

AGENCY FOR INTERNATIONAL DEVELOPMENT WASHINGTON, D. C. 20523 BIBLIOGRAPHIC INPUT SHEET	FOR AID USE ONLY
---	-------------------------

1. SUBJECT CLASSIFICATION	A. PRIMARY Economics
	B. SECONDARY General Economics

2. TITLE AND SUBTITLE A methodology for the evaluation and adjustment of income distribution data
--

3. AUTHOR(S) Szal, R. J.

4. DOCUMENT DATE 1975	5. NUMBER OF PAGES 51p.	6. ARC NUMBER ARC
--------------------------	----------------------------	----------------------

7. REFERENCE ORGANIZATION NAME AND ADDRESS Princeton Woodrow Wilson School, Princeton University, Princeton, New Jersey

8. SUPPLEMENTARY NOTES (Sponsoring Organization, Publisher, Availability) (In Discussion paper no. 54)

9. ABSTRACT

The three main sources of data relating to income distribution in less developed countries are income tax data, census data, and income and expenditure surveys. Of the three, income tax data are the least useful. Taxpayers normally constitute less than 10 percent of the population; their incomes lie in the upper end of the distribution; and declared income can constitute as little as 10 percent of actual income because of evasion, fraud, and prevailing forms of production. Thus peripheral data sources must be used to build up distribution data "from the ground up." But the peripheral data must be examined for its reliability, completeness, and detail before it is used as a means of arriving at accurate income distribution estimates. Survey data must be assessed for its sampling properties and possible sampling errors. Non-sampling errors may be large. Definitions of recipient unit, income, and expenditure must be examined for the possibility that adjustment to another type of unit will facilitate use of indirectly related data as a crosscheck. Expenditure data must be sufficiently detailed to permit its accurate conversion into income data. The possibility that prices facing different income groups have changed at differential rates must also be explored.

10. CONTROL NUMBER PN-AAB-446	11. PRICE OF DOCUMENT
12. DESCRIPTORS	13. PROJECT NUMBER
	14. CONTRACT NUMBER AID/CM/otr-C-73-237 Res.
	15. TYPE OF DOCUMENT

A Methodology For
The Evaluation and Adjustment of
Income Distribution Data

by
Richard J. Szal

Research Program in Economic Development

Woodrow Wilson School
Princeton University
Princeton, New Jersey

Discussion Paper No. 54

October 1975

This research was supported by the U.S. Agency for International Development
under the joint Princeton/Brookings Income Distribution Project.

PREFACE TO THE SERIES OF PAPERS
DEALING WITH DATA OF DEVELOPING COUNTRIES

The possibilities of empirical research on developing countries are hindered by the lack of data sufficient to support such research. In the past decade or so, however, the availability of data sets such as household income and expenditure surveys, labor force surveys, demographic surveys, etc., for developing countries has been increasing by leaps and bounds. Nevertheless, the increase in data is not evenly spread across developing countries. It tends to be related directly to such factors as stage of development or extent of urbanization. Furthermore, the fact that a data set exists for a given country guarantees neither its statistical accuracy nor its comparability through time or across countries.

It is with a view to these problems that the Brookings-Princeton project on income distribution in the developing countries commissioned six papers to be written, each dealing with the data of a geographically proximate group of countries. The papers and their authors are:

Hakchung Choo, "Review of Income Distribution Data: Korea, the Philippines and Taiwan," RPED Discussion Paper #55.

Oey Astra Meesook, "Review of Income Distribution Data: Thailand, Malaysia and Indonesia," RPED Discussion Paper #56.

Indira Rajaraman, "Review of Income Distribution Data: Pakistan, India, Bangladesh and Sri Lanka," RPED Discussion Paper #57.

Adedotun O. Phillips, "Review of Income Distribution Data: Ghana, Kenya, Tanzania and Nigeria," RPED Discussion Paper #58.

Miguel Urrutia, "Review of Income Distribution Data: Colombia, Mexico and Venezuela," RPED Discussion Paper #59.

Carlos Geraldo Langoni, "Review of Income Distribution Data: Brazil," RPED Discussion Paper #60.

The authors were requested to fulfill three requirements: (1) list socio-economic data available in each of the countries, emphasizing especially those that would apply to empirical studies of the size distribution of income, (2) evaluate the statistical accuracy of the more important data, and (3) suggest the more promising areas for future research in light of both data constraints and the particular political and economic situation of the country.

The general findings of these papers lead one to conclude, happily, that there exists a great deal of data which can be used for studying the size distribution of income in the countries. Unfortunately, many of them have not been fully exploited. These conclusions, however, vary by geographic region; in both the African countries and the Latin American countries (excluding Brazil) several data sets must be combined to arrive at a representative national size distribution of income. Most of the data suffer from the usual problems of underestimation of income, seasonal biases in the figures, the exclusion of non-monetary components of income, and changes in definitions and concepts over time. Aside from these more general criticisms most of the authors also find particular shortcomings. In certain countries, such as India and Indonesia, the better data sets deal only with consumption and not total income. In most of the countries, there is a problem with the representativeness of the sample. This occurs for several reasons: (1) many samples have an upper income cutoff and do not collect data from households whose total income lies above the cutoff, (2) lower income households are underrepresented because zero income responses are excluded and the poor, who tend to be more highly mobile geographically, often disappear from samples from one enumeration to the next, and (3) coverage is sometimes limited to only certain specific

geographic areas, e.g., urban or rural, one province, large metropolitan areas, etc.

Taxation data is of limited use, because most developing countries have poorly administered direct taxation systems; there is a great deal of evasion and fraud, relatively high income cutoffs below which no tax is assessed, and little detailed socio-demographic information included in the taxation statistics. It should be pointed out, however, that taxation statistics on individual incomes can be used either as a cross-check on higher income classes derived from other sources or for estimating the frequencies for these classes in cases where another source excludes them from the enumeration.

Another problem arises from the fact that certain countries define geographic units in ways that are not comparable as between countries. In Thailand and Bangladesh, for example, the rather rough equivalents of a province are sanitary districts and thanas, respectively. Over and above all these problems, resource constraints often dictate very small sample sizes leading to relatively large sampling errors.

The present paper is a theoretical discussion of how these shortcomings can be recognized and dealt with. As the paper points out, one cannot hope to perfectly correct for all deficiencies, but only to improve the available estimates.

CONTENTS

PREFACE TO THE SERIES OF PAPERS DEALING WITH DATA OF DEVELOPING COUNTRIES	11
I. SOURCES OF DATA	1
II. RECIPIENT UNITS	5
A. Relationship Between Individual Incomes and Household Incomes	7
B. Estimating Mean Incomes of Open-Ended Income Classes	10
III. CONSISTENCY CHECKS	12
A. National Income Accounts	12
B. Fiscal Data Consistency Checks	16
C. Historical Trends	17
D. Novel Approaches	18
IV. THE CONCEPT OF INCOME	19
A. Home-Produced Agricultural Consumption	20
B. Imputed Rental Income From Owner-Occupied Housing	21
C. Income In-Kind Offered as Remuneration for Employment	24
D. Social Security Contributions	25
E. Unrealized Capital Gains	25
F. Undistributed Corporate Profits	26
G. The Imputed Value of Government Subsidized Services	27
H. Taxes	30
1. Shifting Assumptions and Taxation Incidence	33
V. INCOME DISTRIBUTIONS BASED UPON EXPENDITURE DATA	37
VI. DIFFERENTIAL PRICE TRENDS	38
VII. SEASONALITY OF DATA	40
VIII. SUMMARY	41
FOOTNOTES	43

A METHODOLOGY FOR THE EVALUATION AND ADJUSTMENT
OF INCOME DISTRIBUTION DATA

Things are seldom what they seem,
Skim milk masquerades as cream.

William Schwenk Gilbert
H.M.S. Pinafore, Act I

I. SOURCES OF DATA

There are three principal sources of data directly related to income distribution; income tax data, census data, and income and expenditure surveys. Of the three, tax data tends to be the least useful for studies of income distribution in less developed countries. There are two basic criteria that fiscal data must fulfill if they are to be useful for this purpose: (1) all or a great majority of households declare their income, and (2) declared income constitutes a close approximation of actual income. Less developed countries, in general, meet neither of these criteria. Taxpayers normally constitute less than 10% of the population, all their incomes lie in the upper-tail of the distribution, and because of evasion, fraud, tax administration, and the prevailing forms of production (traditional forms of enterprise), declared income can constitute as little as 10% of actual income. These points are sufficient to dismiss the use solely of fiscal statistics for the study of the entire income distribution. Income tax data can, however, prove beneficial as cross-checks or supplementary data.

The second source of data is the population census usually conducted decennially. Although income data is collected in many censuses, the usual questionnaire deals more with demographic characteristics than with details of income. In the U.S., for example, where data collection has progressed

to high levels of sophistication and efficiency relative to that of developing countries, data is collected on income from wages and salaries, non-farm and farm self-employment, transfer payments, and the catchall "other" sources. The economist studying income distribution would most certainly prefer a further breakdown. An advantage of census data is that many socio-economic characteristics are recorded. Combined with the fact that sample sizes are large, this allows many cross-classifications of the data without sampling error problems. Statistical techniques are usually of high calibre in census enumerations, although it is not clear that the same can be said of enumerators. The main disadvantage of the census is that it is conducted only once a decade, thus separating the time series by wide intervals.

Income and expenditure surveys or consumer budget surveys have only recently begun to concentrate much attention on income. Their main emphasis has been on gathering data relevant to studies of consumption and savings behavior and of consumer price indices. Neither they nor the census are tailor-made to answering questions on income distribution. Although the volume of data collected for each household is usually larger and more relevant to the economist than is the case of the census, sample sizes are smaller. As a result, cell sizes in multiple cross-classifications can become prohibitively small from a statistical point of view.

Aside from the "regularized" income and expenditure surveys, ad hoc surveys are included in this classification. Regardless of their special purpose, their sampling procedures are akin to those of income and expenditure surveys, and they often employ the same sampling frame. Most developing countries originally undertook their income and expenditure surveys as ad hoc and have only recently converted them to an annual,

biennial, or quinquennial basis.

Before proceeding further, it will prove helpful to devote a few sentences to the possible sources of error that can negate, at least partially, the statistical accuracy of data. The sources of error can be classified generally under two heads: (1) sampling errors, and (2) non-sampling errors. Sampling errors are those errors that result from the fact that the entire population has not been enumerated. That is, unless the sampling proportion is 100 percent, one can expect the result of the sample to differ from the true value of a population parameter by some degree, the maximum value of the difference being related inversely to the sampling proportion. Non-sampling errors are the result of inaccurate responses from the persons surveyed. That is, non-sampling errors would result even if the sampling proportion were 100 percent. In general, one normally evaluates sampling techniques to insure that no unusual amount of sampling errors is present in the results of a survey and then creates confidence intervals within which it can be said with surety that the true population value lies. Non-sampling errors are more difficult to deal with, and an assessment of their magnitude is normally arrived at by the comparison of the results of a survey with some other data source of known validity. The theoretical aspects of the former type of error have been investigated thoroughly and not much can be added to this literature. For the most part, this paper concentrates on the identification and adjustment of non-sampling errors.

Returning to the discussion of survey and census data, Fishlow has argued that the amount of non-sampling error in the census is usually less than that found in surveys.¹ Using Brazilian data from both sources, he finds the mean of the survey data to be approximately twice that of the

census data for both rural and urban areas. Two explanations are offered: (1) surveys oversample those with established residences, and (2) the rate of non-response is high in low incomes areas because of the detailed information required by the surveys, especially in the area of expenditures.

Others have attempted to test the consistency of survey statistics with those of national accounts.² The usual conclusion is that income is underreported in the survey. Since the census estimates are lower than those of the survey, they must be underestimated to an even greater extent. Several reasons can be offered to explain the larger degree of sampling and non-sampling errors in census data:

- (1) Respondents in the census feel that a cross-check of their reported income is less probable than is the case for a survey, i.e., they will lie to census enumerators more readily;
- (2) Because most surveys collect detailed expenditure data, the estimate of income may be more accurate since the respondent attempts to make it jibe with expenditures;
- (3) The survey concept of income may be more inclusive than the census concept, e.g., the inclusion of home-produced consumption, imputed rental income, etc.;
- (4) Survey enumerators are usually better trained than those of the census.

These points are offered only as possibilities, and they are not to be interpreted as a complete list. One or more may be operative at a given time. As a result, most researchers conclude that survey estimates are more accurate than those of the census from both the vantage points of sampling error (because of point (3)) and especially response error.

One cannot make a blanket statement that survey data is in all cases preferable to that of the census. Certainly, another factor operating against the choice of census data is the long time lag between enumerations with resultant changes in concepts and data collection techniques. But the ultimate decision must be made on a case-to-case basis in light of particular factors operative in the data collection processes of the country under consideration and the use to which the data will be put.

II. RECIPIENT UNITS

The personal distribution of income can be analyzed from the vantage point of the individual, the family, or the household. The distribution between the latter two recipient units is often more important in the case of a developing than in a developed country. When the extended family type of residential organization exists to any measurable degree, as in the case of developing countries, family income -- that is, the sum of the incomes of the immediate members of a family -- can diverge from household income. It has been found in the past that as development proceeds, the distinction between the immediate family and the household disappears and, thus, also the difference between family and household income.

One would expect a priori that the lower the level of disaggregation of household income, the larger will be measured inequality. The incomes of women, children, and part-time workers tend to lie predominantly in the lower tail of income distribution. The summation of individual incomes to the household level should decrease total variance by adding the relatively lower incomes of women and children to the relatively higher ones of males, and, thus, measured inequality should decrease. The

relationship is, however, heavily dependent upon the variance of household size. An illustration in which the hypothesized relationship does not hold can be taken from Indian data for 1964-65 assembled by the National Council of Applied Economic Research. They find measured inequality as represented by the Gini coefficient to be higher for income distribution by households than for that by individuals. The explanation of this phenomenon lies in the fact that average family size increases with average household income.³ Clearly one can visualize a situation in which all individual incomes are equal, although household incomes differ because the number of income earners differs as between households. In this case inequality of the distribution by individuals is zero while that for households is non-zero and is exactly equal to the inequality of the distribution of income earners per household.

The component of measured inequality introduced by differences in the number of economically active as between households can operate either to increase or decrease inequality. The direction of change is dependent upon the relationship of average household size and average household income. If average household size is directly (inversely) related to average household income, the measured inequality of the household income distribution will be larger (smaller) than that of the income distribution by individuals.

It is possible to decompose total inequality of the household income distribution to that due to differing incomes and that due to differing numbers of household income recipients. The data necessary for this decomposition, other than household income, is the number of income recipients per household. This may or may not be available from the household surveys

in developing countries. Most censuses include a question on this household characteristic.

An adjustment to household income for the number of recipients assumes greater importance for the case of a developing country than for a developed one. The high population growth rates of developing countries lead to a relatively young population. This, along with the existence of the extended family system, means that the size of a household and a family within a household tend to be larger than in the developed countries. The distribution of household size may also have a larger variance.

A. Relationship Between Individual Incomes and Household Incomes

Another aspect of the units of observation is the relationship of the individual and household distributions.⁴ These two types of data should be related, and given a distribution on the basis of one type of unit it should be possible to transform the distribution to one based on the other type of unit. Information on household size and on established earnings patterns for different demographic groups can be used for this purpose. The transformed distribution is, of course, only an estimate, and one can only hope for reasonable accuracy. The benefit of such transformations is that income distributions for two countries each based on a different recipient unit can be compared. The use of such a transformation also can be envisaged in cases where a country has changed the unit of observation over time or has collected data for population subgroups using different concepts of the recipient unit.

The data transformation is possible from either households to individuals or vice versa. It is more preferable to estimate from the household distribution to that of individuals, because with household data there is normally data on the number of individuals in the household, their

ages, and other associated demographic characteristics. The transformation suggested here is one using the concept of a composite recipient unit for assigning income to individuals. To aggregate incomes from the individual to the household level involves detailed information on the extent to which the extended family system exists among income classes or socio-economic groups. Such information is rarely available, and even if it were, a great amount of error would creep into such estimations. Thus, it is preferable to decompose household income to arrive at the individual distribution rather than aggregate incomes to arrive at a household distribution. Since the reason for this transformation is to make household and individual income distributions comparable, one normally has a choice of the one to be transformed.

Total household income is equal to the sum of all male, female, and child incomes of the household. Male incomes are usually larger than those of females or children, and those of females are usually larger than the incomes of children. Thus,

$$THI = \sum_a I_{ma} + \sum_b I_{fb} + \sum_d I_{cd} \quad (1)$$

where THI = total household income

I_{ma} = income of the a^{th} male in the household

I_{fb} = income of the b^{th} female in the household

I_{cd} = income of the d^{th} child in the household

To obtain a relative scale of incomes, one can divide both sides of (1) by the average male income of the household. Equation (1) becomes

$$\frac{\text{THI}}{I_m} = \frac{\Sigma I_{ma}}{I_m} + \frac{\Sigma I_{fb}}{I_m} + \frac{\Sigma I_{cd}}{I_m} \quad (2)$$

Others have termed the value of (2) the "adult male equivalent" household size.⁵ One should recognize that, were there only one male, one female and one child in the household the value of (2) would be

$$\frac{\text{THI}}{I_m} = \frac{I_m}{I_m} + \frac{I_f}{I_m} + \frac{I_c}{I_m}$$

which is equivalent to weighting the male income as one and the female and child incomes as fractions whose sizes depend on age and sex. One could further disaggregate children's incomes to males' and females' obtaining the specific fractions that apply to each group. Once averages of these fractional components are obtained for a society, demographic subgroups, geographic groups, or income classes, it is possible to decompose total household income into its individual components. One cannot define a set of fractions that apply to all countries in general since they are determined by the established income patterns of a specific country. Their identification must be handled on a case by case basis, and it is beneficial to attempt to derive the fractional components for as many socio-economic groups as the household data to be decomposed will accommodate. Income patterns can be widely divergent as between socio-economic groups. An obvious relationship that should hold across countries is that the fraction for rural children's income should be larger than that for urban children. Other well-known relationships should be exploited in the transformation process.

The decomposition of household income next follows a straightforward methodology. One simply sums up the adult male equivalent household

size for a given household, divides this sum into the household's total income, and allocates the total income among the males, females, and children of the household. As an example, assume a male, female and one child comprise a household whose total income is \$1500. Assume further that the adult male equivalent values are determined to be 1, 0.75, and 0.25 for the male, female and child, respectively. The decomposition would result in

$$\text{Adult Male Equivalent} = 1.0 + 0.75 + 0.25 = 2.0$$

$$M = \$1500./2.0 = \$750.00$$

$$F = \$750.(0.75) = \$562.50$$

$$C = \$750.(0.25) = \underline{\$187.50}$$

$$TMI = \$1500.00$$

It is now easier to understand why decomposing household incomes is preferable to aggregating individual incomes in the transformation. When aggregating individual incomes, one must have fairly detailed data on family size and the extent to which extended families exist. Estimates are less precise and bunching of incomes may occur, while the control of total societal income must be exercised. The difficulties are far fewer in the decomposition methodology.

B. Estimating Mean Incomes of Open-Ended Income Classes

When using grouped data, the mean income of an income class is normally assumed to be its mid-point. Such an estimate is based upon the belief that income observations within the interval are more or less evenly distributed across its range. This causes a problem for the lowest and highest income classes, since they are usually open-ended. Furthermore, unless realistic values are chosen for these classes, any measure of inequality based upon the estimates will be incorrect.⁶

For the lowest income interval, the problem is lessened by the fact that one can assume the lower bound of this interval to be zero and then simply use the mid-point as an estimate of the mean. This estimate, however, is always below the true mean because of the curvature of the true distribution function.⁷ Nevertheless, use of the mid-point of the lowest income class should not greatly bias the results, although an adjustment is possible. Since the true mean of this class is always greater than its mid-point, an adjustment that will improve the results is to assume that the mean is equal to $.65x_1$, ($=1.3M_1$) for the lowest income class where x_1 is the upper bound of that class (and M_1 is its mid-point).

Estimation of the mean of the highest income class must be handled in a more sophisticated manner. The upper bound of this class (or the highest observed income) is usually not disclosed in tabulated data, and even if it is, one would err greatly by assuming the mid-point to be the mean. For most distributions of income, the upper brackets can be accurately described by the Pareto-Levy function. That is,

$$N = AY^{-\alpha}$$

or in log form

$$\log N = A - \alpha \log Y$$

where Y = household or individual income

N = number of recipient units having Y income or more

A = constant.

The coefficient α can be determined by the standard practice of fitting the equation to the two highest income classes.⁸ Given a value for α , the mean income of the highest income class is

$$\bar{X}_h = \left(\frac{\alpha}{\alpha-1}\right) (Y_{hl})$$

where

\bar{X}_h = mean of highest income class

Y_{hl} = lower class bound of open-ended class.

It should be clear that the estimates of A and α allow the disaggregation of the upper open-ended class into as many intervals as desired. This is an especially important point for those data that are classified into only a few intervals with a large share of the total number of observations in the upper open-ended class.

III. CONSISTENCY CHECKS

There are a myriad of ways one can check the consistency of income estimates resulting from a survey or a census. The purposes of such checks are to determine areas of discrepancy between the survey or census and the comparison data and what groups are primarily responsible for any found. This section details the better known consistency checks and some inventive techniques that researchers have utilized. For the most part, one must use whatever country specific information is available to create imaginative checks which often prove more valid than the standard techniques.

A. National Income Accounts

The most obvious choice for cross-checking census or survey results is national income accounts. It is normally assumed that the national accounts data are more correct than sample data (although this is questionable for some countries). The concept of income in national accounts does not usually conform to that utilized in the survey or census

and thus either the sample estimates or national accounts data must be modified before comparing them.

Table 1. details the adjustments to national income that must be made in order to obtain total factor income received by households. This concept is comparable to that used in most surveys.

Table 1.

Adjustments to National Income Necessary to
Obtain Total Factor Income Received by Households

National Income

- Income of Institutionalized Individuals
 - Government Property and Enterprise Income
 - Corporate Savings
 - Corporate Income Taxes
 - Employer Contributions to Social Security
 - + Transfer Payments
-

Factor Income Received by Households

Two additional components, business contributions to foundations and gifts, are usually deducted from national income to arrive at a comparable figure, but they are excluded from the foregoing list. Although they may prove to be important components in the developed countries, they are frequently quite small in developing ones and can, thus, be ignored.

The figures for two of the components of Table 1., the income of institutionalized persons and employer contributions to social security, are usually not readily available. One must, therefore, estimate them. Institutionalized persons can represent from two to six percent of the total population dependent upon the country under consideration. In the developing countries, consumption levels of such persons are probably nearer

the national average than is the case in developed countries where they are normally lower than average. Thus, it can be assumed that they each receive an income equal to the national average.⁹

The second component for which figures may not be available is employers' contributions to social security. These figures are part of the total wage fund, while the survey relates to income actually received. It is possible to estimate the payments given the rate schedule and coverage. Coverage rates, however, are not well documented. If an arbitrary estimate must be used, one can judge from the experience of other countries in which the ratio of contributions to total earnings is known. Depending upon coverage, the ratio normally varies between 0.10 and 0.20.

Once these adjustments are completed, the national accounts total can be compared with the sample total income. One almost invariably finds that the sample total is less than that of the national accounts. For example, such a check for Taiwan reveals that the survey income is approximately 20% lower than that of national accounts.¹⁰ The next step is to attempt to define which income classes are responsible for the underestimate of income. One can obtain an idea of the group that are under-reporting by disaggregating the national accounts estimate to its functional sources and matching these with the totals of the sources of income listed in the sample (assuming they are available). If an account such as property income is underestimated by the sample, it is clear that the top quintile or decile is primarily responsible. Similarly, interest income normally accrues to the upper income classes, especially in developing countries. One can devise a series of adjustments from such comparisons.

Another use of national accounts data for a cross-check of the survey estimates is by a sectoral breakdown of income. Since cross-checks necessarily cannot be extremely precise, a sectoral breakdown to agricultural and nonagricultural income will suffice. Kuznets has shown empirically that the average product per worker in agriculture is generally below the countrywide average in developing countries, while in other sectors the per worker average is above that of the country.¹¹ Accordingly, it can be expected that the share of the agricultural sector from national accounts data should be equal to the share estimated by the survey and close to the share of the bottom X percent of the survey distribution, where X is determined by the percentage of the population gainfully employed in agriculture. A higher survey share may indicate either an underestimate of incomes in the upper (1 - X) percent of the distribution or an overestimate of incomes in the lower X percent. The former case is the more likely. Since the bottom X percent of the survey distribution undoubtedly contains some nonagricultural workers, the survey share of income of this X percent must be less than the share of agriculture from national accounts data. Conversely, some agricultural workers will be included in the upper (1 - X) percent of the survey income distribution and, thus, the income share of this group must be greater than the share of nonagricultural sectors in the national accounts. If either condition is violated, income has been incorrectly reported for one of the groups and inequality has been underestimated. When the conditions are met, no conclusion can be drawn since the share values are compatible with the theoretical expectations. One can think of imaginative uses of the data in this area, depending upon the quality and breakdowns of the survey and national accounts data. For example, one might compare the national accounts property income share with the share

of the survey distributions upper decile or top five percent. The survey share should be larger since it includes non-property components of income. Or again, one might sum the proportions of agricultural workers and unskilled, nonagricultural workers, call this X' , and calculate the polar values of the income share of this X' percent from national accounts data. A comparison of the survey's income share for the bottom X' percent will allow conclusions to be drawn about the correctness of the survey estimates.

A final use of national accounts data is in the area of consumption expenditures. The usual survey includes expenditures as well as income. Based upon the relationship of consumption to income, and under the assumption that an individual who incorrectly reports income will also incorrectly report expenditures, one can indirectly check the consistency of the survey's income estimates by comparing expenditure categories in the national accounts and the survey.¹² It is known that the percentage of total expenditure on food, for instance, varies inversely with income, while that on leisure activities varies directly. A difference between the proportion of total expenditures from national accounts and the survey can indicate errors in the survey. Morrison gives the example of expenditure on food representing 50% in national accounts and 55% in the survey, whereas these percentages for health expenditures are 15 and 6, respectively. Such relationships can indicate an underestimate of the concentration of incomes, because either the upper income groups have underreported income and expenditures or an unrepresentative sample has used a disproportionate number of high income recipients. One can envisage, as before, imaginative ways of analyzing expenditure data subject to the constraints imposed by available breakdowns.

B. Fiscal Data Consistency Checks

Although they relate to a very small proportion of a specific

segment of the total income distribution, fiscal data can be used for cross-checking the reported incomes of the survey distribution's upper tail. For the great majority of developing countries, fiscal statistics relate to no more than the upper decile of incomes. An estimate of this proportion for the country whose data is to be analyzed should be made prior to the cross-check. As pointed out earlier, most developing countries have many categories of tax exempt income and substantial amounts of evasion and tax fraud. Thus, fiscal data relate to only a small proportion of income recipients and may be underreported by as much as 90%, depending upon tax laws and their administration and enforcement. When total incomes of the upper decile (or X percent) do not exceed those of the fiscal statistics by factors of from two to ten, one can be sure that survey incomes are underreported. If the extent of underreporting is larger for this group than for survey incomes in general, income inequality is underestimated.

C. Historical Trends

When a number of surveys are available forming a time series one can use general information on historical trends to compare the estimates. Such comparisons are valid, however, only in cases where the concepts and methodologies of data collection have not changed. When concepts have changed, distributions can be compared by taking account of the effects of changes on the survey estimates.

Inconsistencies between the changes in survey results and historical changes in the economy can signal errors. For example, little change in the capital stock or its distribution accompanied by little change in the labor force and miniscule GNP growth signal that the distribution should not change a great deal. Conversely, a significant land reform program should result, ceteris paribus, in an increase in the share

of the lowest one or two quintiles. The results of surveys separated by time should be evaluated in light of changes in the economic system that have occurred between the times of their data collection.

D. Novel Approaches

In cases where data are not sufficient to support more sophisticated cross-checks, peripheral data can be used for this purpose. Two such approaches are those presented by Webb to check the consistency of his estimates of income distribution.¹³ The number of non-commercial cars owned in Peru in 1961 (the year of his estimate) was on the order of 30,000 - 50,000. The cutoff level for car ownership was estimated to be between 50,000 and 70,000 soles in 1961. These figures can be used to estimate the approximate number of individuals having incomes in the upper tail of the distribution.

A second cross-check is based upon the direct relationship found to exist in Peru between the level of provincial income and the rate of sewing machine ownership of households in the province. Webb utilized these findings to test the consistency of his estimations using data gathered by the census bureau of Peru.

There is a certain methodology followed in consistency checks based upon peripheral data sources. One attempts to identify characteristics of the population that can be associated with a range of incomes, but that are not as prevalent among other income groups. It is important to check on the availability of statistics on the characteristics before launching into sophisticated definitions of group-specific characteristics. It appears to be easiest to identify such characteristics for upper income groups, but it is important to not neglect the middle and lower parts of the distribution lest one fall into a trap akin to 'Pareto's law of income

distribution.¹⁴ The characteristics will most probably differ as between countries, and it is impossible to identify a list applicable in all situations.

IV. THE CONCEPT OF INCOME

In studying the concentration of income, one would like to employ a global concept of income. Economists normally define total income as the sum of wages, rent, interest, profits, and transfer payments. Because of the difficulty of measuring each of these components, the concept of income normally utilized in surveys is less than comprehensive. Furthermore, what is meant by income may differ radically from one country to another; changes in definition or methods of data collection may even cause differences between surveys within a given country. To facilitate comparability both within and between countries, it is important to have a clear understanding of components that may differ or change more frequently than others and how one can adjust for them when they are excluded. One might think that such adjustments should be carried out prior to consistency checks. This may or may not be true, and it is dependent upon whether the specific component is included in the data utilized for the cross-check.

There are essentially eight components of income that are frequently included in some, but excluded in other concepts of income. These are (1) home produced consumption, (2) the imputed rental income from owner-occupied housing, (3) income in-kind offered as remuneration for employment, (4) employer and/or employee contributions to social security schemes, (5) unrealized capital gains, (6) undistributed corporate profits, (7) the imputed value of government subsidized services, (8) taxes, The effect of each component on the distribution of income, and how

one might adjust for these components are discussed in the sections that follow.

A. Home-Produced Agricultural Consumption

Home produced agricultural consumption is the most significant non-money income component in the developing countries. With a high proportion of people engaged in agriculture, the magnitude of subsistence income in-kind can be large. If this component of income is not included in the survey data, some type of estimate should be attempted. The exclusion of home produced agricultural consumption increases measured inequality, since it is a larger proportion of total income for those in the lower three quintiles than for the upper two. Two methods for adjusting incomes for this component will be discussed.

The first method of estimating income in-kind is one utilizing the data of agricultural surveys. These surveys, where available, normally list data such as the size of farm, family size, total production (both crop and livestock), and income. One can use this data to obtain relationships between these variables and subsistence income. Although income is recorded, this figure may also exclude income in-kind. A preferable total income figure can be obtained by the use of average market prices over the relevant period to value total production. Using standard regression techniques, relationships between total income and the aforementioned variables can be determined. These estimates can then be used to adjust the survey incomes for the subsistence component. The disadvantages of this technique are its heavy reliance on the accuracy of the agricultural survey data, the use of what may be unrepresentative average market prices to value production, and the fact that one obtains an average income figure

for all agricultural households possessing a given set of socio-economic characteristics.

The second method for adjusting the data is one based upon the distribution of land and livestock, the average value of production per unit of land for given crops, and the average value of output per head for livestock. Data on the distribution of land should be used with other agricultural data to determine the relationship between size of plot and number of sown acres.¹⁵ Given these data, total agricultural income can be estimated by summing total crop income and total livestock income. Total crop and total livestock incomes are themselves the sums of several components. Total crop income is the sum of the products of the average value of crop output per unit of land and the number of units of land sown with each crop. Total livestock income is, likewise, the sum of the products of the average value of output per head and the number of head of each type of livestock. Summing the totals yields estimated total agricultural income of a household. One, next, must determine the proportion of reported income of each household that is derived from agricultural endeavors. If such an estimate cannot be made, the assumption that all of reported income is agricultural income must be relied upon. The difference between estimated and reported agricultural income is an estimate of the home produced consumption of an agricultural household.

B. Imputed Rental Income From Owner-Occupied Housing

Individuals who rent their residences must expend a portion of their income for the rental charges. Those who own their abodes, on the other hand, incur no such expense. Thus, once a residence is owned by its occupant, it bestows a stream of benefits which is non-money income of the owner-occupant. Several methods have been used to estimate the imputed value

of rental income from owner-occupied housing. Two of these methods are presented here.

It should first be pointed out that if a survey collects expenditure data, a question on imputed rental income is often included. The accuracy of respondent estimates, however, is not altogether clear. In the event expenditure data is collected but no question on imputed rental income is included, one can use the expenditure information of renters to estimate the value for non-renters. Stratifying rent actually paid by income levels, one can simply allocate imputed rent by the income level of non-renters and the rent paid within that income class by those who do rent. Such a procedure is based on the assumption that rent is a relatively constant proportion of income within an income class. In fact, this estimate may be more accurate than an estimate of imputed income given by the respondent, since his is most probably based on historical prices with little account taken of general price increases.

A second procedure that can be used to estimate imputed rental income is one utilized by the ECIEL group.¹⁶ They have a distinct advantage in that their data lists the current market value of an owner-occupied home. One percent per month of the current market value is imputed as income to those who own their residences.

Either of these imputation procedures rests upon detailed information. One must either have expenditure data for renters or the current market value of the owner-occupied home. The question arises whether anything can be done in the absence of such information. Assume that the only data available is that relating to incomes and that it is desired to impute rental incomes to homeowners. Although more imprecise than the former methods, a methodology can be devised for imputation. It is first necessary

to determine the average amount of income allocated to rental expenditure for geographically specific areas of the country. The geographical specificity should be to as small an area as feasible given data constraints. One then determines the extent of home ownership within the geographic areas delimited by the data on average rental expenditures. If the percentage of home ownership within an area is greater than 50, an imputation of rental income is added to all incomes of the area. Total income can be written as

$$Y_{ij} = Y'_{ij} + R_{ij} \quad (3)$$

where Y_{ij} = income of the i^{th} recipient unit in the j^{th} geographic area

Y'_{ij} = reported income (before rental imputation) of the i^{th} unit in the j^{th} area

The quantity R_{ij} can be defined as

$$R_{ij} = r_j Y_{ij} \quad (4)$$

where r_j = average percent of income allocated to rental expenditure in the j^{th} geographical area

Substituting (4) into (3) yields

$$Y_{ij} = \frac{Y'_{ij}}{(1-r_j)}$$

which is an estimate of total income after the imputation of rental income from owner-occupied housing. The obvious disadvantage of this procedure is that an overestimate of income results for all income units that are renters, i.e., that actually do pay rent. The logic behind the 50 percent cutoff for the determination of whether an adjustment should or should not be made is that in cases where more than half the income units own their residences, an adjustment, such as that suggested, will decrease the errors of income measurement due to non-monetary components of income.

There is a basic problem with all imputations of income from owner-occupied housing. The problem is a result of the fact that almost never will one know the actual percentage of a home that is owned. Even if "owner-occupiers" can be perfectly identified along with the exact market value of the home, an overestimate of imputed income will result in cases where the occupant does not own the entire residence, i.e., where a mortgage has been only partially paid. In such cases imputed rental income is overestimated by an amount proportional to the share of the residence that is not owned.¹⁷

With all the difficulties and inaccuracies that are part and parcel of the imputation process, one must seriously consider whether the end justifies the means. The decision whether to undertake the imputation must rest on the answer to the question, "Are the adjusted incomes more representative of reality than the unadjusted ones?"

C. Income In-Kind Offered As Remuneration For Employment

Income in-kind as a partial remuneration for employment occurs in both agricultural and non-agricultural sectors. In the main, these payments take the form of free housing and food. The primary groups that receive remuneration in this form are agricultural laborers and domestics. It is difficult to estimate this component of income with any accuracy, and most researchers have relied on a lump sum adjustment to the incomes of those in given professions. Webb, for example, in estimating income in-kind for domestic servants adds 375 soles to the incomes of those employed in Lima, and 300 soles to those employed elsewhere. This estimate is based upon rent for a one-room home and on a low-income food budget in 1961.¹⁸ The important things to note in making an adjustment for income in-kind as remuneration are the types of workers to which the adjustment applies,

the type of accommodations, if any, they receive and the amount of food allocated to them. For a given group of workers, each individual's adjustment is the sum of the values of the latter two components.

D. Social Security Contributions

In most cases, social security contributions are payments by an employee to a national "pension" fund with a matching amount paid by the employer. It is important to identify whether incomes are measured gross or net of the employee payments, and even in cases where they are measured before payment, one should attempt to add the employer's contribution to the employee's income. The employer's contribution is part of the employee's earnings and, thus, should be considered part of his income.

Coverage of social security systems is not usually universal. It is necessary, therefore, to attempt a separation of insured from non-insured. The usual approach to this separation is by working classes, e.g., blue-collar and white-collar, since social security coverage is normally based upon job characteristics. One can further cross-classify the coverage by income class. This decomposition of the workforce can also prove helpful in adjusting earnings for: (1) bonuses, (2) fringe benefits, e.g., health insurance, and, (3) profit sharing. Once the decomposition has been completed, the employer's contributions can be estimated by the published rate schedule. If employee incomes are measured net of contributions, account can also be taken of their payments. The main caution to be observed in this procedure is to see that totals match. Social security payments larger than those recorded as being received by the government are inconsistent with a proper adjustment.

E. Unrealized Capital Gains

An unrealized capital gain is an increase in the value of an asset

prior to the disposal of that asset by sale. Once the asset is sold at the revalued price, the gain becomes realized. Unrealized capital gains never appear as components of income in income distribution data. Due partly to the volatility of asset values and partly to the poor data on asset holdings, estimates of this transfer payment component of income are never attempted. Because unrealized capital gains are disregarded and the prime concern of the adjustments in this section is to obtain consistency in the income concept and data, no method of adjustment will be suggested. One should realize, however, that the exclusion of unrealized capital gains lowers measured inequality, because they accrue mainly to upper income groups.

F. Undistributed Corporate Profits

Undistributed corporate profits are savings of the stockholders of the business. To distribute these savings in the absence of detailed data on stock ownership, one must decide on the groups of income receivers most likely to own the majority of the business. Unlike the case of a developed country where ownership may be more dispersed, corporate stock is held predominantly by the top income groups in developing countries. The decision as to whether the top 0.1 or 20 percent control corporate stock is nevertheless country-specific and must be based on whatever available information exists.

Once the percent of top income earning units that control corporate stock is decided upon, the distribution of undistributed corporate profits can be assumed to be linearly related and proportional to income. That is, total undistributed corporate profits can be allocated on the basis of the proportion of total income of the assumed stockholding group that is received by a given household. These estimates will err for two reasons:

(1) There are most certainly some income units in income groups below the assumed stockholding ones that hold corporate stock, and

(2) The relationship of undistributed corporate profits within the assumed stockholding group to income deviates somewhat from a linear proportional one.

To attempt to achieve greater accuracy, however, involves acquiring unavailable data or using an assumed relationship between income and undistributed profits that is as arbitrary as the linear-proportional one.

G. The Imputed Value of Government Subsidized Services

A discussion of the incidence of government expenditures might appear to be far afield from minor adjustments to income statistics. There are, however, several government expenditures that are akin to privately paid income in-kind with the exception that they are supplied by the government. Examples of these expenditures are low-income housing projects and food budget supplements (such as food stamps). A review of the government budget and poverty programs will reveal the exact categories of expenditures to be considered in each case. The desire in the review is to identify those government expenditures that are designed to operate exclusively on groups in poverty.

There are wide latitudes of choice for imaginative methods of allocating these expenditures. As has been the case in many of the previous discussions, the choice of a method is dependent upon country-specific characteristics and available data. Illustrations will be given of how one might approach the allocation of low-income housing subsidies and food budget supplements.

Data should be available on the number of housing units that are federally subsidized and their location. Hopefully, the geographic breakdowns of these data are compatible with those of available income distribution data. The overwhelming majority of subsidized housing is usually located in urban areas, and most survey data on incomes can be classified by rural and urban and oftentimes geographic breakdowns. On the assumption either that the range of federally subsidized prices for housing units of different sizes is not large or that the federal subsidy per housing unit is constant regardless of size, one can use the average subsidy per unit as the figure to be allocated to subsidized housing residents' incomes. The subsidy is arrived at by estimating the average size of a subsidized unit, the subsidized price of this unit, and the market price for a unit of this size and location. Since federally subsidized housing should theoretically benefit the poorest households in its immediate location,¹⁹ one can (beginning at the bottom and working upward) allocate the average subsidy to the incomes of households in the lowest portion of the income distribution of that location. The number of households to whose incomes the allocation is made should not exceed the number of subsidized housing units. Other information (perhaps informal conversations) may lead one to believe that this allocation procedure is incorrect. The additional information may afford an opportunity to more accurately determine the socio-economic groups to whose incomes the subsidy should be allocated.

The size of a food budget supplement is often based upon family size and/or income level. If information on household or family size is available from the census or survey, it is important to investigate the

relationship between this and the supplement formally specified for the program. Should size data not be available, one must assume that each household is the size of the average one. Information on the relationship between the size of the supplement and income level is important in any event. This information along with the total amount of the supplements permits one to allocate the supplements to households. In the absence of anything to the contrary, one should assume the benefits of the program accrue to the lowest income groups.

The allocation procedure will be illustrated for the cases where household size is and is not available from the census or survey. Assume that the total amount of supplements is known in either case. In the absence of data on household size, the assumption must be made that each household receives the same amount. (This is equivalent to assuming that each household is the size of the average one.) The supplement per household is then equal to the total amount of supplements divided by the number of households. One must, however, estimate the number of households receiving supplements. This can be estimated by resort to the income level cutoff for receiving benefits. The resulting average supplement is then added to the incomes of households below the cutoff line. This procedure has the benefit of a built-in check on actual and allocated totals.

The second case assumes that the food budget supplement program specifies a relationship between family size, income, and the amount of the supplement and that household size is available from the data of the census or survey. One simply allocates the supplements on the basis of household income and family size. Since this procedure has no built-in checks on totals, care must be exercised to assure consistency of the actual and allocated amounts. Seldom will these amounts balance if each eligible

household, determined by income level, is allocated the amount specified by the formal program stipulations. When the amount allocated is less than the actual total expended by the program, one may opt to simply not worry about the remainder -- the balance can be assumed to be part of the administrative costs of the program. On the other hand, the remainder can be allocated among the eligible households. A satisfactory procedure for the allocation is one which apportions the remainder on the basis of the relationship of the household's original supplement to the total allocated supplements. That is, the excess is apportioned to eligible households on the basis of the proportion that the households' original supplement forms of total allocated supplements.²⁰ The primary benefits of this procedure are its ease of application and the fact that it does not alter the relative amounts allocated to eligible households.

The allocation procedure for the remainder suggested has concentrated on the case in which the amount allocated is less than the actual amount spent by the program. One of the options discussed was that of disregarding the remainder. In cases where the allocated amount is larger than the actual amount this option should not be considered, since this is equivalent to creating government expenditures where none, in fact, exist. In the case, the allocation procedure specified above should be used to subtract an amount from each allocation so that allocated and actual expenditures balance.

H. Taxes

Total taxes can be decomposed generally into direct and indirect ones. Direct taxes are composed primarily of individual income taxes, while indirect taxes relate to those assessed on consumption expenditures. Although the effect of these taxes on the concentration of incomes is of

primary concern in adjustments, the regressivity or progressivity of either type is not the point at issue. Income statistics should be measured before the imposition of either tax. A distribution of income net of either or both types of tax does not allow one to assess the inequality of total incomes. Those statistics are preferred which allow maximum flexibility in researchable questions. When data are scarce, however, one must make due with what is available. The aim of the present section is, therefore, to specify methods of adjusting income data that is, in its published form, net of direct and/or indirect taxes.²¹

In adjusting incomes from an after to a before-tax basis, one necessarily cannot be extremely precise. Two individuals with equal incomes may pay different rates of direct and indirect taxes because of differing amounts of either income exempt from taxation or expenditure. The less error one is willing to suffer, the finer the breakdown of the income intervals one should employ. As with all other adjustments, the precision possible is dependent upon the quality and quantity of available data.

Since individual income taxes are by far the largest component of direct taxes, especially in developing countries, taxes such as inheritance or gift taxes can be disregarded. In choosing the ranges of income intervals for which tax rates will be calculated it should be borne in mind that income distribution is normally right-skew, i.e., the majority of the observations lie left of the mean. As a result, the range of an income interval should increase with income. In this initial stage it is also important to note the legally set level of income below which income is tax exempt. This level sets a cutoff below which further calculations are unnecessary.²² It is important to recognize that reported incomes, i.e., those net of direct taxes, usually include income exempted from taxation whereas income before

direct taxes does (if it is considered taxable income). With this information, the following equation can be derived and it can be utilized for adjusting incomes from after-tax to before-tax:²³

$$Y_b = \frac{Y_a - Xt}{1 - t} \quad (5)$$

where:

Y_b = before-tax income

Y_a = after-tax income (reported income)

X = income exempted from taxation

t = tax rate

The procedure for adjusting incomes, given the above equation, is a simple one. Income intervals are defined for before-tax income. The average income exempted from taxation is then calculated for each interval. This calculation is based upon the midpoint income, average family size, and average expenditure pattern of the interval. It is preferable to use average family size cross-classified by income interval should the data exist, but use of the national average or region-specific average will not usually introduce a large bias.²⁴ Subtracting the average exemption from the midpoint income yields average taxable income for the interval. From this, tax rates can be obtained and income net of taxes can be calculated. Adding back the interval-specific exemptions yields reported income net of direct taxes. Intervals of income net of direct taxes can be obtained easily from the results.

The values obtained by the above calculations can be used to adjust the after-tax incomes recorded by the survey. Once the interval in which an individual income figure falls is identified, the values of X and t are known and can be used in equation (5) to determine total income before

tax (Y_p). Although this is admittedly an imprecise procedure, it should help narrow the difference between after-tax and total income for the majority of observations.

There are several ways in which income observations can be adjusted for the exclusion of indirect taxes. All methods are based upon expenditure patterns and the difference between the procedures is determined by the level of disaggregation of the data. The two polar cases are: (1) that in which detailed income recipient unit expenditure data is available, and (2) that in which one must rely on average expenditure patterns for given income groups, e.g., those of low, medium and high income. The methodology of the adjustment is the same regardless of the level of disaggregation of expenditure data.

The amount of indirect taxes paid by a recipient unit is equal to the sum of the products of commodity-specific tax rates and expenditures of the recipient unit on the commodity. In cases where detailed expenditure data are available for each recipient unit this method of estimation is applicable to each unit. In the absence of such detailed data, one must rely on whatever breakdowns of expenditure patterns are available. The latter case corresponds to an allocation of an average amount of indirect taxes to each income receiving unit in a given income group. The average allocation will contain greater inaccuracies, and the degree of inaccuracy is directly related to the variance in expenditure patterns within each income group.

Shifting Assumptions and Taxation Incidence

Shifting, in its conventional usage in public finance theory, refers to the procedure by which the burden of taxation is passed from the place where statutory liability is imposed to the place where the burden

finally comes to rest. Because it is impossible to isolate the particular chain of events that occur in a general equilibrium system, the current approach to determining final incidence attempts to evaluate effects on sources of funds (the incomes received by producers in the taxed sector) and users of funds (the consumption expenditures of individual households).²⁵

The effect of taxes on absolute prices is not of interest in the analysis of incidence and shifting. Since it is the distribution of real incomes that are available for private use that is at issue, the effects on relative prices and relative factor incomes are the foci of attention. The usual analysis assumes perfect competition, price flexibility, and perfect factor mobility. Furthermore, the assumption is employed that factors receive the value of their marginal products. Although these assumptions may not hold in the short run, they are not unreasonable for the longer run in which adjustments to relative factor and output prices can occur.²⁶

Pechman and Okner have pointed out that this approach to incidence analysis suggests certain conclusion for the allocation of tax burdens and shifting assumptions:

1. The individual income tax probably is not shifted since workers and investors do not appear to alter working hours or savings in response to changes in tax rates. Only in the cases of young people and some women do income taxes appear to affect labor force participation or hours worked. These groups account for a relatively small proportion of the total labor supplied. Despite large changes in tax rates and in its composition, the private savings rate has been constant for many years. Based upon the presumption that the imposition of the tax does not change either the demand for, nor the supply of, factors of production and thus

factor proportions remain unchanged, a tax on incomes must be borne by those upon whom it is imposed. Furthermore, the tax does not change relative commodity prices and thus there is no burden on the expenditure side of household budgets.

2. Indirect taxes can be classified as general sales taxes and excise taxes (commodity-specific taxes). A general sales tax is borne by customers in proportion to their total expenditures. Relative prices are unaffected and thus consumption patterns are unaltered. Excise taxes do affect relative prices, thus forming a burden to those who consume the commodities subject to tax. There is no burden on the sources side of income, however, because any labor or capital displaced by the tax ultimately will receive approximately the same income when reemployed in an untaxed industry.

The implication of both these points is that no portion of income, general sales, or excise taxes is shifted and the burden falls on those upon whom it is imposed by statute. The adjustment mechanisms laid out in the previous section do not assume any shift of the taxes. There is, however, one tax which may be significant in the case of a developing country and which may be at least partially shifted. This is the tax on corporate income,

Much discussion, both theoretical and empirical, has centered around the shifting of the corporation income tax. The traditional view has been that the corporation income tax is not shifted in the short run, irrespective of the degree of competition in the affected market. Maximization of profits occurs at that level of output where marginal cost is equal to marginal revenue. Assuming that the firm seeks to maximize profits,

the imposition of a profits tax will alter neither the price nor the output that maximized profits prior to the imposition.

The converse argument is based on the belief that firms do not necessarily attempt to maximize profits. Whether one believes that firms set prices on a cost-plus basis or that firms have a target rate of return, the implication is that the tax must be shifted either to consumers or to workers, or to both. The issue of whether or not such a shift occurs has not yet been resolved.

If one opts for the belief that corporate income taxes are shifted, then a decision must be made vis-à-vis the allocation of total taxes to owners and workers of the firm and consumers of the product. A breakdown that has often been used is Musgrave's $1/3 - 1/3 - 1/3$ rule which apportions a third of the total taxes to owners, workers and consumers. One must then decide how the groups can be identified in income distribution data and how the tax can be allocated within each of the groups.

Owners of the firm are normally in upper income groups, especially in developing countries. As a result, the upper decile or quintile can be chosen as the income group in which all corporate owners are located. Under the assumption that corporate income is closely related to total income for owners, their proportion of total corporate taxes can be allocated on the basis of the proportion each recipient unit's income bears to the total of the quintile or decile.

In the absence of information to the contrary, it can be assumed that all income groups consume the products of corporations. The consumer share of corporate income taxes thus can be allocated to all income groups on the basis of the proportion each recipient unit's expenditures bear to total consumption expenditures.

The major difficulty occurs in apportioning the share of the tax borne by workers. Those to whom the owner's share of corporate taxes have been allocated can be excluded immediately. One would also like to exclude those not employed by corporations, i.e., primarily agricultural workers. Reasoning that agricultural incomes are lower than others, this group can be expected to be concentrated in the lower tail of income distribution. As a result, the lower X percent of the distribution can also be excluded from the allocation procedure, where X is the proportion of the population in rural areas.²⁷ Assume that X is equal to 40 and that corporate owners are concentrated in the upper quintile. The proportion of corporate taxes borne by corporate workers then can be allocated to those comprising the middle and upper-middle quintiles of the distribution. Since the tax borne by workers is related to their incomes, it can be apportioned on the basis of the proportion that each recipient unit's income comprises of the total income in the two quintiles.

V. INCOME DISTRIBUTIONS BASED UPON EXPENDITURE DATA

India is a notable example of a country in which the national household survey collects data on expenditures but not on income.²⁸ In the absence of information on the average propensity to save of different income groups, certain assumptions are necessary in order to change the distribution from one of consumption to one of income.

Total income of each recipient unit is comprised of consumption and savings. Since savings is unknown in the case of expenditure data, total income is unknown. An estimate of savings thus is necessitated. In the usual case, the estimate is obtained by assuming a certain relationship between the average propensity to save (or, its complement, the average

propensity to consume) and total income. Most econometric investigations have found a direct relationship between APS and income (which, of course, translates to an inverse relationship between APC and income).²⁹ Once one values the relationship, an estimate of recipient unit income can be obtained by dividing consumption by the average propensity to consume.

Where data on income and either consumption or savings have been collected for developing countries, it has usually been found that the positive relationship between income and savings is stronger than in developed countries. That is, the great majority of saving is done by upper income groups. It is for this reason that Ranadive assumes that positive savings occur only at the upper levels of income and that at all other levels average savings are approximately equal to zero. Based upon a small urban and rural saving survey undertaken by the NCAER, Ranadive assumes that urban households with annual incomes less than RS. 2000 and rural households with incomes less than RS. 720 have zero net savings.³⁰

Ranadive's assumptions concerning the relationship of the average propensity to save and income relate to decile group rates of savings. It is possible to transform this assumption to one relating to absolute income levels, and the NCAER data that she used would most probably allow a more precise specification of the relationship. In cases where such peripheral data sources are unavailable one may have to rely on informal information or the savings rates that exist in other developing countries at a similar level of development.³¹

VI. DIFFERENTIAL PRICE TRENDS

In assessing changes in the income of a recipient unit over time, account must be taken of the generally rising price level with its consequent effects on real income. This is not a problem in assessing changes

in income distribution over time unless price changes differ as between income groups. Most measures of income inequality are based upon relative standings in the distribution, and the deflation of all incomes by a constant will not alter those relative positions. Previous investigations have found, however, that both marketbaskets for income groups and price changes for commodities vary to an extent great enough to influence relative standings significantly. In India, for example, the wholesale price index of foodgrains rose 93 percent in the eight year period 1961-1969, whereas that for finished manufactures rose only 36 percent during the same period. Since foodgrains comprise a larger proportion of the budget of the poor than of the rich, the former group probably experienced a larger drain on purchasing power. The most ambitious attempt at constructing income-group-specific price indices for rural India has found that, relative to 1960-1961, price indices for the top and bottom 5 percent of income distribution were 172.98 and 191.13, respectively, in 1967-1968.³²

Some countries, such as the Philippines, construct price indices for low, middle, and high income groups. Taiwan is at present beginning the calculation of such indices. Where such statistics are available, they should be utilized to adjust the incomes of the relevant groups. Where they have not been constructed, some attempt nevertheless should be made to convert incomes to real terms. If it is generally true that price increases are larger for the poor than for the rich, income distributions in current dollars can mask significant amounts of inequality.

An important consideration in the construction of group specific price indices is the level of aggregation of commodity groups. Price changes for a category such as cereals may mask significant variation between coarse and fine cereals. One can also draw this analogy for coarse and fine textiles.

The work on India discussed above uses ten commodity groups. All items of non-food other than fuel, light and clothing are, however, lumped together as non-food. Although this category is relatively unimportant for the poor, it does loom large in the budgets of the middle and upper income groups. Any impact of differential price increases in the important sub-groups of the 'non-food' category is lost because of the excessive aggregation. Attempts should be made to disaggregate the commodity grouping to as low a level as is feasible. Some variation will always be lost in the construction of such indices, but one would like to keep the loss to a minimum.

VII. SEASONALITY OF DATA

Timing of data collection is an important factor influencing the results of a sample survey of income. Data are usually collected in a given week, month, or quarter of the year and the time of this collection can influence markedly the estimates of both income and inequality. For example, Korean data on household income indicates that the inequality of the distribution in the rural sector is higher than/in the urban. This result is, of course, exactly opposite the usual expectation. It is attributed to the fact that income is derived from averaging the monthly incomes of January through March, the off-season in agriculture.³³ During the off-season, unemployment can be expected to be higher, and thus the spread of incomes larger, in agriculture. Without a corresponding off-season in the industrial sector, one might expect this bias to creep into the results. It should be recognized that a bias toward underestimating inequality can also result from surveys taken during the peak season.

In the absence of data relating to periods other than that to which the biased data relates it is virtually impossible to adjust incomes for cyclical bias and one is forced to accept and note transitory inequality.

The best one could hope to do is to estimate average yearly earnings by occupation and match this with the occupations listed by respondents in the survey. To estimate average earnings, agricultural surveys may be used for the rural sector and industrial surveys or censuses of manufacturing for the urban sector. Such a procedure can be used to develop a floor income for each occupation below which recipient unit incomes should not fall. Since a great number of assumptions and arbitrary estimations are involved in this procedure, however, it is not clear that the end result justifies the means. It is perhaps preferable to attempt to obtain data and an estimate of inequality from other sources so as to obtain a gauge for judging the amount of bias in the data.

VIII. SUMMARY

The previous discussion has assumed that either census or household survey data are available. A preference for the latter over the former also was expected. There are, however, countries in which these data still do not exist or are so poor in quality as to render them useless. In such cases, peripheral data sources must be used to build a distribution from the ground up. One can apply certain of the techniques discussed in this paper for the estimation of some components of income. Complete methodologies for estimating income distribution are contained in Morrison and Webb.³⁴

Before one proceeds to an analysis of income distribution, it is useful to address the following questions.

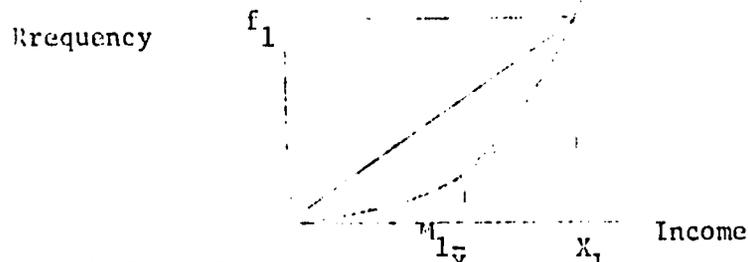
1. What are the existing sources of data directly related to income distribution within the country under consideration?
2. Is there a sufficient amount of detail in these data to allow specialized analyses?

3. What are the sampling properties of the data collection procedure, and do they lead to unbiased results?
4. Are there any strong reasons to believe that non-sampling errors of the data are large?
5. Can peripheral data sources -- that is, ones not containing recipient unit data on income explicitly -- be used in conjunction with the data directly related to income distribution for specialized analyses?
6. What concept of recipient unit is utilized in the collection of directly related data and should an adjustment to another type of unit be undertaken?
7. Is the directly related data consistent with data of known reliability?
8. What is the concept of income employed, and should adjustments be undertaken to achieve a more global concept?
9. If the recipient unit data is other than income, e.g., expenditure, is there sufficient information to allow this concept to be transformed to total income?
10. Have the prices facing various income groups changed at differential rates, and if they have, how might one account for the effect of these differential changes on growth rates of real income?
11. Do the time series data jibe with the theoretically expected results of historical facts?

Some of these questions have not been treated in as much depth as they warrant, whereas others have occupied perhaps more space than they should have. Other papers that list and evaluate the data sources of specific countries may deal with the underrepresented questions in more depth.

FOOTNOTES

1. Comments of Albert Fishlow on 'Data Problems and Measurement of Income Distribution in Developing Countries,' during a research workshop held in conjunction with Brookings-Princeton project on Income Distribution in Developing Countries, Princeton, N.J., October 16, 1973.
2. It is normally assumed that national accounts data is more accurate than either survey or census data. For an example of such a consistency check, see Jo-Lo Han, The Reliability of Income and Expenditure Data Collected in Taiwan's Three Household Income Surveys (Taipei, Taiwan, The Institute of Economics, unpublished, August 1973).
3. See P.K. Bardhan, The Pattern of Income Distribution in India: A Review (I.B.R.D., mimeo, June 1973), pp. 3-5.
4. The term household will be used as interchangeable with family throughout this section.
5. See Martin Bronfenbrenner, Income Distribution Theory (Aldine-Atherton, 1971), p. 37.
6. Oey Astra Meesook, 'Review of Income Distribution Data. Thailand, Malaysia and Taiwan,' Princeton University Research Program in Economic Development, Discussion Paper No. 60, 1975, p. 2-3.
7. Assuming a lower bound of zero and upper bound of x_1 for the lowest income class, the mid-point of the lowest income interval can be shown graphically to be less than the actual mean of the interval:



8. This method results in an estimation of the Pareto-Levy coefficient which is an interpolation based on only two income classes. Alternatively, one can improve the estimate by using more than two classes as long as they lie above the model income. For applications of this procedure, see Shail Jain, 'Size Distribution of Income: Compilation of Data,' (Washington: World Bank, 1974), or Lin Lin Lean, 'The Pattern of Income Distribution in West Malaysia 1957-1970,' (Geneva: ILO, 1974) or Richard O. Vada, 'Changes in the Size Distribution of Income in Postwar Japan,' (Geneva: ILO, 1974).
9. This and the other approximations discussed in this section are taken from Christian Morrison, Note on a Methodology for Estimating Income Distribution, (I.B.R.D., mimeo, 1973).

10. See Han, The Reliability of Income, op. cit.
11. See Simon Kuznets, "Quantitative Aspects of the Economic Growth of Nations, II Industrial Distribution of National Product and Labor Force," Economic Development and Cultural Change, Supplement, (July 1957).
12. One must be careful to check whether the national accounts data is based upon the survey. See Morrison, Note on a Methodology, p. 11.
13. See Richard C. Webb, The Distribution of Income in Peru, Research Program in Economic Development Discussion Paper No. 26, (Woodrow Wilson School, Princeton University, September 1972).
14. According to Pareto, the upper brackets of income distribution can be accurately represented by the frequency function:

$$X = AY^{-\alpha} \text{ or in log form } \log X = A - \alpha \log Y$$

where Y = income

X = proportion of income receivers with income equal to or greater than Y

A, α = empirical constants

The Pareto distribution has often been used to characterize income distribution. In fact, it is only valid for incomes above the mode even in cases where it is accurate.

15. Two methods for estimating home-produced consumption are specified in the paper. The second one is based upon the distributions of land and livestock, the average value of production per unit of land for given crops, and the average value of output per head of livestock. The equation to be used for estimating total agricultural income is

$$X_i = \sum_j L_i s_{ij} c_j + \sum_k h_{ik} m_k \text{ for all } i$$

X_i = estimated total agricultural income of the i^{th} household

L_i = size of plot of the i^{th} household

s_{ij} = proportion of total plot sown with the j^{th} crop

c_j = average national (or regional or provincial) value of output per unit of land sown with the j^{th} crop

h_{ik} = number of head of the k^{th} type of livestock of i^{th} household

m_k = average value of output per head of the k^{th} type of livestock.

16. The ECIEL group is one concerned with Latin American economic integration. Latin American research institutes and the Brookings Institution collaborate on studies related to the problems of and prospects for economic integration.
17. In cases where occupants do not own 100 percent of their residences the imputed value of rent is overestimated. In such cases, the desired value of imputed income is

$$II_i = \alpha (V_i) (Q_i)$$

where II_i = imputed rental income of the i^{th} recipient unit

α = percentage of value of owner-occupied home imputed as income

V_i = value of home for the i^{th} income unit

Q_i = percentage of home owned by the i^{th} income unit and $0 \leq Q_i \leq 1$.

In the absence of information on Q_i it is assumed equal to unity, with a resulting overestimate of II_i of¹

$$\text{OVERESTIMATE} = \alpha(V_i) (1-Q_i) \quad (Q_i < 1).$$

18. See Webb, The Distribution of Income in Peru, p. 35.
19. Oftentimes information is available on deviations from this theoretical principle. For example, civil servants in India are given housing subsidies. The civil servants are not usually members of the lowest income groups.
20. See Appendix for the methodology.
21. Direct taxes are normally expected to be progressive, while indirect taxes, because consumption is a larger proportion of the incomes of the poor than of the rich, are usually regressive. The total effect of the fiscal system and of each type of tax are discussed in R.H. Bird, and L. de Mulf, Fiscal Incidence Studies in Developing Countries: Survey and Critique, (International Monetary Fund, Fiscal Affairs Department, mimeo, February 1974).
22. It should be noted that some incomes in the income interval just below the cutoff may have paid some taxes. Income net of direct taxes falls below the cutoff, but that gross of taxes does not. It may be necessary to adjust incomes in the first interval below the cutoff.
23. See Appendix 1 for the derivation of the equation.

24. In most cases, the variance in family size can be expected to be small. There may be a large difference in rural vs. urban averages, however, and some account should be taken of this.
25. See Richard A. Musgrave, The Theory of Public Finance (McGraw Hill, 1959), pp. 205-31.
26. For a complete discussion of the points in this paragraph, see Joseph A. Pecorman, and Benjamin A. Okner, Who Bears the Tax Burden? (The Brookings Institution, 1974), pp. 27-9.
27. One would prefer to use the proportion of the total working population employed in agriculture. This proportion, however, may either not be available or may be inappropriate for use in the context of an LDC, since it may not include family members (primarily children) whose employment goes unrecorded.
28. This is true of the National Sample Survey (NSS), although the National Council of Applied Economic Research (NCAER) has undertaken ad hoc samples in which income is explicitly included. The main difficulty with the data of the NCAER is that the sample sizes are usually quite small leading to large sampling errors.
29. Friedman's consumption function is a notable exception to these findings, although his relates to permanent income and consumption. Friedman's consumption function, of course, leads to constant average and marginal propensities to consume and save throughout the income distribution. See Milton Friedman, A Theory of the Consumption Function, (Princeton University Press, 1957).
30. A second set of assumptions uses income levels of Rs. 300 and Rs. 1200 as breakeven levels of income and consumption. See K.R. Kanadive, 'The Equality of Income in India,' Bulletin of the Oxford University Institute of Economics and Statistics, Vol. 27 (May 1965).
31. Examples of other methodologies that have been used to adjust the Indian data are discussed in Bardhan, The Pattern of Income, op.cit.
32. See A. Vaidyanathan, Some Aspects of Inequalities in Living Standards in Rural India, (paper presented to the Indian Statistical Institute, Seminar on Income Distribution, February 1973).
33. See Hakchung Choo, 'Review of Income Distribution Data - Korea, The Philippines and Taiwan,' Princeton University Research Program in Economic Development, Discussion Paper No. 55, 1975.
34. See Morrison, Note on Methodology, op. cit. and Webb, The Distribution of Income in Peru, op. cit.