



## Compiling Social Accounting Matrices

Scott McDonald

### Abstract

*These guidelines assume that a National Accounts Matrices (NAM) has been created from national accounts data and then converted into a macro-SAM. Hence the guidelines concentrate on the transition from a macro-SAM to a detailed prior SAM; prior in the sense that mathematical techniques have not been applied to reconcile the SAM. The method advocated is that of successive disaggregation, first by adding details about the commodity and activity accounts, then the tax, factor, and institutional accounts. The process is agnostic with respect to the mathematical techniques used to reconcile the SAM although the method implicitly advocated is based on information theory and links to the entropy metric as a mathematical technique.*



## Table of Contents

|                                                                  |    |
|------------------------------------------------------------------|----|
| 1. Introduction.....                                             | 3  |
| Data Sources .....                                               | 6  |
| 2. A General Statement of Difficulties .....                     | 7  |
| Prices and the Information in a SAM .....                        | 10 |
| 3. Classification of Accounts .....                              | 12 |
| 4. Control Totals .....                                          | 16 |
| Use of Control Totals .....                                      | 17 |
| 5. Disaggregating Commodity and Activity Accounts.....           | 19 |
| Commodity and Activity Classifications.....                      | 19 |
| Why Supply and Use Tables? .....                                 | 20 |
| 5. Adding Supply and Use Table Data .....                        | 22 |
| 6. Disaggregating Commodity and Activity Tax Accounts.....       | 26 |
| Taxes on Commodities/Products.....                               | 27 |
| Taxes on Production .....                                        | 29 |
| Taxes on Income and Wealth.....                                  | 30 |
| 7. Disaggregating Factor Accounts .....                          | 31 |
| Mixed Income Accounts.....                                       | 31 |
| Gross Operating Surplus (GOS).....                               | 32 |
| Labour Accounts .....                                            | 33 |
| 8. Disaggregating Household Accounts .....                       | 36 |
| Plan A: Access to ‘Official’ Household Survey Data.....          | 36 |
| Plan B: No Access to ‘Official’ Household Survey Data .....      | 39 |
| 9. Disaggregating Other Institutional Accounts.....              | 41 |
| Non-Profit Institutions Serving Households (NPISH).....          | 41 |
| Enterprise Accounts.....                                         | 41 |
| Government Accounts .....                                        | 42 |
| Investment Account.....                                          | 42 |
| 10. Further Disaggregating Activity and Commodity Accounts ..... | 44 |
| Adding Paired Commodity and Activity Accounts .....              | 44 |
| Adding Unpaired Commodity and Activity Accounts .....            | 45 |
| References .....                                                 | 47 |

## **1. Introduction**

The guidelines outlined in this paper provide individuals, who are developing a detailed Social Accounting Matrix (SAM), with suggestions about progressing from a macro-SAM to a (full) SAM. The guidelines assume that the available national accounts data for a country have been compiled by following the conventions in the System of National Accounts (SNA) and that the macro-SAM is complete but may not be fully consistent and aligns with the national accounts. The implicit presumption when undertaking this approach is that a top-down method is being followed.

It is a truism when developing a SAM that “[I]f we put all these initial estimates together [for a SAM] we reach the familiar situation: an incomplete and inconsistent social accounting matrix” (Stone, 1977, p xxiv). This should alert compilers of SAMs to two important requirements: the SAM must be complete, i.e., all transactions must be included, and the SAM must be consistent, i.e., all transactions must be reconciled. A SAM can be consistent but incomplete: the data in the macro-SAM should provide control totals that help to satisfy the complete condition and avoid a distorted SAM. However, it is important to recognise that transactions between accounts and themselves, i.e., entries of the principal diagonal, may not be recorded in the macro-SAM. But when disaggregating accounts in the macro-SAM such transactions need to be recorded if the SAM is to be complete.

‘Aggregate’ national accounts data are often readily available, and therefore compiling a macro-SAM is overwhelmingly a case of rearranging the national accounts data into a matrix/single-entry format. The data for disaggregating the accounts is typically less easily acquired. In an ideal world the compiler will have access to the raw data from surveys and censuses, detailed information on the tax revenues realised by different instruments from different accounts and Supply and Use tables that are reconciled with the national accounts’ aggregates. Reality is typically less compliant.

There will be gaps in the data that need to be filled in if the SAM is to be complete. There are three fundamental methods for identifying gaps: first, the knowledge that some transactions in a SAM should always be null, i.e., entries that will always be zero, is important information; second, an understanding of the economic relationships in a country can identify cells in the SAM that should contain non-zero values, e.g., government

*Guidelines for Compiling a SAM*

social/welfare/unemployment schemes that transfer income to households are ubiquitous; and third, the row and column totals for the SAM should equate so that marked differences between incomes and expenditures by agents can be used to locate possible missing transactions. A common error is to assume that the default value for cells is zero when in fact the opposite assumption may be better: a lack of readily available data does not mean there have been no transactions.

This document is not a ‘cookbook of recipes’, if only because countries are different. However, the guidelines assume a general and commonly used structure for the SAM, although the structure departs from that outlined in the SNA<sup>1</sup> by collapsing the ‘Allocation of Primary Income’, ‘Secondary Distribution of Income’ and ‘Uses of Income’ accounts. This involves no loss of information since it depends only on the reorganisation of the data.

The approach advocated is that of successive disaggregation. Start with highly aggregated data and progressively disaggregate the data. However, disaggregation requires determining the accounts that will be disaggregated and identifying the individual accounts. Typically, the accounts that will be disaggregated are those for commodities and activities, factors, e.g., multiple types of labour and capital, institutions, e.g., households and enterprises, and tax instruments, but decisions will always be case specific.

At each stage of the process of disaggregation a decision needs to be made about whether the SAM is to be reconciled and if so the choice of approach to reconciliation. One option is to treat the previously reconciled stage as known and certain except for the accounts being disaggregated; this can be achieved using various metrics, e.g., Modified RAS, Entropy, etc. A second approach is to assume that a subset of transactions from the previous stage are known and certain; this can be achieved using various metrics. A third approach is to assume that the transactions or a subset of transactions from the previous stage are priors with varying levels of reliability, and the accounts being disaggregated are priors with lower levels of reliability, i.e., assume that the next stage of in the disaggregation adds some information, perhaps limited, about the transactions from the previous stage; this needs a metric that can attach error bounds to the prior estimates, e.g., Entropy, Stone-Byron, etc. A fourth approach is to reconcile the SAM at each stage of the process, then use the additional information to

---

<sup>1</sup> See paragraphs 28.71 to 28.75 of the System of National Accounts.

*Guidelines for Compiling a SAM*

revise the prior and re-evaluate the estimated SAM; this process may be repeated until the prior does not generate any unrealistic transactions, although the objective is not to achieve a prior that is the same as the reconciled SAM it helps generate. Then the next stage of the disaggregation process is undertaken by extending the prior to including the next set of disaggregated accounts, which is then reconciled and the prior adjusted. This continues until all the disaggregations have been achieved and then the reconciled SAM is the final SAM. A fifth approach is to estimate a prior with all the disaggregation, with each nonzero cell in the SAM being assigned error bounds, and then reconciling the SAM. And sixth, a hybrid mix of these options may be used.

With modern computers and algorithms, the choice of options is now flexible. In the 1960s and early 1970s RAS, and variants, solved using iterative methods were justifiably popular, although they are still being used, e.g., by the Global Trade Analysis Project (GTAP, <https://www.gtap.agecon.purdue.edu/>). Subsequently, statistical methods using the Stone-Byron method became practical with the ability to invert very large matrices. Since the 1990s there has been a steady flow of publications advocating different solution algorithms with an emphasis on solution speed and problem size; the logic adopted in this course is that speed and problem size are only relevant if insights from information theory are embedded in the algorithms. Hence, this course develops estimation techniques that use entropy metrics. It is argued that the logic of information theory is central to the development of the RAS and Stone-Byron method.

There is unlikely to be a single approach that is correct for estimating all SAMs. The choices, and options, will be influenced by the available data and the preferences of team estimating the SAM.

But whichever option is chosen it is important when compiling the prior estimates that the compiler records the reliability of the priors. Prior estimates are point estimates measured with some degree of error: not all priors are equally reliable. In practice estimates of reliability are likely to be subjective probabilities based on knowledge about the source data, relative importance of each transaction, the role of each transaction in the SAM, e.g., taxes may be deemed important, and the subjective judgement of the person/s closest to the data. As such the evaluation of the reliability of the priors will be rooted in Bayesian statistics. It is suggested that a simple system is used to record estimates of the reliability of the priors:

*Guidelines for Compiling a SAM*

scoring reliability using a scale of, say, 1 to 5 (with 1 the most reliable) and an assumption about the distribution, e.g., normal or uniform.

The guidelines below use a SAM template as the basis for the suggestions.

Data Sources

The compilation of a SAM will require data from several different sources. The main sources will be

1. National Accounts,
2. Supply and Use Tables,
3. Production surveys,
4. Household Income and Expenditure surveys
  - a. Official
  - b. Alternatives,
5. Labour Force surveys
  - a. Official
  - b. Alternatives,
6. Government accounts
  - a. Taxes
  - b. Transfers,
7. Enterprise accounts,
8. Trade accounts.

Before spending time delving into the detailed data in the selected data sources, it is important to evaluate the sources. The aggregate national accounts data will have been evaluated when compiling the NAM and/or macro-SAM. For the survey data the priority is the questionnaire, which will identify the data collected/recorded, and then the year (the nearer to the year for the SAM the better). Government accounts always seem to be limited, so complementary data should be sought. Time spent collecting and reviewing data at the beginning is often rewarded, but the process of compilation will always identify gaps.

## **2. A General Statement of Difficulties**

The difficulties encountered when moving from a macro-SAM to a (micro) SAM have similar properties; appreciating these properties in advance can save a lot of time later. All these difficulties arise from three underlying causes: imperfect information, which can never be avoided, incomplete information, e.g., missing data, and data inconsistencies of various forms.

The difficulties can be listed under 10 headings in two groups:

### **1. Data Inconsistencies**

- a. Inconsistencies due to revisions to the national accounts data used to compile the macro-SAM.
- b. Supply and Use tables (SUT) that are not for the same year as the reconciled national accounts or for the same year but prior to revisions of the national accounts.
- c. Data from different years to the macro-SAM due to survey and census schedules.

### **2. Imperfect and Incomplete information**

- a. Data from surveys and censuses that were conducted for purposes other than disaggregating national accounts data.
- b. Segmenting the mixed income accounts.
- c. Limited information about compensation of employees with different attributes by activities and numbers of employees with different attributes by activities.
- d. Limited information about expenditure patterns (by commodities, taxes, savings, and transfers – domestic and international) by different households.
- e. Limited information about income sources (by factor earnings, transfers – domestic and international) by different households.
- f. Limited information about tax revenues from taxes on commodities, activities, factors, institutions.

*Guidelines for Compiling a SAM*

- g. Decisions by national accounts agencies to label data as confidential.

These difficulties may be of greater or lesser severity, but they are pervasive. Since they cannot be avoided it is necessary to develop robust strategies that can be used to overcome these difficulties. The first is to recognise that the difficulties exist; the second is to embrace the needs for completeness, i.e., avoid assuming that mathematical techniques are geared to completeness; the third is to appreciate the usefulness of control totals and that more (reliable) control totals are better than fewer; and fourth, and far from the least important, is the role of the column coefficients for economic analysis.

The importance of embracing the need for completeness lies in accepting that (a) if evidence indicates that a cell should be reporting a transaction, then a ‘poor’ nonzero prior estimate is infinitely superior to assuming there is no transaction, and (b) using all available information (objective and subjective) to help derive the prior estimate is appropriate. Subsequently, *ex-ante* information may be used to direct efforts to improve the prior SAM, in particular certain mathematical techniques use optimisation routines that yield estimates of the marginal improvement in the objective function from ‘improvements’ in prior estimates. This information can be used, iteratively, to direct the search for improvements in prior estimates.<sup>2</sup>

A detailed macro-SAM can provide a relatively large number of control totals. Most of these control totals relate to sub-matrices, including vectors, in the SAM, although some will be scalars, i.e., where individual cells in the macro-SAM map to individual cells in the SAM, e.g., the internal and external balances. Start with the recognition that a control total for a sub-matrix will equal the sum of the transactions reported in the cells of the sub-matrix. Hence the transactions in a sub-matrix after disaggregation can be expressed as a coefficient/share multiplied by the control total. This implies that one approach to developing prior estimates for transactions in a sub-matrix can begin by estimating shares and then multiplying the shares by the control total. In extreme cases there may only be a single control total for an  $i*j$  submatrix, where  $i > 1$  and  $j > 1$ , but often it is possible to develop additional control totals.

Consider the data reported in Table 2.1, which relates to a hypothetical submatrix that reports TVs for the consumption of commodities  $c_1$  to  $c_4$  by households  $h_a$  to  $h_c$ . The control total for the total value of consumption by households reported in the macro-SAM is

---

<sup>2</sup> The documentation for the SAMEST codes provides guidance on one way to access such information.

100 – the ‘total’:‘total’ cell. Ignoring for now some complexities that will be addressed in subsequent sections, it is possible to develop a general approach.

Assume, for now, that a SUT reconciled to the national accounts is available, in which case there are four additional control totals in the column ‘total’, i.e., there is more information. Assume further, for now, that there is a household expenditure survey, which, since it is a survey, not a census, means that the data do not encompass all households. However, given a survey with a random stratified sample and a classification of households, the survey data can yield estimates of the total consumption expenditures by each household; these will not equal 100 – the total consumption expenditure – but will yield estimates of the shares of consumption expenditures. Multiplying those share estimates by the total of consumption expenditure will produce prior estimates of the column totals.<sup>3</sup>

**Table 2.1 Consumption by Households**

|              | <i>h a</i> | <i>h b</i> | <i>h c</i> | <i>total</i> |
|--------------|------------|------------|------------|--------------|
| <i>c_1</i>   | 2.13       | 0.50       | 18.36      | <b>20.98</b> |
| <i>c_2</i>   | 2.27       | 0.70       | 10.55      | <b>13.51</b> |
| <i>c_3</i>   | 15.82      | 18.69      | 8.14       | <b>42.65</b> |
| <i>c_4</i>   | 6.90       | 12.09      | 3.86       | <b>22.85</b> |
| <i>total</i> | 27.13      | 31.97      | 40.91      | <b>100.0</b> |

In addition, the survey data can provide estimates of the patterns of expenditures by different households on different commodities. Again, the totals will not equate, but expenditure shares by household can be derived from the survey data, and these can be used to derive prior estimates of the TVs for each commodity by each household. Moreover, the survey data may be adequate to estimate standard errors for the patterns of expenditure by each household: more information.

The information content of the TVs in such a submatrix is not identical. It may be reasonable to argue that the value of total consumption expenditure is most reliable with the

<sup>3</sup> The alternative of using the survey weights to gross up the value is an option but requires assuming the weights correctly gross up to national levels of household consumption expenditure. However, the weights should/can be used to ensure that any divergence from a truly random stratified sample is compensated.

### *Guidelines for Compiling a SAM*

total consumption expenditures by commodity the next most reliable. However, if the data from the SUT are from a year in which a new SUT database was compiled from primary data, rather than an updated SUT, then the estimates may well be equally reliable. Similarly, estimates of total consumption expenditures by households may be considered more reliable than estimates of expenditures on individual commodities. Even if the estimates of reliability of the prior estimates are subjective (Bayesian), they can be used to inform mathematical estimation methods.

Inevitably the process is more complicated in practice, but the general principles can be followed and provide a method for deriving prior TV estimates from limited information. This illustration does identify several considerations

1. how can the column categories be defined,
2. how can the row categories be defined,
3. are the account definitions in the expenditure survey the same as those in the SUT,
4. if the share estimates are derived directly from the expenditure survey, i.e., without first estimating household consumption expenditures, do the prior TVs differ,
5. estimating control totals, even if they are measured with error, is important.<sup>4</sup>

These and other considerations will be addressed in subsequent sections.

### Prices and the Information in a SAM

Before looking at the various stages in detail it is important to emphasise that the assumed objective of the exercise is the development of a SAM to support economic models where the price formation process is deemed important. The theory of social accounting demonstrates that the (average) prices in the rows of a SAM are determined by the (average) costs defined by the information in the columns of a SAM. The Law of One Price (LOOP) in a SAM means that the TVs reported in a column can be defined as  $p_i * q_{ij}$  and the column sum as  $p_j * Q_j$ , which will equal the row sum ( $p_i * Q_i$ ). Hence

---

<sup>4</sup> This is especially so where the information about the individual transactions is limited, e.g., inter household transfers, government transfer to domestic institutions, etc. (see below for further on these sub matrices).

*Guidelines for Compiling a SAM*

$$p_j = \sum_i p_i * \left( \frac{q_{ij}}{Q_j} \right) = \sum_i p_i * a_{ij}$$

and therefore, the column coefficients,  $a_{ij}$ , are (value) quantity weights that define (average) costs,  $p_j$ , as the weighted sum of the prices of inputs,  $p_i$ . In a circular system everything depends on everything else, but in circumstances where prices matter the role of the relative shares of expenditures by accounts is of utmost importance.

### 3. Classification of Accounts

An important stage in the development of a SAM is the determination of the appropriate accounts. The basic structure of the SAM is illustrated in Table 3.1. As this makes evident, the transition from a macro-SAM to a SAM involves potential disaggregation of the broad groups of accounts. Typically, the disaggregation of the commodity and activity accounts would initially follow the detail provided by the SUT<sup>5</sup>: this may not be adequate and may subsequently require revision, e.g., if the focus of the analyses is on power systems of electricity generation and only one generation technology is recorded the activity accounting for ‘electricity’ generation may need to be segmented between, *inter alia*, coal, gas, nuclear, wind, etc., technologies.

It is, however, suggested that such segmentation of the commodity and activity accounts is delayed until other accounts have been disaggregated. This keeps the distance between published data and estimated data to a minimum when disaggregating other accounts.

The decisions about factor and household accounts need to be undertaken simultaneously. An important facet of a SAM is that it encompasses a social dimension that is intimately linked to issue of income distribution, i.e., the functional distribution of incomes. The logic is simple: incomes to households from factors depend on the factor endowments of households with different ‘types’ of households owning differing quantities of different ‘types’ of labour – skilled, unskilled, university educated and other ‘types’ of labour – and ‘types’ and quantities of capital and land. Changes in economic circumstances and policies impact differentially on the returns to different factors and therefore on the factor incomes received by households. A formal statement of this is provided by the conditions required for distributional invariance (Pyatt and Round, 2012). The authors show that “if we are primarily interested in the distribution of income among households and other institutions for which current expenditure is endogenous then, as a **rule of thumb**, we should be looking for an SAM for which  $\min([f], [c]) \geq [h]$ , where  $f$  is the number of factors,  $c$  the number of commodities and  $h$  the number of households] since, beyond a certain point, the further

---

<sup>5</sup> Examples exist where the commodity/activity accounts chosen are the same across a range of countries, e.g., the NEXUS SAMs by IFPRI (<https://www.ifpri.org/project/nexus-project/>); I regard this as poor practice. NB: GTAP uses common sector and factor accounts but this is in the context of a global model where country specific considerations need to be subsumed to wider considerations.

*Guidelines for Compiling a SAM*

disaggregation of households is not a priority because any gain in [the rank for  $h$ ] is contingent on the [number of factor and commodity accounts] and not on [the number of household accounts]. It therefore follows from these arguments that [there is] a case for SAMs which achieve some proximate balance between [ $f$ ], [ $c$ ], and [ $h$ ], depending on the ultimate purpose of the exercise.” (Pyatt and Round, 2012, p 266, emphasis added). In contexts where the number of commodities does not equal the number of activities, the number of activities should replace the number of commodities.

A conclusion from this, when building a SAM, is that since the number of activities/commodities is typically larger than the number of factors or households that the most likely constraint on the number of households in a SAM will be the number of factors accounts **IF** the distribution of income, i.e., between households, is an important consideration.<sup>6</sup> And **IF** the functional distribution of income, i.e., between factors, is important then the number of activities should at least equal the number of factors.

It is likely that access to details about factors will be a more binding constraint, and that there will be more information about different ‘types’ of labour than other primary factors.

If the decisions about numbers of factors and households starts from the perspective of the household accounts, there are some simple rules of thumb that may provide useful guidance. Starting from the perspective of the cost of living from the expenditure side suggests that criteria that reflect differences in preferences may be an appropriate starting point; in this context the incomes of households may be a secondary criterion. For instance, in developing countries the rural v urban distinction may be an important criterion as may the distinctions based on considerations such as ethnicity, religion, e.g., Hindus/Janis and Muslims in India and eating of meat, gender and educational attainment of the ‘head’ of household<sup>7</sup>, with income levels representing a lower order criterion.

---

<sup>6</sup> There was a body of CGE literature that advocated for ‘real’ households in SAMs – in one such SAM there were c7,000 representative/real households in a SAM for South Africa. This was supposed to provide greater information about income distribution, but the SAM had few factor accounts and a small fraction of the number commodity accounts compared to the number of households. Interest in this approach seems to have waned greatly although the papers may still be available.

<sup>7</sup> All known CGE and multi sector models employ, implicitly, the unitary household model. When the term ‘head of household’ is used it means the person responsible for making decisions. Regrettably, too many household surveys assume the eldest male is the ‘head of household’; it such cases classifying households by the gender of the ‘head of household’ limits female ‘headed’ households to households with no adult males. (There are exceptions, e.g., Ethiopia’s surveys seek to identify the prime decision maker.)

*Guidelines for Compiling a SAM*

Such an approach is likely to have implications for the factor accounts and hence sources of income to households. For rural households' employment in agriculture is likely to be relevant, while if gender is used in the household classification it would be appropriate to distinguish between labour types using different genders and if educational attainment is used then educational attainment might be appropriate distinction for labour types.

It is not possible to define concrete rules for the classifications chosen for households and factors. Moreover, even if it was, it is doubtful that the requisite data would always be available. The important point is to recognise the issue and make the classifications thoughtfully rather than mechanically.

**Table 3.1 Structure of a SAM**

|                      | Commodities      | Margins       | Activities              | Factors                        | Households                  | Enterprises                         | Government               | Investment             | Rest of World                | Account Total                |
|----------------------|------------------|---------------|-------------------------|--------------------------------|-----------------------------|-------------------------------------|--------------------------|------------------------|------------------------------|------------------------------|
| <b>Commodities</b>   |                  | Margins       | (Combined) USE Matrix   | 0                              | Household Consumption       |                                     | Government Expenditure   | Investment Expenditure | Exports of G&S               | Commodity Demand             |
| <b>Margins</b>       | Margins          | 0             | 0                       | 0                              | 0                           | 0                                   | 0                        | 0                      | 0                            | Margin Demand                |
| <b>Activities</b>    | Production       | 0             | 0                       | 0                              | 0                           | 0                                   | 0                        | 0                      |                              | Production                   |
| <b>Factors</b>       | 0                | 0             | Remuneration of Factors | 0                              | 0                           | 0                                   | 0                        | 0                      | Factor Income from RoW       | Incomes to Factors           |
| <b>Households</b>    | 0                | 0             | 0                       | Distribution of Factor Incomes | Inter Household Transfers   | Distribution of Income              | Transfers to Households  | 0                      | Remittances from RoW         | Household Income             |
| <b>Enterprises</b>   | 0                | 0             | 0                       | Distribution of Factor Incomes |                             |                                     | Transfers to Enterprises | 0                      | Income from RoW              | Enterprise Income            |
| <b>Government</b>    | Commodity Taxes  | 0             | Production Taxes        | Factor Taxes                   | Income Tax & other payments | Income Tax & Distribution of Income |                          | 0                      | Transfers from RoW           | Government Income            |
| <b>Savings</b>       | 0                | 0             |                         | Depreciation                   | Household Savings           | Enterprise Savings                  | Government Savings       | Stock Changes          | Capital Acc Balance          | Savings                      |
| <b>Rest of World</b> | Imports of G&S   | 0             | 0                       | Payments to RoW                | Remittances to RoW          | Payments to RoW                     | Transfers to RoW         |                        | 0                            | Imports of G&S and transfers |
| <b>Totals</b>        | Commodity Supply | Margin Supply | Cost of Production      | Expenditure on Factors         | Household Expenditure       | Enterprise Expenditure              | Government Expenditure   | Investment Expenditure | Exports of G&S and transfers |                              |

#### 4. Control Totals

The idea of control totals is very simple although the practice of using control totals is more complex, which perhaps explains why many developers of SAMs only use a limited number of them.<sup>8</sup>

The idea of a control total is that it provides an estimate of the sum of the transactions in a sub-matrix of dimensions  $n*m$ , where  $n$  and  $m$  are greater than or equal to 1, of the SAM, i.e., scalars, vectors and matrices. Estimates of control totals can be measured with or without error. In its simplest form control totals can be known values, i.e., estimates without error, of the row and column totals for the SAM; this is the case with the simplest application of the RAS reconciliation method. An elaboration of this might be where known values exist for row and column totals **and** specific cells; this is the case with a modified RAS reconciliation method; adding such information is known to greatly improve the results of RAS. Another elaboration of modified RAS is a reorganisation of the SAM so that there are known row and column totals for selected (sub) vectors of the SAM, e.g., total current consumptions by households ( $C$ ), government ( $G$ ), investment ( $I$ ), imports ( $M$ ) and exports ( $X$ ). See Module M4 for an exploration of some aspects of the RAS reconciliation method.

The fact that control totals are likely to be measured with error has long been recognised. Since we only have estimates of the transactions in the body of a prior SAM, it is arguably illogical to assume that we have estimates of control totals that are measured without error. Hence it was long recognised that reconciliation methods should be adapted to allow for errors in the prior estimates and/or the control totals (see Stephan, 1942; Allen and Lecomber, 1975). This was often achieved by attaching standard errors to the prior estimates of cells and control totals, but limited information means that quantifying standard errors is often subjective, which most econometricians are reluctant to endorse (“ it is impossible to find any fully operational basis for quantifying this information.” Lecomber, 1975, p5).

The lessons for developing control totals for SAM estimation are

1. more control totals are better than fewer,

---

<sup>8</sup> For instance, the GTAP database documentation (<https://www.gtap.agecon.purdue.edu/databases/v12/>) only identifies five control totals for the within region transactions.

### Guidelines for Compiling a SAM

2. estimates of the standard errors for priors and control totals are important,
3. estimates of standard errors may need to be subjective, and
4. the accuracy of an estimated SAM, as a reflection of economic transactions, depends strongly on the information in the prior estimates.

The macro-SAM, as developed in Module M2, can provide a series of important control totals. As a micro-SAM is compiled other control totals should also be sought.

### Use of Control Totals

Control totals have always been part of the process of estimating a SAM, regardless of the estimation method used. In many cases the control totals are simple row and column totals that may or may not be organised so that the SAM is consistent with selected macroeconomic aggregates, e.g., *C*, *G*, *I*, *X*, *M* and *GDP*, irrespective of the estimation or reconciliation method. The SAMEST estimation method developed in Modules M5, M6 and M7 extends the use of control totals to any sub matrix of the SAM and allows for all control total to be measured with error, with or without standard errors of known distribution. Thus, control totals are a key feature of all mathematical estimation and reconciliation methods used when compiling SAMs. These uses of control totals are examined in Modules M4, M5, M6 and M7, but controls are also useful when developing a prior SAM for estimation and reconciliation.

Any matrix of transactions ( $T_{ij}$ ) can be expressed as a scalar ( $x$ ) multiplied by a matrix of coefficients/shares ( $S_{ij}$ ), i.e.,

$$S_{ij} = T_{ij} / x .$$

If there is a matrix of transactions in a different period ( $T_{ij}^*$ ) and a control total for the total values of transactions for the matrix for the period of the SAM ( $x$ ), then a matrix of coefficients ( $S_{ij}^*$ ) can be used to estimate transactions for the period of the SAM ( $T_{ij}$ ),

$$T_{ij} = \frac{T_{ij}^*}{\sum_{i,j} T_{ij}^*} . x = S_{ij}^* . x$$

and  $x$  is a control total measured with or without error and  $T_{ij}$  is a prior estimate of any (sub) matrix of the prior SAM.

### Guidelines for Compiling a SAM

These prior estimates of matrices will always be measured with error since they depend on the assumption that the coefficients for the previous period ( $S_{ij}^*$ ) are identical to those for the period of the SAM ( $S_{ij}$ ). This is an unavoidable situation.

An art of the process of estimating a SAM is identifying when the structural information in  $S_{ij}^*$  is a reasonable approximation to  $S_{ij}$ .

In this general formulation there is no consideration of whether coefficients should be derived using row or column totals. More generally the coefficients can be derived as

$$S_{ij}^{R*} = \frac{T_{ij}^*}{\sum_j T_{ij}^*} \quad OR \quad S_{ij}^{C*} = \frac{T_{ij}^*}{\sum_i T_{ij}^*}$$

where  $S_{ij}^{R*}$  are row coefficients that define shares of income to the account  $i$ , and  $S_{ij}^{C*}$  are column coefficients that define shares of expenditures by the account  $j$ . In these cases, the control totals will be for row and column totals, although error bounds may differ.

It is highly likely that for any SAM development process both row and column coefficients will be used. But, as a rule, priority should be given to the column coefficients because they are the key determinants of price systems in SAMs.

## 5. Disaggregating Commodity and Activity Accounts

The first stage of a disaggregation of the commodity and activity accounts is typically determined by the available Supply and Use Table (SUT) data. The presumption in this section is that an official SUT is available for a year close to the year chosen for the macro-SAM; in practice it would be expected that the year for which the official SUT was available would influence choice of year with allowance for the importance of choosing a ‘normal’ year. In addition, it is assumed that the SUT is presented with a combined Use table, i.e., commodity demand is not segmented by domestically produced and imported commodities<sup>9</sup>.

In this section certain issues are reserved for later consideration, these include

1. suppressed information about the composition of commodity outputs by activities,
2. missing data points due to ‘confidentiality’, e.g., sugar production,
3. aggregated data because of ‘limited’ competition, e.g., food processing,
4. estimating imports Use matrix.

An implicit presumption is that the full sets of commodity and activity accounts will be carried over from the SUT in the SAM. There may be reasons why it is better to aggregate the commodity and activity accounts in the SUT, because of very large amounts of data in the SUT or the scale at which other data are available or as part of a sequential process. This would produce a smaller SAM that may be more tractable. In practice this can be a time efficient option. In section 7 issues associated with disaggregating commodity and activity accounts are discussed.

### Commodity and Activity Classifications

The starting point for account classification will almost always be the commodity and activity accounts used in the available SUT. Most national accounts agencies use the ‘principal product’ classification scheme wherein enterprises/establishments are assigned to an activity group on the basis of their principal product, e.g., a metal manufacturer whose largest output,

---

<sup>9</sup> Imports Use matrices are typically only produced for years for which IOT will be produced (every 10 years or so), and even then, are often not easily available. (Apocryphal information indicates that while national account statisticians are confident about the combined Use data they are less confident about the domestic/import split. Given the raw data this is not surprising.)

### *Guidelines for Compiling a SAM*

by value, is aluminum, will be assigned to the activity group to which the product aluminum is assigned, e.g., non-ferrous products.

It is possible/probable that the classification used in the official SUT does not match the desired set of commodities and activities; this raises the question of when in the process to disaggregate the commodity and activity accounts. It is suggested that this is delayed until after the factor, institution and tax accounts have been disaggregated, because much of the data used for further disaggregation is likely to be reported at the level of commodity and activity accounts used in the SUT.

### Why Supply and Use Tables?

‘Modern’ methods for deriving Input-Output Tables (IOT) use mathematical techniques and associated assumptions to derive IOT from SUT (see Miller and Blair, 2009, pp 184-242; UN, 1999, pp 75-103; Eurostat, 2009, pp 293-364). These methods involve a degree of redefinition of the commodity and activity accounts, which requires that ALL transactions in a SAM must be re-quantified to remain consistent with the revised definitions for commodities and activities. The degrees of adjustment in the TV estimates depend upon the chosen control totals and mathematical method, e.g., C(ommodity) by C(ommodity) with Hybrid Technology Assumption (TA) IOT, or A(ctivity) by A(ctivity) with Commodity TA IOT, or C by C with Industry TA IOT, etc. All known methods are irreversible without access to the detailed data used and (the almost) unavoidable manual adjustments implemented. This means that, in theory, every change in the accounts in the SAM should first be implemented in a SUT based SAM before converting to an IOT based SAM.<sup>10</sup>

A major reason for working with a SUT based SAM is that national accounts agencies report data using the commodity/product and activity/industry categories used in the SUT. Hence, for instance, Labour Force Surveys report the quantities of types of labour used by activities, which means the definitions of labour and activities are consistent with the data in the Use table. Similarly, tax revenues on products/commodities are reported for the

---

<sup>10</sup> This fact is often ignored by compilers of IOT based SAMs under the implicit presumption that the differences in estimated TVs between ‘natural’ definitions in SUT and those in IOT are small enough to be ignored without evidence to justify the presumption. The GTAP database is an example. The problem is compounded by the fact that the choice of commodity or activity definition for the accounts is not reported.

*Guidelines for Compiling a SAM*

commodities as classified in the Supply table. This makes the process of adding information less confusing and reduces any tendency to treat account definitions as malleable.

A secondary reason is that decisions made by the compiler of the SAM about the mathematical method used to move from an SUT base to an IOT base are embedded in the data. A simple summary of the choices needed to move from SUT to IOT involves

1. choose the commodity or activity accounts for the IOT, i.e., choose the applied account totals,<sup>11</sup>
2. select the technology assumption to apply, e.g., pure commodity, pure activity, a hybrid technology, by-product, etc.,<sup>12</sup>
3. linearly transform the SUT based data to generate an IOT, i.e., by elimination the off-diagonal TVs in the supply/make matrix.

The presumptions embedded in the process are then imposed on the user of the SAM, whereas a SUT based SAM allows the user to choose between imposing a mathematical transformation or modeling directly from the SUT based SAM.

---

<sup>11</sup> So called commodity by commodity or industry/activity by activity IOT.

<sup>12</sup> The technology assumption essentially relates to the technologies used to produce secondary commodities. It defines whether the same commodity produced by any activity uses 'pure' commodity technologies, all commodities produced by an activity using 'pure' activity technologies, a mix (many hybrids) of the two 'pure' assumption, or secondary commodities are by-products.

## 5. Adding Supply and Use Table Data

Adding SUT data to a macro-SAM is straightforward: the commodity and activity accounts are disaggregated, the intermediate inputs use and domestic production submatrices are populated, the margin, product tax and import accounts are populated for the commodity accounts and the factor and tax revenue accounts for activities are populated. The extent of the information that can be added depends on the detail in the SUT. For the first SAM it is strongly recommended that a ‘smallish’ SUT is used, say less than 50 commodities and activities.<sup>13</sup>

It is important to note that published SUT will have been produced using reconciliation and/or estimation methods (see Eurostat, 2008, Chapter 8). Hence it is appropriate to regard all values in published SUT as measured with error.<sup>14</sup>

There are typically two simple cases and two more complex cases, which are dealt with in turn below.

### *SUT for the Chosen Year: The Simple Cases*

The first case involves a SUT that is fully reconciled with aggregates reported in the macro-SAM. This will produce a first SAM that is fully consistent with the macro-SAM; hence, job done, and we can move onto the next stage.

But real-world data are rarely that compliant. The second simple case arises when the SUT and macro-SAM are for the same year, but the aggregate national accounts data have been revised but the SUT has not been revised. This is not uncommon. The differences in the aggregate data from the SUT and macro-SAM will, in this case, usually be ‘very’ small, and it can reasonably be assumed that the structural information in the SUT is a good reflection of the economy for that year.<sup>15</sup> In this case it may be adequate to adjust the column and row totals for the accounts (commodity, activity, factors and taxes) e.g., use shares from the SUT and

<sup>13</sup> I doubt this advice will be listened to, but I live in hope!!!

<sup>14</sup> The EU for instance advocates an Economic Data Processing (EDP) method (<https://ec.europa.eu/eurostat/documents/1015035/2041329/Procedures-for-development-and-implementation-of-methodology-underpinning-EDP-data.pdf>). The extent to which and how this method reduces inconsistencies in SUT is unclear to me; it appears to me to be more about organising the workflow than anything else, which is good practice.

<sup>15</sup> This will not always be so. The energy (oil) price rises of the 1970s and the energy (gas) price rises of 2021/2 are examples of years where the structural information may not be a good reflection of a ‘normal’ year. The COVID pandemic is a similar case.

### *Guidelines for Compiling a SAM*

totals from the macro-SAM (see the sub section below) and use a simple mathematical reconciliation method, e.g., RAS (see Module 4).

In this case the combination of the SUT macro-SAM will provide important control totals, e.g., total value of final demand by commodity by all domestic institutions, total values of gross outputs and factor payments by each activity, etc. These control totals will be important when moving on to disaggregate the household and factor accounts.

But these are point estimates of the true values. It is reasonable to argue that the national accounts agency is likely to have more information than the compiler, and therefore that the prior (point) estimates have relatively narrow error bounds. However, without additional information there is no evidence about the distributions of the error bounds.

### *SUT for an Adjacent Year: Using Share data for Aggregates*

The second simple case arises when the SUT and macro-SAM are for adjacent but ‘normal’ years AND the aggregate totals from the SUT do not depart markedly from those in the macro-SAM; say less than 5%. If the differences in the aggregate data are small, it may be reasonable to assume that the structural information in the SUT is a reasonable reflection of the cost structure of the economy for the year of the macro-SAM; in these cases extending the ‘control’ totals using national accounts data to revise the **control totals** for some, or all, of gross outputs, total intermediate inputs, value added and remuneration of labour together with additional data to give controls for commodities, e.g., imports, exports, final demand, etc., may be enough to use a mathematical estimation method (see Module 2 5, 6 and 7).

Again, these are point estimates of true values. While the national accounts agency is likely to have more information than the compiler about transactions for the SUT used they will not have included data for the year of the SAM. Hence, the prior (point) estimates will have wider error bounds than when the SUT was for the year of the SAM, although there is unlikely to be robust evidence about the distributions of the error bounds. Important additional information will be about individual, and major, transactions for the year of the SAM (see Module M4.5 on modified RAS).

### *A Complete but ‘Distant’ SUT: Using Coefficients*

The first of the less simple cases arises when there have been marked structural changes, e.g., production technologies, between the dates for the SUT and macro-SAM and/or several years,

*Guidelines for Compiling a SAM*

say 5 or more, but the SUT is complete. In this case there is a need to collect more data to help inform the prior. Assuming the SUT will provide indicative structural information about cost shares, the objective will be to find data that can increase the number of control totals; a surprising amount of useful data can be found in the national accounts. The following are suggestions:

1. gross outputs by activity or groups of activities,
2. value of total intermediate expenditures by activity or groups of activities,
3. value of primary input expenditures by activity or groups of activities,
4. value of expenditures on groups of intermediate inputs by activity or groups of activities,
5. values of domestic production by commodity, or groups of commodities,
6. value imports and exports by commodity, or groups of commodities, and
7. values of commodity expenditure by final demand by commodity, or groups of commodities.

The objective is to generate a series of prior estimates based on shares and control totals, like those illustrated in Table 2.1. The prior estimates for TVs can then be used, with the control totals, in a mathematical reconciliation method.

Deriving prior estimates. The ‘dated’ SUT can provide column or row coefficients/shares for purchases by activities (intermediate inputs and primary inputs), institutions (final demand) and trade (exports). Priority should be given to column coefficients, i.e., the cost shares. Then the extra data should provide a series of column totals, e.g., intermediate inputs and gross outputs, that when multiplied by the coefficients, will produce prior expenditure TVs, e.g., the TVs for expenditures on factors and intermediate inputs.

However, the priors will be best estimates that are known to be wrong. If published national accounts aggregates are available and used as control totals, it is arguably reasonable to assume that the control totals are more reliable than the individual TVs, i.e., the error bounds on the control totals will be smaller than those on the individual cells. Ideally, information about the standard errors and distribution would be known although this is

*Guidelines for Compiling a SAM*

unlikely. The one possible exception will be for remuneration of labour; this is discussed in section 7 when looking to disaggregate the factor accounts.

*Incomplete SUT*

The second less simple case arises from an increasing tendency by national accounts agencies to suppress detail about domestic commodity production by activities, i.e., the off diagonal elements in the Supply matrix. In the main, this means that although the account totals can be identified for domestic production of commodities by activity the output mix by activity is withheld. In the worst-case information about an activity is completely missing; one example is the withholding of information about sugar processing activity.

The first option is to estimate the value of the principal commodity produced by each activity, i.e., populate the principal diagonal; this reduces the problem to estimating the values on non-principal production. An examination the data can often lead to identification of cells that should have zero entries, e.g., cells where the output and commodity supply are identical so that the entry on the principal diagonal is known and other row and/or column entries are zero, so those can be populated. It may also be possible to work out the total values of non-principal commodities produced by each activity and the amounts of principal commodities produced by activities as non-principal commodities, which produces some extra control totals. Then crude estimates can be made for these cells, which means the error bounds are wide and of unknown distribution. Thereafter the problem is left to a mathematical estimation method.

Better prior estimates of the off-diagonal transactions may be obtained from previous, or rarely, subsequent SUT. One possibility is that a full Supply table was published to support the last set of IOT that were derived and reported. If such a table exists it can provide share estimates to populate the Supply matrix. Similarly, a Supply table for a similar, but different, economy may be usable to derive estimates of the off-diagonal matrix. But the error bounds will be even wider.

**REST THE TEMPATION TO USE DATA THAT HAVE BEEN ESTIMATED TO CONSTRUCT A PREVIOUS SAM FOR THE COUNTRY. IF THE PREVIOUS SAM WAS COMPILED PROFESSIONALLY, YOU SHOULD BE ABLE TO REPLICTE THE EXERICISE AND USE THOSE DATA SOURCES FOR YOUR OWN EFFORTS.**

## 6. Disaggregating Commodity and Activity Tax Accounts

Tax accounts are important components of a SAM. While domestic costs of production may be presented as technologically driven and consumer preferences driven by tastes, taxes are policy instruments that are, at least in part, driven by legislative fiat. Given the price definitions built into a SAM, taxes are instruments that can be used to change relative prices and thereby change incentives, e.g., import duties, energy and emission taxes, excise taxes on alcohol and tobacco, etc. Moreover, taxes are necessitated by the need to fund the operations of government and the provision of public services. Consequently, adjustments to tax instruments are important in nearly all analyses conducted using SAM data.

The recording of tax instruments in a SAM has several features that need to be considered. These include:

1. a SAM records transactions, and therefore records the revenue realised by tax instruments on specific transactions (not the rates),
2. different tax instruments are levied in different ways, e.g., *ad valorem*, specific (per unit quantity), share of taxable income,
3. published tax rates may not be consistent with associated tax revenues, in fact it should be expected that realised revenues are less than implied revenues,
4. data on tax revenues linked to specific transactions are difficult to extract,
5. recorded tax revenues should not include extraneous information, e.g., tariff equivalents typically include rents that accrue to agents other than the government<sup>16</sup>, and
6. tax instruments, and how they are applied, differ across countries.

So, although data on tax revenues are very important, they are difficult to acquire and disentangle. In these circumstances a strategy is needed. One such strategy has the following steps:

---

<sup>16</sup> The use of tariff equivalents rather than tariff revenues AND rents is distortionary: it imposes a presumption that rents accrue to the government which unusual. GTAP uses tariff equivalents and avoids the problem of rents using its construct of the regional household, but as soon as users include domestic institutions with 'active' behavioural relationships tariff equivalents are an issue.

### Guidelines for Compiling a SAM

1. identify the major tax instruments, e.g., national accounts data often publish the total revenues by different tax instruments as part of the government accounts,
2. identify how the different tax instruments are levied, e.g., *ad valorem* vv specific/quantity,
3. collect data on applied tax rates for each instrument,
4. determine the tax instruments to be included in the SAM
  - a. segment between taxes that are levied in different ways, e.g., *ad valorem* vv specific,
  - b. avoid aggregating revenues that are levied in different ways,
  - c. make sure the taxes relevant to policy choices are separate, e.g., energy and emissions taxes,
  - d. there will nearly always be an aggregate of tax instruments not separately identified (if omitted the SAM will be incomplete/distorted);<sup>17</sup>
5. ensure the total revenues by each of the selected tax instruments are control totals and assume these control totals are measured with little error.

This strategy provides a means by which tax revenues can be estimated in discrete submatrices with reasonable control totals (see below).

### Taxes on Commodities/Products

Taxes on commodities are often the most difficult to disentangle and yet they are arguably the most important because of their role in the transition from basic to purchaser prices. But the strategy outlined above does provide a route through this ‘jungle’.

A SUT will/must provide at least one account for ‘taxes (revenues) on products’, and the government accounts almost always provide information of government revenues from different tax instruments. If these data points are combined in a submatrix, similar to Table 6.1, then row and column control totals are identified.

---

<sup>17</sup> It may be appropriate to assume that such aggregations of tax instruments should be fixed in policy analyses, i.e., treated as fixed distortions.

*Guidelines for Compiling a SAM*

For this hypothetical economy 4 major tax instruments have been identified, i.e., VAT, Import duties, Excise tax, Fuel tax, and one residual instrument, i.e., other taxes. In this example it may be the case that VAT is levied (overwhelmingly) on final demand and *ad valorem*, that the fuel tax is only levied on fuel (c\_2) and the excise and fuel taxes are specific. The column control totals ('taxes on products') are from the SUT and the row control totals ('total revenue') are from the national/government accounts. The creation of prior estimates then reduces to getting estimates of the transactions in italics.

**Table 6.1 Taxes on Products**

|                          | <b>c_1</b> | <b>c_2</b> | <b>c_3</b> | <b>c_4</b> | <b>Total revenue</b> |
|--------------------------|------------|------------|------------|------------|----------------------|
| VAT                      | <i>49</i>  | <i>35</i>  | <i>1</i>   | <i>40</i>  | <b>125</b>           |
| Import duties            | <i>4</i>   | <i>8</i>   | <i>24</i>  | <i>5</i>   | <b>41</b>            |
| Excise tax               | <i>16</i>  | <i>43</i>  | <i>0</i>   | <i>22</i>  | <b>81</b>            |
| Fuel tax                 | <i>0</i>   | <i>50</i>  | <i>0</i>   | <i>0</i>   | <b>50</b>            |
| Other taxes              | <i>2</i>   | <i>5</i>   | <i>3</i>   | <i>4</i>   | <b>14</b>            |
| <b>Taxes on Products</b> | <b>71</b>  | <b>141</b> | <b>28</b>  | <b>71</b>  |                      |

It is possible to reduce the scale of the problems when estimating priors. A good starting point is the identification of zeros, e.g., the fuel tax can only be levied on fuels, some excise tax revenues identify the commodities they are associated with, e.g., alcohol, import duties can only be levied on imports where there are imports, and services are often duty free, etc. The next step would be to use data on applied rates that can be applied to the basic price value of the commodity; typically, the prior estimates will sum to more than the revenue so scaling them to the row control will be appropriate to help the mathematical techniques that will be needed.

For all instruments it is important to identify who in fact pays the tax. For instance, some taxes are rebated on the purchases of intermediate inputs, with VAT being the obvious case, but tax rates on energy inputs often differ according to the purchaser.

If a tax is levied at different rates on intermediate commodities than final commodities, then it is important to segment such tax instruments. Tax on energy products typically fall into this category of tax instruments to separate out these instruments, especially if subsequent analyses will consider energy and/or environmental policy issues.

### *Guidelines for Compiling a SAM*

The process is laborious, but important. The principle to follow is that of using all the information that you can find, noting that ZEROs are important pieces of information since a zero in a prior will be a zero in the solution. Note also that if you set a nonzero transaction to zero in the prior it will be zero in the solution; hence if the evidence indicates that tax revenues are realised on a transaction the prior for the cell must be nonzero, however unreliable the estimated prior.

### Taxes on Production

Taxes on production often appear simpler to derive because the instruments are less complex. There are two obvious distinctions that need to be recognised; taxes levied on the gross output and taxes levied on factor use. These are commonly reported as a row of taxes on production and, sometimes, a row of subsidies on production. If possible, these are best kept separate.

### *Taxes on Activities*

These aggregate taxes on production have various definitions and the applied definitions are not always obvious. The first distinction is between taxes and subsidies, and control totals for revenues, from the taxes, and expenditures,<sup>18</sup> on subsidies, need to be collected.

The second distinction relates to the treatment of ‘social contributions’ for which the SNA advocates a specific approach

*“It is worth explaining in some detail here the way social contributions are recorded in the SNA. Although employers normally pay social contributions on behalf of their employees directly to the social insurance schemes, in the SNA these payments are treated as if they were made to employees who then make payments to social insurance schemes. In terms of the accounts, this means that they first appear as a component of compensation of employees in the use side of the generation of income account of employers and the resource side of the allocation of primary income account of households (adjusted for external flows in compensation of employees). They are then recorded as uses in the secondary distribution of income account of households (and possibly of the rest of the world), and as resources of the*

---

<sup>18</sup> Typically defined as negative expenditures.

*Guidelines for Compiling a SAM*

*sectors managing social insurance schemes. All employers' social contributions follow this route. This way of recording transactions as if they followed another course is often called "rerouting" (SNA, para 2.96).*

This means that compensation of employees by activities should be gross of all social contributions: if the SNA conventions have been followed. If a NAM has been compiled it should be evident if the conventions have been followed.

The third distinction relates to whether there is any special treatment of activities within an economy: these would normally be expected to appear as subsidies but not always.

The fourth distinction involves determining whether there are any specific taxes or subsidies levied on factors used by an activity, these are sometimes referred to as 'taxes on factor use'. Evidently the distinction between taxes/subsidies levied on activities and those levied on factors used by activities can be opaque. With both these possibilities it is important to recognise that the rates must be activity specific since otherwise they are effectively taxes/subsidies on the factors alone. Some agricultural policy support instruments in developed economies have been assumed to operate via factor use subsidies but, in these cases, subsequent model behavioural relationships will need to define how the subsidies operate. The SAM is only concerned with the TVs/payments.

The final distinction is between taxes and subsidies that are applied to intermediate inputs used by an activity. These are distinct from taxes on the intermediate input as a commodity; again, it is important to recognise that the rates must be activity specific since otherwise they are effectively taxes/subsidies on the commodity.

### Taxes on Income and Wealth

These taxes are levied on institutions and the issues they raise are dealt with when discussing the disaggregation of institutional accounts.

## 7. Disaggregating Factor Accounts

The classification of factor accounts in a SAM has been discussed above, so here attention is focused on practicalities and the approach is generic. It is implicitly assumed that the compiler has access to information about factors used by activities; from surveys of agriculture/manufacturing/services and/or labour force surveys and/or household income and expenditure surveys. Ideally the available data will include information about payments by activities to factors and the quantities of factors used by each activity.

The section starts by addressing the issues raised by the category mixed income whose component parts must be disentangled. This is difficult since the category only exists because national accounts agencies are admitting lack of information. Next is the issues presented by the GOS category; this encompassed returns for capital and entrepreneurial services and any self-employment incomes not included in mixed income. The last part looks at labour services.

### Mixed Income Accounts

This category is always a problem; not least because it exists to address a difficulty that national accounts agencies have (implicitly) placed in the box labelled “too difficult”: **BUT it must be addressed**. Overwhelmingly it refers to situations where compensation for the labour services of the self-employed and gross operation surplus (GOS) cannot be readily segmented.<sup>19</sup> If data on labour use is available, then one approach is to use data on labour qualities and compensation of employees to infer the share of mixed income assigned to self-employed labour with the residual assigned to GOS. But GOS then must cover payments for capital and land and natural resource services.

Special problems exist for land. Land is overwhelmingly used by agriculture (and forestry) and hence part of mixed income payments by agriculture will be attributable to land, as should some of the GOS recorded for agriculture. One option is to impute the payments for land services based on the rental value of land, noting that some farmland is rented, for which rental payments should be in GOS, while other farmland is owner occupied with implicit

---

<sup>19</sup> Compensation of employees should not be part of ‘mixed’ income.

### *Guidelines for Compiling a SAM*

‘payments’ included in mixed incomes.<sup>20</sup> A particular problem is this the possibility that imputed payments for land and self-employed labour services<sup>21</sup> exceed the value of mixed income, which implies negative payments for capital services: the residual from mixed income. Arguably, however, rental payments to landowners should be recorded as part of GOS not mixed income, and therefore some of the payments for land services should be assigned to GOS.

Land area data should also be collected. If more than one agricultural activity is recorded these data will be useful when calibrating models using the SAM data. But note that the implicit rental prices of land are likely to vary, often widely, across activities.

The discussion about payments for land services above reflects conditions found in archetypal developed economies where tenancy arrangements are based on fixed annual payments. But in many economies the tenure arrangements for land are less transparent, e.g., share cropping, which makes it less straightforward to assign TVs for the provision of land services.

**The issue of payments for land services MUST be addressed, so compilers of SAMs will be forced to make some strong assumptions.**

Natural resources typically relate to extractive activities, i.e., mining, quarrying, etc., and payments for these services should not be included in mixed incomes. In such cases the returns to the natural resources will be embedded in GOS; but care is needed since returns to natural resource extraction can be realised as royalties paid by an incorporated business corporation, linked to extractive activity, to the government. In this case GOS passes to the business corporations inclusive of royalties and the return to the natural resource is in the share of GOS going to the beneficial owner/government.

### Gross Operating Surplus (GOS)

Gross Operating Surplus is ambiguous in practice. In theory it is straightforward, it records payments for capital services plus returns to entrepreneurship plus any economic rents (from imperfect competition) accruing to an activity gross of any depreciation of capital. The

<sup>20</sup> The simplification in classical economics of all labour owned by workers, all capital owned by entrepreneurs and all land by landlords (gentry) avoids these complications.

<sup>21</sup> Defining self-employed labour services in agriculture is difficult because of the important role played by unpaid family members.

### *Guidelines for Compiling a SAM*

distinction between GOS and Net Operative Surplus (NOS) being the economic depreciation of capital used by the activity. Alternatively, it can be defined as all payments to factors not included in mixed income and the compensation of employees, in which case it is, implicitly, inferred as a residual after accounting for all other costs incurred by an activity.

It almost always enters a SAM as a residual. It can become a more interesting account if the types of capital are segmented, e.g., buildings, powers stations, vehicles, livestock capital, etc., and/or economic rents are inferred. The concept of fixed and variable capital is useful here in that it makes a distinction between capital services that are fixed to specific activities and those that are mobile between activities, and aligns with economic theory. But there are issues with the units required for the measurement of capital items that make such distinctions tricky.

If the types of capital are segmented it is appropriate to segment the Gross Fixed Capital Formation (GFCF) account because the composition of expenditures to produce different types of capital will differ.<sup>22</sup> If the SAM data are used in comparative static CGE models the segmentation of the GFCF account may not be important, but if the data will/may be used in a dynamic formulation the distinction is important.

### Labour Accounts

A pervasive problem with the labour accounts that can be encountered relates to the classification of labour types in surveys and national accounts. The SNA endorses the International Labour Office (ILO) classification of labour types, see Table 7.1 where ISCO-88 is the latest set of categories<sup>23</sup>: this classification is occupation focused, i.e., it is based on job descriptions. This may not be ideal because many whole economy models are concerned with the reallocation of labour across activities where a skills based schema may be more appropriate.

In analytical models, a primary concern is labour mobility between activities. But categories like ‘Skilled agricultural and fishery workers’ may be overly concentrated in a limited range of activities, so ultimately consideration needs to be given to appropriate

---

<sup>22</sup> The columns for GFCF identify the costs associated with the production of each type of capital; it is expected that the production costs differ by type of capital.

<sup>23</sup> <https://ilostat.ilo.org/resources/concepts-and-definitions/classification-occupation/> provides details.

alternative classifications. In practice the options will be constrained by the available data, and a sequential approach may be needed, e.g., first follow the ILO classifications, and then further disaggregate using additional data.<sup>24</sup>

**Table 7.1 ILO Labour Categories**

| Broad skill level                  | ISCO-08                                               | ISCO-88                                              |
|------------------------------------|-------------------------------------------------------|------------------------------------------------------|
| <b>Skill levels 3 and 4 (high)</b> | 1. Managers                                           | 1. Legislators, senior officials and managers        |
|                                    | 2. Professionals                                      | 2. Professionals                                     |
|                                    | 3. Technicians and associate professionals            | 3. Technicians and associate professionals           |
| <b>Skill level 2 (medium)</b>      | 4. Clerical support workers                           | 4. Clerks                                            |
|                                    | 5. Service and sales workers                          | 5. Service workers and shop and market sales workers |
|                                    | 6. Skilled agricultural, forestry and fishery workers | 6. Skilled agricultural and fishery workers          |
|                                    | 7. Craft and related trades workers                   | 7. Craft and related trades workers                  |
| <b>Skill level 1 (low)</b>         | 8. Plant and machine operators, and assemblers        | 8. Plant and machine operators and assemblers        |
|                                    | 9. Elementary occupations                             | 9. Elementary occupations                            |
| <b>Armed forces</b>                | 0. Armed forces occupations                           | 0. Armed forces                                      |
| <b>Not elsewhere classified</b>    | X. Not elsewhere classified                           | X. Not elsewhere classified                          |

The derivation of a submatrix of prior estimates of payments by activities to different labour types can proceed by the coefficient/share method outlined in section 4, except in this instance the column control totals are known from the compensation of labour data in the Use matrix and row totals need estimation. The steps taken when estimating these priors will depend heavily on the available data. If a Labour Force Survey and an Income and Expenditure Survey exist, above and beyond the information available from surveys/censuses of manufacturing, services, agriculture etc., it is probable/possible that the data sources are complementary. In these cases, the estimated prior TVs will have been drawn from census/survey data so in addition to the point estimates, estimates of standard errors and distribution may be available.

<sup>24</sup> Alternatively, analytical models can use a more elaborate set of behavioural relationships to offset the issues presented by the ILO classifications.

*Guidelines for Compiling a SAM*

Labour Force Survey data will often adopt the ILO classification of labour types that will support estimating a submatrix of payments to each (ILO) labour type by each activity. In survey data, the reported transactions will need to be scaled to match the totals of expenditure on labour by each activity<sup>25</sup> before using coefficients and control totals to derive prior estimates. Income data in an Income and Expenditure Survey often report household incomes from sale of labour services to different activities and sometimes include data by ILO labour types and or skill attributes of household members. These data can be gross substitutes for Labour Force Survey data or complements. Moreover, Income and Expenditure Survey data often report on the characteristics of household members in terms of educational attainment and other indicators of skills; these data can provide a method to inform transitioning from ILO labour classifications to a skill-based classification.

While collecting transactions data for labour use by activity attention should also focus on collecting estimates of the quantities of labour of each type used by activities; these data, in the form of a satellite account, are very useful when using a SAM to calibrate a model. However, it is important to recognise that the implicit wage rates for each type of labour are likely to vary, often widely, across activities. Hence using the labour quantity data to derive mechanically estimates of transactions values is unlikely to be good practice.

---

<sup>25</sup> While surveys should be based on random stratified samples that perfectly represent the population this is rarely achieved. So, weights are often available to adjust for sample biases.

## 8. Disaggregating Household Accounts

When SAMs are intended to bring out issues of income distribution the household accounts are usually the prime focus of attention, although all too often the ‘rule of thumb’ from the distributional invariance theorem is set aside (see section 3 above). In all SAMs it is important to remember that TVs are point estimates that are intended to summarise the transactions between two accounts that can be defined as representative agents, i.e., some form of average of the individual agents that constitute the representative agents. Hence the individual agents would report a range of different transactions. The aim then in defining representative agents is to define an aggregate that expresses how on average the individual agents in a group might behave. Hence the process involves identifying Representative Household Groups (RHG),

Among the features that need consideration are the characteristics that reflect differences in preferences (column entries), e.g., gender (of ‘head’ of RHG), race, religion, religion, education, incomes, etc., and differences in endowments (row entries), e.g., wealth and physical asset ownership, human capital, etc. Hence it is a balancing act that seeks a balanced representation of the determinants of expenditures and income sources.

Consequently, the determination of the RHG depends upon knowledge about the economy for which the SAM is being compiled; there is no single solution, and incomes levels may often not be the most important determinants.

### Plan A: Access to ‘Official’ Household Survey Data

The optimum data source will be the raw data from an official household (income and expenditure) survey for the year of the SAM.<sup>26</sup> The year of the household survey will often be different to the surveys/censuses of manufacturing, agriculture, services etc., and a labour force survey, so shares/coefficients will be needed. Other household surveys that have been collected for various purposes may be available. Because the purposes for which the surveys were undertaken are very rarely to provide data for a SAM, it is important to carefully review the questionnaires used **before** spending time accessing the data.

---

<sup>26</sup> Official household surveys are often primarily focused on the collection of data used to construct weights for indices, e.g., CPI, RPI. Hence, in some countries, e.g., India in the late 20<sup>th</sup> century, the income side is incomplete.

*Guidelines for Compiling a SAM*

At the risk of repetition, assume that data from a household expenditure survey are available. First, the reported transactions will need to be scaled to match the totals of expenditure on commodities, taxes, savings, etc., by each household.<sup>27</sup> Then coefficients and control totals can be used to derive prior estimates of TVs by each RHG. The starting point will usually be aggregates for each RHG – total value of consumption (shares of C), income taxes, savings, transfers. Then deriving estimates of the shares of total consumption expenditures by each RHG and shares of each RHG’s expenditure on commodities as matrix of prior estimates of consumption expenditure like that in Table 6.1. In this example, see section 2, the presumption was that shares of total household consumption expenditures were first derived and then used to estimates expenditures on commodities.<sup>28</sup> The row controls will come from the Use matrix, and might therefore be deemed reasonably reliable (‘narrow’ error bounds), the values of total consumption by each RHG may be deemed somewhat less reliable (less ‘narrow’ error bounds) and the individual TVs the least reliable.

The same process for deriving estimates for payments of income tax, savings and international transfers by the RHGs can be followed. The matter of non-tax payments by RHG to the government will depend upon whether the non-tax transactions between RHGs and the government have been recorded gross or net: if gross then the same process can be applied to estimate non-tax payments by RHGs to the government.

**Table 8.1 Consumption Estimates by Households**

|                     | <i>h a</i> | <i>h b</i> | <i>h c</i> | <i>total</i> |
|---------------------|------------|------------|------------|--------------|
| <i>c_1</i>          | 2.13       | 0.50       | 18.36      | <b>20.98</b> |
| <i>c_2</i>          | 2.27       | 0.70       | 10.55      | <b>13.51</b> |
| <i>c_3</i>          | 15.82      | 18.69      | 8.14       | <b>42.65</b> |
| <i>c_4</i>          | 6.90       | 12.09      | 3.86       | <b>22.85</b> |
| <b><i>total</i></b> | 27.13      | 31.97      | 40.91      | <b>100.0</b> |

<sup>27</sup> While surveys should be based on random stratified samples that perfectly represent the population this is rarely achieved. So, weights are often available to adjust for sample biases.

<sup>28</sup> An alternative approach would be to derive estimates of the distribution of consumption across the whole submatrix, i.e., the sum of the shares in the submatrix would equal one rather than the sum of the shares in each column equal to one.

*Guidelines for Compiling a SAM*

The most difficult component is the estimation of inter-household transfers, in part because they are usually netted out in the national accounts<sup>29</sup> and in part because the data are often less reliable. If the raw survey data are available, or the published summaries report inter-household transfers, then these can be used to derive initial estimates using the shares method: when doing so it is important to note the limited information for control totals. Plan B for these transfers is addressed after considering the income side for RHGs.

Most commonly there are 5 sources of income to RHGs. The first is the direct distribution of factor incomes, which will be dominated by returns to RHGs from the sale of labour services plus mixed income. All labour and mixed<sup>30</sup> income should go to the RHGs.<sup>31</sup> The difficulty is with returns to capital, usually after depreciation, because a substantial proportion of the returns to capital are paid to corporations that subsequently distribute these as dividends to the ultimate owners (RHGs). If a NAM has been constructed this should provide an indication of the proportions of operating surplus distributed directly and indirectly; even so strong assumptions may be required because survey data are often vague about the exact source of returns to capital arriving in RHGs directly as opposed to indirectly.

**NB: If a SAM contains the presumption that all returns to capital assets (after taxes, depreciation, etc.) are paid directly to corporations it will always be wrong unless the compiler explicitly justifies the allocation. Returns to capital assets owned by unincorporated businesses, e.g., self-employed, most farmers, etc., will be paid directly to RHGs. It is highly likely that the less developed an economy is the greater the proportion of returns to capital assets that are paid directly to RHGs.**

The estimates of indirect distribution, via corporations, face the same difficulties. The estimates will probably have appreciable error bounds.<sup>32</sup>

The third component, government transfers to RHGs, depends on whether the transfers have been recorded gross or net: if gross then the same process can be applied to estimate non-tax payments by RHGs to the government. If the available data are for net transfers, then

<sup>29</sup> In this context of this section national accounts can be seen to have a single RHG.

<sup>30</sup> It is implicitly assumed here that mixed income is from self-employment as opposed to returns to corporations.

<sup>31</sup> Note that the data on mixed income that were allocated as part the development of the factor accounts still has a role as a control total even after the mixed income account has been eliminated.

<sup>32</sup> Prior estimates are required. Note that zero prior estimates will always lead to zero TVs in the final SAM.

*Guidelines for Compiling a SAM*

a decision is needed about whether they should be recorded on the income or expenditure side. There are arguments for the income side since it is more likely that the net transfers will be positive. The estimates will probably have ‘largish’ error bounds.<sup>33</sup>

The fourth component is inter-household transfers and depends on whether the transfers have been recorded gross or net: if gross then the same process can be applied to estimate non-tax payments by RHGs to the government. If the available data are for net transfers, then a decision is needed about whether they should be recorded on the income or expenditure side. There are arguments for the income side since it is more likely that the net transfers will be positive. The estimates will probably have ‘largish’ error bounds.<sup>34</sup>

The final component is remittances for the rest of the world can be important; economies such as Ethiopia and Kerala (in India) receive large remittances. These should be recorded in the national accounts, so the share approach can be adopted.

Plan B: No Access to ‘Official’ Household Survey Data

Plan B will typically start with a recognition that the number of RHGs will be less than desired and the RHG categories will be less than ideal and will start with a search for alternative data sources. There are no simple solutions to this problem, and it may prove terminal with respect to disaggregating the household accounts.

There will often be summary data published by the statistical agency derived from the underlying survey data; these data can often be used to derive control totals although they are unlikely to be for the desired RHGs.

In many countries there are other surveys of households that collect expenditure and income data for various research activities. These surveys may be by national research institutes or by international collaborations. A substantial data source for developing countries is the World Bank’s Living Standards Measurement Study (LSMS) (<https://www.worldbank.org/en/programs/lms>) for which data are available from the World Bank’s Microdata Library (<https://microdata.worldbank.org/collections/lms>).

---

<sup>33</sup> Prior estimates are required. Note that zero prior estimates will always lead to zero TVs in the final SAM

<sup>34</sup> Prior estimates are required. Note that zero prior estimates will always lead to zero TVs in the final SAM

*Guidelines for Compiling a SAM*

Estimates derived from these other sources may have ‘largish’ error bounds, although the LSMS data will often be relatively reliable.<sup>35</sup>

---

<sup>35</sup> Prior estimates are required. Note that zero prior estimates will always lead to zero TVs in the final SAM

## **9. Disaggregating Other Institutional Accounts**

The other institutional accounts are typically more straightforward than the household accounts, if only because less detail is required.

### Non-Profit Institutions Serving Households (NPISH)

In the SUT the NPISH should be a separate account. If they are, then the expenditure on commodities will be reported and other expenditures and sources of income will be available from the macro-SAM. The most likely issue with incomes will be incomes from RHGs for which there will be a control total but little information on the amounts from each RHG. Some estimates may be available from the household income and expenditure survey.

If the NPISH is not a separate account in the SUT, the expenditure on commodities will, usually, be reported as part of expenditure on commodities by the household account. Ideally this would be separated out.

Since NPISHs may include charities and aid agencies they can be important agents in lesser developed economies.

### Enterprise Accounts

The process of compiling a NAM may segment corporations as financial and non-financial corporations. There are however circumstances where further disaggregating the non-financial corporations account may be important. If the government receives substantial amounts of income from the non-financial corporations account this implies that the government is the beneficial owner of a substantial share of the non-financial corporations, e.g., parastatals and the armed forces in Egypt. It may not be reasonable to assume that the government and private corporations have the same behavioural relationships. In such cases it can be a good idea to have two accounts for non-financial corporations: private and public corporations.

This can also be important because of its implications for transactions by the government account: the government is likely to fund investments in public corporations but may also run a deficit, i.e., the government may be a borrower, in its role in the government account, and a saver, in its role in the public corporations' account.

### *Guidelines for Compiling a SAM*

Incomes to enterprises will come from the ownership of capital and land assets, both domestic and overseas, that accrue to the enterprises as the beneficial owners. The enterprises will typically pay corporation taxes and retain earnings for investment. They also need to pay dividends to the ultimate owners of the enterprises, i.e., other domestic institutions and overseas owners. Typically, they do not purchase commodities.

The difficulties with the enterprises are twofold. First, determining the proportions of net income to capital and land assets paid to the enterprises and RHG. And second, determining the proportions of distributed earnings paid to domestic and overseas institutions. Control totals are available from the macro-SAM but data on the proportionate distributions is usually limited. Nevertheless, prior estimates are required.<sup>36</sup>

### Government Accounts

In some cases, it may be argued that segmenting the government account between central and local government. This can be appropriate if local government has a marked degree of autonomy, but it is usually difficult to define the different spheres of policy determination.

On the expenditure side: the consumption data will invariably come from the Use matrix, with the government's savings/borrowings coming from the macro-SAM along with international transfers. The tricky issue of government transfers to other domestic institutions has been covered in previous sections. On the income side: the main sources of income are taxes (see section 6 above), government owned enterprises and international transfers.

### Investment Account

The investment account, in the format advocated, is concerned with expenditure on commodities for investment purposes with a minor account to record changes in stocks. Investment expenditures are reported in the SUT so total investment expenditure by commodity is straightforward.

But it needs to be recognised that a single column of investment data is only (logically) consistent with the presumption of a single type of capital factor. In essence the investment expenditure column is a production function where commodities are combined to produce

---

<sup>36</sup> Prior estimates are required. Note that zero prior estimates will always lead to zero TVs in the final SAM.

*Guidelines for Compiling a SAM*

new capital goods, which means there is a single price of capital goods, i.e., the capital goods are homogenous. The implicit presumption is clearly restrictive, e.g., investments in livestock and nuclear power plants use the same technology!!

Relatively little effort has been expended in literature on disaggregating the capital factor account. If the capital factor account is disaggregated, e.g., into fixed and variable/mobile capital, then the investment account should be disaggregated to reflect the different costs of production, i.e., prices, of different types of capital.

The estimates may be crude but this needs to be balanced against the information loss associated with the presumption that all capital goods are homogenous.

## **10. Further Disaggregating Activity and Commodity Accounts**

The process developed so far implicitly presumes that the commodity and activity accounts in the SUT are the commodity and activity accounts that will be used for the SAM. If the SUT were aggregated before generating the SAM, the information gained as part of that process can be used to restart the process using the full SUT database. This may seem time consuming, but the smaller SAM can help to provide additional control totals that can guide the development of the tax, factor and institutional accounts; in practice this can be an efficient approach, especially when compiling a first SAM.

In this section it is assumed that the full SUT database has been included in the SAM. The object then is to develop additional commodity and activity accounts; such a process may be necessary if commodities and activities that are the focus of a study have been aggregated in the SUT, e.g., food processing in EUROSTAT data, agriculture in countries where it is an important source of income/employment/etc. The approach suggested adopts the procedure of deriving estimates of shares of control totals to derive prior estimates of transactions.

Disaggregating accounts in published SUT inevitably requires making prior estimates for which the reliability of the estimates may be relatively low, i.e., there may be large error bounds.

### Adding Paired Commodity and Activity Accounts

Assume for instance that the establishments included in an activity that can be defined by reference to a principal commodity, e.g., the sugar industry and commodity are subsumed within the accounts for an 'other food processing'. Starting from the industry perspective the first objective would be to get estimates of the gross output of the sugar industry, the mix of outputs produced by the sugar industry, the total value of intermediate and primary inputs used. The objective being to establish a series of control totals. These data may be found from data in Census/Survey of Manufacturing but may have been suppressed on the grounds of confidentiality. In that case the source of data may be company accounts, since if the data have been suppressed it is likely to have been because there are a limited number of firms producing the product.

### *Guidelines for Compiling a SAM*

Once the control totals have been established it is then necessary to derive estimates of the shares of different intermediate inputs. Census data often present data for selected major input costs, e.g., energy, that can provide additional control totals as might company accounts. Thereafter, without additional data, it will become necessary to make some potentially strong assumptions about the shares of individual intermediate inputs; one common assumption is that the shares of expenditures in the control totals are the same as those for the activity from which the new activity is being disaggregated. Another is to find data from a ‘similar’ economy for which the data on the new industry have not been suppressed.

Then attention needs to turn to the commodity side. Estimates of the output mix at basic process are needed; these should be in a Census/Survey of Manufacturing but may have been suppressed on the grounds of confidentiality. In that case the source of data may be company accounts. Thereafter estimate of imports (*cif*) and exports (*fob*) are need, together with estimates of realised tax revenues, including import duties. This will provide control totals for the demand side, which then requires estimates that split the commodity accounts.

### Adding Unpaired Commodity and Activity Accounts

The SUT format allows the sets of commodity and activity accounts to diverge, i.e., not rigidly follow the principal product classification. Examples of this might be agriculture where the activity accounts are intended to represent different types of farms, e.g., arable and livestock, different land types (hill and lowland), etc., and electricity generation where different technologies can be used to produce electricity, e.g., coal, gas, nuclear, solar, wind, etc. In these cases, the data availability is likely to be more limited and the required assumptions may be strong.

However, there are often detailed databases for agriculture that are collected to inform domestic support policies in developed economies or understand the economic circumstances found in developing country agriculture. These make take the form of surveys and/or censuses that, for developing countries, may be complemented by data in the Income and Expenditure Surveys for households. Income and Expenditure Surveys for households can also often provide detailed data on the demand for commodities that can be used to guide the disaggregation of the commodity accounts.

*Guidelines for Compiling a SAM*

Whatever the data sources the approach is likely to remain the same: seek control totals and derive share estimates.

When the objective is to identify the cost structures associated with producing the same commodity the available data are likely to be less fulsome. Estimates of the outputs of the commodities by different technologies may be available, which can guide segmenting the activity accounts in the Supply table. Because the commodity account is not split on the demand side, the determinants of the wedges between basis and purchaser price valuations are unchanged. Major difficulties can arise when seeking information on the cost structures associated with different technologies. One source of information is recognition that different technologies may use different discrete inputs, e.g., coal fired vv wind powered electricity generation, which can guide evaluation of the logic of the assigned input expenditures. There may be some data in Surveys and Censuses, but it is likely that information will need to be found in the records of companies, regulators and informal agencies.

## References

- Allen, R.I.G., and Lecomber, J.R.C., (1975). 'Some Tests of a Generalised Version of RAS', in Allen R.I.G. and Gosling, W.F., (eds) *Estimating and Projecting Input-Output Coefficients*. London: Input-Output Publishing.
- EC (2003). Handbook on Social Accounting Matrices and Labour Accounts: Population and social conditions 3/2003/E/N 23. (Leadership group SAM) (working Group: <http://forum.europa.eu.int/Members/irc/dsis/employ/home>)
- Eurostat (2008). *Manual of Supply, Use and Input-Output Table*. Methodologies and Working papers. <https://ec.europa.eu/eurostat>
- Lecomber, J.R.C., (1975). 'A Critiques of Methods of Adjusting, Updating and Projecting Matrices', in Allen R.I.G. and Gosling, W.F., (eds) *Estimating and Projecting Input-Output Coefficients*. London: Input-Output Publishing.
- Pyatt, G., (1989). The Method of Apportionment and Accounting Multipliers', *Journal of Policy Modelling*, 11(1), 111-130.
- Pyatt, G. & Round, J.I., (2012). 'Distributional Invariance and the Design of SAMs', *Economic Systems Research*, 24:3, 251-273
- Stephan, F.F., (1942). 'An iterative Method of adjusting Sample Frequency Tables when Marginal Totals are Known', *Annals of Mathematical Statistics*, Vol 13.
- Stone, R. 1976. "The development of economic data systems. Social Accounting for Development Planning with special reference to Sri Lanka." In: *Social Accounting for Development Planning with Special Reference to Sri Lanka* (G. Pyatt et al., eds). Cambridge: Cambridge University Press
- UN, (1999). 'Handbook of Input-Output Table Compilation and Analysis'. *Handbook of National Accounting, Studies in Methods, Series F, No. 74*. United Nations; New York.