India

Sources:
Swamy 1967, Tables 4, 5 and 6 for deciles, Table 3A (in appendix) for Ginis
India, NCAER 1972, Table on p.26
Paukert 1973, Table 6 p.104-105
Jain 1975
Cromwell 1977, Table 1
Dowling and Soo 1983, Table 2 in the appendix
Lecaillon et al. 1984. Table 4 p. 26-27
van Ginneken and Park 1984
Fields 1989
Datt 1999, Table 1
Deininger & Squire, World Bank 2004
Sen and Himanshu, 2004
World Bank India Database

Surveys:
National Sample Survey (round 3 to 55) 1951-2000
The survey was conducted for the first time in 1951, but starting from 1972-73 the National Sample Survey Organisation (NSSO) adopted a program of quinquennial surveys on consumer expenditure that would provide time series of household consumer expenditure data. Besides the quinquennial surveys or the so called thick samples, expenditure surveys are conducted for the period between the successive quinquennial rounds using a much smaller or so called thin sample, with somewhat different sampling designs. The last thick rounds were round 43 (1987-1988), round 50 (1993-1994) and round 55 (1999-2000).

The sample households for consumption inquiries are selected on the basis of probability proportional to population. A two-stage stratified sample with the urban and rural sectors separately divided into a number of strata is selected. Villages are the first stage units in the rural areas, stratified on the basis of similarity with respect to population density or crop pattern. Blocks are the fist stage units in the urban areas, stratified on the basis of population sizes of towns and cities. The second stage units are households in both sectors. The villages and blocks are obtained from different censuses that have been used as the sampling frame. NSS covers only private households; homeless persons and institutional population are excluded. In Round 55, the whole of the Indian Union was covered except (i) Ladakh & Kargil districts of Jammu & Kashmir, (ii) interior villages of Nagaland situated beyond 5 kilometers of a bus route & (iii) villages of Andaman & Nicobar Islands remaining inaccessible throughout the year. The data are collected in sub-rounds. In round 55 there were four sub-rounds, each with a duration of three months.

In round 55 a major change took place in the survey as the recall periods for consumption items were changed. Instead of a uniform 30 days recall period for all items, mixed recall

periods were adapted. Food items, pan, tobacco and intoxicants were asked for last 7 and last 30 days; fuel and light, miscellaneous goods and services including non-institutional medical, rents and taxes for last 30 days; and educational, medical (institutional), clothing, bedding, footwear and durable goods for last 365 days. The 7 days recall periods for food increased the estimates of food consumption by 30%, but to keep a better comparability with earlier surveys, the 30 days recall periods were used for the consumption aggregates. It appears however that the inclusion of both 7 and 30 days questions on food has resulted in "contamination" of the 30 days estimates and weakened the comparability with previous estimates. In addition, the 365 days recall period for less frequent items made the distribution of consumer expenditure more equal. This happened since the longer reference period also made poor households report durables whereas rich household reported less of these due to recall problems. This pulled up the bottom tail, increasing the share of the poorest 40% by nearly 10%. The fall in inequality in 2000 as reported by Sen and Himanshu (2004) is not due to comparability problems with other sources but due to the methodological differences in the survey (in fact they report almost exactly the same estimates as the World Bank India Database for earlier periods).

The consumption aggregate includes consumption (in value) of goods and services, out of home-grown stock, out of monetary purchases, out of receipts in exchange of goods and services, out of gifts, loans etc. The items are separated in a very detailed list of expenditures with details on food, medical care, education, housing, utilities and other items.

In Swamy (1967) all-India distributions were derived by pooling rural and urban distributions. Polynomial regressions were used to derive share of end-groups, using cumulated proportions. According to Swamy (1967) the weakness of the data is changes in reference or survey periods, changes in the valuation of retained produce and exclusion of certain items, e.g. imputed rent.

The data reported by the World Bank India Database have been corrected for fire wood prices which in the original survey was assumed to be constant over the rounds. This does not have a big impact on the Gini coefficients. The methodology for deriving the estimates is based on parametric Lorenz curves. In the World Bank India Database the urban means are given in rural prices whereas they are given in urban prices in Datt (1999). Otherwise these two sources are comparable.

The consumption aggregate used by Deininger & Squire (2004) does not include durables and imputed rents. This is probably also the case in the World Bank India Database and Datt (1999).

All India Household Survey of Income, Saving and Consumer Expenditure 1956, 1964, 1968, 1975 (also called All India Consumer Expenditure Survey)

This survey is conducted by the National Council of Applied Economic Research. WIDER has documentation that mostly applies to the 1968 survey. The coverage was almost national with some smaller areas and islands excluded; still 98% of the population is reported to be covered. A multistage sample design, featuring over-sampling of the

relatively high income households was adopted for both urban and rural areas. The effect of over-sampling of high income households was later offset by assigning to each sample household a suitable weight being the inverse of the inclusion probability. The sample size was 2550 households in urban areas and 2380 in rural areas. Even if household is defined as the statistical unit, the definition is narrow: the household members are required to be related by blood, marriage or adoption and to have been living together for no less than 6 months of the reference period to be considered as living in the same household.

The income concept reported for 1968 is disposable income; the income taxes have been calculated separately and deducted from the gross incomes. Only income tax from non-agricultural income has been deducted. Gross income is defined as the sum of salaries and wages, self-employment income from agricultural and non-agricultural activity, rents from land and property, interest and dividends on financial investments, pensions and regular and irregular contributions. In addition, imputed rental income from owner occupied dwelling and imputed income of own labour input in agriculture have been taken as a part of household income. The reference period is the agricultural year 1967-1968 (i.e. July 1967 to 30th June 1968).

The 1975 survey covered 5125 households and was collected July 1975 to June 1976. In van Ginneken and Park (1984) it is reported that the survey did not account for the 600000 homeless households. In their study the population data were adjusted assuming that the homeless belonged to the lowest income class.

National Sample Survey 2009-10, 68th Round: Detailed information available in the document.....