DATA SOURCES ON INCOME DISTRIBUTION IN BANGLADESH, INDIA, PAKISTAN AND SRI LANKA: AN EVALUATION

BY INDIRA RAJARAMAN*

This paper examines the data base available in four South Asian countries, India, Pakistan, Bangladesh, and Sri Lanka, for the examination of trends in real inequality and poverty. Taking the position that sample surveys of household income and consumption are the only really adequate bases on which size distributions of income for a less developed country can be constructed, the paper examines in Section I the reliability of the surveys available in the four countries. Section II evaluates available price data. Section III looks at directions for future development of data collection. The conclusion is reached that sample surveys regularly conducted in these countries do not provide a particularly good basis for this type of analysis. Needed alterations include permitting access to the primary data (or redesign of published tabulations to meet the needs of this type of analysis), use of *per capita* rather than total household income and consumption, better coverage of regions and occupations, and exploitation of the price data implicit in the survey data collected. Further, the surveys themselves need to be overhauled, especially with regard to timing of interviews. The paper concludes with a short discussion of alternatives to estimates of inequality that can be used to measure absolute deprivation, such as the QUAC stick for identifying nutritional insufficiency.

INTRODUCTORY

This paper examines, for the four South Asian countries under review, the data base available in each for the construction of trends in real inequality and poverty. The countries will not be dealt with separately in sequence; instead, the treatment of all four will thematically be united under criteria of evaluation of quite general applicability. Comments will be provided alongside on studies that have been done using these data in either primary or tabulated form, though the focus throughout will be on the quality of the underlying data rather than on the results of the studies based on them.

Sample surveys of household income and consumption are the only really adequate bases on which size distributions of income for a less developed country can be constructed. The alternative of using income-tax statistics is not available in such countries where income taxes are typically not levied on the rural sector where the majority of the population lives, and where only a small fraction of urban incomes are subject to taxation. The further alternative of piecing together income distributions from data on wages and salaries, and data on the distribution of agricultural land and other property together with figures on the average productivity of such assets is tedious and at best incomplete. The considerable self-employed sector consisting of rural artisans, urban small business, and traders both urban and rural is left out of such a reckoning—unless data on this sector is

*Indian Institute of Management, Bangalore 560027, India.

The author is grateful to the Woodrow Wilson School, Princeton University, for their hospitality during the 1973-74 academic year at which time the research for this paper was done.

Prof. John Mellor of Cornell University, Messrs. James Harrison, Nicholas Krafft, and Ralph Hofmeister of the International Bank for Reconstruction and Development, Mr. W. Tilakaratne of the International Monetary Fund, and Messrs. S. C. Chaudhri and G. R. Ayachit of the National Sample Survey Organization, Delhi, provided generous access to their materials. The author is deeply indebted to them.

available from a sample survey. Further, the wage data must have a coverage large enough to take care of seasonal, sectoral (urban/rural, agricultural/non-agricultural), geographical, and skill variation, and be accompanied by figures on the incidence of yearly employment at each of these wage-levels (abstracting for the moment from the problem of intra-sectoral, intra-regional, intra-skill variation in the availability of such employment). Lastly, and most importantly, the three groups of wage-earners, the self-employed, and those deriving income from property (including self-cultivated agricultural land) are by no means mutually exclusive, with many workers drawing support from more than one such class of activity, so that the final size distribution of income is an amalgam rather than a summation of its various parts.

Thus, this paper will deal exclusively with surveys of household income and consumption, which are conducted systematically in all four countries with varying periodicity. In Section I a careful assessment is made of the reliability of these surveys. A comprehensive listing and evaluation of other relevant but piece-meal sources of data, such as those mentioned in the preceding paragraph, is available elsewhere (Rajaraman, 1974).

For any estimation of trends in the absolute condition of the poorest percentiles of the population, or of the degree of relative inequality in real as opposed to nominal terms, a second crucial input must be price data. Section II explores the options available on this front. At the end of both sections it will be seen that the outlook for reliably estimating past trends is not especially good in any of the countries. The shortcomings in some cases are such that the direction of bias can be established, but in studying trends it is important to know in addition if the degree of bias itself has changed.

In Section III directions for future data collection are looked at.

Section I: Sample Surveys of Household Income and Consumption

This section deals with those surveys of household income and consumption that are conducted by the official statistical apparatus of the countries on a systematic basis. These are the most comprehensive sources available for an estimation of trends. Other surveys have been conducted, but they are more limited in their coverage and have usually been done on a one-shot basis so that they can yield estimates only for a point in time; these surveys are covered in the source cited earlier (Rajaraman, 1974).

The assessment that follows is done at two levels. At the first level, if the primary data will be available, the relevant considerations are the reliability of the sampling frame used, the sample design adopted, and the concepts used in the formulation of the questionnaire. At the second and more likely level, if the primary data from the surveys are not available, there is the additional consideration of the quality and coverage of the tabulated output.

1. *Year of First Survey; Periodicity*

Surveys of household income and consumption were conducted in the rural sector of Bangladesh and Pakistan, separately in each wing of what was then one

country, as early as 1959–60, but the first nation-wide survey was held in 1963–64. After the next such survey in 1966-67, complete surveys have been conducted every single year.[1] In the eastern wing the disruptions of 1970–71 which led to the emergence of Bangladesh as a separate nation put a temporary halt to the series.[2] Thus, because all the surveys done so far in what is now Bangladesh were conducted when it was still a part of Pakistan, they will be referred to throughout as Pakistani surveys; the countries will be referred to as if they are three in number rather than four.

In India, sample surveys of household consumption have been conducted every year since 1950 in the course of the yearly rounds of the National Sample Survey (NSS). Starting with the 29th round (1974–75)[3] however, these surveys will not be a part of every round; they may be held as infrequently as once in five years. In some of the consumption surveys, the questionnaire used was designed to collect information on income as well. The income data so collected is of doubtful value (as is that collected in the Pakistani surveys, and for much the same reasons, but more on that further along), and the practice of collecting income data was discontinued with the 26th round survey.

The first Sri Lanka survey of household income and consumption was conducted in 1953; since then, they have been conducted every ten years, in 1963 and again in 1973.[4]

The most desirable periodicity for surveys of this sort is yearly for at least two reasons. Samples from adjacent surveys can be pooled if individually they are not sufficiently large to provide reliable estimates. Further, it is important for any study of trends that the points of comparison, the survey years, be "normal," free of climatic excesses, war and so on. Since the normality of any year is difficult to predict in advance, it is useful to have yearly surveys from which to choose points of comparison.

2. *Sample Design*

(a) *Sample selection.* A reliable frame for sample selection is provided in India and Pakistan by the population censuses conducted in each country at the beginning of each decade, although some distortion might be involved at the end of any decade if patterns of natural increase and migration are different from those predicted. In Sri Lanka, where there is universal rice rationing, the household lists maintained by the government for the purpose provide an up-to-date sampling frame.

Sample selection in all the surveys typically involves a one to three stage selection procedure from each of the several strata into which the country is first

[1]The series is entitled the *Quarterly Survey of Current Economic Conditions* and is conducted by the Central Statistical Office. The "Sample Survey of Household Income, Expenditure, Savings and Liabilities" is a part of the Quarterly Survey, the other part being the "Sample Survey of Pakistan's Labour Force and its Selected Characteristics" which collects data on employment. Details on the latter are available in Rajaraman, 1974.

[2]Efforts were made in 1974 by the Bangladesh government to get the surveys going again; whether these efforts were successful or not could not be ascertained.

[3]Some of the early rounds were held two or more to the year; this accounts for the the fact that the 29th round was held in the twenty-fourth year after the start of the series.

[4]The "Sample Survey of Ceylon's Consumer Finances" is conducted by the Central Bank of Ceylon in association with the department of Census and Statistics.

divided. The foremost stratification is by sector: rural and urban in the case of the Pakistani and Indian surveys, rural, urban and estate in Sri Lanka. Within each sector, there is a first-stage random selection from each sub-stratum of larger sampling units, for example villages in the rural sector, from among which the final set of sample households is randomly picked. This is the typical procedure used in all the surveys, with minor variations such as in the number of stages.[5]

The sampling fraction in the 1970–71 Pakistani survey was 1/400 in the urban sector and 1/2,000 in the rural, and appears to have been approximately the same since the 1966-67 survey. The fractions for the Sri Lanka surveys were approximately 1/1,400 in 1953 and 1/400 in 1963, uniformly for all sectors. (Information on the 1973 survey was not available.) Sampling fractions in India vary widely from state to state, since sample selection is done separately for each. In general, the sampling fraction varies inversely with the size in terms of population of the state, and has increased steadily over the years for all the states.[6]

On the whole, the procedures used for sample selection appear well-conceived and are in all probability well-executed. One problem, however, arising out of the use of the "household" as the ultimate sampling unit, is that in the urban areas of these countries a considerable segment of the population is houseless and not readily identifiable in terms of households. The bias against the inclusion of this population is serious and must be kept in mind whenever these data are used to estimate urban inequality, or more especially, the extent of urban poverty.

Another bias, one frequently noted, has to do with the upper-income households, both rural and urban; samples obtained seem to under-represent these. This bias, however, arises more due to non-response than out of any fault of sample selection. The problem here is that the degree of bias can change randomly over time. For surveys of urban incomes, a correction factor can be obtained from income-tax sources. But for consumption data such correction is much harder to do, and with rural data for either income or consumption the option is not available at all.

(b) *Survey period.* The Pakistani and Indian surveys have a year-long survey period usually going from July to June.[7] Each sample household is visited and interviewed only once in the course of the year; the date of survey is different for each.

The Sri Lanka surveys are designed differently. All sample households are surveyed at the same time. The surveys are held in April which marks the end of

[5]Occasionally, the sample design calls for a stratification of the households falling in the sampling units selected at the first stage; this is done when it is felt necessary to ensure that the final set of sample households will give equal representation to each of a certain type of household. Aside from this, the only major departure from the procedure described was in the 1953 Sri Lanka survey which was based on cluster sampling.

[6]There were a few discontinuities in the steady increase. For the rural sector there was a doubling in sample size with the 8th, 16th and 18th rounds. For the urban sector, sample size was doubled in the 13th, 14th and 18th rounds.

[7]This has been the case in India ever since the 14th round (1958–59). Before that each round varied in duration from three months (the 2nd round) to nine months (the 8th and 13th rounds). Even after the 14th round there have been a few exceptions; details are available in Rajaraman 1974.

one of the two major agricultural seasons in the country.[8] Both these factors, the timing and the simultaneous coverage, are an improvement on the non-simultaneous, year-round design of the Pakistani and Indian surveys; the reasons will be given in the next part of this section.

3. *Questionnaire*

(a) *Reference period.* The foremost point of interest in the questionnaire, to be looked at in conjunction with the timing of the survey, is the period of reference with respect to which the data on income or consumption are collected.

There are two conflicting considerations here. Since the average respondent in these surveys cannot readily provide totals for what he earns or consumes, and since these totals must therefore be arrived at through a careful accounting of the components of each, the period of reference must be short enough so that such information can reliably be recalled. At the same time, the reference period must be long enough so that the figures show what is usual for the household, without seasonal or other distortions.

With consumption data, the second consideration is not as important as with income data. Consumption is much less subject than income to seasonal variation over the year. Thus, for the collection of consumption data, reference periods as short as the month preceding the date of survey (as in the Pakistani and Indian surveys) or even as short as a week (the Sri Lanka surveys) are quite adequate—periods that would be entirely too short for the collection of reliable income information. The shorter period of reference makes for greater accuracy of recall so that in general consumption data are more easily and reliably collected than income data. The timing of survey, however, is important because the seasonal element in consumption is not entirely negligible. If, as in the Sri Lanka surveys, the households have all been surveyed at the same time, the direction of the seasonal bias would be the same for all the households, so that the distribution among them would not be distorted unduly. But where the date of survey varies from household to household so that the reference month is different for each, some part of the variation in consumption figures must be attributed to seasonal variation and the data must first be de-seasonalized.[9]

For income data on the other hand, a reference period of a month—used for example in all but one of those of the Indian surveys that collected income data[10]—simply would not suffice. However, the reference period used in the

[8]The Maha season goes from Sept./Oct. to March/April so that the survey would have to be held in April (as in 1963) so as to come at the end of the season. The 1973 survey, however, seems to have been held in the months of January and February of 1973 (Ceylon, 1974). The timing of the 1953 survey could not be ascertained.

[9]The de-seasonalization could be done by quarter in Pakistan where the survey year is divided into four, each quarter being assigned one of the four independent sub-samples in the form of which the total sample is selected. There is a corresponding division of the survey in India into "sub-rounds."

[10]Income data were collected in the course of the consumer expenditure surveys between the 10th and 14th rounds, and again between the 19th and 25th rounds. The exception referred to here is the 25th round survey in which a reference period of the year before date of survey was used. Income data were also collected in the course of the agricultural labour surveys of the 11th and 12th rounds (1956–57) and 18th round (1963–64). These surveys are not covered in this paper because they were not comprehensive surveys designed to survey the entire population; they are covered however in Rajaraman, 1974.

Pakistani surveys of the *year* preceding the date of survey[11] is entirely too long for reliable recall. The conflict here between the requirements of accurate recall and avoidance of seasonal distortion cannot really be resolved except by providing for repeated—two or three—visits to each household in the course of the year. Most of all, the timing of these visits would have to be such as to fall at the end of a natural accounting period, such as a crop season for the rural sector, and all sample households would have to be covered simultaneously. The need for this is self-evident in the preponderant agricultural sector; agricultural receipts and disbursements are strongly associated with the crop season taken as a whole, and recall of these figures cannot be expected to last much beyond the end of the season. In light of this, it is not clear what possible value the income data collected in the Pakistani surveys could have. Since the dates of survey were spread evenly over the year, most households would have been visited in the middle of a crop season, and the year preceding the date of survey would have consisted, most awkwardly, of one season and bits of two others. The trouble with the kind of unreliability thus introduced into the data collected is that the direction of bias is impossible to ascertain. It would vary from respondent to respondent according to whether, when asked to guess at total income earned in a period that cuts across normal accounting units, each one tends to overestimate or underestimate. For this reason the income data from these surveys cannot be considered a reliable basis on which to estimate inequality.

The Sri Lanka survey[12] by contrast is vastly superior in its timing, although there is still the drawback that each household was visited only once. All sample households were surveyed at the end of one of the two major crop seasons, which also coincided with the end of the financial year. Two periods of reference were used, one a year long, and the other two months long. Neither can be seen to be particularly satisfactory, illustrating once again the need for repeated visits.[13]

(b) *Concepts.* In the Pakistani[14] and Indian surveys, the household is the unit in terms of which consumption and income data are collected. In the Sri Lanka surveys, however, the recipient unit in terms of which income data are recorded is the individual "income receiver."[15] This is a somewhat problematical concept in a context where the household still retains its function as a producing unit. To ascribe the entire income of such a unit to one individual would be wrong conceptually, and yet that is what would have to be done in practice. Fortunately, where the primary data are available, the option still remains of treating the income of the "income receiver" in the household as that of the household.

The concept of income used, as also that of consumption, varies from country to country and sometimes from survey to survey within a country; once again, it is necessary to have access to the primary data so as to adjust the totals where this is

[11]Four sources of income in the Pakistani survey, viz. wages and salaries, pensions, boarders and lodgers, and professional, have a reference period of a month. In the case of seasonally invariant receipts this would be all right, but many of these sources, agricultural wages in particular, are not.

[12] See footnote 8.

[13]In the 1973 survey the two periods were of six-month and two-month duration.

[14]The Pakistani questionnaire referred to here is the one used after the 1964 revisions; the old questionnaire is discussed in Bergan, 1967.

[15]This concept was used in all three surveys; since the 1973 questionnaire was not available for inspection, however, some of the other concepts may well have been defined differently for that survey.

required to maintain uniformity or for other reasons. For example, the Sri Lanka definition of income includes receipts from sales of assets on the grounds that they add to spending power; these receipts would have to be deducted so as to get an income total more in line with common usage. Again, consumption is defined in the Pakistani and Sri Lanka surveys to include expenditure on births, deaths, marriages and litigation; such expenditures because of their infrequent occurrence can cause distortions and are best deleted. These consumption items along with some others were included in earlier rounds of the Indian surveys and were dropped for the later rounds; figures from the earlier rounds would have to be adjusted down so as to provide a uniform basis for a study of trends.[16]

The questionnaires used in all three countries have far better provision for recording the item-wise composition of consumption than income. The listing of all possible items of consumption in the questionnaire is an important aid to accurate recall, and is yet another factor in the greater reliability of the consumption figures from the surveys. The absence of provision for item-wise listing in the income section of the questionnaire is particularly notable in the case of that problem component—imputed income, from home-produced goods or from free collection. In both urban and rural areas, a considerable part of fuel consumption and even food is obtained through free collection particularly by poorer households, and in rural areas farm and garden produce is a significant element in income. (A third source of income in kind, government consumption subsidies, is important in the case only of Sri Lanka, where the government distributes free rice to non-income-taxpayers.) In the absence of item-wise listing to aid recall, there is sure to be an underestimation of income obtained from such sources.[17]

From the foregoing, it is clear that the consumption figures provided by the sample surveys are likely to be much more reliable than the income figures. They are more carefully recorded by item with much better provision for the recording of the imputed value of freely collected and home-produced goods. This is especially important because it ensures that the consumption levels of poorer and rural households will not be underestimated. Most importantly, because of the inherently lesser seasonality in consumption, the consumption figures are not affected by the timing of the surveys, nor by the shorter periods of reference used; the latter, if anything, is a factor making for greater reliability.

4. *Kinds of Breakdown Possible*

In any study of income or consumption inequality it is important not merely to obtain Lorenz curves from the data, but to be able to obtain the characteristics of the population falling in the different deciles, or to assess the degree of absolute or relative deprivation separately for different subgroups of the population.

A breakdown by region is quite possibly the most important kind of breakdown needed. In all the three cases the rural/urban breakdown is available right away (and in Sri Lanka a three-way rural/urban/estate breakdown). Sample

[16]Up to the 18th round, the consumption total included expenditure on house construction for example.

[17]In the Sri Lanka questionnaire there is no provision either for recording item-wise costs incurred in the earning of income; it is thus impossible to tell how carefully the totals for cost and net income were arrived at.

selection is done separately for each sector and a sufficiently large sample size chosen so that estimates may be obtained separately for each. In Pakistan, the sample design further ensured a breakdown by region so that estimates could be made separately for rural West Pakistan and rural East Pakistan, and similarly for the urban sector. The surveys however are not designed to provide reliable estimates at further levels of regional disaggregation.

In India, regional breakdowns for both rural and urban sectors are absolutely necessary; because of the size and diversity of the country, estimates of inequality or poverty for the country taken as a whole conceal more than they reveal. However the sample surveys were not designed to provide reliable estimates at the level of the individual states until the 13th round (1957–58).[18] This revises the effective availability of the series by at least seven years.

In Sri Lanka, regional breakdowns for each sector are not designed for[19] but then again they are not quite as necessary in such a small country.

The next most useful kind of breakdown to have is by occupation or industry. The necessary data are collected in all three countries. In the Indian surveys in particular, there is provision for the recording of information to a very great level of detail. Each sample household is classified by a six-digit code, three for industry and three for occupation, according to whichever activity provides the major part of its income.

Other information collected in all the surveys includes data on demographic variables, as well as on the education levels of members of the sample households.

5. *Availability of Primary Data: Evaluation of the Tabulated Output.*

The primary data from the surveys are in most cases extremely difficult if not impossible to obtain. Even if access is granted by the government organizations concerned, the data are frequently simply not available. Storage facilities are typically inadequate, and the data are often destroyed after the officially prescribed tabulations have been done on them. This seems to be the case for example in Pakistan.[20] Further, since the reports on all the Pakistani surveys after 1966-67 were issued after the breaking away of Bangladesh,[21] no tabulations have been provided for these years on the data collected in the eastern wing. If these primary data were destroyed together with those of the western wing after the tabulations done on the latter, official sample survey data for Bangladesh would be available for no year later than 1966-67. In India, data gathered in the surveys are not destroyed after the official tabulations on them, but because of a severe shortage of storage facilities, data from earlier surveys are often destroyed

[18]The option of pooling samples from adjacent rounds is of course always open.

[19]Regional estimates are possible however because the different sectors seem to have little geographical overlap.

[20]Private communication, subject to correction. The author would be happy to be advised otherwise.

[21]The 1967–68 survey has not been tabulated, but the surveys done in 1968–69, 1969–70 and 1970–71 have. The reports on these were issued between February and May 1973; the secession of Bangladesh was completed in December 1971.

to make way for data from more recent surveys.[22] As for Sri Lanka, nothing could be learned about the availability of the primary data from the surveys.

Most of the studies that have been done on income or consumption distribution in these countries have been based on the tabulations provided by the government organizations responsible for the surveys. Unfortunately, these tabulations leave a great deal to be desired.

The Pakistani reports[23] for example provide no tabulations whatever by class of consumption expenditure. The consumption data are indeed used in the reports but only to provide figures of average consumption or of the composition of consumption by class of income. Income, then, is the variable by which the households are ranked and grouped into classes, which is unfortunate in view of the unreliability of the income figures collected.

Further, the ranking is done in terms of total household income, not *per capita* household income. The case for ranking by the *per capita* figures rather than by household totals is self-evident; the affluence or poverty of any household is a function of both the total size of the pie and the number of people among whom the pie has to be divided. Even granting economies of scale in consumption, so that ten people living on a monthly income of Rs.1,000 are better off than one person on Rs. 100, the disparity between the two households would by no means approach what is suggested by the total figures. Since the tabulated figures show a strong positive correlation between household size and household monthly income, the Lorenz curve obtained from a ranking by household income would show lesser inequality than one obtained from a ranking by the *per capita* figure.[24] For classes of monthly household income then, the Pakistani reports provide (separately for rural and urban sectors, and for 1963–64 and 1966–67 for both East and West Pakistan), the distribution of earners[25] by employment status and by occupation. A breakdown is also available of income by source for each class.

The tabulations provided[26] of saving by income class seem to bear out a point made earlier with respect to the imputed value of freely collected and home-produced goods, that there might well be an underestimation of these in the estimation of income (but not in the estimation of consumption). The figures show substantial dissaving, especially by lower income households, to an extent that cannot possibly exist on a yearly basis. In some months of the year, say towards the end of the crop season in rural areas, there might be temporary dissaving, which is offset as soon as the harvest comes in. Or some households might in a particular year undergo net dissaving. But the kind of average annual dissaving indicated can only mean that income is uniformly underestimated. (The alternative explanation, a general overestimation of consumption, is quite implausible.) The extent of

[22]For rounds after the 19th, however, primary data collected have been centrally stored by the National Sample Survey Organization.

[23]In addition to those mentioned in footnote 21, reports are available for the 1963–64 and 1966–67 surveys, and for the partial survey conducted in 1965.

[24]It also follows that after a *per capita* ranking, the Lorenz curve for deciles of *households* would show greater equality than that for deciles of population.

[25]"Earners" include unpaid family helpers defined to mean any family member who worked for fifteen hours or more during the week preceding the date of enquiry for the family enterprise without pay or profit.

[26]In the 1970–71 report, for example, tables 2 and 8.

implicit dissaving is higher in the rural sector, where non-monetary income is more important, than in the urban.

Of the studies that have been done on the basis of the survey data (Bergan, 1967; Azfar, 1973; Khandker, 1973; Chaudhry, 1973), some adjust for the under-representation of high-income households, but no adjustments are made for biases in the reporting of income of those households actually covered. None of the studies questions the basic reliability of the income figures, or suggests the use of the consumption figures instead. All of them, of course, except for Bergan, were limited to the tabulated output from the surveys, and he did not use the data at his disposal to derive estimates of consumption inequality for the population.

The tabulated output of the Indian surveys, on the other hand, do provide frequency distributions of the population by consumption class.[27] Further, these distributions are provided after ranking by both *per capita* figures and household totals. Published reports are available on all rounds up to the 19th. In addition, draft reports are available on some of the later surveys.[28]

State-wise estimates are available starting with the 13th round (1957–58).[29] None of the reports however provide any further breakdowns at the state level, not even by occupation. This is a very major shortcoming since it is particularly important for policy purposes to have a profile by occupation class of the different population deciles.[30] Thus the published reports, while they are good as far as they go, do not go very far. Further, there are problems of comparability with the published data because of changes made over the years, referred to earlier, in the definition of consumption, and because of changes made over the years in the geographical definition of many states. These figures have, nevertheless, been used in the few studies that have addressed the question of inequality trends at state level (Vaidyanathan, 1971; Bardhan, 1973). The only other state study of trends, based on the primary data rather than the published reports, is that for Punjab (Rajaraman, 1974a).

Figures of inequality at the national level are far less interesting, given the size and diversity of the country; it is the inter- and intra-regional differences that are operationally useful in such a context. Several studies have been done at the all-India level, however. Some of these have tried to construct income distributions for the country on the basis of the published consumption distributions, with adjustments for saving and taxes. Since there are no reliable data on the distribution of saving by level of consumption, however, the adjustments have been more or less arbitrary. There are thus as many estimates as authors (S. Swamy, 1967; Mukherjee and Chatterjee, 1967; Ranadive, 1971; Ojha and Bhatt, 1971; Ahmed and Bhattacharya, 1972).

The estimate of total national consumption obtainable from the NSS surveys

[27]The reports on the rural labour surveys mentioned in footnote 10 are the only reports that provide tabulations by income.

[28]They are available for the surveys conducted in the 20th, 22nd, 23rd and 25th rounds.

[29]The size of the sample in that round was too small however for state level estimates to be really reliable. But sample size was doubled in the next round for the urban sector, and in the 16th round for the rural, so that state-wise estimates for these and subsequent rounds should be more reliable. See also footnote 6.

[30]Even at the all-India level, an occupational breakdown is provided in only one report, for the 19th round.

and that obtainable from national accounts sources have been seen to have diverged in the decade of the sixties; the former has fallen increasingly short of the latter (Dandekar and Rath, 1971; Vaidyanathan, 1971). By contrast the two sources had been more or less in agreement in the decade of the fifties (Mahalanobis Committee). The cause of the emerging divergence is not clear; the issue of which source is the more correct remains unresolved. Even if there is a downward bias in the survey data, there is the question as to whether it arises out of a uniform underestimation of consumption of all households, or out of an under-representation of the more affluent households. The latter is much the more probable; the problem, referred to earlier, exists in all three countries. There are several reasons why the better-off household might choose to withhold information on consumption or income. There are, on the other hand, no immediately apparent reasons why there might be a general underestimation of consumption for all sample households. It was seen earlier that, given the structure of the surveys in these countries, underestimation was a distinct possibility in the recording of income though *not* of consumption. Thus, to the extent that there is a downward bias in the consumption data from the Indian surveys, it must arise solely from non-response on the part of the more affluent households in the sample. This possibility should be kept in mind by the data user.

In the Sri Lanka reports, as in the Pakistani reports, frequency distributions of the population are provided by class of income alone. The consumption figures are not used for this purpose—a major drawback.

The income distributions are provided in terms of two kinds of recipient units: "income receivers," and the household.[31] The tabulations that have gained the widest currency and are quoted in summary documents are those done in terms of income receivers. The conceptual problem with the use of such a recipient unit were dealt with earlier. Further a distribution of income by income receiver can only provide one element of what goes into the final distribution of income among the population, the other element being the distribution of income receivers among households. The alternative distribution by household, however, has the drawback that the ranking is done, as in the Pakistani case, by total household income.[32] Tabulations are provided on the figures obtained from both the two-month reference period and the year-long period.

Frequency distributions of income defined and classified in the above ways are available by sector.[33] For each sector, each class of income receiver is broken down by occupation and by ethnic group.

The official published reports on the surveys are thus detailed and comprehensive, but their usefulness is somewhat limited by the uncertain reliability of the income figures. The more reliable consumption figures are not used except in the tabulations of consumption by class of household income. Once again, as in the Pakistani case, these tables show an overall net dissaving.[34]

[31]Termed "spending units."

[32]As in the Pakistani case, household size increases with total income; see Ceylon, 1963, part I, table 46.

[33]The report on the 1953 sector did not provide estimates by sector.

[34]See for example Ceylon, 1963, part I, table 89.

The greatest failing of the tabulated output in all three cases, however, is that figures of inequality are provided only in nominal terms. The change over a period in real terms could be quite different from that in nominal terms if different classes of the population have faced different price rises, as is most usually the case. The sample surveys themselves are the best source of data on prices, which are implicit in the item-wise figures of quantity and value recorded in the consumption sections of the questionnaires. The failure to tabulate these data is, therefore, their most serious shortcoming.

Section II: Price Data

In the consumption sections of the questionnaires used in all the countries, price data are implicit in the quantity and value figures recorded against every item. Sometimes quantity data are not recorded for certain items, for example, non-food items in the case of Sri Lanka, services in the case of India. In the case of the Pakistani questionnaire, quantity data are recorded against only a few key items in each commodity group. Even in the Pakistani case, however, the surveys are the most rich and detailed source available of data on consumer prices. In addition, the item-wise figures of value could be used to obtain weighting diagrams for the construction of price indices separately for different classes, by income or consumption, of the population. Price indices could also be constructed cross-sectionally, for different regions with respect to one another, or for the rural sector with respect to the urban, so that cross-sectional comparisons too could be done in real rather than in nominal terms.

That these data from the surveys have not been tabulated is, therefore the most serious shortcoming of the survey reports. In the case of India, price data from the surveys have been tabulated from time to time, though not systematically and usually for only a few items like foodgrains. Further, the price data, where provided, have always been for the country taken as a whole, and never for the states taken separately. There have been some useful price tabulations at state level, however, by individuals with access to the primary data (Bhattacharya and Chatterjee, 1970; Iyengar and Bhattacharya, 1965; Rajaraman, 1974a), although the availability of these is too scattered and piecemeal to substitute for a systematic provision of price data in the survey reports. In the case of Pakistan and Sri Lanka, the reports on the surveys do not present any price data at all.

Since the primary data collected in the surveys cannot be assumed to be readily accessible to the research worker, other sources of price data must be looked for. A crude indicator of the direction of the price adjustment needed for trends in inequality is provided by wholesale price data which are usually available annually by broad commodity groupings if not by item. If food item prices in general have risen more than non-food prices in general, it can be assumed that the poor have faced a higher price rise than the rich, and that the trend towards inequality/equality in nominal terms understates/overstates the real trend. Some attempts have been made to assess the real trend in this manner for India, though not for the other two countries. In India, however, wholesale prices are available only at the national level, so that attempts to use them to assess trends in real inequality have been confined to the all-India estimates (Vaidyanathan, 1971).

Wholesale prices are, however, very crude approximations to consumer prices, but not too much published data on consumer prices exist. In Pakistan, there are no sources of consumer price data for rural areas whatever.[35] For urban areas, retail price figures for selected items are available for major cities alone—Karachi, Lahore, Peshawar and Rawalpindi in West Pakistan, and Daçca, Chittagong, Khulna and Rajshahi in East Pakistan. A constructed price index is published for clerical (white collar) workers in the above urban centres, and another for industrial (blue collar) workers in the major industrial cities of Karachi and Lahore in the West, and Narayanganj, Sialkot and Chittagong in the East.[36] Thus, the advantages of processing the data from the sample surveys can be seen to be overwhelming: consumer price data would then be available for rural areas, for which there is none at the moment, and urban price data would be available with far better geographical coverage than what is presently available.

The situation for India is again a little better. A consumer price index for agricultural labourers is published regularly; what is more, this index is available separately for each state.[37] The index is based on price data collected round the year from a fixed set of village markets. The field work is done by the NSS but as a separate operation quite distinct from the sample surveys. The coverage in terms of sample villages is by no means as good as that of the sample surveys[38] but in the absence of any tabulated figures from the latter, it does provide a valuable index of price movements for the rural poor in each state. Nothing comparable is available for the rural affluent. If the actual prices that enter into the agricultural labourers' index were published, they could be used in conjunction with the consumption pattern data by consumption class, published in the reports on the sample surveys by state, to obtain fractile-specific price indices for each state. But the actual item-wise price quotations are published only for the country as a whole and not for each state taken individually. As for the urban sector, there are two published indices as in the Pakistani case, one for the working class for about twenty industrial towns, and another for the middle class for major urban centres.

For Sri Lanka, published data on consumer prices are very meagre. No price data are available for the rural sector whatever. For the urban sector, there is just the Colombo Consumer Price Index, for the city of Colombo but for no particular section within it. In Sri Lanka there is the additional problem that there has been a rationing system for some of the most important consumption goods like rice, wheat and sugar with an open market and controlled price for each rationed item. The Colombo index is acknowledged to pay inadequate attention to this

[35]The study (Bose 1968) of rural trends in real wages for East Pakistan describes the price data used as having been obtained from several locations in East Pakistan, but does not specify that they were all locations in large urban centers.

[36]The series is being continued for those centers in the eastern wing by the Bangladesh Bureau of Statistics.

[37]It is issued by the Labour Bureau, Ministry of Labour.

[38]Thus, price data for the index are collected every year from a fixed set of villages at the rate of one price quotation for each item per village (taken every month from a fixed market). The villages were selected during the sixteenth round at the rate of one per stratum; by contrast, the sample survey of the 16th round had an average of nine villages per stratum per subsample, and there were two subsamples. Further, the size of the sample in surveys after the 16th has been much bigger.

problem.[39] The quantity and value data from the sample surveys on the other hand would immediately yield a weighted average of the two prices for each household.

To conclude, while enough published data exist with which to be able to tell in the very broadest terms whether a trend in inequality in nominal terms over- or underestimates the "true" trend, very little more can be done with presently available published data. The one exception is in India; here, a reliable price index is available by state for agricultural labourers who generally comprise the poorest sections of the rural population, so that real trends in poverty if not in inequality can be carefully assessed—to the extent that comparable figures in nominal terms are available at state-level in published form to start with. The primary data from the sample surveys themselves are the best source of the price data needed, but they have unfortunately remained untabulated for the most part.

Section III: Alternative Directions for Data Collection

From the preceding two sections, the outlook for gathering adequate and reliable data from the sample surveys regularly conducted in these countries has not emerged as particularly good. Further, the outlook will not improve substantially unless the primary data are made freely available or unless there is a complete overhaul of the kinds of tabulations done on these for the official reports: frequency distributions of the sample population must be provided by class of consumption even where income data are collected, the ranking for these must be done in terms of *per capita* rather than total household figures, consumption totals must be adjusted so as to be comparable over the years in terms of what they include and exclude, better provision must be made than presently exists for breaking down the sample by region and occupation, and lastly, the price data implicit in the data collected must be fully exploited. Beyond that, there is a need to overhaul the surveys themselves, especially if income distributions are needed in addition to consumption distributions. The kinds of changes needed however, such as the need for carefully-timed simultaneous coverage of all sample households, run strongly counter to requirements of administrative and budgetary convenience: with non-simultaneous coverage a smaller force of investigators can be kept fully employed the year round. Again, in the case of India, for example, survey timings would have to be different for different parts of the country even within the agricultural sector because of variation in the number and duration of crop seasons. All these factors make for administrative difficulties in the design and execution of the surveys, but they need to be taken into account if reliable income data are to be collected. Even if all these changes were to be implemented, however, problems would remain. The most intractable is that of wilful concealment or non-response on the part of the affluent sample household. Although this problem affects all surveys whether of income or consumption, it is probable that it is worse where the collection of income data is involved.

Independent surveys may of course be conducted by the interested research worker along the recommended lines, but they are very large and labour-intensive

[39]See for example Ceylon 1974, p. 201.

operations and are particularly costly when the manpower that has been assembled and trained is to be disbanded after one survey.[40]

Sample surveys of household income and consumption are unavoidable if estimates of inequality are needed but there are alternatives available if what is needed is not a comprehensive measure of relative distribution so much as the incidence of absolute deprivation within the population as a whole or subgroups of it. Income and consumption surveys are themselves of course an excellent source of the latter kinds of information, but alternatives must be sought in view of the extreme difficulty involved in conducting them on the one hand, and in getting the needed information out of those conducted on the other.

An alternative that appears particularly promising is a method used by the United Nations Relief Operation in Dacca (UNROD).[41] The method is essentially a simple way, using anthropometric measurements, of identifying those segments of the population suffering nutritional deprivation. In the case of children, to which group the UNROD surveys were confined, the anthropometric measurements can further be simplified to the Quaker Arm Circumference method (the QUAC stick) developed by the Quaker Service Committee in Nigeria; the method substitutes the measurement of arm circumference for weight. Thus, arm circumference measurements are taken along with measurements of height and the ratio of the two, compared to a norm established for the country or region, provides an index of nutritional sufficiency. The advantages of conducting a survey along these lines are many. Unlike the sample surveys of income and consumption, the investigators do not have to undergo a long period of training; the measurements can be taken very quickly, and do not require the use of heavy or expensive equipment. Most of all, there are no problems of recall lapse; the respondent does not have to do anything and merely has to submit to a few very simple measurements. It is not clear if the QUAC stick itself can be used on adults, though even a survey confined to children, as in the case of the UNROD studies, can provide good indices of the relative position of different occupation groups or regions vis-a-vis one another.

Any measure of the kind must lose something in sensitivity, though it is hard to determine without delving into the appropriate nutritional or medical literature just where the direction of error may lie. If the target population of the nutritionally deprived so identified is merely a subset of the desired or true target population, the problem is not as serious as if the target population identified cuts across the "true" level and includes many who are not really nutritionally deprived at the same time that it excludes many who are. Regardless of this particular measure, however, further investigations into the use of anthropometric measurements and into their use as indices of poverty are absolutely necessary. Ultimately, the redressing of the most extreme aspects of absolute deprivation is the most urgent need of an income redistribution program, and surveys that

[40]That is why there are no comprehensive surveys done by organizations other than by the official statistical machinery of these countries.

[41]Two surveys were done by UNROD in 1972, the first in June, and a follow up in December on the same families. These surveys, incidentally, are the only source of any kind available on what happened in income distribution terms in Bangladesh in the turbulent months following its independence.

determine the incidence and location of nutritional deprivation quickly and reasonably accurately are an indispensable prerequisite for the formulation of such programs of action.

References

Ahmed, M. and Bhattacharya, N., "Size Distribution of *Per Capita* Personal Income in India." *Economic and Political Weekly*, Special Number, August 1972, pp. 1581–1588.

Azfar, J. "The Distribution of Income in Pakistan 1966–67." *Pakistan Economic and Social Review*, Spring 1973, pp. 45–46.

Bardhan, P. K., "On the Incidence of Poverty in Rural India of the Sixties," *Economic and Political Weekly*, Annual Number, February 1973.

Bergan, A., "Personal Income Distribution and Personal Savings in Pakistan: 1963–64." *Pakistan Development Review*, Summer 1967, No. 2, Vol. VII, p. 169.

Bhattacharya, N. and Chatterjee, G. S., "Between States Variation in Consumer Prices and *Per Capita* Household Consumption in Rural India." Calcutta, 1970. (Mimeographed).

Bose, S. R., "Trend of Real Income of the Rural Poor in East Pakistan, 1949–66." *Pakistan Development Review*, Autumn 1968, Appendix B.

Ceylon, Central Bank of, Department of Economic Research. *Survey of Ceylon's Consumer Finances 1963*. Colombo: Central Bank of Ceylon, 1964.

Ceylon, Central Bank of, *Annual Report of the Monetary Board to the Minister of Finance for the Year 1973*. Colombo: Central Bank of Ceylon, 1974.

Chaudhry, M. G., "Rural Income Distribution in Pakistan in the Green Revolution Perspective." *Pakistan Development Review*, Autumn 1973.

Dandekar, V. M. and Rath, N., *Poverty in India*. Poona: Indian School of Political Economy, 1971.

India, Government of, Cabinet Secretariat, Central Statistical Organization. Reports of the National Sample Survey Consumer Expenditure Surveys. New Delhi: Government of India, various numbers.

India, Government of, Planning Commission. *Report of the Committee on Distribution of Income and Levels of Living (Mahalanobis Committee)*. New Delhi: Government of India, 1964.

Iyengar, N. S. and Bhattacharya, N., "On the Effects of Differentials in Consumer Price Index on Measures of Income Inequality." *Sankhya*, Series B, XXVII, September 1965, pp. 53–56.

Khandker, R. H., "Distribution of Income and Wealth in Pakistan." *Pakistan Economic and Social Review*, Spring 1973, pp. 1–36.

Mukherjee, M. and Chatterjee, G. S., "Trends in Distribution of National Income 1950–51 to 1965–66." *Economic and Political Weekly*, 15 July 1967, pp. 1259–1268.

Ojha, P. D. and Bhatt, V. V., "Pattern of Income Distribution in India 1953–55 to 1961–64." Paper presented at the seminar on income distribution at the Indian Statistical Institute, New Delhi, February 1971. (Mimeographed).

Pakistan, Government of, Ministry of Finance, Planning and Development, Statistical Division. Household Income and Expenditure Survey. Islamabad: Government of Pakistan, various issues.

Rajaraman, I., (a) "Poverty, Inequality and Economic Growth: Rural Punjab 1960–61 to 1970–71." Princeton, N. J.: Princeton University, Woodrow Wilson School RPED Discussion Paper 45, 1974.

———, "Data Sources on Income Distribution in Bangladesh, India, Pakistan and Sri Lanka: An Evaluation." Paper presented at the Princeton–Brookings Income Distribution Seminar, Princeton, N. J., September 1974. (Mimeographed).

Ranadive, K. R., "Distribution of Income: Trends Since Planning." Paper presented at the seminar on income distribution at the Indian Statistical Institute, New Delhi, February 1971. (Mimeographed).

Swamy, S., "Distribution of Income in India." *Economic and Political Weekly*, Annual Number, February 1967, pp. 289–292.

Vaidyanathan, A., "Some Aspects of Inequalities in Living Standards in Rural India." Paper presented at the seminar on income distribution at the Indian Statistical Institute, New Delhi, February 1971. (Mimeographed).