

# GUIDELINES TO THE CONSTRUCTION OF A SOCIAL ACCOUNTING MATRIX

BY STEVEN J. KEUNING AND WILLEM A. DE RUIJTER<sup>1</sup>

*Institute of Social Studies and TEBODIN Consulting Engineers, The Hague*

The increasing number of countries for which a Social Accounting Matrix (SAM) has been compiled testifies to the usefulness of this integrated data framework. Considerable resources are always involved in the construction of a SAM, for it provides a comprehensive description of an economy with emphasis on distributive aspects. This means that, unlike other data systems, incomes and expenditures of several categories of households and their relation to the production structure, the balance of payments and transactions by other institutions are shown.

However, apart from this minimum requirement, no standardized concepts and guidelines for SAM construction are as yet available. Although a SAM should stay as close as possible to the specific (institutional) reality of the economy it describes, some general remarks as to its design and compilation are in order. This paper represents a first attempt in that direction. After a general introduction to SAMs, each stage of the construction process is reviewed in turn.

The construction process begins with the overall design of the system and various options are discussed. This section includes a schematic representation of a fairly extensive SAM. Next, the sources for the SAM need to be identified, and a provisional checklist is given here. After an overview of considerations regarding the choice of a reference year, the topic of classification in the SAM is reviewed in detail. Finally, the paper describes how the different data sets might be integrated and reconciled for consistency.

The guidelines may also aid in designing a time schedule and in organizing the work when constructing a SAM.

## 1. INTRODUCTION

It is more than a decade since the first Social Accounting Matrices (SAMs) were constructed. Both the development and the application of this accounting framework arose from a growing dissatisfaction with the existing practice of national accounting, particularly its exclusive emphasis on measuring economic growth.<sup>2</sup> After it had become apparent that economic growth *per se* is no guarantee for an increase in living standards of all population groups (not to mention a sufficient condition for the eradication of poverty), more information on distributional issues was called for. Although the study of inequality started much earlier, as is evidenced by the long history of a summary statistic like the Gini coefficient, the explicit linkage with growth issues is of a relatively recent nature.

<sup>1</sup>Our practical experience was gained in participating in the construction of two subsequent, independently built, SAMs for Indonesia and in setting up a structure to compile a second SAM for Sri Lanka. We are greatly indebted to Roger Downey and the staff of the Central Bureau of Statistics in Jakarta for their ideas and encouragement. Of course, our gratitude extends to colleagues and referees who gave useful comments on an earlier version, and in fact to all those who have constructed a SAM: their experiences served as our example. Responsibility for the views expressed here lies solely with the authors.

<sup>2</sup>The study by Pyatt and Thorbecke (1976) is generally considered as the first comprehensive description of the SAM framework, including a justification of its design. Soon afterwards a book was published which contained a completely worked out example applied to the case of Sri Lanka (Pyatt and Roe, 1977).

Stone (1985) points out that the topic of distribution, and in particular the distribution among households, of income, consumption and wealth, was not yet exhaustively covered in the revised System of National Accounts (SNA) as published by the United Nations in 1968. This omission was remedied to some extent in a report containing provisional guidelines on statistics in this area (United Nations, 1977). Even the relation with all kinds of social and demographic statistics has already been worked out (United Nations, 1975). These theoretical developments are, however, hardly reflected in the national accounts statistics which at present appear throughout the world. Developing countries, in particular, tend to publish only consolidated income, outlay and capital finance accounts, distinguishing at most a few aggregated institutions as prescribed by the SNA. Until recently more detailed information within this system was available only for the production accounts, in the form of Input-Output tables (I-O). Perhaps the popularity of the Input-Output framework explains why the SAM, which can be considered as an extension of an I-O table, originated from research for a pragmatic data system in which both macro-economic aggregates (the growth indicators) and distribution and redistribution (through taxes and such) could be recorded, and thus integrated.

A SAM can be defined as a numerical representation of the economic cycle with emphasis on distributive aspects. As in the complete System of National Accounts (United Nations, 1968, Table 2.1) and in the I-O framework, transactions in a particular year appear in a matrix format, showing receipts on the rows and outlays in the columns (see Table 1 in section 2 below). Briefly, a SAM shows how sectoral value added accrues to production factors and their institutional owners; how these incomes, corrected for net current transfers, are spent; and how expenditures on commodities lead to sectoral production and value added. The "leakages" from this cycle, for example in the form of payments abroad or savings, are also shown. In turn, capital finance may then be linked to savings, thereby presenting a glimpse of the dynamics in an economy.

The essence of a SAM lies in its comprehensive recording of inter-relationships at the meso-level. First of all, this means a disaggregation of the household sector and usually also of the various categories of value added. Secondly, primary inputs into production and final (household) demand are linked. But tracing distributional mechanisms should go even further, since various goods and services may not be produced by a uniform technology throughout the country concerned, which is in turn related to income distribution. Or, apparently homogeneous commodities may be traded in different markets at different prices for consumption by specific population groups (e.g. subsistence production by farmers). Consequently, the commodity and industry classification changes as well. This also implies that, contrary to the SNA, achieving international comparability is *not* a main purpose of SAM construction. Because of its direct relationship to national (and possibly sub-national) planning and policy-making, a SAM should stay close to the institutional reality of the geographical area under study. Besides, a SAM is always constructed by means of integration of diverse statistics at the meso-level, employing almost all available basic data which refer to a certain period, so that the results may not agree with a straightforward disaggregation of national accounts totals. Finally, a SAM always has a matrix format

because of its emphasis on the identification of source and use of all transactions.

Summarizing, a SAM in our view serves as an alternative for traditional Input-Output tables, but as a supplement to traditional national accounts statistics which remain necessary, if only for the sake of a summary overview of the economic situation and for international comparisons.<sup>3</sup> In turn, a SAM should ideally be complemented by satellite accounts, containing:

- (a) a decomposition of most SAM values into prices (including wage rates, tax rates and so on) and volumes (consumption, employment, etc.),
- (b) other (non-monetary) socio-economic indicators such as household composition, other demographic data, intake of nutrients, housing situation, health conditions and access to education,
- (c) stocks underlying the SAM-flows like population (size and educational background), capital stock (land, livestock, industrial capacity and housing), foreign debt, equity ownership and durable goods possession, and
- (d) a re-routing of some of the SAM-flows (e.g. for the study of the incidence of public expenditures these are, wherever possible, allocated to the beneficiaries).

The information in these supplementary tables should then be consistent with the SAM values. This will be worked out below. The complete data set could be tentatively labelled: a System of Socio-economic Accounts (SSA).<sup>4</sup>

Gradually, more researchers and policy makers are becoming convinced that the combination of data in a SAM permits a better analysis of the occurrence of poverty and inequality in living conditions, both as such and as factors hindering economic growth. The increase in the number of countries for which a SAM has already been compiled also testifies to this. However, considerable resources are always involved in such an exercise. These costs would be reduced if a manual for the construction of SAMs were to become available. Moreover, since the choices made at an early stage largely fix the options later on, it is preferable to evaluate the implications of various construction methods and to form an idea about possible problems en route before one starts. Otherwise, decisions that seemed sensible at the beginning may backfire at a later stage.

This paper does not provide an elaborate blueprint of the construction process; it only argues that a number of stages can be distinguished, and also contains some observations about them. In each phase a great variety of problems can occur. Obviously, the kinds of problems and their seriousness differ from one country to another, depending on the availability and quality of data and on the wishes of policy makers with respect to classifications and other characteristics. Nevertheless, the sequence of tasks tends to follow a regular pattern.<sup>5</sup>

<sup>3</sup>Refer also to van Bochove and van Tuinen (1986), whose ideas about the structure of the next SNA, consisting of a general purpose core supplemented by special modules, are in essence complementary to the proposal in this paper: to construct, at regular intervals, a System of Socio-economic Accounts, in which a SAM serves as the core.

<sup>4</sup>See the Indonesian SSAs for an example (Downey, 1984; BPS, 1982; BPS/ISS, 1986; and Keuning and de Ruijter, forthcoming).

<sup>5</sup>In several cases the construction process has been documented to some extent; see e.g. Pyatt and Row (1977), Eckaus *et al.* (1981), Downey, Keuning and staff of BPS (1982), Pyatt and Round (1984), Webster (1985) and Greenfield (1985). King's (1985) introduction to the concept of SAMs also includes a few remarks about this.

Before continuing, it should be noted that a SAM is meant to fit into the existing national statistical and planning infrastructure. That is to say that, first, a SAM is typically built on the basis of data which are already available. Thus, there is no need for costly and time-consuming new sample surveys, provided that some information about household incomes and expenditures and inter-industry demands has been gathered. Considering that these data are essential for economic policy, they definitely ought to be collected, if they are not yet available. An advantageous side-effect of the integration of various statistics into a comprehensive framework is the detection of data gaps and inconsistencies at the meso-level. This feeds back into a streamlining of coverage, definitions, survey methodologies and classifications, thereby improving the comparability of separate sources and the overall quality of statistics.<sup>6</sup> In a number of cases this side-effect has become increasingly important. SAMs have proved to be expedient tools for comparing inconsistent data sets. Quite often national accounts, I-O tables and budget surveys are not at all compatible, which hampers the design and evaluation of socio-economic policies. Evidently, the more detail that is included into a SAM, the more inconsistencies can manifest themselves. On the other hand, the time needed for constructing a SAM expands very rapidly relative to the total number of accounts.

Secondly, the social accounting framework is flexible enough to incorporate country-specific features and planning priorities, for international comparability is not the main issue. Even so, the conventions laid down in the SNA usually serve as a frame of reference. Thus, national priorities are primarily reflected in the classification of institutions, production factors, activities and the like.

Naturally, the uses to which the SAM will be put are also important. These can vary from tax incidence studies (mostly in industrialized countries) to income distribution monitoring and sectoral manpower planning (mostly in developing countries). SAMs may also serve to provide base year data needed for a (general equilibrium) government policy simulation model.

The compilation of a SAM is here divided into eight steps or phases (see Figure 1).<sup>7</sup> In practice, the distinctions between these steps are not very clear, and sometimes the results of an earlier stage are re-adjusted again in order to circumvent a snag later on. Possibilities to do so are of course enhanced by the use of computers. The rapid development of both hard- and software in the last decade has undoubtedly influenced both the size and accuracy of SAMs.

The stages are discussed below in more or less chronological order.

## 2. OVERALL DESIGN OF THE SYSTEM

A SAM must always contain detailed information about the incomes and outlays of institutions (household groups, companies and the government and relevant accounts for the rest of the world) and about the production structure

<sup>6</sup>One might consider e.g. improving household survey questionnaires by inserting a standard module with several questions which enable a clear socio-economic identification of households.

<sup>7</sup>This flow chart is not typical to SAM construction and serves mainly as a device for a time schedule. Besides, this paper is structured around it. The phases were originally designed by Roger Downey for the first Indonesian SAM and are worked out here by the authors.

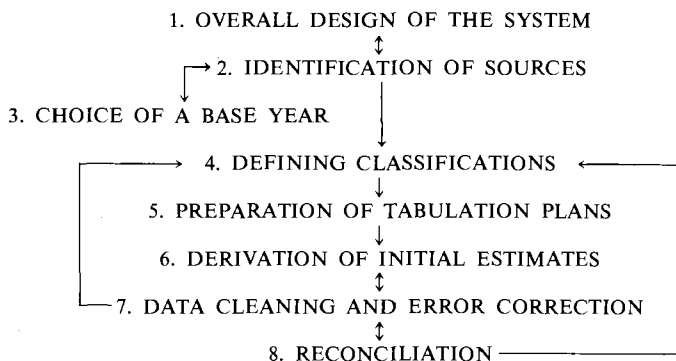


Figure 1. Flow Chart of SAM Construction

(e.g. in an Input-Output table). The rest of the design depends on national socio-economic structure, policy needs and availability of data and resources. Table 1 presents an example of a fairly extensive SAM.<sup>8</sup> The flows recorded in Table 1 are listed in more detail in Appendix A.

Some of the options for the design of a comprehensive framework are:

- a. *Inclusion of factor accounts.* In some cases, value added from business activities is not allocated first to all kinds of production factors, and subsequently to the owners, but directly to household groups and other institutions. However, it is preferable not to skip over this link, if only to permit the estimation of employment composition and the functional income distribution. Besides, multiple income sources of households are best revealed with the help of factor accounts. In general, more insight into demand and supply of production factors facilitates research on how capital and labour markets operate. Depreciation allowances may be treated separately and channelled directly to the companies' capital account (cf. Table 1).
- b. *Distinction between production activity and commodity accounts.* This enables correct treatment of joint production and by-products. In, for example, analysis of the impact of technical change on income distribution, specification of various production activities (technologies) producing the same type of commodity is required (see Khan and Thorbecke, 1986). In many developing countries, various commodities are made by means of a number of quite distinct technologies which coexist for a long time. A well-known example is the formal-informal dichotomy. The above distinction within a SAM is therefore essential for an assessment of the employment and income-generating role of the informal sector.<sup>9</sup>

<sup>8</sup>Trade and transport margins (TTM) are here included both in all commodity supplies (registered at purchasers' prices) and in trade and transport supply. This could be avoided by booking trade and transport margins to one (or more) separate row(s) where total margins appear with a minus sign in the column(s) for trade and transport activities (so that the sum of the additional row(s) equals zero).

<sup>9</sup>In addition, explicit treatment of the informal sector requires the distinction of own-account workers from employees, of unincorporated capital from corporate capital and of household enterprise from limited liability companies (cf. Keuning, 1985b).

TABLE 1  
A SCHEMATIC REPRESENTATION OF A FAIRLY EXTENSIVE SOCIAL ACCOUNTING MATRIX

|                        |               | Outlays |                              | Institutions (current)                             |                                     | Institutions (capital)          |   | Production activities (current)        | Production activities (capital)        | Commodities   |                               | Financial claims | Total                      |
|------------------------|---------------|---------|------------------------------|--|-------------------------------------|---------------------------------|---|--|--|---|-------------------------------|------------------|----------------------------|
|                        |               |         |                              | National   | Rest of World                       | National                        | Rest of World                             |  |  | Domestic  | Imported                      |                  |                            |
| Incomes                |               | Wants   | Factors of Production        |  |                                     | Indirect taxes                  |   |  |  |   |                               |                  |                            |
| Wants                  |               |         |                              | wants satisfaction households                      |                                     |                                 |   |  |  |   |                               |                  | demand for wants           |
| Factors of Production  |               |         |                              |  | factor incomes from abroad          |                                 |   | gross value added allocation           |  |   |                               |                  | gross factor incomes       |
| Institutions (current) | National      |         | allocation of factor incomes | inter institutional transfers                      | current transfers from abroad       | government indirect tax incomes |   |  |  | fictive TTM <sup>1</sup> and taxes on own consumption |                               |                  | net income distribution    |
|                        | Rest of World |         | factor incomes to abroad     | current transfers to abroad                        |                                     |                                 |   |  |  |   | imports (c.i.f.)              |                  | current payments to abroad |
| Indirect taxes         |               |         |                              | extra net indirect taxes on government consumption | extra net indirect taxes on exports |                                 | extra net indirect taxes on stock changes | non-commodity net indirect taxes, etc. | extra net indirect taxes on investment | net indirect taxes on domestic commodities            | net indirect taxes on imports |                  | net indirect taxes         |

|                                 |               |                              |                         |                               |                                     |                    |   |                                |                               |                           |   |                              |  |
|---------------------------------|---------------|------------------------------|-------------------------|-------------------------------|-------------------------------------|--------------------|---|--------------------------------|-------------------------------|---------------------------|---|------------------------------|--|
| Institutions (capital)          | National      |                              | depreciation allowances | savings                       |                                     |                    | existing non-financial asset transactions | existing asset sales to abroad |                               |                           |   | borrowing etc.               | finance of gross accumulation                            |
|                                 | Rest of World |                              |                         |                               | balance of payments current deficit |                    | existing asset purchases from abroad      |                                |                               |                           |   | lending to abroad            | capital payments to abroad                               |
| Production activities (current) |               |                              |                         |                               |                                     |                    |   |                                |                               | domestic commodity output |   |                              | output of domestic production                            |
| Production activities (capital) |               |                              |                         |                               |                                     |                    | fixed investment allocation               |                                |                               |                           |   |                              | capacity expansion                                       |
| Commodities                     | Domestic      | household consumption demand |                         | government consumption demand | exports                             |                    | net stock increase                        |                                | intermediate demand           | fixed investment demand   | domestic trade and transport                            | domestic trade and transport | demand for domestic commodities <sup>1</sup> (TTM twice) |
|                                 | Imported      | household consumption demand |                         | government consumption demand |                                     |                    | net stock increase                        |                                | intermediate demand           | fixed investment demand   |   |                              | demand for imports                                       |
| Financial claims                |               |                              |                         |                               |                                     |                    | lending etc.                              | borrowing from abroad          |                               |                           |   |                              | increase in assets                                       |
| Total                           |               | supply of wants              | gross factor payments   | net expenditure distribution  | current receipts from abroad        | net indirect taxes | gross accumulation                        | capital receipts from abroad   | inputs in domestic production | capacity expansion        | supply of domestic commodities <sup>1</sup> (TTM twice) | supply of imports            | increase in liabilities                                  |

<sup>1</sup>Trade and transport margins (TTM) are included both in all commodity supplies (registered at purchasers' prices) and in trade and transport supply.

In addition, the broad range of government functions becomes more clearly visible if total public expenditures are first assigned to expenditure programmes (general administration, education, irrigation etc.) and then to commodities (not shown in Table 1). This breakdown offers the opportunity to study income distribution effects of alternative budget allocations.

- c. *Separate accounts for domestically made and imported commodities.* These shed light on differences in the destination of similar goods of domestic and foreign manufacture. A next step is to study which institutions put the greatest burden, directly and indirectly, on the balance of payments.
- d. *Inclusion of so-called wants accounts.* Fulfilment of household needs (first, basic needs for food, shelter, clothing, education and medical services, and then supernumerary wants) appears in a special submatrix (either in the SAM or in a satellite table), and commodities that satisfy each need are also shown. Typically, a number of commodities can fulfil the same need (e.g. nutrition), but this bundle differs by socio-economic group (cf. food consumption patterns in rural and urban areas). Therefore, wants accounts provide a clearer picture of the (relative) well-being of households and enable a.o. a comparison with more "common" poverty indicators. If the SAM is applied to a model, grouping commodities in this way also facilitates the estimation of nested demand systems.
- e. *Inclusion of flow-of-funds accounts.* For a thorough understanding of economic dynamics it is crucial to know how savings are channelled through financial intermediaries and used for capital accumulation. A flow-of-funds block in a SAM can lift a tip of the veil here (see Table 1). Simultaneously it may lead to a better estimate of household savings which are notoriously difficult to assess. On the other hand, much data on monetary flows are required and such information is quite often not readily available. Gathering it will lengthen the time span needed to finish a SAM. Therefore, these accounts are frequently deleted.<sup>10</sup>

As to fixed capital accumulation, a SAM should show not only who invests and what kind of asset is added, but also in which production sector capacity is expanded. This implies that institutions' investment expenditures are channelled through the production activities in which the investment is made to the commodities which are demanded for this purpose. This is also shown in Table 1.

It would be even more ideal, but presently hardly feasible, to insert opening and closing wealth balances and revaluation accounts by institution (see Pyatt and Thorbecke, 1976, Table 4). Besides this, changes in stocks belonging to the national common good, like natural resources and environmental quality, ought to be recorded in a supplementary table which is part of the System of Socio-economic Accounts. To date, resource limitations and data problems have retarded progress in this direction.

<sup>10</sup>Exceptions are the SAM for Botswana (Greenfield, 1985) and for Ecuador (Vos, forthcoming).



- f. *Valuation of commodity sales*, either at purchasers' values, or at producers' values or at (approximate) basic values.<sup>11</sup> Some advocate that basic values be used, particularly if trade and transport margins and indirect tax rates differ significantly by category of purchaser (United Nations, 1973). This applies to economies with substantial own-account production for consumption, primitive physical infrastructure, or a system of value added taxes (with drawbacks on exports). However, even in these cases it is advisable also to distinguish these taxes and margins by the group of commodities to which they apply. On the other hand, the study of economic behaviour as a function of market prices requires that transactions are shown at purchasers' values (cf. Pyatt and Round, 1984, section 5.3).

Another difficulty with the basic value approach is the collection of primary data. Purchasers generally know only about the prices they paid, which is naturally the market price. As a consequence, commodity sales are often valued at purchasers' values. Indirect taxes and distribution margins are merely shown by commodity.

Nevertheless, it is easy to correct for differential duties by category of buyer and to record output of production activities at approximate basic values in the same table (see Table 1 and Appendix A).

- g. *Inclusion of subsidiary (non-monetary) accounts*. Little can be derived from a SAM *per se*. At the least, estimates of the size of each household group are needed for the computation of *per capita* incomes and expenditures. Likewise, a decomposition of wages into estimates of employment and wage rates is quite illuminating. More generally, it is useful to supplement the SAM with four sets of tables:
- quantities and prices underlying the value transactions in the SAM,
  - other (non-monetary) socio-economic indicators which are related to SAM values,
  - stocks underlying the flows in the SAM, and
  - some SAM-flows recorded in a slightly different way.

Computation of physical volumes and prices for commodity supply and demand is indispensable if household consumption is analyzed, if a SAM is to serve as a data base for a price-endogenous model or if changes in two subsequent SAMs are analyzed. An easy way out is to select a quantity unit such that the base year price equals one. It goes without saying that this solution impedes the presentation of recognizable quantities in later years, thereby unduly distracting those readers not involved in constructing the SAM. On the other hand, quantities of some "commodities" cannot be reduced to a meaningful common denominator (e.g. transport equipment which includes both bicycles and airplanes). In that case, the above-mentioned method has to be applied, and estimation of

<sup>11</sup>These values are defined in the SNA (United Nations, 1968) and also discussed in Greenfield and Fell (1979). Basic values exclude (a) trade and transport costs from producer (or importer) to consumer, and (b) all commodity taxes on outputs as well as inputs. Producers' values exclude only trade and transport margins; when those margins are included, transactions are recorded at purchasers' values (or market values).

a price index in later years is the best one can hope for. There are other multifarious commodities, like vegetables, which can still be expressed in one volume unit, as long as the price per kilogram (or meter etc.) of the principal constituents does not diverge too much. Caloric value is a suitable unit for measuring staple-food quantities. The value of by-products should be converted into main product volumes with the help of the main product price (cf. Keuning, 1986). If not all demand is expressed in volumes, it is worthwhile to trace at least the quantities of food consumed by households and of the nutrients taken in.

Grootaert (1982) sketches an expedient matrix in which flows of production factors (population, land, capital) are shown from the supply side (institutional owners) and from the demand side (production activities).

Other socio-economic indicators to be presented in satellite tables concern household composition, data related to family planning activities, housing situation, health condition, access to education and so on. Some of these indicators are related to household consumption expenditures and should be consistent with that information in the SAM.

The third category of subsidiary data comprises stocks: population size and educational background by socio-economic group, distribution of wealth (land, livestock, education, durables, real estate, production capacity and financial assets) and monetary indicators (money supply, outstanding credits and time deposits). Since wealth is a crucial determinant of income, recording changes in the distribution of assets enables a better explanation of shifts in income distribution.

In cases where flow data for the SAM are not available, they can sometimes be derived from stock estimates for two subsequent years. Moreover, part of the allocation of household incomes and consumption expenditures over socio-economic groups might be based on asset ownership. Possession of durables provides a more reliable indication of expenditures by household group than current purchases. Imputation of rents for owner-occupied houses is often done haphazardly, especially in (rural) areas where almost everybody owns his place of residence. Information about housing quality, size and facilities can then give clues about the allotment of imputed receipts and outlays for shelter (Downey, 1984). Keuning (1984) demonstrates that relying on survey respondents' statements regarding revenues from food crops may lead to underestimation, not only of total agricultural incomes but also of the degree of inequality between those incomes. Large farmers tend to underrecord their receipts to a much greater extent than small farmers, as became evident from computations employing statistics on land ownership, tenancy arrangements, cropping patterns and yields. Finally, asset possession can also be instrumental in assessing, by approximation, the distribution of household savings.

The last set of satellite tables refers to a different way of recording some of the transactions in a SAM. A familiar example concerns the allocation of part of public expenditures (for education, health, etc.) to

the beneficiaries. If these transfers in kind were to be shown in the SAM properly, that matrix would lose its function as a transparent overview of actual (monetary) transactions. Besides, these imputed "special purpose transfers" should be left out if the SAM is used in an analysis based on the assumption of fixed coefficients. On the other hand, the usefulness of public incidence studies prompts the inclusion of the required information in one or more tables appended to the SAM.

Another subsidiary table might contain a breakdown of (current) transfers by type (e.g. property income, direct taxes, social security, social assistance, other transfers), as recommended by the SNA.

If all these supplementary tables are made consistent with the SAM values, one can speak of a System of Socio-economic Accounts (SSA). It is worth mentioning that if this is done, many social indicators become an integrated part of the system, thereby enhancing their usefulness for policy-making and planning (see Bull (1978) on this point).

- h. *Regionalization*. A complete specification of transactions within and between various geographical areas within one common boundary amounts to the construction of a series of SAMs plus their interlinkages. Distinguishing regions within a SAM may enhance both its realism (homogeneity!) and its usefulness (study of inequality between regions). However, it will certainly mean a manifold increase of the workload as well. Particularly, interregional linkages are difficult to trace since statistical sources are usually absent (see also Pyatt and Round, 1985). An intermediate solution is to distinguish several regions when classifying the most important variables in a (nation-wide) SAM (cf. section 5 below).

### 3. IDENTIFICATION OF SOURCES

This phase interacts with the previous one and with the next two. In principle all available socio-economic statistics can (and in fact should) be used, as long as they meet two modest requirements: (1) the information should cover a year rather close to the SAM reference year; and (2) it must be possible to classify the raw data in accordance with the taxonomies applied in the SAM. It goes without saying that SAM builders can only make the best use of various sources if they have access to basic data (see de Ruijter (1985) for an example referring to Sri Lanka).

Because a SAM can also be seen as extension of an Input-Output (I-O) matrix, such a table usually serves as a fruitful starting-point. If a recent I-O table is not available, it has to be constructed or updated to become part of the SAM.<sup>12</sup> A limitation of most existing I-O tables is that production activities are not distinguished according to the type of technology used (to show whether e.g. both labour intensive and capital intensive technology is used in a given sector).

<sup>12</sup>A standard reference work on I-O tables is published by the United Nations (1973), while Skolka (ed.) (1983) gives a recent overview of national practices and special problems in the compilation of I-O tables.

However, compiling a new I-O table is quite time-consuming, so that when one is on hand, SAM builders usually accept it with its shortcomings.<sup>13</sup>

If an I-O matrix is available, the main tasks which remain are:

- a. Linking primary incomes and final demand (mapping factor incomes to household incomes, and mapping household incomes, after correction for transfers, to consumption expenditures).
- b. Disaggregating primary incomes (by factor type) and part of final demand, namely household consumption expenditures (by household group) and fixed capital formation (by sector in which the investment takes place and possibly by investing institution). In addition, the destination of imports has to be sorted out.
- c. Collecting supplementary information on savings, interinstitutional transfers (taxes, dividends, government subsidies and grants to private institutions, transfers between household groups and the like), current transactions with the rest of the world not shown by the balance of trade (factor services, interest payments, emigrant remittances) and, ideally, the flow of funds.

Commonly, the supplementary data can be obtained from a variety of sources. Minimally needed are:

- a. *National Accounts*, these being the natural source for a preliminary estimate of national aggregates. If the I-O table is not incorporated in the national accounts for the same year and estimates for the same variable vary, one is inclined to trust the former, since it was built up in more detail (assuming that both data sources have been compiled in an equally solid way). Evidently, the applicability of the national accounts is greatly enhanced if they give more details. Like the I-O table, national accounts serve as a useful benchmark, but they are not, in our opinion, sacrosanct.
- b. *Demographic data* (e.g. the number of households and the population in each socio-economic group, preferably supplemented by more detailed information on family composition). A special population survey or census may be available; otherwise, this information is derived from household budget surveys or some other multipurpose household survey.
- c. *Survey data on wages and entrepreneurial incomes*, arranged by household group and sector of activity. Hopefully, wages and employment can be cross-classified by type of labourer (e.g. skilled/unskilled, male/female, young/old, urban/rural) and branch of industry on the one hand, and by household group and type of labourer on the other. In that case, the SAM can distinguish factor accounts. A labour force survey may have been organized to collect this information. Most household budget surveys also enable a crude estimation of incomes. A population survey or census may yield insights into labour incomes, or at least employment by household group and by production sector, which can be combined with other

<sup>13</sup>In this paper it is assumed that an I-O table exists and that it is used in the SAM. Nevertheless, constructing a SAM and an I-O table simultaneously is preferable; the disaggregation and interlinkage of household demand and primary incomes may lead to improvements in the I-O table. Once an I-O table has been finished, alterations are more cumbersome.

data on wage rates by branch of industry and labour type. Moreover, general establishment surveys, industrial surveys, agricultural surveys and the like usually include questions about the incomes of employers and employees.

Statistical yearbooks, establishment surveys, production sector overviews, reports by government departments and other agencies on relevant industries, public enterprise accounts and so on, may all be consulted for an approximation of the distribution of sectoral profits between corporate (private, public, foreign) and unincorporated (household) enterprise.<sup>14</sup> Only the latter accrue directly to households.

These sources lead to sectoral value added totals which are typically not consistent with the I-O table and may therefore be used only to allocate the I-O aggregates, unless there are reasons to suspect that the I-O figures are wrong.<sup>15</sup> If employment and wage rates or, in general, group sizes and *per capita* incomes originate from different sources, they are combined before apportioning incomes to classes.

- d. A *household budget survey*, showing in particular consumption (purchased as well as own-produced) by commodity and household group. If quantities of food intake have also been recorded, estimation of nutrition conditions in each socio-economic class is certainly worth the extra effort. Budget survey questionnaires frequently contain a few additional questions in order to provide figures on (types of) income and savings. Unfortunately, such savings estimates tend to be rather unreliable.

Here again, the I-O table typically gives the more credible consumption values by commodity (though own production and waste might not have been treated correctly). These are then allocated to household groups in accordance with the distribution of the corresponding expenditures. For each group these expenditures are computed as *per capita* consumption, derived from the budget survey, times the group size, data on which may originate from another source.

A complication may arise when the commodity classifications in the budget survey questionnaire and in the I-O table do not coincide. The former source is typically dovetailed to categories of household wants, while the latter is more closely linked to the production system. Ideally, the commodity classification in the SAM distinguishes homogeneous categories of wants and the I-O table classification is converted to this. Anyhow, the SAM classification will have to represent an intersection of both existing taxonomies.

- e. *Government statistics*, which serve various purposes: 1. to find out who contributed to direct tax and other (central and local) government receipts (fees, fines, etc.), 2. to apportion government transfers (including interest payments on public debt) to various private incomes, 3. to unravel the incidence of public expenditures (education, health, others), if possible,

<sup>14</sup>Keuning (1985b, Appendix B) contains an overview of estimation procedures for the distribution of profits.

<sup>15</sup>Obviously a more integrated reconciliation procedure is applied when the I-O and SAM are constructed simultaneously.

and 4. to obtain a better insight into the destination of public investment in particular and into the influence of the state on the economy in general.

Public enterprise and parastatal organizations fulfil a specific function in the economy and their incomes and outlays should therefore be separated from the general government accounts.

Broad distinctions between categories of institutions paying into or receiving from the exchequer can normally be made with the help of government accounts, but for an insight into e.g. tax payments by household group, additional assumptions tend to be necessary (although the household budget survey may also contain information about taxes). Average income in each class roughly determines the liability to income taxation. These proportions can be adjusted for differential degrees of tax evasion, if any information is available on that subject (e.g. from a micro study). Next, these amounts multiplied by the group size can be scaled until their sum agrees with what actually ended up in the public purse. The distribution of property taxes could follow some indicator of asset possession (again corrected for tax evasion). For other legal charges an equal contribution *per capita* can be assumed if no other indicators can be found. Local levies can often be considered as indirect taxes. Even then it should be verified whether they were indeed included in the I-O matrix.

In dealing with government transfers, subsidies to business deserve special attention. If they are meant to lower the output price, they should be moved to the I-O table's row for negative indirect taxes.

Scholarships are allocated on the basis of conditions on which they are given. For instance, if parental incomes are taken into account, the distribution of students in groups with an average income below the upperbound serves as the yardstick. In some cases, all government outlays on education (possibly even including depreciation on capital stock) are imputed to household groups on the basis of the number of students (by school type) in each socio-economic category. If the backgrounds of people treated in public hospitals are known, the associated health expenditures can be assigned. Public transport expenditures, if measured by a budget survey, provide a clue to the distribution of a possible government (investment) subsidy in this area (as far as this has not been included in the I-O table).

A tricky issue arises in handling social security benefits. According to the United Nations' guidelines on income distribution statistics, contributed premiums should not be subtracted from salaries and other primary incomes, but instead considered as part of salaries before being transferred from employees to another (government) institution taking charge of the money. The benefits are then treated as an interinstitutional transfer from this fund to the unfit, the unemployed, the pensioners, etc. If an employer or the government pays social security benefits from its own purse, these should first be imputed as implicit wages and then booked as a transfer from employee households to the social security institution. Finally, the real benefits are then recorded as a transfer from

the latter to the receiving households (cf. Appendix A and United Nations (1968, 7.17)).<sup>16</sup> All this is less of a problem in many developing countries, where the social security system has not yet matured.

f. *Itemized balance-of-payments data*, as can be found in the national accounts, Central Bank statistics or the IMF yearbooks. After carefully checking which entries have already been included in the I-O table (e.g. trade in non-factor services!) and which method of recording the flows (timing!) has been used in each of the sources, the other rest-of-the-world transactions can be allotted to the accounts where they belong. This concerns:

1. factor payments like direct investment income (profit remittances) and border workers' incomes,
2. current transfers like other investment income (interest on public debt and private portfolio investment), property income (not included elsewhere), other goods, services and income (if not included in the I-O table) and possibly unrequited transfers, and
3. capital transfers (which appear separately only when a flow-of-funds account has been included).

g. *For a flow-of-funds block, financial data*, usually collected by the Central Bank, are indispensable. In some countries regular surveys of the financial sector are undertaken. Furthermore, in some Ministries of Planning or comparable agencies, at least some know-how concerning the sectors in which capacity is expanded may exist (only the commodity composition of fixed investment demand is recorded in the I-O table).

If no nation-wide source for certain information is available, even individual company accounts and micro-studies can be useful. Combined with some common-sense notions about the representativeness of the results, they indicate at least the order of magnitude of the variable concerned (e.g. land rents, interhousehold transfers, emigrants' remittances, dividends etc.).

#### 4. CHOICE OF A REFERENCE YEAR

After identifying data sources, and with bearing in mind recent fluctuations in economic conditions, a reference year for the SAM should be chosen. The chosen year cannot be too recent, for processing of surveys takes a while. On the other hand, the more recent a SAM is, the more relevant it will be. Nevertheless, a certain (or even large) degree of pragmatism cannot be avoided since the designated year should be covered in one or more major data sets (e.g. an I-O table or household budget survey). As a rule of thumb, less than ten years and ideally less than five years should lapse between the vintage of a SAM and the date of its completion. Commonly, not all main sources relate to this reference period, which means that commodity flows must be corrected by means of price and quantity indices, money transfers are scaled with the help of inflation rates, population estimates are adjusted etc.<sup>17</sup>

<sup>16</sup>These and other issues relating to accounts for households are reviewed in Ruggles and Ruggles (1986).

<sup>17</sup>Eckaus et al. (1981) describe the updating of an I-O table by using a modified RAS-technique and price and quantity indices.

## 5. DEFINING CLASSIFICATIONS

This phase is vital for the uses which can be made of the SAM. Conclusions regarding the degree of inequality and poverty depend very much on how a population has been subdivided, as within-group variations are seldomly registered.<sup>18</sup> Besides, policy designed for certain target groups can only be monitored when the groups are separated out in the statistics. Furthermore, a SAM can contribute at an earlier stage to debates on whether such special treatment of certain groups will be justifiable and effective.<sup>19</sup>

The level of aggregation deserves general consideration. It is easy enough to show less detail than is available in a finished SAM even on the back of an envelope, while a further subdivision at that stage implies repeating a great deal of the work. Since analysts and policy-makers alike always want more details than have been included, it is better to start at a level which is as detailed as data reliability, confidentiality issues, sample sizes and resources for building the SAM permit (here the use of computers is of course very helpful). A further advantage is that the cause of possible errors and inconsistencies is easier to trace. Moreover, if a second SAM, in constant prices, is ever calibrated, better deflators are available at a more disaggregated level. On the other hand, a casual reader of a SAM publication should not be drowned in an enormous amount of numbers, and therefore several matrices, at different levels of aggregation, should be available.

The degree of homogeneity also plays an obvious role in the design of classifications. For instance, all small farmers can be lumped together in one category if their living conditions and socio-economic behaviour are similar throughout the country, while it may be necessary to split a less numerous group such as the urban poor into several segments because their income sources and expenditure patterns differ significantly.

Finally, for those parts of the SAM which result from the combination of two or more sources, the classifications cannot be more extensive than the intersection of taxonomies derivable from each of the sources. Here we refer to the basic data and not so much to the publications which typically tabulate survey results in a rather aggregated way.

When two sources use different nomenclatures for the same phenomenon, a so-called classification conversion has to be drafted. This means combining subgroups of each of the classifications in such a way that a number of completely overlapping classes results. Information contained in both sources can then be compared with respect to the new groups which make up the taxonomy to be used in the SAM. A common case is the linkage of an I-O table to a consumers' expenditure survey on the one hand and to a labour income survey on the other.

<sup>18</sup>An improvement in this respect would be to state not only average figures by class but also the variances. Such a statistic is also quite functional in the reconciliation process (see e.g. Stone (1981, ch. 8)). Unfortunately, the amount of additional calculations involved appears prohibitive if they must be done by hand.

<sup>19</sup>In a few SAMs, such target groups, which otherwise are small, have been distinguished: e.g. in Swaziland, where a specific type of land and the households living on it were shown (Pyatt and Round, 1977); and in Mexico, where several public enterprises were shown (Pleskovic and Trevino, 1985).



A classification conversion is certainly indispensable if the I–O table does not have an industry-commodity format. But even if it does, the I–O columns may not refer to exactly the same industries as the labour income survey and the I–O rows may refer to commodities which are defined differently in the household expenditure survey. The overlap requirement should not be applied too rigorously to minor products, though, in order to prevent the number of new groups from shrinking too much. The preparation of a conversion implies scrutinization of questionnaires, survey manuals and work files to discover how commodities, industries and so on have been classified in each data source.

If intended for use in a SAM, the classification of every account should meet certain requirements, viz. it should:

- a. distinguish groups which are relatively homogeneous with regard to the main characteristics (decisions to be taken) in the account under consideration,
- b. correctly reproduce the variety within society,
- c. be composed of groups which are recognizable for policy purposes and useful for socio-economic analyses, i.e. special target groups should be identified (for some time to come),
- d. be based on comparatively stable characteristics which can be measured easily, reliably and by means of only a few questions, and
- e. be derivable from (a combination of) existing data sources.

It is remarkable that a household classification based on income or expenditure brackets does not satisfy any of these desiderata—except perhaps the second. Consider for instance a heterogeneous group like the poorest segment (say, 10 percent) of society. This segment tends to include households headed by landless farmers and casual agricultural labourers as well as by urban informal sector workers and rural unemployed females. Policies aiming at improving conditions for these groups have to and will be very different, the number of households in this segment depends on many incidental factors and many useful surveys do not even ask exhaustive questions about incomes and expenditures. Instead, as the above list suggests, qualitative criteria, such as place of residence and profession and the employment status of the main earner, should be used. This can be combined with data on possession of unsaleable or infrequently sold productive assets like agricultural land, education or even an (inherited) large connection. To summarize, data on income sources, and not on income size, are appropriate to capture causes of continuing inequality between households.

Each account of Table 1 can be disaggregated.<sup>20</sup> A couple of broad, standard groups are distinguished in almost every SAM, but subdivisions are much less uniform. A common approach is to start with selecting the most appropriate classification criteria and then delineate segments which are not too small and relatively homogeneous with respect to the adopted criterion. The main sets of classifications and several criteria for subdivisions are listed in Appendix B. A few relevant criteria should be applied simultaneously, but if each criterion

<sup>20</sup>Classification of institutions and production activities in the capital account need not be exactly the same as in the current account. There may be no good reason to distinguish banks from other business in the current account (except when one wants to trace interest flows very carefully), but in the capital account financial institutions obviously have a distinct role to play.

determines several constituencies, their total number easily exceeds what the sample size and resources can manage. In this crucial phase policy choices are therefore indispensable.

A possible strategy is to select those characteristics which have the greatest policy relevance and which cause substantial and structural segmentation of unit incomes and outlays. Sometimes statistical methods like analysis of variance or the Theil index are used to detect which characteristics are able to explain (in a statistical sense) most of the observed overall inequality. This can be expedient if it is otherwise difficult to choose, but one must realize that these methods do not provide an economic interpretation, and that it is also a bit hazardous to base the choice of classification criteria upon data from a single year.

Furthermore, it is often not necessary to maintain complete cross-classifications when several criteria are being combined. Distinguishing urban households on the basis of size of land owned does not make much sense, for instance. On the other hand, this is typically an important yardstick for stratification of rural farmers. In practice, taxonomies are built using a tree structure. If households are first broken down according to location, rural households may be subdivided into agricultural and non-agricultural, based on their main source of income, while in urban areas there are few farmers, so that main employment status or main occupation of the chief breadwinner is considered a more significant cause of social disparities. Next, rural agricultural households are decomposed into e.g. food crop farmers, grouped by size of land owned (landless and small, medium and large owners), and other agricultural households, and so forth.

We shall not list the most appropriate taxonomies for each account, for, as has been stressed before, the peculiar characteristics of the country (region) in question should always be incorporated. In practice, operationalization of the above-mentioned criteria and the resulting classifications have been as diverse as the economies to which they refer (see the list of references for examples). Ideally, the construction of a SAM stimulates a discussion about standard national classifications. Some of these may already be available, e.g. when established international conventions are followed. The classification of households in particular has not yet crystallized, though (cf. United Nations, 1977; Downey, 1984; Stone, 1985). Once the taxonomies are fixed, a standard module containing the (few) questions required to apply them can be inserted into all relevant statistical surveys. This will greatly facilitate future work on SAM construction and intertemporal comparisons.

However, without taking back what has been said above, a warning is due here. Standard classifications inevitably lead to stereotypes. In order to prevent stereotypes from becoming stigmas, regular evaluation of whether the classifications continue to be valid is called for.

The design of classifications usually proceeds in several steps. Once again it is important to realize that a SAM is made by combining various existing data sources. The steps are:

- I. defining desirable classifications,
- II. taking stock of relevant questionnaires and data processing procedures,
- III. confronting desirable and possible classifications,

- IV. designing schemes for conversions between classifications from different sources (thereby possibly modifying the classifications),
- V. listing provisional classifications,
- VI. filling the cells of the SAM, evaluation of results, preliminary reconciliation of sub-matrices, and
- VII. deciding on final classifications.

## 6. PREPARATION OF TABULATION PLANS

After the first four phases one knows what kind of SAM submatrices (and subsidiary tables) have to be filled and what their row and column entries are. A proper way to proceed in this phase is to divide the SAM into blocks (e.g. allocation of sectoral value added to production factors, allocation of factor incomes to institutions, interinstitutional transfers, household consumption demand, allocation of imports, interindustry transactions, capital accounts, government accounts, rest-of-the-world accounts and flow-of-funds accounts). Subsequently, the sources for filling the cells in each block are identified.

Many submatrices can be fully estimated only with the help of unpublished information. Mimeographed sources need to be consulted and survey results retabulated. After studying questionnaires and survey manuals a tabulation plan is drafted (see e.g. Keuning, 1985a). This entails the following steps:

- I. *Making a flow chart of possible answers to questions which determine the classifications.* This procedure is meant to ensure that in a later stage, each record is assigned to exactly one group. The treatment of non-response deserves special attention because it is easily overlooked. It is also useful to know how the answers have been edited (how extreme values were corrected or response gaps removed) and coded. In any case it seems advisable to have a "safety net", in the form of a separate category labelled "unclassified", for each taxonomy, if only to get a correct idea about the total number of elements.
- II. *Programming the above flow chart.* This is necessary only if the survey data are stored in a computer; if not one might consider storing them first.
- III. *Preparing a list of cross-tabulations.* This list can be more extensive than is strictly necessary for the SAM. Additional tables often serve as a useful tool in detecting causes of errors or as a guideline for correcting unreliable parts (besides giving valuable information in themselves). For example, a table showing the distribution of durable goods possession serves to correct the distribution of expenditures on durables.
- IV. *Preparing the framework for the tables,* in addition to an indication, if still necessary, of questions and answers which cause an individual record to appear in a certain row/column/cell of each table concerned. Especially when a computer is used, it is advisable to tabulate rather excessively, e.g. by including population estimates by household group in each table even when annual expenditure totals and *per capita* outlays have also been printed. Mistakes are more easily traced if a few cross-checks on the data are on hand.
- V. *Programming the tabulation plan* (see step II above).

## 7. DERIVATION OF INITIAL ESTIMATES FROM SEPARATE DATA SETS

In this stage and the next one, the emphasis lies on filling the separate blocks, without integrating them into a SAM as yet. The computer retabulates the raw data from the surveys, data already available (e.g. the I-O table) are scrutinized (e.g. the treatment of the interest margin of banks), and data which are lacking are estimated provisionally, for instance on the basis of micro-studies or even ratios in other comparable countries. In some cases sufficient data to fill a matrix are not available and applying ratios, RAS methods, etc. does not make much sense. This may be the case with interhousehold transfers. Obtaining estimates of total transfers paid and received by each socio-economic group (without knowing who paid whom) is a tedious task. If that is all one can get, the problem may be solved by extending the SAM framework by inserting a dummy account (labelled interhousehold transfers) to which these transfers are paid (row wise) and from which they are distributed (column wise). Unlike in a transfer matrix, one then does not have to show both the remitting and the receiving institution by transaction.

Lack of data is only one of the complicating factors; another is an abundance of contradictory figures for some cells (e.g. several items on the balance of payments may be calculated by more than one department without any attempt at consistency). Differences in definitions, timing and sampling methods will also manifest themselves.

Splitting money flows into volumes and prices is frequently frustrated by the enormous versatility of commodity unit values, due to differences in quality, location and timing of the transaction and characteristics of the parties involved. Ideally, each commodity should fetch only one price, but in practice this would lead to an unmanageable number of commodities. In some cases the total quantity of supply (domestic production plus imports) can be traced and the average price is then simply computed as value divided by volume.

Unfortunately, it is also conceivable that non-sampling errors are found in a survey which has already been processed. Paying attention to distributional issues also implies a rigorous test on the basic data, for the obscuring law of large numbers no longer applies.

## 8. DATA CLEANING, ERROR CORRECTION AND RETABULATIONS

As soon as initial estimates are available, attention should, first and foremost, be directed towards the reliability of the values of the classifying variables. A first test is on the number of elements in each group: does it seem reasonable? How should possible incomplete coverage of the survey be corrected? Are there not too many elements that fall into the "unclassified" category? Evidently, the likelihood of inaccuracies is smaller if the classification is based on non-numerical criteria (e.g. main occupation is typically reported in a more reliable way than total income). At this stage one should check the credibility of extreme values, which is also a useful test if a tabulation shows other peculiarities. Suspicious outcomes may lead to a second tabulation in which constraints are tightened or inserted.

It is normally easier to detect mistakes than to correct them. Revisions are relatively straightforward if there is a systematic inconsistency, e.g. if the valuation of own production and its consumption diverges, or rents of owner-occupied housing are taken to be zero if the enumerator has not been able to think of a reasonable imputation. Programming errors are also human. Another possibility is to compare preliminary SAM values and related (non-monetary) information at this stage (see the discussion in section 2g).

There may be a few other inaccuracies which immediately suggest their own improvements, but in survey tabulations often the best solution one can devise is to delete the questionable records. That will evidently also alter the inflation factors from sample to population figures. Hopefully, most of the outliers will have disappeared after this stage. On the other hand, it is commonly known that homeless persons as well as the very rich are normally not covered by household surveys. Although these omissions can be partially remedied with the help of other pieces of information (e.g. corporate dividend payments, interest on time deposits), the general well-being of these population extremes can only be guessed. As a consequence, a SAM (or any other income distribution statistic, for that matter) tends to present a conservative image of inequality and poverty.

Next, the sources which refer to the same block in the SAM are confronted. It may be that in order to arrive at sound estimates, classes should now be combined or even rearranged and reconciliation should take place at a somewhat less disaggregated level. The removal of discrepancies requires that the strong and weak points in each data base are fairly well known. Roughly speaking, nation-wide surveys produce more reliable data than micro-studies, integrated sources (like I-O tables) are more credible than unrelated surveys (like a household budget survey), and more detailed integrated information (like, once again, I-O tables) is more convincing than integrated aggregates (like national accounts).<sup>21</sup>

There is a trade-off between (a) complementing each block in an internally consistent way before integrating all of them into one SAM and (b) leaving all initial estimates unchanged before reconciling them simultaneously in one exercise. The latter choice is preferable, but takes more time. It may also require the use of a computer package, e.g. following Byron's (1978) method. On blocks for which a sequence of observations is available, van der Ploeg (1982) should be consulted. Finally, if hardly more than row and column totals can be made up, some kind of modified RAS-method (Bacharach (1970), Allen and Gossling (eds.) (1975)) constitutes the last resort.

## 9. FINAL RECONCILIATION

Once each building block is filled, the overall reconciliation of the SAM starts. First of all, a uniform estimate of the number of households and of the population in each group is required. This in turn may lead to adjustment of household incomes and expenditures while maintaining relative *per capita* or per household figures for the moment. Next, some blocks are considered to be more

<sup>21</sup>See also the discussion in section 3 of usable sources.

reliable than others. This of course depends on the sources from which they have been derived. For instance, procedures to compute household consumption, based on an I-O table and a budget survey, may lead to plausible results while savings estimates are still weak. Quite often various sources for labour incomes exist, but the allocation of corporate and noncorporate profits to households poses a much bigger problem. Interhousehold transfers are typically ill-documented, although most of the time substantial amounts are involved. Particularly in the informal sector, many entrepreneurs do not own their means of production. In so far as renting out capital goods is not included in the I-O table as a business service, the rent payments still need to be settled as interhousehold transfers. Moreover, remittances to non-inhabitant old-aged parents and disabled family members and to children attending college should be included here. Of course, the accounting constraints may also be called upon: the fact that total receipts must equal total outlays on each account implies that per row and column one item can be computed residually. If this does not solve the problem, simple rules of thumb and common sense sometimes provide a way out. Illustrative in this regard is using the spread of durable goods possession as an indicator for the distribution of savings among households.

Once one has initial estimates for all the cells, the reconciliation can be done by hand, or, preferably, with the help of a little mathematics. Worth mentioning is a linear programming method, as it minimizes the largest adjustment needed to remove the discrepancies, subject to a number of constraints on a reasonable range for some parameter values and the relation of some variables to each other (Pyatt and Round, 1984). Stone (1981, ch. 8) and Byron (1978) discuss a solution in which one considers certain cells to be more credible than others (because of the type of data source or economic reasoning), and this relative reliability is then inversely related to the maximum adjustment in the reconciliation algorithm. Teekens and Louter (1985) achieve consistency for the Ecuadorian SAM by means of a two-step approach in which a quadratic loss function is minimized two times subject to the consistency requirements. For purposes of clarity and analysis it is advisable to aim at complete consistency, i.e. by removing all errors and omissions (and thereby clearly documenting how this is done!). Evidently, consistency does not guarantee accuracy, but the SAM builders probably know better than the SAM users how to allocate an omission. If possible, in some parts point estimates could be supplemented by range estimates.

There is still some debate on whether it is methodologically more sound to start from data at a disaggregated level and then confront them with previously computed aggregates, or to consider the totals (e.g. from the national accounts) as sacrosanct and break these down. Since national accounts have to become available soon after the end of the year to which they refer, they are typically based on less information than has been incorporated in a SAM (this may even apply to subsequent annual revisions of national accounts figures). So we would be inclined to choose the former option. In that case SAMs may be used to revise the national accounts aggregates instead of the other way round. Clearly, the degree of inconsistency which has arisen from the original estimation provides an important feedback to the national statistical agency. In this sense, not only the construction of the SAM itself, but also the interaction between improvement

of basic statistics and compiling a SAM (say every five years), is an iterative process.

## 10. CONCLUDING REMARKS

In addition to a solid methodology, the organization of the work is a vital factor in the process of constructing a SAM. Not only because a SAM is made by combining data from different sources (e.g. the Central Bank, the Central Statistical Office and various Ministries), but also because the institutionalization of the compilation of SAMs should constitute an important objective. These notions have consequences both for persons involved in constructing a SAM and for the location of the work. Furthermore, a SAM is meant to be a tool for designing socio-economic policies and planning at various levels. This implies that it should not be compiled by statisticians alone. Sufficient input from planners and policy-makers is required to ensure that the SAM caters for their needs. Economists are needed to evaluate the (intermediate) results. Evidently, not all people may be involved to the same extent in all phases (for instance, defining classifications typically requires input from many sides, but preparation of a tabulation plan can best be done by just the statisticians).

Above we have tried to initiate a discussion which may lead to more standardized guidelines for SAM construction. Although a detailed blueprint can never be developed, in view of the importance of including country-specific features, the work sequence tends to follow a roughly similar pattern for all SAMs. Besides, many of the snags hit along the way are also standard (although the solutions may be less uniform). Evidently, there is still scope for improvement of the overview presented above. In our view, further development is called for, considering the usefulness of the SAM-framework on the one hand and the large investment of time and resources in building a SAM on the other.

There is one more reason for a wider publication of SAM construction procedures. Until now, SAMs for developing countries have almost always been built by teams of experts from developed countries, with the help of local statisticians. Obviously, the SAM methodology will only be firmly rooted in countries concerned when SAM construction is institutionalized within a national agency, preferably the national statistical office.<sup>22</sup> That will require more transfer of know-how than has been achieved in the past: hopefully, this paper provides a modest first step in that direction.

### APPENDIX A. FLOWS RECORDED IN A FAIRLY EXTENSIVE SAM (SEE TABLE 1)

Incomes of wants account:

from national institutions (households) current account:

households' total consumption expenditures, allocated to the type of want which they fulfil.

<sup>22</sup>The first Indonesian SAM, referring to 1975, was mainly built by foreign consultants, although the final report (BPS, 1982) was written by staff of the Indonesian Central Bureau of Statistics (BPS). In the second Indonesian SAM (BPS/ISS, 1986), the input of foreign consultants was rather limited. At the moment, BPS staff in Jakarta is constructing the third Indonesian SAM, referring to 1985.

Incomes of factors of production account:

from rest of world current account:

factor income from abroad, such as income of national border workers;

from production activities current account:

gross sectoral value added (at factor costs), allocated to the contributing factors of production.

Incomes of national institutions current account:

from factors of production account:

net national income by factor of production, distributed among the entitled institutions;

from national institutions current account:

interinstitutional current transfers, like remittances between household groups (e.g. student allowances, migrant remittances, informal interest payments), transfers from companies to households (e.g. dividends, interest on deposited household savings, pensions and other social security benefits paid from a specially administered fund, other insurance claims), transfers from government to households (e.g. social assistance grants, emergency aid), transfers from households to companies (e.g. interest on mortgage loans, insurance premiums net of operating costs of insurance companies (treated as household consumption of insurance services), pension and other social security premiums paid to a specially administered fund), transfers between companies (e.g. interest on deposits and on credits, insurance claims and net premiums), transfers from government to companies (e.g. interest on domestic public debt), transfers from households to government (e.g. direct taxes on personal incomes and wealth, fees, fines and penalties), transfers from companies to government (e.g. corporation taxes and distributed net profits of public enterprise), and transfers between public authorities (e.g. from central to local government);

from rest of world current account:

current transfers from abroad, like remittances of emigrant workers, salaries of local employees of foreign embassies, interest receipts on portfolio investments abroad, and government income from visas issued abroad;

from indirect taxes account:

total indirect taxes minus subsidies, received by the government;

from domestic commodities account:

fictive trade and transport margins and net indirect taxes on own-account consumption of production, imputed to ensure that all household consumption of production of a commodity is valued at the same (purchasers') price, and subsequently allocated to consumers of own-account products (these fictive mark-ups are obviously not collected by traders, transporters and the government respectively).

Incomes of rest of world current account:

from factors of production account:

factor income to abroad, such as profit remittances of foreign-owned domestic companies (preferably not including their retained earnings which should be booked as foreign-owned companies' savings);



from national institutions' current account:

current transfers to abroad, like interest on foreign debt, interest payments on foreign portfolio investments, royalties for use of patents, copy-rights and trade-marks, remittances of immigrant workers, salaries of local employees of embassies abroad and levies for foreign visas issued to nationals;

from imported commodities account:

imports of goods and non-factor services (including expenditures of domestic tourists abroad), valued at prices which cover costs of production, insurance and freight.

Incomes of indirect taxes account:

from national institutions (government) current account:

special net indirect taxes on government consumption, e.g. special rebates (bearing a negative sign):

from rest of world current account:

special net indirect taxes on exports, e.g. export taxes or duty drawbacks (the latter bearing a negative sign);

from national institutions capital account:

special net indirect taxes on stock changes, e.g. a correction if commodities are (partly) not taxed when they are added to stocks;

from production activities current account:

non-commodity net indirect taxes, like fees, licenses and penalties, which are not proportional to the commodity output and are paid by companies prior to compensating the production factors;

from production activities capital account:

special net indirect taxes on fixed investment by investing production activity, e.g. special rebates (bearing a negative sign);

from domestic commodities account:

indirect taxes minus subsidies on domestic commodities, recorded as if a uniform rate applied to each commodity sold, independent of its use (cf. the other types of (negative) incomes in this account and the incomes of households from the domestic commodities account);

from imported commodities account:

domestic indirect taxes and import duties minus subsidies on imports, recorded as if a uniform rate applied to each commodity sold, independent of its use.

Incomes of national institutions capital account:

from factors of production:

allowances for the depreciation of capital goods, being part of company retained earnings (an alternative would be to assign these provisions directly to the production activities capital account, thereby separating the allocation of net investments and replacement investments);

from national institutions current account:

household savings, company retained earnings and government budget surplus on current account;

from national institutions capital account:

transactions in existing non-financial assets, like land and other intangible assets (net sales of second-hand capital goods could be included here as well);

from rest of world capital account:

sales of existing non-financial assets, like land and other intangible assets, to abroad;

from financial claims account:

(net) incurrence of liabilities, like trade credit, loans, bonds and equity securities.

Incomes of rest of world capital account:

from rest of world current account:

the deficit on the current account of the national balance of payments (in case of a surplus this amount appears with a negative sign);

from national institutions capital account:

purchases of existing non-financial assets, like land and other intangible assets, from abroad;

from financial claims account:

lending to abroad, specified by type of financial claim.

Incomes of production activities current account:

from domestic commodities account;

make-matrix, containing domestic commodity output, valued at approximate basic values, whereby more than one production activity (technology level) may produce the same commodity and a production activity (e.g. mining) may produce more than one commodity (e.g. minerals and electricity, for own use and for sale).

Incomes of production activities capital account:

from national institutions capital account:

gross fixed capital formation, specified by the institution which invests and by the production activity in which capacity expansion takes place.

Incomes of domestic commodities account (at purchasers' values):

from wants accounts:

household consumption expenditures, showing which domestically produced commodities fulfil each type of want;

from national institutions (government) current account:

domestically produced government consumption expenditures (possibly shown as being channelled through a separate account for various government expenditure programs);

from rest of world current account:

exports of goods and non-factor services (including expenditures of foreign tourists), valued at f.o.b. prices minus special net indirect taxes on exports;

from national institutions capital account:

net increase in domestically produced commodity stocks, owned by each institution;

from production activities current account:  
 demand for domestically produced intermediate inputs by each production activity;

from production activities capital account:  
 fixed investment, specified by type of domestically produced capital good and by production activity in which this capital good will be used;

from domestic commodities account:  
 services concerning domestic trade and transport of domestically produced commodities;

from imported commodities account:  
 services concerning domestic trade and transport of imported commodities.

Incomes of imported commodities account (at purchasers' values):

from wants account:  
 household consumption expenditures, showing which imported commodities fulfil each type of want;

from national institutions (government) current account:  
 imported government consumption expenditures (possibly shown as being channelled through a separate account for various government expenditure programs);

from national institutions capital account:  
 net increase in imported stocks, owned by each institution;

from production sectors current account:  
 demand for imported intermediate inputs by each production activity;

from production sectors capital account:  
 fixed investment, specified by type of imported capital good and by production activity in which this capital good will be used.

Incomes of financial claims account:

from national institutions capital account:  
 (net) acquisition of financial assets, like currency, bank deposits, bonds and equity securities;

from rest of world capital account:  
 borrowing from abroad, specified by type of financial claim.

## APPENDIX B. MAIN SETS OF CLASSIFICATIONS

Standard classes in each account of a fairly extensive SAM (cf. Table 1) are listed below. By major group various criteria for further subdivision are given in more or less descending order of importance; these criteria are separated by a semi-colon, but when examples of subgroups are given instead of a criterion, these are separated by a slash:

### I. *Wants*

- a. basic needs: food (or, more specifically, various kinds of nutrients)/shelter/clothing/education/medical services, and
- b. supernumerary wants: a dispensable way of fulfilling each of the basic needs/other goods/other services.

- II. *Factors of production*
  - a. labour: occupation (distinguishing skill, level and occupational category); branch of industry; employment status; location; education; sex; age; type and size of the firm (see III.b “companies” below); job importance (distinguishing primary/secondary jobs, full-time/part-time jobs, permanent/seasonal/casual jobs),
  - b. land and other natural resources: land type; size of the holding; location; type of resource, and
  - c. capital: net operating surplus/depreciation allowance; ownership (distinguishing national/foreign and private/public); legal status of the firm; type of capital good.
- III. *Institutions*
  - a. households: asset ownership; location; size and composition; characteristics of the head or main earner (distinguishing e.g. main employment status, main occupation, main branch of industry, educational attainment, age, sex, main language, race (tribal) kinship, religion and political affiliation); household/private non-profit institution,
  - b. companies: ownership (distinguishing national/foreign and private public); legal status; one-establishment unit/multi-establishment unit; financial/non-financial; key industry/other industries; pension fund/life insurance company/other social security fund/other business,
  - c. government: central/local; purpose, and
  - d. institutions in the rest of the world: within/outside the customs union, if applicable; other regions/foreign countries, in the case of a regional SAM.
- IV. *Indirect Taxes/Subsidies*
  - a. indirect tax: value added tax/general sales tax/excise tax/customs duty/differential exchange rate levy/special purpose levy/illegal fee, and
  - b. subsidy: type of subsidy.
- V. *Public Expenditure Programmes*
  - a. current: government administration/defence, law and security/education/health/family planning/social welfare/environment/business development/other, and
  - b. capital: in agriculture/transport infrastructure/other infrastructure/housing/education/health/drinking water/other.
- VI. *Commodities/Production Activities*
  - a. commodities: (two digit) ICGS (United Nations, 1974) adapted to local conditions; tradable/non-tradable; domestic/imported; way of acquisition (distinguishing purchased/own produced/wage in kind/gift); principal price determining factor; fulfilling similar needs; income elasticity; price elasticity; subject to similar tax rates; perishability, and
  - b. production activities: (two digit) ISIC (United Nations, 1959) adapted to local conditions; cost structure; demand structure; output flexibility; scale of operations; technology level; institutional organization (distinguishing formal/informal, national/foreign owned and private/public); location; (potential) contribution to GDP; (potential) contribution to

target group income; (potential) contribution to employment; (potential) contribution to exports; price setting mechanism (degree of monopoly power); degree of unionization; contribution to environmental damage.

## VII. *Financial claims*

- a. currency: domestic/foreign,
- b. non-money international reserves: gold/IMF position etc., and
- c. other financial assets: degree of liquidity; nature of the financial instrument (distinguishing e.g. claim/equity); issuing institution (see III "Institutions" above); type of capital market (distinguishing e.g. formal/informal); maturity; object of financing.

## REFERENCES

- Allen, R. I. G. and Gossling, W. F. (eds.) *Estimating and Projecting Input-Output Coefficients*, Input-Output Publishing Company, London, 1975.
- Bacharach, M., *Biproportional Matrices and Input-Output Change*, Cambridge University Press, Cambridge, 1970.
- Biro Pusat Statistik (BPS), *Social Accounting Matrix Indonesia 1975*, Vol. I and Vol. II, BPS, Jakarta, 1982.
- Biro Pusat Statistik (BPS) and Institute of Social Studies (ISS), *Social Accounting Matrix Indonesia 1980*, Vol. I and Vol. II, BPS, Jakarta, 1986.
- Bull, R., Social Accounts, Social Indicators and Development Planning, *Development and Change*, 9(4), 613-630, 1978.
- Byron, R. P., The Estimation of Large Social Account Matrices, *Journal of Royal Statistical Society, A*, 141(3), 359-367, 1978.
- Downey, R. A., *Indonesian Inequality: Integrated National Accounting of Who Gets What*, Ph.D. thesis, Cornell University, 1984.
- Downey, R. A., Keuning, S. J., and staff of the Central Bureau of Statistics, Jakarta, *An Indonesian SAM*, Consultant Report to Research Project 672-49, World Bank, 1982.
- Eckaus, R. S., McCarthy, F. D., and Mohie-Eldin, A., A Social Accounting Matrix for Egypt, 1976, *Journal of Development Economics*, 9, 183-203, 1981.
- Greenfield, C. C., A Social Accounting Matrix for Botswana, 1974-75, in Pyatt, G. and Round, J. I. (eds.), *Social Accounting Matrices: A Basis for Planning*, World Bank, Washington, 1985.
- Greenfield, C. C. and Fell, H. A., The Estimation of Price Effects in a Social Accounting Matrix, *The Review of Income and Wealth*, 25(1), 65-81, 1979.
- Grootaert, C., *The Labor Market and Social Accounting: A Framework of Data Presentation*, Living Standards Measurement Study Working Paper No. 17, World Bank, Washington, 1982.
- Hayden, C. and Round, J. I., Developments in Social Accounting Methods as Applied to the Analysis of Income Distribution and Employment Issues, *World Development*, 10(6), 451-465, 1982.
- International Labour Office, *International Standard Classification of Occupations*, ILO, Geneva, 1957.
- Keuning, S. J., Farm Size, Land Use and Profitability of Food Crops in Indonesia, *Bulletin of Indonesian Economic Studies*, 20(1), 58-83, 1984.
- , *Retabulating the Household Budget Survey 1980*, project Modelling the Indonesian SAM Working Paper Series No. 2, ISS, The Hague, 1985a.
- , Segmented Development and the Way Profits Go: The Case of Indonesia, *The Review of Income and Wealth*, 31(4), 375-397, 1985b.
- , *Integrated Estimates for Quantities, Prices and Nutritional Values: A Data Base for a Price-endogenous Model*, project Modelling the Indonesian SAM Working Paper Series No. 11, ISS, The Hague, 1986.
- Keuning, S. J. and de Ruijter, W., Estimating for the Distribution of Socio-economic Welfare in Indonesia, in: Alarcon, J. V., van Heemst, J., Keuning, S. J., de Ruijter, W., and Vos, R. (eds.), *The Social Accounting Framework for Development*, ISS, The Hague, forthcoming.
- Khan, H. A. and Thorbecke, E., *Macroeconomic Effects and Diffusion of Alternative Technologies within a Social Accounting Matrix Framework: the Case of Indonesia*, International Labour Office, Geneva, 1986.
- King, B. B., What is a SAM?, in Pyatt, G. and Round, J. I., (eds.), *Social Accounting Matrices: A Basis for Planning*, World Bank, Washington, 1985.

- Pleskovic, B. and Trevino, G., *The Use of a Social Accounting Matrix Framework for Public Sector Analysis: the Case Study of Mexico*, ICPE Monograph Series No. 17, International Center for Public Enterprises in Developing Countries, Ljubljana, 1985.
- Ploeg, F. van der, Reliability and the Adjustment of Sequences of Large Economic Accounting Matrices, *Journal of Royal Statistical Society, A*, 145(2), 169-194, 1982.
- Pyatt, G., Roe, A. R., and associates, *Social Accounting for Development Planning with special reference to Sri Lanka*, Cambridge University Press, London, 1977.
- Pyatt, G. and Round, J. I., Social Accounting Matrices for Development Planning. *The Review of Income and Wealth*, 23(4), 339-364, 1977.
- , *Improving the Macro-economic Data Base, A SAM for Malaysia 1970*, World Bank Staff Working Papers No. 646, World Bank, Washington, 1984.
- , Regional Accounts in a SAM Framework, in Pyatt, G. and Round, J. I. (eds.), *Social Accounting Matrices: A Basis for Planning*, World Bank, Washington, 1985.
- Pyatt, G. and Thorbecke, E., *Planning Techniques for a Better Future*, International Labour Office, Geneva, 1976.
- Ruggles, R. and Ruggles, N. D., The Integration of Macro and Micro Data for the Household Sector, *The Review of Income and Wealth*, 32(3), 245-276, 1986.
- Ruijter, W. A. de, *A Proposal for the Building of a Social Accounting Matrix for Sri Lanka*, The Hague, mimeographed, 1985.
- Ruijter, W. A. de, *Constructing Social Accounting Matrices: Choices and Consequences*, with Senanayake, R. and Sikipala, N., Institute for Policy Studies, Colombo, Sri Lanka, 1986.
- Skolka, J. V. (ed.), *Compilation of Input-Output Tables*, Springer Verlag, Berlin, 1983.
- Stone, R., *Aspects of Economic and Social Modelling*, Librairie Droz, Geneva, 1981.
- , The Disaggregation of the Household Sector in the National Accounts, in Pyatt, G. and Round, J. I. (eds.), *Social Accounting Matrices: A Basis for Planning*, World Bank, Washington, 1985.
- Teekens, R. and Louter, A. S., *Note on the Consistency of the Social Accounting Matrix*, ISS-PREALC Working Paper No. 18, ISS, The Hague, 1985.
- United Nations, *International Standard Industrial Classification of All Economic Activities*, UN, New York, 1959.
- , *A System of National Accounts*, Studies in Methods, Series F, No. 2, Rev. 3, UN, New York, 1968.
- , *Input-Output Tables and Analysis*, Studies in Methods, Series F, No. 14, Rev. 1, UN, New York, 1973.
- , *Draft International Standard Classification of all Goods and Services*, UN, New York, 1974.
- , *Towards a System of Social and Demographic Statistics*, Studies in Methods, Series F, No. 18, New York, UN, 1975.
- , *Provisional Guidelines on Statistics of the Distribution of Income, Consumption and Accumulation of Households*, Studies in Methods, Series M, No. 61, UN, New York, 1977.
- van Bochove, C. A. and van Tuinen, H. K., Flexibility in the Next SNA: The Case for an Institutional Core, *The Review of Income and Wealth*, 32(2), 127-154, 1986.
- Vos, R., Social Accounts and Capital Accumulation, in Alarcon, J. V., van Heemst, J., Keuning, S. J., de Ruijter, W., and Vos, R. (eds.), *The Social Accounting Framework for Development*, ISS, The Hague, forthcoming.
- Webster, S. J., A Social Accounting Matrix for Swaziland, 1971-72, in Pyatt, G. and Round, J. I., *Social Accounting Matrices: A Basis for Planning*, World Bank, Washington, 1985.