INTERNATIONAL WORKSHOP ON OFFICIAL DATA

11 July, 2013

An international workshop on official Indian statistics was organised jointly by the Sampling and Official Statistics Unit (SOSU), Indian Statistical Institute (ISI) Kolkata and the Central Statistics Office, Industrial Statistics Wing, Government of India at ISI Kolkata on 11 July 2013. The day-long workshop was co-sponsored by the International Growth Centre, (IGC) India Central Programme. There were around 50 participants from government bodies such as the Central Statistical Organisation (CSO), National Sample Survey Organisation (NSSO), Micro, Small and Medium Enterprises (MSME) and Council of Scientific and Industrial Research (CSIR); international organisations such as the International Monetary Fund (IMF) and World Bank; academic researchers from Boston University, Yale University, Jawaharlal Nehru University (JNU) and University of California, Berkeley; and corporate users of Indian data.

Prasanta Pathak (Head, SOSU, ISI Kolkata) made introductory remarks. He said that the purpose of the workshop is to identify and bridge the gap between the officials involved in data generation and the users of data. The unique event was the first of its kind and there are plans of holding such workshops at other places as well in the future. Another objective is to popularise the use of official India data so as to justify the expense and effort that goes into generating the data. **Bimal Roy (Director, ISI Kolkata)** welcomed the participants to the workshop. He said that the workshop reiterates the commitment of the institute to working with the government on projects of national importance.

Session I: Perspectives of the data users

This session was chaired by **Pronab Sen (National Statistical Commission; IGC India Central).** He noted that in the context of the Indian statistical system, user interaction has been a one way conversation, and nothing much has come out of it. The workshop is an opportunity to have a more meaningful interaction between researchers and academics where the researchers can put down the nature of the problems that they face when using Indian statistics, and for officials to explain why data is collected in the way that it is. The complaints of data users are generally of three types: (i) Data is not appropriate for the purpose for which it has been collected on account of errors, data being out of date etc., (ii) Data is not suitable for purposes other than the ones for which it has been collected (common complaint by academics), and (iii) Lack of data on correlates that would allow for more analytical work.

With respect to the first problem, there is indeed a need for dialogue. The data officials are only generating the data and have no time to analytically examine it. Hence, problems come up in using the data and it is important that these be fed back into the system so as to enable a re-engineering of the processes. The second problem is more complex; academics have a research idea in mind, and the data they require may not be compatible with what is currently being collected for those purposes. For instance, the original purpose of the NSS consumption surveys and National Income Accounts was to feed into the consumption models of Plans. In the 60s, researchers began to use them to compute poverty and by the 80s, this became their dominant use. If there is a change in the basket, there are protests by academics that inter-temporal comparability is being destroyed. As a result, this data has now become useless for the planning model and they are too far removed from

the structure of the economy. With regard to the third problem outlined above, there is no formal forum through which new research issues can come to notice. User interactions are dominated by the government and the government demands descriptive data – a lot of resources go into satisfying these demands.

The objective of the workshop is to assess the appropriateness of current data, and to think about how to re-jig the processes with respect to existing data, and what new data is required. In doing so, one should keep in mind that the NSS is already over-burdened and academics should think of data processes that would satisfy more than one need.

Pranab Bardhan (University of California, Berkeley) delivered the keynote address based on his experiences with Indian statistics. He focused on data quantity and quality issues. He said that Indian statistics was a reputed enterprise at one time, and an example for many other developing countries and that this reputation is now in shambles. Data issues are hurting analyses of some of the basic elements of the Indian economy – growth, poverty and inequality. For example, with respect to *growth*, he said that even though services now constitute nearly 58% of national income, there is no separate price index for services yet. Further, he also pointed out problems with the Economic Census (EC) such as undercounting, erratic frequency, lack of training of enumerators etc. In the context of manufacturing, the fact that NSS and the Annual Survey of Industries (ASI) do not follow a common sampling framework causes problems. Moreover, we do not have information on the linkages between the organised and unorganised sectors, or on the quality of employment in the informal sector. Finally, we do not factor in depreciation of natural capital in growth accounting.

With regard to *poverty*, he pointed out that the data problems relate to the consumption expenditure figures and appropriate price indices for the poor. There are large and increasing discrepancies between NSS estimate and that from National Accounts statistics. Non-response errors are large and increasing. The price index is not covering the self-employed poor. Non-income indicators of poverty can be estimated from the National Family Health Survey (NFHS), which has a large sample but is not panel data.

With regard to the analysis of *inequality*, the usual Gini coefficients are misleading, as there is no income data, except for that produced by the National Council of Applied Economic Research (NCAER) and there are some doubts about that data as well. More important than the inequality of outcome is the inequality of opportunity, which depends on inequality of land, education, locality and social status. We have very little data on educational inequality by households. Even though intergenerational mobility is crucial for understanding the dynamic nature of Indian inequality, neither the NSS nor NFHS give panel data following the same households, and neither gives the parents' job history.

Finally, he recommended that the National Statistical Commission be made a permanent and independent statutory body, and the various statistical units of the Ministries should report autonomously to the Commission. Immediate steps need to be taken by the whole statistics community to address issues of shortage and poor quality of staff, inadequate resources and the general low priority given to statistics in government circles.

Next, **T.N. Srinivasan (Yale University)** delivered a lecture on the issue. He said that the blame game in the context of the recent Uttarakhand tragedy is a familiar story – the Met. Department said that they had provided forecasts, but the government said that they were not specifiable or actionable, and not useful for advance response. Similarly, the issue is that the data is too general for a response from policy. But in a sense, it is difficult to be more specific in advance. He outlined various issues with official data in the context of the CSO, NSS, industrial statistics, employment data, price statistics and poverty, and savings and investment statistics. According to him, the savings and investments statistics are among the weakest in the National Accounts Statistics. The household sector in these data in fact should be called "Sector other than public and private corporate sectors". The data for this misnamed sector are residuals and as such absorb the errors in measurement in the estimates for the other two sectors.

He further discussed two issues arising from the *'Statistical Federalism'* in India: Firstly, in NSS, two independent samples – Central and the state – are canvassed, so that states can do their own analyses on issues of interest to them. However, this is turning out to be an enormous waste of resources as most states do not make use of their samples. Secondly, while the CSO is responsible for National Accounts Statistics estimation of the country as a whole, they only publish the aggregate State Domestic Product (SDP) data series, but do not analyse the series and look at the problems.

He said that the Ministry of Statistics and Programme Implementation should put together and publish periodically a volume on *historical statistics* of India that would be a comprehensive coverage, and would include pre-independence data. In his view, the availability of historical statistics would enable more cliometric research on India.

He emphasised the importance of *experimentation*, for example, in terms of the size and shape of plots for cross-cutting experiments, choice of reference periods for household expenditure surveys, composition and length of questionnaires etc. – as was done in the early years of NSS. States should be used as laboratories to improve coverage, quality, timeliness, and reliability of data collection. The rapid changes in Information and Communication Technology (ICT) should be exploited.

He concluded with a couple of suggestions: (i) Sub-round estimates of the NSS can be used to generate seasonal adjustments, and could also be integrated into the quarterly GDP estimates. A panel feature could be introduced by revisiting for several rounds same set of villages and urban blocks if not households of one of the independent samples. (ii) Price indices and other data need to take into account changes in the quality of goods and services produced and used.

Dilip Mookherjee (Boston University; IGC India Central) talked about the research and policy issues pertaining to the Small and Medium Enterprises (SME) sector, and data problems in this context. He said that there are specific problems with generation of productive employment opportunities in the manufacturing sector on account of which growth acceleration in per capita GDP has not resulted in significant benefits for the bottom half of the population. To test possible causes in terms of governance factors or market/ technology access, data is required on firms at the firm and plant level, on an annual basis, combined with their external environment.

As the ASI excludes the unorganised sector entirely, the NSSO Employment/ Unemployment Surveys (EUS) are meant to represent the unorganised and organised sector employment, but they are undertaken once in five years and do not have a panel feature. Also, there are issues of

inconsistency of sample frame and questionnaires with ASI. He also outlined issues with the EC such as undercounting of very small firms. The need to 'stitch' together the ASI data with NSSO or EC is important to be able to say something about the importance of small enterprises or the unorganised sector, or when trying to compare productivity.

In conclusion, he said that there is a need for greater dialogue between the users and generators of data so that the users can ask questions regarding the data in the short run, and provide feedback to help design future surveys in the long run.

Michael Gechter (Boston University) presented key questions from his on-going research on the effect of size-based regulations on the distribution of enterprise size. The researchers plan to use NSSO and ASI data on organised and unorganised manufacturing and raised some initial questions regarding both datasets – for instance, he asked how national estimates are produced from the NSSO surveys for comparisons with the EC. He also said that the researchers would like to find additional data sources on small and medium enterprises such as employment data from the Ministry of Labour and Employment.

Sharon Buteau (IFMR Research) spoke about the currently available MSME data and unmet requirements. She said that the mission of the Small Enterprise Finance Sector (SEFC) at IFMR is to study how financial markets can foster growth of MSMEs and the role that MSMEs play in poverty alleviation. Important research topics in the context of MSMEs are access to finance, human capital and markets. She also pointed out major data gaps for research and policy in the MSME sector. While data is available on the amount of credit given, Non-Performing Assets (NPA) etc., data is lacking on time taken for credit processing, rejections, withdrawals, and so on, cost of obtaining credit from different financial institutions, and on collateral requirements. In terms of employment, data is available on the percentage of employment in MSMEs across rural, urban and state sectors, but not on the type and quality of employment, family businesses and the MSME labour force.

Session 2: Economic and Employment Statistics

This session was chaired by **V.K. Arora (CSO)** and **B.N. Goldar (JNU).** It involved presentations by officials from the CSO on the EC, ASI and Unemployment Surveys. Goldar raised a few issues, which he wanted the presenters to shed light on. These issues included deficiencies in the sampling frame, how to make better use of the state sample data, and whether it is feasible to combine state and central statistics to get more robust estimates. He further commented on the issue of comparability of data - employment and unemployment data is collected by various agencies such as ASI, NSSO etc.; however, often the data is not easily comparable. He raised the issue of limited resources and whether new surveys may be a burden on the existing resources. Lastly, he went on to talk about the utility of panel data and whether it will be possible to have panel data in household surveys.

Sunil Jain (CSO) made a presentation on the EC. He spoke about the meaning and relevance of the EC, international practices, other economic censuses in Asia, the procedure followed in the sixth EC and issues relating to India. He outlined the objectives of the sixth EC, which include providing information on operational variables (activity-wise for all establishments), providing data at lower geographical levels, like tehsils and villages, and on the number of workers in the unorganised sector. He said that the initial figures for the EC are likely to be released by December 2013, the All-India report by October 2014 and the final completion of the EC will be by March 2015. Finally, he suggested that the EC should have a periodicity of five years, so that it can be used for planning

purposes. Also, there should be an effort to create awareness about the EC among socio-economic ministries.

Bivas Chaudhuri (CSO) presented on the ASI data. ASI is the principal source of industrial statistics in India, and is collected by the NSSO. It provides statistical information to assess and evaluate, objectively and realistically, the changes in the growth, composition and structure of the organised manufacturing sector. The survey techniques are broadly based on the recommendations of the United Nations Statistics Division (UNSD), and international comparability is possible. A legal guarantee of the confidentiality of the data is given to the factories, and any unit that refuses to furnish ASI units or provides false information can be prosecuted. He described the reference periods of surveys, schedules of enquiry, scope and coverage, frame and sampling design of the ASI, as well as the industrial and product classification. He noted that the time lag in the release of data has been reduced considerably in the past few years. Moreover, certain new initiatives have been implemented such as e-schedules and an ASI web portal that includes frame updation, data collection, processing and dissemination. ASI has rich time series data spanning over 25 years. He also described how data is disseminated and how it can be obtained.

He highlighted a few shortcomings of the data, for example: (i) The Collection of Statistics Act prohibits disclosure of data relating to individual factories (ii) Sampling design is changing over time (iii) Industrial and product classifications are updated as per the UN recommendations, so concordance is a challenging task (iv) All estimates, especially for quantity and value figures for any particular item consumed and produced, are subject to Statistical Error as these are estimated on the basis of a selected sample.

Finally, for controlling non-sampling errors, he recommended record-based surveys, experienced field investigators, training of field investigators, close monitoring of field work, and data validation and scrutiny.

B.K. Giri (CSO) discussed issues in Employment and Unemployment Surveys (EUS). The major sources are the population census, EUS of NSSO, recent EUS of the Labour Bureau, and MSME, Ministry of Agriculture etc. for their respective domains. He said that there are differences in the estimates from the Population Censuses and the NSSO EUS for various reasons, and these have been addressed by the Expert Committee on Data Divergence. He also said that there are differences in the estimates from the Population Censuses and the NSSO EUS for various reasons, and that these have been addressed by the Expert Committee on Data Divergence.

He outlined the changes in concepts and definitions, and their impacts on the estimates. For instance, since the 61st Round of the NSS, information on voluntary participation without remuneration in the production of goods and services was added. These are not considered as economic activities in the NSS surveys. In terms of measurement, information on two subsidiary activities was collected in the 55th Round of the NSS to understand the extent of labour mobility.

Other issues discussed were the feasibility of providing absolute estimates in household surveys with reference to employment and unemployment indicators, usability of EUS for providing estimates of jobs, recording of data on earning from employment, additional categories for status classification, measuring informal sector employment, and definitions of decent work, green jobs and child labour.

Session III: Panel Discussion on future agenda for official data

The final session was chaired by **T.C.A. Anant (MOSPI, Government of India).** The users and generators of data discussed plans for the future for official data.

T.C.A. Anant began the discussion by highlighted the need for innovative solutions for two major problems - firstly, the official statistical system does not address the need for panel data to answer dynamic questions, and secondly, there are various constraints on the official statistical system in terms of resources, skilled manpower etc. He noted that the data already being produced is voluminous, and is stretching the scarce resources available. He further said that the panellists should suggest better synergies of available processes and ways to economising the use of existing resources, in order to be able to respond to the needs of the users of data.

Manisha G. Singh (CSIR) suggested using technology to improve data collection. For instance, GPS-enabled tablets can be given to enumerators that will show how much time they spend in the field interviewing respondents. This is something that is already being done by corporates. **Sandip Mitra (ISI Kolkata)** said that the time of data collection should be specified in NSS. Data required for behavioural economics studies such as on aspirations should be incorporated in surveys. Also, an understanding of formal and informal sectors is very important in the Indian context and data collection should take this into account.

Other important points raised during the discussion were the trade-off between resources and reliability; the quality of data; insufficient data on the rich and associated problems with capturing inequalities; lack of state-level capital stock information that is required for sub-regional growth analysis, limited utility of yearly surveys such as the consumer expenditure survey to researchers; and inability of official statistics to capture the dynamic nature of the economy. The possibility of creating pseudo-panels to capture dynamic incentives was also explored.

Concluding Remarks

In conclusion, Pronab Sen said that given the resources, we need to change the way we do business. He noted that the NSSO has been doing the same things in the same way – this cannot carry on going forward.

We need to consider the needs of policymakers as well as the needs of academics, which are often conflicting, and prioritise. It is important to re-visit some questions with regard to the type of data being generated. For example, why not discontinue household consumption surveys if they are not being used? We need to have a debate on the need for panel data vis-à-vis cross sectional data. He further said that this workshop was a good beginning, and we need more precise interactions on the areas of focus and the nature of surveys. However, we cannot give everyone what they ask for as the system does not permit this. The changes that are required need to be brought about in a collective manner.

T.C.A. Anant said that e-governance efforts are being made. Usable statistics can now be extracted. However, this is at early stages. For example, there is a partnership with the corporate affairs ministry and datasets on company returns can be mined online. There is an on-going exercise in partnership with the department of revenue for creation of an excise database. Finally, he said that there is also a need to manage expectations and keep feasibility in mind.

Nachiketa Chattopadhyay (ISI Kolkata) thanked all the participants and closed the workshop.