PoS
Lecture Note (Contact hours 2)
The Making of a Statistical Database in Early Postcolonial India

Background Information: At this stage of the course we are going to study the making of the statistical database in India at the behest of Prof. P.C. Mahalanobis and the ISI. I am going to circulate an essay by Arunabh Ghosh which, in my view, is self-explanatory. Kindly read the essay and treat this note as providing a set of background information. The following is extracted from one of my projects that studies certain anticipations of the 'Big Data' phenomenon in three inter-connected developments of the mid-20th century, i.e., the emergence of algorithm as the science of computation, the birth of the idea of an integrated 'national economy' as a central object of postcolonial governmentality, and the triggering of the Cold War 'Big Science' initiatives that required data management at a cosmic scale. The project identifies the Indian Statistical Institute (ISI) in Calcutta, Gokhale Institute of Politics and Economics in Pune and the Planning Commission in Delhi in the 1950s as sites where these three inventions interfaced in 1950s. Taking the year 1951 as a vital moment in the history of data in India, the project further shows how the ISI emerged as a key player in the science of government. In short, the project attempts an Indian history of big data, and thus it contributes to the critical literature on the subject that appears to privilege context and circumstance over historical antecedents to the overwhelming accumulation and circulation of big data in the current conjuncture. You may contact me if you are interested in knowing more about the project.

Decades before the introduction of the census operations in India, the colonial government insisted the provincial governments to publish relevant statistical accounts in their Annual Reports. Between 1840 and 1865, such accounts were made uniform and published in the form of the Statistical Abstract of British India. In 1862, the British Government in India constituted a Statistical Committee at the behest of which the robust Statistical Abstract of British India came out in 1868. Containing returns of provincial administrations, this publication became a regular annual ritual of the British government until 1923. Between 1860s and 1880s, various branches of the imperial government in India established statistical wings. Thus, successive famines in 1870s prompted the Agriculture Department to come up with Agricultural Statistics of British India in 1886. In 1862, the Finance Department established a statistical branch, which in 1895 morphed into a Statistical Bureau with a subsequent addition of a commercial intelligence wing to it in 1905.

Until the inter-war period however, official statistical enterprises in India outside the census operation remained scattered, uncoordinated and largely decentralized. In 1925, the Economic Enquiry Committee was instituted which consisted of Sir M. Visveswarayya as Chairman, and Pt. Harkishen Kaul and Prof. Burnett Hurst as members. The objective of the said Committee was to 'examine the material at present available for framing an estimate of the economic income of the various classes of the people of British India, to report on its adequacy, and to make recommendations as to the best manner which it may be supplemented, and as to the lines on which a general economic survey should be carried out, with an estimate of the expenditure involved in giving effect to such recommendations' (The Legislative Assembly Proceedings, 22 January, 1925). The Committee was also asked to review the 'question of adequacy of the statistical data available and the desirability and possibility of supplementing it, and of undertaking an economic enquiry' (MOSPI, undated: http://www.mospi.gov.in/141-historical-perspective). In addition, the Legislative Assembly desired that the same Committee be asked to conduct the Economic and the Taxation Enquiries. Eventually, the Government appointed two separate Committees to undertake the task. The Committee's recommendation was that both central and provincial governments should come under one central authority, a central statistical bureau of

sort, with provincial branches to 'provide a common purpose and a central thinking office on the subject of statistics' (MOSPI, undated: http://www.mospi.gov.in/141-historical-perspective).

Nothing very significant happened before the end of the Second World War. In 1945, Government of India appointed an inter departmental committee which again recommended the establishment of a) a central coordinating statistical office, b) statistical bureaus in provincial headquarters, c) a statistical cadre. After Independence, P.C. Mahalanobis was appointed as the statistical advisor to the government in 1949 and in the same year, a micro statistical unit was inaugurated in the cabinet secretariat which in 1951 emerged as the Central Statistical Organization (CSO) being entrusted with the work of coordination with various organizations producing statistics. In 1949, a national income committee was set up with Mahalanobis as its Chairperson and V. K.R.V. Rao and D.R. Gadgil as members to estimate national income. The Committee was externally advised by a distinguished panel comprising J.R.N. Stone (Cambridge), Simone Kuznets (University of Pennsylvania, who devised national income accounts in the US) and J.B.D. Derksen (the UN Statistical Office). It is at this point of trying to quantify national income with an aim to figure out the pattern of income distribution [which would form the basis of a general revision and rationalization of the existing tax structure], and ultimately to predict the trend of economic growth in the country that the key functionaries of India's statecraft began to make an inventory of the existing economic database, infrastructures, and started taking into account the kind of statistical and infrastructural apparatus they required to think of an Indian economy. Thus, while setting up the Committee, the Government's resolution (No. 15(33)-P/49, dated the 4th August, 1949) mentioned the following:

> The Government of India have been giving consideration for some time to the inadequacy of the factual data available for the formulation of economic policies. One important gap is the absence of authoritative estimates of the national income and its various components. The Government of India have accordingly decided to set up a committee to advise how best this gap could be filled up (GOI-NIC 1951: 1).

The National Income Committee encountered three major problems that appeared to the Committee members to be unique to the Indian condition when compared with the advanced industrialized countries who had initiated the process of measuring their national economies a number of years ago. The first was what the Committee called the 'problem of measurability':

> …when calculating the value of output, one normally proceeds on the assumption that the bulk of the commodities and services produced in the country are exchanged for money. In the case of India, however, a considerable portion of output does not come into the market at all, being either consumed by the producers themselves or bartered for other commodities and services. The problem of imputation of value thus arises and takes on significantly large proportions in some sectors of the economy (GOI-NIC 1951: 12).

The second major problem that the Committee encountered emanated from India's 'comparative lack of differentiation in economic functioning':

> While it is true that we have a sector in our economy that is as differentiated and subject to modern income classification as in the west, it is also true that a major portion of our economy consists of household enterprises, simultaneously and without differentiation performing

> functions which would normally fall under different industrial categories. Thus, sizable groups among agricultural producers pursue other occupations in other industries, often in urban places or at any rate outside their domicile. Hence the customary classification of national income by industrial origin cannot be taken except as a rough approximation to a classification of distinct groups in the population, whose main income is derived from a single industry (GOI-NIC 1951: 13).

Hence, the Committee felt the genuine need for a 'substantial revision of the industrial classification and a much greater emphasis than is customary in the west upon social groupings connected with the character of the enterprise rather than with industry would not be more useful for India' (GOI-NIC:13) that called for 'a great deal of analytical work' and a reworking of the existing classificatory schema 'developed in the west'. A third round of problems concerned with 'the non-availability of statistical data for the estimation of income and related accounts in India' (GOI-NIC:14), for which fresh surveys outside the scope of the measurement of national income appeared necessary. Over all, the task before the National Income Committee was to create units of measurement and equivalence to be able to place the Indian case in a comparative scale in the competitive arena of nation states. Ideologically, thus invented economy appeared to carry the very essence of the newly liberated nation—the amphitheatre of governmental action. Understandably, such a project needed to be backed by a credible process of survey that would bring into being a community of producers and consumers within a overarching national frame.

On 18 December 1949, the Prime Minister expressed his desire that a sample survey be organized 'covering the whole country to collect essential information'. Mahalanobis drafted a proposal for the NSS and handed over the same to the Finance Minister C.D. Deshmukh on whose advice the National Income Committee finally recommended the use of the NSS data for national income estimation. Starting in 1950-51, this multipurpose sample survey became the biggest and the most comprehensive sampling enquiry in the contemporary world. As a round-the-year and continuous data collecting and data processing machinery, the NSS was tasked to collect reliable data concerning production, consumption and various other kinds of data to better comprehend the emerging trends in the national economy. The Annual Report of the ISI for the year 1950-51 gives a comprehensive picture of the first round of NSS data collection:

> The whole of rural India has been split up into 156 strata, the formation of each stratum depending on geographical contiguity and homogeneity in topographical characters. Wherever the necessary data have been available, each stratum has been further divided into four sub-strata according to the population of the village comprising them. A total of 1833 villages have been selected for survey, and the quota of each stratum has been made proportional to its population. In each village a random sub-sample of the households has been studied in respect of the principal occupation, and the households comprising the sub-sample have been divided into 'agricultural' and 'non-agricultural' groups, those engaged on a primarily agricultural enterprise being treated as agricultural and the rest as non-agricultural. Further sub-sample have been drawn in both of these groups for detailed examination of general demographic and economic characteristics, production and cost data in the enterprises concerned and consumers' expenditure in the domestic field (ISIAR 1950-51, 2).

Already in 1950, in a paper Mahalanobis stated that 'with an accepted level of precision the costs of [random] sample surveys are only about ten percent of that of a complete enumeration (cf.

Ghosh 2016, 6). His long-term interlocutor Ronald. A. Fisher mentioned in addition that in complex and vast countries such as China and India, sample surveys were much more scientific and economically viable mode of statistical operation (Ghosh 2016).

Along with its routine work, the NSS also began to undertake occasional and special surveys on the pressing social and economic issues affecting the country. Thus, apart from expanding the scale of the sample survey, the NSS in ISI conducted a number of surveys on behalf of various ministries. Thus, in a couple of years of its operation, the NSS produced impressive ad hoc surveys on the displaced persons [due to Partition] in the states of West Bengal and Bombay (for Ministry of Rehabilitation), collected information on country-wide basis for the Press Commission (for Ministry of I&B), surveyed the magnitude of unemployment in Calcutta and other metropolitan areas (for Planning Commission), came up with a study of the housing of the working class (for Ministry of Works, Housing and Supply), and so on. Soon, the NSS became one of the most credible databases for social and economic research, the macro-economic and demographic policies of the Government. Within six years of its operation, the NSS became so robust that the American statistician and the pioneer of the statistical quality management Edward Deming wrote:

> No country, developed, under-developed, or over-developed, has such a wealth of information about its people as India has in respect to expenditure, savings, time lost through sickness, employment, unemployment, agricultural production, industrial production. We in this country, though accustomed to work in large sample surveys were aghast at Mahalanobis' plan for the national sample surveys of India. Their complexity and scope seemed beyond the bounds of possibility, if not beyond anyone else's imagination, but they took hold and grew (cf. Rudra 1996: 204).

Needless to say, such a permanent and accumulative work of data collection required the setting up of a permanent bureaucracratic, scribal, field level and mechanical installation at multiple levels. A look at the Annual Reports of the ISI in 1950s gives you a sense of how such an installation came into existence over a short period of a decade. Mahalanobis' induction to the core of the Nehruvian statecraft on the other hand brought into being a promising unification of official statistical initiatives with the research tradition in descriptive statistics of the ISI. The Institute began to receive liberal funding from the government that enabled it to quickly elevated its esteem within India. In addition, Mahalanobis' global intellectual collaborations made it a nodal point in thinking about the new sciences of the state and the new data regime for planning national development outside the well-known western set-up. On 9 December 1956, Chinese Premier Zhou Enlai visited ISI to specifically understand the modalities of the NSS operation in India. When he was visiting the NSS wing of the ISI, the following conversation took place between Enlai and Mahalanobis:

> Zhou Enlai: Which are the countries most advanced in statistics? Are you in touch with them?
> Mahalanobis: UK, USA and USSR. We are in touch with all three and we accept from each what we find useful. We have found, however, that in applied work, it is not desirable to copy from any country. *In India, we are trying to adopt and develop the methods to suit our own needs.*
> Zhou Enlai: [nodding his head vigorously]: Yes, yes. One group of Chinese statisticians will soon come here. I want them to see everything in detail. We want to learn from you… (Mahalanobis to Pitambar Pant,

16 December 1956, File No. 90 pp. 11-12, PCMMMA, ISI, emphasis ours, also see Arunabh Ghosh, 2016).

'In India, we are trying to adopt and develop the methods to suit our own needs'. Exactly four years down the line, Mahalanobis introduced a new method of analysing the NSS data called the 'Fractile Graphical Analysis', which C. R. Rao (1973, 480) describes as 'a semi-parametric method for comparison of two samples' to 'find a disaggregated measure of changes in income inequalities or disparities in consumption over time, a measure that reflected what was happening to the levels of living of different income groups over time' (Srinivasan 1996: 242-43). This mode of analysis was seen as a decisive shift of theoretical perspective from aggregate measures such as Gini Ratio to enable Mahalanobis to compare 'the distribution of total consumption expenditure from NSS data for the eighth round (July 1954-March 1955) when prices were low, with the distribution for the 16th round (July 1960-June 1961) when prices were high' (Srinivasan 1996: 244). His analysis revealed that 'with an increase in prices, the distribution of expenditure on consumption did become more equal' (Mahalanobis 1975: 1166, cf. Srinivasan 1996: 245).

Mahalanobis also devised a technique to address 'one of the long-standing peculiarities' of the Indian labour question [when compared with the advanced Western economies], i.e., most of the workers are involved in domestic home-based industries in which they did not have tangible jobs to lose to be registered as unemployed, even though they might not be having gainful work. To address this and to standardize data in this field, the 9th round of the NSS collected data on unemployment and underemployment 'on hours of work done per week and number of days of gainful work during 30 days preceding the day the interviews were conducted' (Srinivasan 1996: 246).

In a sense, the NSS brought into being a community of producers and consumers as an economy articulated as the very essence of the liberated Indian nation. The task of nation-building revealed itself as the task of bringing producers and consumers into a single framework for governmental interventions in the form of national planning—one that would render economy with anthropomorphic features (Deshpande 1993). Thus, the postcolonial economy became a site in which decolonization (from enslavement to liberation) was to be performed in complex exchange between the empty-homogenous time of the nation and the dense-heterogenous time-space of governmentality. Interesting enough, the Big Data moment arrives at a moment in the postcolony [this was around the time of the so-called GDP revolution (Samaddar 2018)], when the conception of the economy is estranged from the collective conception of the nation.

The emergence of the NSS overlapped with the emergence and continuation of the operations of a host of specialized public data collection institutions such as the Registrar General and Census Commissioner of India (under the Ministry of Home Affairs, mandated to organize decennial census operations, and linguistic surveys), Archaeological Survey of India, Botanical Survey of India, Forest Survey of India, Geological Survey of India, National Institute of Oceanography, Zoological Survey of India and so forth. In addition, the Reserve Bank of India and various public-sector banks, the Securities and Exchange Board of India, the National Crime Records Bureau, etc., produced impressive corpuses of sector-specific data (Samaddar 2018). In this context, the contribution of the NSS was to produce large scale and continuous data (as opposed to the discrete and periodic data collected by the census operations) on the government and market initiatives to imagine the national population as a mass of consumers as hinged on the state-capital nexus (Samaddar 2018). A look at such initiatives may enable us trace the history of state capability of the management of public data in India prior to the arrival of the Big Data moment.