wide variety of Statistics that are being collected in the Country, it would be difficult for the Commission, all the members of which are part time, to concentrate on the problems facing the Commission. It is true that it was set up parallel to the one in the United Kingdom, but for the size of a country like India, the Chairperson and the Members NSC should have better support and service conditions (Rao, 2013).

Unfortunately, since the time the 1999 International Statistical Review paper of Ghosh et al. appeared, much change has taken place with regard to the statisticians as well. Some have retired and the younger colleagues, like most young statisticians elsewhere in the world, work in both theory and applications, but are not interested in working on official statistics.

In this paper, we concern ourselves first with the problem of divergence between consumption figures as estimated by NAS and NSS, and an assessment on the quality of their current estimates. Because we are statisticians, we feel more competent to comment on NSS rather than NAS, but we do have comments and suggestions for both.

We start with NSS. Bardhan (2012) had pointed out in his lecture that India's very large informal sector may be one of the main reasons for inaccurate estimates of GDP. Analysing this a little, we find that much of NAS data is not based on samples, but it does use information supplied by surveys of NSSO for the huge informal sector. Any error in the NSS estimate due to the informal sector will be transmitted to NAS, but NSS remains the primary agent. It, therefore, makes sense to begin by analysing the estimate of NSS for the informal sector. The problem with the informal sector is that, on one hand, it is very scattered and on the other, it is somewhat lumpy, i.e. several informal establishments will grow up in the same location. If surveys of NSSO miss one, it will miss several others. Since the informal sector is believed to contribute about half of the total GDP, we are talking of a huge collection of scattered clusters of informal establishments. Even though NSS adopts a well planned sampling design for the surveys on unorganized sector, it is the frame based on the Economic Census that misses several units. Also, the Gross Value Added Per Worker (GVAPW) is likely to get underestimated, leading to an under estimate of GDP. It is known that the Economic Census having been conducted earlier becomes old and since then the Informal sector like the whole Indian Economy has increased a lot. NAS uses the Economic Census to estimate the number of workers in different informal sector industries which include even stone breaking and these estimates may not be accurate. Moreover, the Economic Census was done by

reluctant teachers during their holidays and inexperienced persons with a rather small remuneration. It is felt that the quality is not good.

We know from past experience that standard designs with net work sampling which takes into account the fact of clustering, tries to capture the whole cluster of establishments in a locality. Network sampling design has been shown to provide better estimates than conventional design as seen in Section 6 (Chaudhury et al., 2004). Unfortunately, this design is too complex to be used in large scale surveys for the whole country. We, therefore, proposed a simpler method in Section 7 of the paper.

A major new statistical fact is that India now has panchayats distributed all over the country with villages (wards) coming under some panchayat (municipality). Panchayats are statutorily responsible for maintaining a list of the informal establishments located in the villages constituting the panchayat. There is some incentive to maintaining a complete list, since the listed establishments pay a licence fee to the panchayat. Therefore, in principle, India has a frame of all informal establishments. This fact can be used in several ways.

We suggested two approaches: In the multistage, stratified sampling scheme of NSS, panchayats should be the last but one stage, just before the villages. Then a ready list of all establishments would be available. We expect none to be missed and thus a major source of underestimation would be removed.

Alternatively as a second choice, if such an approach is not followed, it would make sense to use available lists of informal establishments in the panchayats containing a chosen village under the NSS scheme. We believe that NSS estimate would only be strengthened, if panchayats are sampled and all the establishments in the selected panchayats are visited by investigators. We hope that any sampling design would gain by the above suggestions, which introduce an element of complete enumeration (i.e. census) in the sampling design. Otherwise NSS may not be able to cope up with the size and scatter and clustering of informal establishments over a huge country. Initially, the above suggestions may be tested on a small scale via pilot surveys.

We shall now digress a little from the issue of frames. It is known that the NSS estimates given in Table 1, are arrived at as the product of the estimates of annual per capita consumption expenditure obtained from the HCES and the population projections based on the Population Census (Registrar General of India (RGI), 1996). The product is obtained separately for the rural and urban populations and the sum is taken as the estimate for

total household consumption expenditure of the domestic economy (Expert Group, 2005). Further, it is also known that RGI's projections fall short of the actual figures (cf. Census of India, Population Projections, Chapter 4, Statement 12). Thus, use of correct numbers also could improve the NSS figures to some extent.

Our personal feeling is that, given India's low population growth rate in the current years, the estimated populations will be close to the last census figure so as to be acceptable. Of course, this assumes that the last census figure itself was reliable. Underlying the above methodology are certain demographic assumptions. A direct estimate of population based on a reliable data on births and deaths is not available. If our suggestion of using panchayats is taken, we may need to know the total number of panchayats. This record would be easier to maintain since the number of panchayats would be smaller than the number of people and the former will change more slowly.

There are other reasons for underestimation in the NSS estimates which we have touched upon earlier. Nowadays, respondents are reluctant to give their time for supplying data. It will help NSS to make response to NSS a statutory duty. The Collection of Statistics Act, 2008 takes care of this to some extent.

Sample surveys may not be able to cope up with all the issues relating to the under estimation in the huge unorganized sector. We draw attention of our readers to current inference in high dimensional problems, where standard methods do not apply. During the past few years there were many challenging practical problems in bio informatics, image analysis, astronomy, climate predictions etc. (see Bogdan et al. (2011)) and Neuvial and Roquain (2012).

As we have seen from the previous sections, in the calculation of GDP, unorganized sector plays a considerable role. Thus, the statistician is faced with the problem of estimating a relatively large number  $p$  of parameters such as Number of Workers and Number of Enterprises (by activity category), Emoluments, Gross Value Added Per Worker and Per Enterprise, Fixed Assets, Operating Expenses, Receipts Per Enterprise, just to list a few based on relatively less number of  $n$  units selected. Thus, in a particular Round of a Survey on Unorganised sector a multitude of parameters are involved (with further duplication of the whole set of parameters for all States and UTs to be estimated) on the basis of a relatively few sample of  $n$  units because of the sparsity. With this similarity in mind, our speculative assertion then, is: There could be an analogy between inference about the unorganized sector with 'many parameters' in the model (much larger than or nearly as large as  $n$ ). And under this analogy, sample surveys may fail to do well and

census methods may be needed partially. It is in this context that we have suggested census by Panchayat members for the localities at one of the stages of sampling.

Next, regarding the controversy between PFCE of NAS and HCES of NSS, as per the Report of the Expert Group on Non Sampling Errors (2005), the NSS estimates adjusted for the 'notional' elements and the NAS estimates adjusted for prices closes the gap and the detailed comparison of the two sets of adjusted estimates shows that the divergence between the estimates is due to a few specific sub-groups of food and non-food items. Both Minhas (1988) and Lal et al. (2001) have attributed the non-cooperation from the affluent households to be the main reason for the downward bias. The above arguments suggest that NAS must be seriously biased on the higher side. We recall Minhas's trenchant comment on the NAS, that the increase in NAS consumption expenditure between 1987-88 and 1993-94 has been shown to be so inconsistent with changes in other national income aggregates. Also as per the Report of The Expert Group on Estimation of Proportion and Number of Poor, constituted by the Planning Commission (1993) which we have referred to earlier NSS remains the best available source of assessing poverty incidence.

We admire a lot the careful scrutiny and check with other sources of data as well as internal evidence, but nonetheless bring up completely new points which do suggest the NSS figure may be an underestimate in spite of the good quality of work of NSS. Like Minhas and his colleagues and Sundaram and Tendulkar, we also think the NAS figure could also be wrong, but in a different direction and that it is likely to be an overestimate. We thus conclude that we have two estimates with biases in different directions. None is correct. NSS estimate seems to be a substantial under estimate, while NAS estimate a substantial over estimate. Thus, an average is likely to be less in error than either of them, even though it is based on two wrong estimates because such a mean value may be meaningful. That such a mean would be extremely useful was also echoed in Bhalla (2002) and Deaton (2001). However one may add that the mean could be that of 'adjusted' NSS estimate and the 'adjusted' NAS estimate where the corresponding adjustments are made with respect to NPISH's, notional elements, imputed components, prices etc. With the cluster sampling approach suggested, unorganized sector gets substantially covered and the PFCE component with these adjustments tends to be closer to NSS estimates.

An Associate Editor suggested that a weighted average of NAS and NSS estimates using an empirical Bayes approach could be a step forward. We have suggested a pilot based survey by the NSS to explore the possibility

of cluster sampling as an intermediate stage and once the NSC carries out such pilot, we shall be in a position to construct a) a simple(or 'adjusted') average of NAS and NSS estimates, b) an empirical Bayes type weighted average of NAS and NSS estimates suggested by the Associate Editor and (c) estimates based on the proposed cluster sampling scheme of the present paper and then make a comparison and arrive at a suitable solution which would improve the GDP statistics of the country. Since, availability of data from such a pilot survey either by the NSC or some recognized agency takes some time , our focus of the present paper has been to contribute various alternatives.

Such discrepancies as mentioned above are also seen in the Service Sector. Based on the NAS data, we have the distribution of GDP by industrial activity. An alternative direct estimate of GVA by industry can also be calculated based on the NSS data. Modulo, sampling and non sampling errors we can expect them to be close. However, using the NSS data on 63rd Survey (2006-07) on NSS data, Ashish Kumar et al. (2010) note that the extent of divergence between the two sets of estimates is quite high for each of the five categories they have studied.

It is the vision, foresight, experience and tenacity of Mahalanobis that made NSS a world class large sample survey organization and no wonder it is now a 64 year old successful phenomenon (cf. Rao 2012) which is a candidate for 'Large Data analysis'. To gauge this success, we quote the Report of the NCEUS referred to earlier, which mentions that "*India has built up, over the years, an elaborate and competent statistical system both in terms of institutions as well as professional manpower*". Within the ILO, and beyond it, India's 'Delhi Group' is recognized as a leader in setting standards for informal economy statistics.

*Acknowledgments.* The authors wish to thank an Associate Editor and the Referees for their comments and inputs which improved the earlier versions substantially.

## References

- ADHIKARI, A.K., BANDYOPADHYAYA, S., CHAUDHURI SAHU, R., RAO, T.J. and ROY, J. (1998) Statistical System for Decentralised Planning, Proceedings of the Seminar on Decentralisation: Devolution and Participation, Government of west Bengal, Calcutta.
- KUMAR, A., MANNA, G.C. and LAKSHMI, G.S. (2010) Value Added by the Service Sector in India: A Comparative Analysis Based on NSS 63 rd Round Survey and National Accounts Statistics, Proceedings of National Seminar on NSS 63rd Round Survey Results, 2-3 June,2010, Government of India.
- BARDHAN, P (2012) The State of Economic Statistics: Data Quantity and Quality Issues, S.K.Chakravarti Memorial Lecture , Eighth International Calcutta Triennial Symposium, December 27–30, 2012.

- BOGDAN, M., CHAKRABARTI, A., FROMMLET, F. and GHOSH, J.K. (2011). Asymptotic Bayes-Optimality Under Sparsity Of some Multiple Testing Procedures. *Annals of Statistics* **39**, 1551–1579.
- BHALLA, S. (2002) Imagine There's no Country: Poverty, Inequality, and Growth in the era of Globalisation, Institute of International Economics.
- CHAUDHURY, A., BOSE, M. and GHOSH, J.K. (2004). An Application of Adaptive Sampling to Estimate Highly Localized Population Segments. *Jour. Stat. Plann. and Inf.* **121**, 175–189.
- DEATON, A (2001). Counting World Poor: Problems and Possible Solutions. *The World Bank Research Observer* **16**, 125–147.
- GARNER, T.I., MCCLELLAND, R. and PASSERO, W. (2009) Strengths and Weaknesses of the Consumer Expenditure Survey from a BLS Perspective, Bureau of Labour Statistics.
- GHOSH, J.K. (1998). Emerging Priorities in the Indian Statistical System, Publication on the Occasion of Golden Jubilee of Indian Independence, Department of Statistics, Ministry of Planning and Programme Implementation, 9–12.
- GHOSH, J.K. (1999) 1. Strengthening Local Government in Madhya Pradesh, India. Project Report, Indian Statistical Institute, Kolkata.
- GHOSH, J.K. and MAITI, P. (2004) The Indian Statistical System at Cross Roads-an Appraisal of Past, Present and Future. Presented at IMS /ASAs joint mini meeting on current trends in survey sampling and official statistics. January 1–3.
- GHOSH, J.K., MAITI, P., RAO, T.J. and SINHA, B.K. (1998). Evolution of Statistics in India. *Int. Stat. Rev.* **67**, 13–34.
- GOLAN, A (1998) Maximum Entropy, Likelihood and Uncertainty: A Comparison, In G.J. Erickson, J.T.Rychert and C.R. Smith (eds.), Maximum Entropy and Bayesian Methods, Kulwer Academic Publishers, Dordrecht.
- GOVERNMENT OF INDIA (1993) Report of the Expert Group on Proportion and Number of Poor, Planning Commission.
- GOVERNMENT OF INDIA (2003) Report of the Expert Group on non Sampling Errors, Ministry of Statistics and Programme Implementation.
- GOVERNMENT OF INDIA (2005) Report of the Expert Group on Cross-validation study of estimates of private consumption expenditure available from household survey and national accounts, Sarvekshana, Vol.XXV(4) and XXVI(1), Issue No.88, 1–70.
- GOVERNMENT OF INDIA (2008) Report of the Group for Examining Discrepancy in PFCE estimates from NSSO Consumer Expenditure Data and Estimates Compiled by National Accounts Division, Central Statistical Organisation, Ministry of Statistics and Programme Implementation.
- GOVERNMENT OF INDIA (2012) Report of the Committee on Unorganised Sector Statistics. National statistical commission.
- GOVERNMENT OF INDIA (2014) Report on Basic Statistics for Local Level Development(BSLLD), Central Statistical Office, Ministry of Statistics and Programme Implementation.
- HAN, S (2011) The Final Consumption Expenditure of Households in Korea, Economic Statistics Department, Bank of Korea.
- KOLLI, R. and HAZRA, S. (2005) Estimation of informal sector contribution in the Net Domestic Product -Indian Experience. Expert group on Informal Sector Statistics (Delhi Group), 29-31 March.
- LAL, D., MOHAN, R. and NATARAJAN, I. (2001) Economic Reforms and Poverty Alleviation A Tale of Two Surveys, Economic and Political Weekly, Vol. XXXVI, 12, 1017–28.

- MAHALANOBIS, P.C. and SEN, S.B. (1954). On some aspects of the Indian National Sample Survey. *Bull. Int. Stat. Inst.* **34**, 2, 3–14.
- MAITI, P. (2007) Development of Statistical Information System for Decentralised Planning in the District of Howrah, West Bengal, India. Project Report, Indian Statistical Institute, Kolakta and Howrah Zilla Parishad, Howrah. *Preserved in Prasanta Chandra Mahalanobis Memorial Museum/Archive, Indian Statistical Institute, Kolkata.*
- MINHAS, B.S. (1988) Validation of large scale sample survey data: Case of NSS estimates of household consumer expenditure, *Sankhya*, B, 50,Pt. 3( Suppl.), 1–63.
- MINHAS, B.S. and KANSAL, S.M. (1990). Firmness, fluidity and margins of uncertainty in the national accounts estimates of private consumption expenditure in the 1980. *The Journal of Income and Wealth* **12**, 1, 92–102.
- MINHAS, B.S., KANSAL, S.M., KUMAR, J. and JOSHI, P.D. (1986). On reliability of the available estimates of private consumption expenditure in India. *The Journal of Income and Wealth* **9**, 2, 71–93.
- MURTHY, M.N. (1977) Use of empirical studies in evaluating sample designs for estimating frequency distributions, Proceedings of the 41st Session of the International Statistical Institute, New Delhi, India, 191–211.
- NEUVIAL, P. and ROQUAIN, E. (2012). On False discovery rate thresholding for classification under sparsity. *Annals of Statistics* **40**, 2572–2600.
- O.E.C.D. (2002) Measuring the Non-Observed Economy - a Handbook, O.E.C.D.
- RAJARAMAN, I. (2012) Core issues in a statistical system, Statistics and Development Issues, In Majumder et al.(eds.), Mittal Publications. New Delhi, India, 25-38.
- RAO, T.J. (2010a). Official Statistics in India: The past and the present. *Jour. Off. Stat.* **26**, 215–231.
- RAO, T.J. (2010b) Influence of C.R.Rao on theory and practice of sample surveys, In Proceedings of the International Conference of Frontiers of Interface Between Statistics and Sciences, 2, 663–673.
- RAO, T.J. (2012) Remembering Mahalanobis, Memorial Session Lecture, Eighth International Calcutta Triennial Symposium, 27-30 December, 2012.
- RAO, T.J. (2013) National Statistical Commission and Indian Official Statistics, *Resonance*, 1062–1072.
- RAVALLION, M. (2000) Should poverty measures be anchored to the National Accounts?, *Economic and Political Weekly*, Aug.26- Sept.2, 2000, 3245-52.
- ROBILLIARD, A.-S. and ROBINSON, S. (2003). Reconciling household surveys and national accounts data using a cross entropy estimation method. *Rev. Income and Wealth, Ser.* **49**, 395–406.
- SRIVASTAVA, A.K., RAI, A. and RAMASUBRAMANIAN, V. (2003) On reliability of estimates of inequality in distributions derived from sample survey data, Statistics, Division, Working Paper Series, F.A.O.
- SUNDARAM, K. and TENDULKAR, S.D. (2001) NAS-NSS Estimates of private consumption for poverty estimation A disaggregated comparison for 1993-94.
- THOMPSON, S. (2002). *Sampling*. Wiley, N.Y.
- ZELLNER, A. (1988). Optimal information processing and Bayes theorem. *Amer.Stat.* **42**, 278–84.
- ZELLNER, A. (1990). *Maximum entropy and bayesian methods*. Kulwer, Dordrecht, W.T. GRANDY and L.H. SHICK (eds.), p. 17–31.

P. MAITI

INDIAN STATISTICAL INSTITUTE, KOLKATA  
AND MEMBER WORKING GROUP  
69TH AND 73RD ROUNDS OF NSS  
KOLKATA, INDIA  
E-mail: pulakesh@isical.ac.in

T. J. RAO

INDIAN STATISTICAL INSTITUTE CRRAO  
AIMSCS  
HYDERABAD, INDIA  
E-mail: tjrao@hotmail.com

J. K. GHOSH

INDIAN STATISTICAL INSTITUTE,  
PURDUE UNIVERSITY  
WEST LAFAYETTE, IN 47907, USA  
E-mail: ghosh@purdue.edu  
jayantag1@gmail.com

Paper received: 3 June 2013.