SAMEST: A User Guide (v2)

Scott McDonald and Sherman Robinson

This guide is for samest_5

DRAFT: June 2023

Abstract

*Social Accounting Matrices (SAM) can be used to present the data used in any empirical whole economy model from parsimonious macroeconomic models to large scale general equilibrium models. An important difference between SAMs and Supply and Use Tables (SUT) and Input-Output Tables (IOT) is the inclusion as standard of multiple institutional, factor and tax accounts that facilitate the use of SAMs as databases for the exploration of the 'social' and economic dimensions of economic issues. When using SAMs as databases it is important to recognise three features of any SAM. First, the datapoints – transaction values (TV) – are point estimates, i.e., measured with error. Second, the TVs must be complete **and** consistent, i.e., consistency (equality of incomes and expenditure for all accounts) is not adequate. And third, when a SAM is used to calibrate a mathematical economic model the cause of most simulation results that are hard to explain can be found in the data. Estimating SAMs is a subset of the more general problem of estimating matrices of data.*

Inevitably this document is not a completely comprehensive or perfect User Guide to the SAMEST programme. Similarly, the documentation within the programme will not be completely comprehensive or perfect. Ultimately the only comprehensive and perfect description of the SAMEST programme is the GAMS code.

We welcome corrections to the code and documentation and suggestions for possible improvements.

SAMEST: A User Guide

Table of Contents

1. Introduction

This document is a user guide for the cross-entropy method for estimating Social Accounting Matrices. The presumption throughout this document, and embedded in the SAMEST code, is that the user is developing a SAM that is compliant with the System of National Accounts (SNA). Specifically, it is assumed that the SAM will be limited to the SNA production boundary, the price system will be that defined in the SNA, and that the prior SAM is based on a Supply and Use Table (SUT) that is consistent with the SNA. A standard macro-SAM format is assumed: the format reflects data typically found in published national accounts and SUT, e.g., SUT often only identifies total NET indirect taxes paid on each commodity and activity. In the default settings it is assumed that the published point estimates in the SUT and macro-SAM are estimated with error, i.e., that the addition of new information may alter the point estimates in the SUT and macro-SAM. These assumptions are made solely to make this User Guide more transparent and simplify exposition: experienced users will find it relatively straightforward to relax these assumptions.

The presumption for SAMEST that the SAM is estimated in Supply and Use Table (SUT) format, i.e., that the sub matrix of domestic production of commodities (products) by activities (industries)[1][2] is not a diagonal matrix, derives from the SNA. SAMEST can be, and has been, used to estimate SAMs that go beyond the SNA production boundary and adopt different price systems, and can (regrettably!!!), be used to estimate SAMs based on IOT data. Appendix A2 provides notes on how SAMEST can be adapted to estimate SUT and SAMs that go beyond the SNA production boundary.

The estimation method can be used with limited data, but the greater the effort applied in the development of the prior SAM, and supporting data, the more reliable will be the resultant SAM. It is critical to observe that the cross-entropy estimation programme can only produce a consistent SAM: if the prior SAM is not complete, in the sense that nonzero estimates exist for every cell of the prior SAM that represent 'real' transactions in the

---

[1] Activities/industries are often classified using the International Standard Industrial Classification (ISIC) scheme wherein 'firms' are assigned to activities based on the commodity/product they produce with the greatest value, i.e., the principal product of the 'firm'.

[2] It is common practice to distinguish between 'establishments' and 'enterprises' where, crudely 'establishments' can be defined as individual factories etc., and 'enterprises' as groups of 'establishments' operated under centralised control. Crudely, again, activities/industries can be defined as collections of 'enterprises'/'establishments' for whom the principal product is the same.

economy, the database will be distorted. But, if the prior estimate is complete, the cross-entropy programme can be used to direct attention to transactions for which improved prior estimates would be beneficial.

The step-by-step approach advocated presumes that the estimation process begins with a least restrictive set of conditions, i.e., minimal constraints and (perhaps) wide error bounds, and then progressively adds information in the form of extra accounts in the prior SAM, additional control totals and refined error term specifications. By starting with minimal constraints, the approach minimises the 'difficulty' in finding a first unrefined solution. By progressively adding control totals and refining the error specifications the programme can refine the solution. With each set of additional data and constraints an analysis of the results provides increasing amounts of information about the final SAM; this information, especially the marginal values for the solution value of each transaction, indicates the potential benefits available from refining the prior estimate for that transaction.

The User Guide presumes that the interface to GAMS used is GAMS Studio; this will be evident from the selected screen shots. The use of multiple INCLUDE files presumes the user will always create a reference file combined with making use of the search and replace options across multiple files, run the programme with GDX creation and use the Pin View facilities to view more than one file in the editor window.

There are five example databases: country_BTN_sm, country_BTN_mid, and country_BTN_cgemod1, which use recent data for Bhutan that was developed by Arndt Feuerbacher, and country_ZA_mid and country_ZA_big, which uses older data for South Africa that was developed under the leadership of Cecilia Punt. The original data for Bhutan and South Africa were aggregated into final SAMs for each country and then perturbed with a random uniform distribution or random triangular distribution or normal distribution. GAMS codes are provided to compare the SAMs produced with SAMEST against the originals and to generate different prior SAMs and macro-SAMs with different parameters controlling the random distributions.

The rest of this document is organised as follows. The next section is a simple 'Getting Started' guide that provides minimal detail but (perhaps) enough to get started. This is followed by a section (3) that details the data requirements and then a section (4) that outlines

an estimation strategy. The fifth section provides an overview of the organisation of model and directory structure. designed to assist the user with navigating the programme; this is not a substitute for creating and using reference files and the Search utility. The next two sections address technical matters: first the assignment of the range of control totals that the model can use, and second provides information about setting standard errors and support sets.[3] The eighth section gives an overview of the contents of the various include file used by the programme; more patient users may choose to review this section earlier in the process. The final section offers some closing comments. There are two appendices; the first provides a brief overview of the SAMEST programme's history, while the second addresses some operational issues (1) the estimation of SUT in the absence of 'official' SUT and (2) estimating SAMs that go beyond the SNA production boundary.

---

[3]   For details about the standard errors and support sets see the technical document SAM_Estimation_CE.pdf.

2. Getting Started

This User Guide is a long document that provides detailed information about many of the options that can be used to configure SAMEST. Hence it is intended to be a reference manual for the SAMEST programme rather than a detailed instruction manual for estimating a SAM.

This section is based on recognising that many users will want to start estimating a SAM before they have read this guide.

Nearly ALL the configuration of the programme is implemented through the data and settings contained in an Excel data file; so, the first task is to compile the Excel file based on the files provide with the example applications. The selection of the Excel data file used by the programme is determined by several $SETGLOBAL commands. Similarly, a $SETGLOBAL command is used to label output files. This is illustrated by the code extract from samest_5.gms.

```
*~~~~~~ Substitution to control Dataload
*~~ Identify the Excel workbook
$SETGLOBAL samest_data    country_BTN_sm
*~~ Identify the location of the Prior SAM (data_in if the prior is
 in Excel) else a GDX file
$SETGLOBAL p_sam_in       data_in
*~~ *~~ Identify the location of the EXOGENOUS Macro SAM (data_in if
the prior is in Excel) or GDX file
$SETGLOBAL macsam_in      data_in
*~~~~~~ Substitution to Assign labels for project specific outputs
$SETGLOBAL dstats         country_BTN_sm
```

In most instances the prior SAM and the macro-SAM are available from the Excel workbook; the $SETGLOBAL p_sam_in and the $SETGLOBAL macsam_in commands allow for the possibility the prior SAM and macro-SAM are 'supplied' from a GDX file.

Unless changes are made to the programme, users should not need to edit the master programme file and only TWO of the various include files. If you make changes to the code, these should be documented.

Prior SAM

A prior SAM is essential, and it is important that considerable effort is devoted to its derivation. Optimally the prior will be based on an official SUT, although the programme can readily estimate an SUT given prior estimates for the components and control totals, since this should ensure a clear link between the prior SAM and the published national accounts. It is recommended that when first using SAMEST that the prior SAM is limited to the information from the SUT plus aggregates from the national accounts sufficient to fill out the lower right quadrant of the SUT sufficient to produce a simple SAM consistent with capturing the full circular flow.

Prior Control Totals

Possible prior control totals include targets for the row and column totals (in the worksheet Row-Col), selected aggregate control totals from the macro-SAM (in worksheet Macro-SAM) and the other aggregates (in the worksheet Agg_totals). The row and column totals are essential while the other control totals may be optional.

The default settings for SAMEST assume that the row and column control totals are taken from the worksheet Row_Col.

Assigning Error Support Sets

The assignment of error support sets is in the worksheet `err_values`, as is the assignment of 'general' standard errors, for which there is a default set, `alt0`, and two option sets, `alt1` and `alt2`. The default for the error support sets is uninformative priors, i.e., 7, for all controls, i.e., the RowCol, Micro-SAM, transaction values and the aggregate controls.

These represent reasonable starting points for first running SAMEST with a different prior-SAM. A suggested estimation approach starts this setting with then progressively adds information for the standard errors and adjusts the support sets.

Choosing Options

The worksheet 'controls' allows the user to customise 'data' and 'model' options in Excel. The 'data' options relate to the scaling of the data under the maintained assumption that all data are recorded in the same units; the default is manual scaling (`samautoscal = 0`), in

which case `samscalman` needs to be set and its value will be determined by the units used to record transactions. Autoscaling is an option.

The 'model' options relate to the choice available to control how selected operations are implemented. For a first solve it is recommended to use the default settings.

3.   **Data Requirements**

Data requirements for SAMEST can be summarised under four categories: 'Macro SAM', 'Prior SAM', 'Other Aggregate Control Totals' and 'Standard Errors and Support Sets'. The first three require substantial effort BEFORE using the SAMEST programme. While compiling these data the user should also record estimates about the reliability of the point estimates that will inform the calibration of standard errors and guide decisions about the support sets; this information may be in the form of classical statistics, e.g., variances, or subjective judgements. However, the process would normally be iterative with the output from SAMEST used to identify datapoints in the macro-SAM and prior SAM that justify re-estimation while reviewing estimates about the reliability of the point estimates.

The example applications of SAMEST assume that the prior SAM is based on a Supply and Use Table for a year 'close' to the year chosen for the SAM and that the macro-SAM is based on national accounts data for the year chosen for the SAM.

Prior Macro-SAM

The 'default' structure of the macro-SAM used in SAMEST reflects typical data from national accounts and SUT, see Table 3.1. Assuming the SUT is for the year chosen for the SAM, the cells highlighted (in yellow) will be directly available from the SUT. The scalar estimates for the other cells in the macro-SAM should be available from the national accounts; all the cells for which there are descriptions should include estimates unless there is strong evidence that no such transactions have taken place. The cells in Table 3.1. with a '0' entry should be zero based on economic logic. The cells with ?? entries are more problematic: economic logic, together with certain case specific assumptions, indicates that some of these cells may require entries. For instance, the cell commodities:enterprises may include entries if the user decides to allocate NPISH and/or NGOs in a wider classification of enterprises, while the cell savings:activities may record depreciation estimates by activities if they are known.

The default macro-SAM assumes that estimates are known for all cells in Table 3.1 that have descriptions and are not, by accounting definitions, net zero.

Table 3.1   Structure of macro-SAM for SAMEST

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

The process of compiling a prior macro-SAM may be important even if the prior macro-SAM is not used to provide control totals for SAMEST.

A SAM must be complete AND consistent. SAMEST, and all other mathematical estimation and balancing routines, can ensure consistency but they cannot ensure completeness. While compiling a prior macro-SAM it is critically important to identify ALL cells within the final SAM that must include non-zero estimates if the SAM is to be complete even if the estimated net transactions in the macro-SAM are zero. The cells in the macro-SAM highlighted in GREEN will be net zero, e.g., if the government accounts include central and local government, income to local from central government will be identical to expenditure by central to local government and therefore not impact on the values of other transactions in respective rows and columns. But inter institutional transactions are normal/common so estimates of the gross values of such transactions are important control totals when estimating inter institutional transactions.

This leads to an important, but often neglected, conclusion. If economic logic and/or evidence indicates that a transaction has taken place then a non-zero prior estimate is essential for completeness, even if there is minimal empirical evidence as to the magnitude of a transaction. An unreliable and erroneous estimate of a real transaction is better than a zero, since a zero estimate will be unambiguously wrong..

The prior macro-SAM is accessed from Excel, in the worksheet 'Macro-SAM', and does not require changes to the code unless the set of additional controls totals is extended. The option to access the macro-SAM from a GDX file is provided.

Prior SAM

The structure for the prior and final SAM must follow that for the macro-SAM except that the broad categories of accounts, e.g., commodities and households, will be disaggregated. The sum of the transactions within sub-matrices of the final (post solution) SAM corresponding to scalars in the macro-SAM will equal the values in the final (post solution) macro-SAM. Decisions on the detailed account structure will always be case specific. In most cases the base commodity and activity accounts (aggregates) will be those in a SUT, while the other accounts will reflect the institutional structure of the economy, e.g., tax instruments, factor,

and household types. The choices should be informed by knowledge of the economy in question, the issues to be addressed using the SAM and economic logic.

If the SAM is to be used for applied policy analysis it is important to recognise that the account structure should reflect agents with 'common' behavioural characteristics, e.g., male and female 'headed' households may be expected to have different utility functions. It is also important to maintain 'balance' across the disaggregation of different groups of accounts, e.g., a single factor account for labour will provide very limited information about the impacts of labour markets for different household groups (see Pyatt and Round, 2012).

The accuracy/reliability of point estimates in a prior SAM cannot be over emphasised. The SAMEST software can provide information about datapoints in the prior SAM where greater 'accuracy' would improve the statistical performance of SAMEST, but the more detailed and accurate the prior SAM the more accurate, *ceteris paribus*, will be the estimated SAM. But the SAMEST code can never render a complete final SAM if prior estimates of SAM entries are missing; SAMEST will render a such a SAM consistent but ONLY at the expense of unquantifiable distortions in the estimated transactions. This means that inaccurate prior estimates for 'real' world transactions are preferable to decisions to record zero prior point estimates; then SAMEST may produce a complete and consistent SAM.

The prior SAM is accessed from Excel, in the worksheet 'Prior-SAM', and does not require changes to the code. An option to access the prior SAM from a GDX file is provided.

Other Aggregate Control Totals

The prior macro-SAM, if used, contains estimates for a subset of economic aggregates. The default specification for the prior macro-SAM can be easily augmented with estimates of other aggregate control totals, e.g., GDP, incomes to government from different tax instruments, etc. The choice of appropriate additional aggregate control totals is case specific, depending in part on the availability of data and economic structure. This is especially the case for tax instruments where separating trade taxes from domestic commodity taxes is important as is separating tax instruments by incidence is critical, e.g., separation *ad valorem* VAT taxes from quantity excise taxes.

The SAMEST programme contains generic code that allows for the inclusion of a wide range of additional aggregate control totals. A selection of standard additional control totals is coded in the distributed version of SAMEST, but users should expect to extend the set of control totals applied.

It may not be possible, initially, to identify all the relevant control totals. However, the SAMEST code readily allows for the addition of control totals and thence re-estimation of the SAM.

The control totals are accessed from Excel, in the worksheet 'Agg_totals', and does not require changes to the code unless the set of additional control totals is extended.

Standard Errors and Support Sets

Information about the reliability of point estimates in a prior SAM, or prior macro-SAM, may scarce but is necessary. For instance, the Stone-Byron method for estimating a SAM requires estimates of the variance to be attached to each point estimate; acquiring such data is very difficult. Without such information the estimation process will have limited accuracy.

The SAMEST programme allows users to include known and reliable estimates of variances for each point estimate if they are known. It also allows the user to apply subjective estimates of reliability of each point estimate in the form of standard errors; such subjective or empirical estimates are important. The SAMEST programme also allows the user to define the distribution of the error terms through the selection the support sets. The standard errors and support sets are selected and controlled from Excel unless the user chooses to modify the codes.

The Standard Errors and Support Sets are accessed from Excel, in the worksheet 'err_values'. For the standard errors, *stderrs*, the *alt0* values are defaults that are hard coded and therefore only included in the worksheet, 'err_values', for information. For the support sets, *esupp*, the recommended number of elements is 7. Changes in the standard errors and support sets should be based on evidence.

4.   An Estimation Strategy

Developments in computing power and solver algorithms mean that SAM estimation problems that were large and/or ill-conditioned and hence difficult to solve in the past are now considered small and easy to solve. This combined with a tendency to assume that a bigger SAM is necessarily a better SAM, seems to encourage a process whereby a large prior SAM is developed and then mechanically rendered consistent using some form of estimation/balancing software.

But if the production of ever larger prior SAMs means that the point estimates for transactions become less reliable a larger SAM may not be a better SAM. The estimation strategy outlined here was not devised with the intention of producing bigger SAMs, rather it is intended to assist in producing more reliable SAMs.

The principle adopted is of sequential disaggregation and estimation where after each round of estimation the prior SAM is reviewed and revised because as information is added to the prior SAM, so the information content of the other point estimates changes. The sequence advocated, assuming SUT are available for the year of the SAM, is

1. SAM as an extended SUT;

2. SAM with extended commodity and activity accounts;

3. SAM with disaggregated indirect tax accounts;

4. SAM with disaggregated factor accounts;

5. SAM with disaggregated household accounts;

6. SAM with disaggregated incorporated business enterprise accounts (optional);

7. SAM with disaggregated government accounts (optional);

8. SAM with disaggregated investment (and savings) accounts (optional).

With each step the prior SAM is reviewed and revised as necessary, but the point estimates are still treated as measured with error although the error bounds may change. Strategies for SAM estimation where there is no SUT, or the SUT is for a different year, are outlined in an appendix.

Step 1:

If SUT have been compiled and used, as intended in the SNA, to benchmark the national accounts, then the data are likely to be moderately reliable and therefore probably represent a best point of departure. Moreover, if the national accounts have been fully reconciled the aggregate inter-institutional transactions in the lower right quadrant of a SAM should produce a complete and consistent SAM with aggregated indirect tax, factor, household, enterprise, and savings-investment accounts.

There are several reasons the SAM may be inconsistent and measured with error, although it should be complete. First, the year chosen for the SAM may not the year for which the SUT was compiled; reasons for this may include that the SUT was compiled for an atypical year. Second, the SUT for the chosen year may have been 'statistically' updated rather than estimated from primary data. Third, revisions to the national accounts may have been substantial and cast doubt on the data and totals in the SUT. Fourth, data for the Supply table of the SUT may have been withheld. And fifth, only a Use table may have been produced.

If the prior SAM is not consistent, then the SAM as an extended SUT should be estimated using SAMEST. Note that if the SAM is inconsistent then, by definition, the point estimates in the prior SAM have been measured with error. If no additional data are available, the estimated SAM would be the final stage. But if additional data are available extending the prior SAM changes the information content of the point estimates in the original prior SAM and therefore the estimated SAM should still be regarded as measured with error.

If a SUT is available but for a different year to the chosen year the prior SAM for step 1 will be inconsistent and may be incomplete, especially if the time gap is substantial. In these cases, the process essentially involves deriving a prior SUT using activity controls, e.g., gross output, compensation of employees, etc., from the national accounts, estimates of trade values and domestic final demand, using coefficients from the prior SUT. Inevitably the error bounds will need to be larger than when an official SUT is available for the chosen year.

If an SUT is unavailable, then an SUT must be estimated. For guidance see appendix A2.

Step 2:

If the commodity and activity accounts in the SUT are NOT adequate for the purposes for which the SAM will be generated, the next step is to disaggregate selected commodity and activity accounts. This is often the case for agriculture, which may not be a concern in developed economies where agriculture accounts for a small share of GDP and employment, but for less developed economies a single agriculture account may be inadequate. A new prior SAM is required using data for the commodities and activities to be disaggregated and extra control totals derived from the previously reconciled SAM, noting that these control totals will be measured with error. A new SAM can then be estimated using SAMEST; this SAM should be reviewed with errors corrected by adjusting the prior SAM and re-estimating.

Step 3:

Steps 3 to 8 should adopt the same strategy: collect additional data, generate a new prior SAM, re-estimate the SAM, review the SAM, re-estimate as necessary.

Details about the transaction values of indirect tax instruments are important determinants of the price system in all SAMs and are especially important if a SAM is to be used in a price driven model, e.g., a CGE model. The price system in the SNA means that the indirect tax data do not require disaggregation of the factor and institutional accounts, so this can be a convenient stage to address the issues. It is important to separate taxes from subsidies and distinguish between taxes based on the agents who pay each tax instrument, e g., VAT and excise taxes, and the basis on which a tax is levied, e.g., *ad valorem* or quantities. The tax data should report the transaction values, i.e., the revenues raised by each tax instrument, exclusive of any economic rents; for instance, estimates of tariff equivalent values should distinguish between the tax revenue component and economic rent component with recipients of economic rents identified. Detailed data on tax revenues by commodity, and activity, can be difficult to acquire, but their importance is difficult to overstate.

It is unlikely that data can be obtained for all applied tax instruments. It is appropriate to include tax accounts that encompass all tax revenues not explicitly accounted for so as to preserve the information content associated with identified tax instruments. This can be achieved by including 'catch-all', e.g., other commodity and activity taxes.

Step 4:

Data for prior disaggregation of the factor (labour) accounts often comes from Labour Force Surveys or extended Household surveys. Rarely are SUT, Labour Force or Household surveys produced for the same year due to capacity constraints; hence the prior SAMs will be measured with error and be inconsistent.

A SAM with many labour accounts but one representative household group (RHG) will provide no detail about income distribution. It is recommended that numbers of factor/labour and RHG accounts follow the guidelines produced by Pyatt and Round (2012).

Step 5:

Data for prior disaggregation of the household accounts typically come from Household Income and Expenditure Surveys. Rarely are SUT and Household surveys produced for the same year due to capacity constraints; hence the prior SAMs will be measured with error and be inconsistent.

A SAM with many RHG accounts but one labour account will provide no detail about the functional distribution of income. It is recommended that numbers of factor/labour and RHG accounts follow the guidelines produced by Pyatt and Round (2012).

Step 6:

Steps 6 to 8 can be considered optional. Splitting the enterprise account can be important if the government is the beneficial owner of a substantial share of enterprises, and/or the enterprise accounts are used to account for NPISH and/or HGOs.

Step 7:

Segmenting the government account between, say, central and local government may be important especially in countries with federal structures or other forms of devolved decision making.

Step 8:

Additional information about investment accounts can be important, especially for dynamic models; ideally this may take the form of an activity specific investment matrix that identifies gross fixed capital formation by investing industries, or, alternatively, an investment matrix

that identifies gross fixed capital formation, i.e., the commodity composition, for each type of capital good.

5.   Programme Organisation

The programme is distributed as a single ZIP archive that includes all the programme files together with selected illustrative examples. The default settings for the illustrative examples assume a least restrictive set of conditions.

The ZIP archive should be extracted to a single directory.
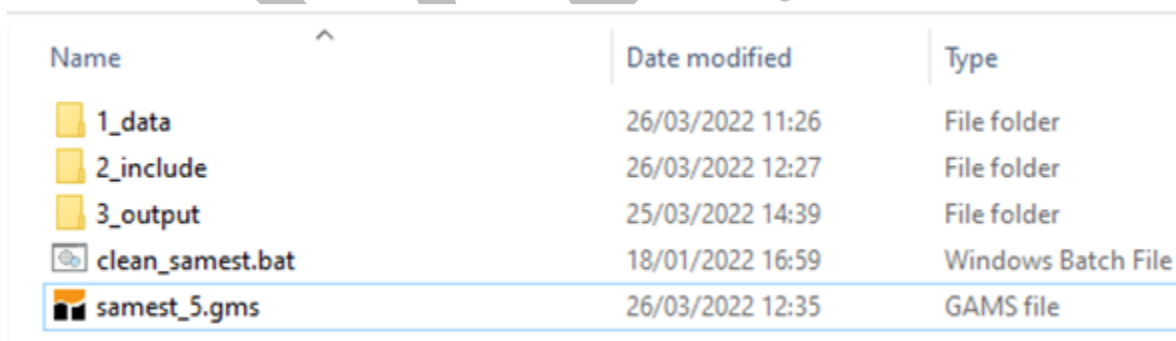
<u>Directory Structure</u>

The directory structure embedded in the Zip archive contains three sub-directories, the `samest5.gms` programme and a batch file[4] that can be used to tidy-up the directories by deleting 'temporary' working files (see Figure 5.1).

1_data

This sub-directory contains one or more Excel files each of which contains the data for one SAM estimation project. It also contains the file `samest_dataload.inc` that converts the Excel data into GDX format and controls the loading of data into the programme.

Figure 5.1   Directory Structure – SAMEST Directory

Name	Date modified	Type
1_data	26/03/2022 11:26	File folder
2_include	26/03/2022 12:27	File folder
3_output	25/03/2022 14:39	File folder
clean_samest.bat	18/01/2022 16:59	Windows Batch File
samest_5.gms	26/03/2022 12:35	GAMS file

2_include

This sub-directory contains the various files called by the programme. The files are organised so that each contains a series of related instructions. Most of these files should not require editing but as explained below, the files `samest_msam_endog.inc` and `samest_aggtotal.inc` may require editing to include project specific changes. If it is

---

4   In the ZIP archive the batch file is labelled as `clean_samest.dat`; the dat suffix needs changing to bat.

necessary to edit these files it is suggested that they should be renamed, e.g.,

`samest_mactotal_**.inc`, so that a record of adjustments is maintained.

3_output

This sub-directory will contain the results from estimation processes – the programme
provides a method for ensuring unique labels for each projects output. In the ZIP archive this
directory contains a single file – `output.dat` – that serves to maintain the directory if it is
otherwise empty when it is zipped.

Figure 5.2   SAMEST in GAMS Studio
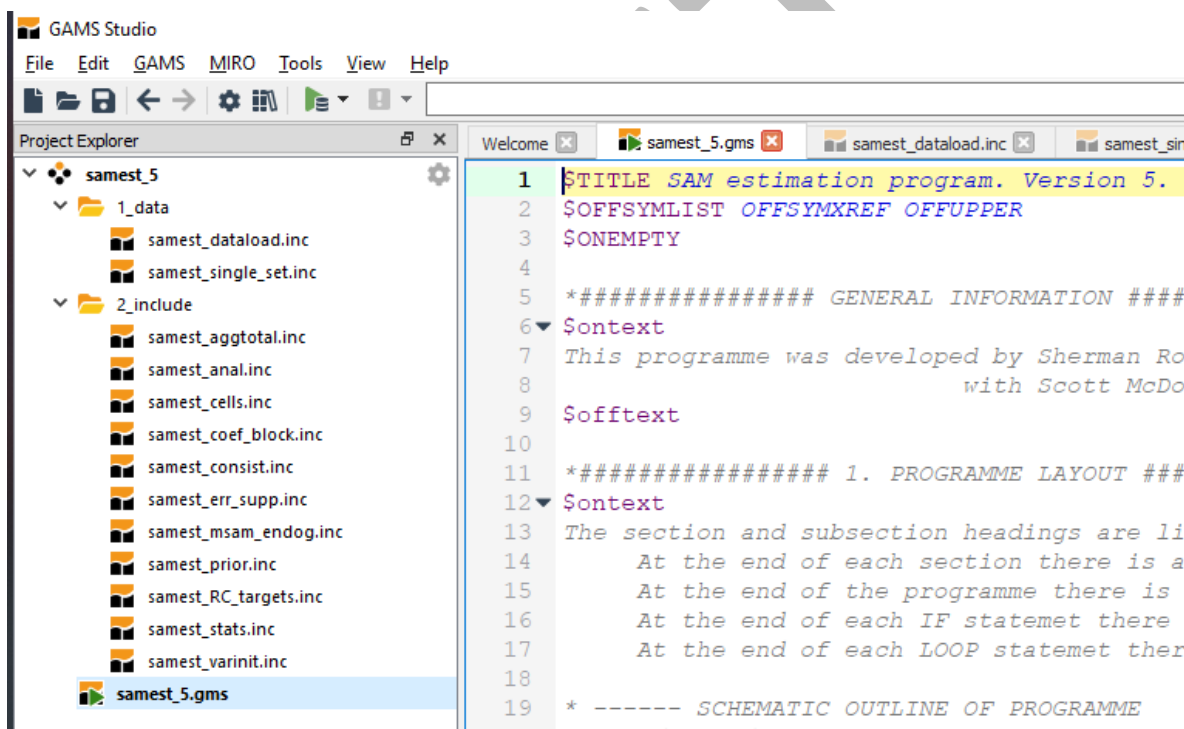
Programme in GAMS Studio

Figure 5.2 illustrates the programme in GAMS Studio with all the files visible. For most
applications it would not be usual to open all the files. However, when first exploring the
programme, it can be helpful to open all the files, as in Figure 5.2, in part because the
search/find function can be set to search all files in the project. It is also recommended that
when running the programme that the `Run/Compile with GDX Creation`
(`F10/Shift+F10` respectively) options are chosen AND that a reference file is generated.

Figure 5.3 Excel Database – Layout Worksheet (standard)

	Data Type	Name	Location	Row dimension rdim	Column dimension cdim	Total dimension dim
1	LAYOUT					
2						
3						
4						
5						
6	dset	sac	sets!A6	1		
7	dset	ss	sets!D6	1		
8	dset	c	sets!G6	1		
9	dset	m	sets!H6	1		
10	dset	a	sets!I6	1		
11	dset	f	sets!J6	1		
12	dset	l	sets!K6	1		
13	dset	k	sets!L6	1		
14	dset	lnd	sets!M6	1		
15	dset	h	sets!N6	1		
16	dset	e	sets!O6	1		
17	dset	g	sets!P6	1		
18	dset	gtind	sets!Q6	1		
19	dset	gtdir	sets!R6	1		
20	dset	i	sets!S6	1		
21	dset	w	sets!T6	1		
22	dset	emrgn	sets!W6	1		
23	dset	etax	sets!X6	1		
24	dset	d_cons	controls!A8	1		
25	dset	m_cons	controls!A29	1		
26	par	m_controls	controls!A28	1	1	
27	par	d_controls	controls!A7	1	1	
28	par	P_SAM	Prior-SAM!A4	1	1	
29	dset	macro	macro!A6	1		
30	dset	macro2	macro!D6	1		
31	dset	mt	macro!e5		1	
32	par	macromat	macro!d5	1	1	
33	par	macsam_exog	macro-sam!a6	1	1	
34	par	macscale	macro-sam!B5			
35	dset	sig	err_values!A7	1		
36	dset	alt	err_values!B6		1	
37	dset	esuppset	err_values!A13	1		
38	dset	nsam	SAM-elements!C5		1	
39	par	stderrs	err_values!A6	1	1	
40	par	esupp	err_values!A15	1		
41	par	stderrs_MSAM	macro-sam!A34	1	1	
42	par	rowcol_tot	row-col!A8	1	1	
43	par	cmac	sam-blocks!A7	1	1	
44	par	sigmac	sam-blocks!A31	1	1	
45	par	samopt_tv	SAM-elements!B5	2	1	
46						

Database Structure

The database contains details about the sets and data plus several parameters that control the operation of the programme. The information in this file is case specific, except the Layout sheet. The various worksheets are explained below.

Sets Worksheet

The set *sac*, and its subsets *c, m, a, f, l, k, lnd, h, e, g, gtind, gtdire, i*, and *w*, require specifying according to the dimensions of the prior SAM. Macro-SAM accounts, *ss*, can be left as they are or modified according to the dimensions of the macro-SAM. NB: multiple different tax

instruments, etc., can be included in the macro-SAM but the programme includes a facility to control aggregates additional to those in the macro-SAM. It is recommended that controls for specific tax instruments are NOY included in the macro-SAM.

If the sets in the macro-SAM are changed, it is necessary to modify the set map_ss_sac in the file samest_msam_endog.inc.

Table 5.1   **Sets Worksheet – Bhutan Small (BTN1)**

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Model Specific Sets												
2		Macro-sam account names					cmdty	mrgn	actvty	fctr	lbr	kptl	land
3	Global Set												
4	GAMS Name	Description		Macro-sam accounts	Description		Commodities	Margins	Activities	Factors	Labour Factors	Capital Factors	land factor
5	sac			ss			c	m	a	f	l	k	lnd
6	cagric	Agriculture		cmdty	Commodities		cagric	marg	aagric	flab	flab	fcap	flnd
7	cmine	Mining		mrgn	Domestic margins		cmine		amine	fcap			
8	cfood	Food processing		actvty	Activities		cfood		afood	flnd			
9	clmanu	Light Manufacturing		fmix	Mixed Income		clmanu		almanu				
10	chmanu	Heavy Manufacturing		lbr	Labour		chmanu		ahmanu				
11	celect	Electricity		kptl	Capital		celect		aelect				
12	ccons	Consntruction		land	Land		ccons		acons				
13	ctrade	Trade and hospitality		hhld	Households		ctrade		atrade				
14	coserv	Other Services		entrp	Enterprises		coserv		aoserv				
15	marg	Margins		govtn	Government								
16	aagric	Agriculture		itax	Indirect taxes								
17	amine	Mining		dtax	Direct taxes								
18	afood	Food processing		invst	Invest_Saving								
19	almanu	Light Manufacturing		wrld	Rest of the World								
20	ahmanu	Heavy Manufacturing		m_sam_tot	Account totals								
21	aelect	Electricity											
22	acons	Consntruction											

Err_values Worksheet

This worksheet provides data for the standard errors, sig_**, and the error support sets, esupp_**. The standard errors for 'alt0' are fixed, while those for 'alt1' and 'alt2' can be user defined. The default values for the error support sets are 7, see also conerrsupp in the controls worksheet; this is the uninformative prior setting.

Table 5.2   Err_values Worksheet

	A	B	C	D	E	F	G	H	I	J	K
1	Default Values for Errors										
2											
3	These parameters are used to set prior standard errors - sigmay_RC, sigmay_MAC, and sigmay_TV										
4	Default prior standard errors for various SAM components. Choice is determined by value of consigm.										
5											
6	stderrs	Alt0 values are default and cannot be changed. Alt1 and alt2 values are set here.									
7		alt0	alt1	alt2							
8	sig_RC	0.25	0.15	0.1		sac	row/col sums				
9	sig_AGG	0.05	0.05	0.03			macro aggregates				
10	sig_TV	0.25	0.15	0.1		sac, sacp	cell values or coefficients				
11	sig_MSAM	0.05	0.05	0.03		ss	macro SAM aggregates				
12											
13											
14	Use these error support set numbers. DEFAULT values are 7. If NOT assigned 3 or 5 or 7 the programme will ABORT with an error message.										
15	Esupp	No. elements		Description							
16	esupp_RC	7		Errors on column/row sums							
17	esupp_AGG	7		Errors on aggregate control totals							
18	esupp_TV	7		Errors on cell values/coefficients							
19	esupp_MSAM	7		Errors on standard macro-SAM aggregates							
20											
21											

Controls Worksheet

This worksheet allows the user to set various controls within the programme without having to find where the controls are implemented.

The Data controls, `d_control`, are concerned with scaling the data to improve algorithm performance. The default is manual scaling, i.e., `samautoscal` set to 0.0 and `samscal` set the scaling factor, e.g., 100. For autoscaling, `samautoscal` is set to 1, with the default being that 95%, `scalprop = 0.95`, of the values are less than 100, `scaltarg = 100`, and `samscal` set to a low value, e.g., 0.001. In many cases this is a good starting point, but with difficult to solve problems adjusting `scalprop` and `scaltarg` can help the programme. If needing to adjust the settings, it is often best to adjust `scalprop` between 0.9 and 0.975; it is not a science.

With large and difficult to solve problems adjusting the scaling factors can greatly reduce the time taken to find a solution and even to find a solution.

The model controls, m_control, are where settings in the model are selected. In most cases it is useful to start with the default settings and then adjust the settings as the estimation process is refined.

Table 5.3 Controls Worksheet

		base		Descriptions
1	Controls for Cross Entropy Code			
2				
3	These values are used to initialise various parameters that control the operation of the Cross-Entropy programme			
4				
5	D(ata)_Controls			
6	d_cons			
7		base		Descriptions
8	cutoff0	0		maximum value for small absolute values after scaling (programme aborts if GT 1.00
9	samscal	1000		SAM scaling factor and initial value for autoscaling
10	scaltarg	100		target level for auto scaling
11	scalprop	0.95		proportion of NON zero elements that must be below the target level
12	samautoscal	0		0 - manual scaling; 1 - autoscaling,
13	samscalman	1000		SAM scaling factor for manual scaling
14				
15				
16				
17				
18				
19				
20	M(odel)_Controls			
21	m_cons			
22		base		
23	con_target	3		1 = mean Row and Col; 2 = Col totals ; 3 = Row total
24	con_sigm	0		0 - default standard errrors; 1 - alt1 values; 2 - alt2 values
25	con_opt_blocks	0		0 - default; 1 - specify special treatment for blocks of the SAM
26	con_opt_agg	0		0 - default; 1 - specify special treatment for aggregate constraints
27	con_opt_RC_exog	0		0 - default - specify rows and columns from Row-Col worksheet; 1 - totals from P_SAM;
28	con_opt_tv	0		0 - default; 1 - specify special treatment for individual elements of SAM
29	con_opt_msam	0		0 - default - values from exogenous Macro-SAM; 1 - Macro-SAM controls calculated from P_SAM
30	con_macsam	0		0 - use macsam_exog; 1 - use macsam_endog
31				

Table 5.4 Model Controls

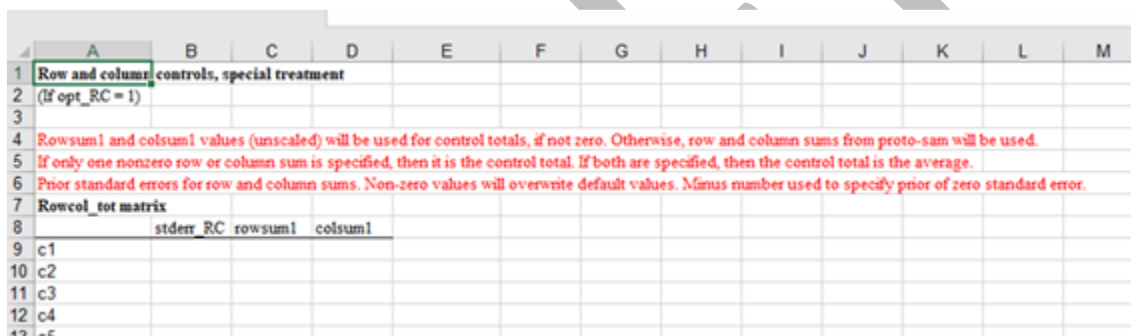| Control	Default	Explanation
con_target	1	1 = mean Row and Col; 2 = Col totals ; 3 = Row total
con_sigm	0	0 - default standard errors; 1 - alt1 values; 2 - alt2 values
con_opt_blocks	0	0 - default; 1 - specify special treatment for blocks of the SAM
con_opt_agg	0	0 - default; 1 - specify special treatment for aggregates
con_opt_RC_exog	0	0 - default - specify rows and columns from Row-Col worksheet; 1 - totals from P_SAM;
con_opt_tv	0	0 - default; 1 - specify special treatment for individual elements of SAM
con_opt_msam	0	0 - default - values from exogenous Macro-SAM; 1 - Macro-SAM controls calculated from P_SAM
con_macsam	0	0 - use macsam_exog; 1 - use macsam_endog

Row-Col

This worksheet provides the option to provide the row and column controls independently from the prior SAM. This can be good practice if the row and column totals in the prior SAM differ widely, e.g., demand for tobacco and alcohol vv supply of tobacco and alcohol.

The default standard errors, from the `err_values` worksheet, can be overwritten by setting the values in the column `stderr_RC`.

The use of exogenous estimates for row and column total requires `opt_RC = 0`. The use of endogenous estimates for row and column totals, i.e., from `P_SAM`, requires `opt_RC = 1`.

Table 5.5   Row-Col Worksheet

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Row and column controls, special treatment												
2	(If opt_RC = 1)												
3													
4	Rowsum1 and colsum1 values (unscaled) will be used for control totals, if not zero. Otherwise, row and column sums from proto-sam will be used.												
5	If only one nonzero row or column sum is specified, then it is the control total. If both are specified, then the control total is the average.												
6	Prior standard errors for row and column sums. Non-zero values will overwrite default values. Minus number used to specify prior of zero standard error.												
7	Rowcol_tot matrix												
8		stderr_RC	rowsum1	colsum1									
9	c1												
10	c2												
11	c3												
12	c4												
13	c5												

Macro-SAM

This worksheet contains a prior macro-SAM. Ideally the macro-SAM will be complete and consistent although the estimates may be measured with error; it is not required that the macro-SAM is complete and consistent. The units in the macro-SAM, `macsam_exog`, should be same as those in the prior SAM, `P_SAM`, and this is the default assumption, i.e., `macscale = 0`, for the automatic scaling

The standard errors, `stderrs_msam`, are the default values from the worksheet 'err-values' or those set in this worksheet. If the macro-SAM is complete and consistent it is reasonable to expect that the standard errors will be smaller than if the macro-SAM is not complete and consistent

Table 5.6   Macro-SAM Worksheet

Other Aggregate Controls

This worksheet defines a series of aggregates that the user can define. Column A defines the set, `aggr`, of available aggregates that described in column B; the elements of `aggr` used, `aggr2`, are assigned by entering a 1 in column E (`aggr_inc`).

The target value is set in column F – it must be in the same units as the `P_SAM` since it is scaled with the same scaling factor (`samscal`). The special standard errors, `stderr_mac`, are set in column G, otherwise default standard errors are applied.

The list of available other aggregates can be readily extended. Three steps are required.

1. Add labels for the additional aggregates to column A of the worksheet AGG_totals and repeat in column D.

2. Add details for the target, column F, the standard errors, column G (note including the target requires a 1 in column E)

3. In the file `samest_aggtotal.inc` extend the assignments for the parameter `aggagg` to encompass the new aggregate control.

The selection of other aggregate controls will likely be case specific. But care needs to be taken with these controls to ensure that the conventions for sets etc., used in `samest_aggtotal.inc` remain consistent with those elsewhere in the programme.

Table 5.7   **AGG_totals Worksheet**

	A	B	C	D	E	F	G
1	AGGREGATE control totals			Aggregate controls actually used and std errors (if opt_A			
2							
3				If value or standard error are not specified, defaults are			
4	Potential aggregate controls			Negative entry for std-err is interpreted as specifiying ze			
5							
6	GAMS Name	Description		AGGMAT			
7	aggr			aggr	aggr_inc	value_agg	stderr_agg
8	t2_gdp_exp	Total GDP expenditure at purchaser prices		t2_gdp_exp	1	179341.3	0.05
9	t2_gdp_fc	Total GDP at factor cost		t2_gdp_fc			
10	t2_p_con	Total private consumption		t2_p_con			
11	t2_g_con	Total government consumption		t2_g_con			
12	t2_i_con	Total investment consumption		t2_i_con			
13	t2_x_con_fob	Total exports fob		t2_x_con_fob			
14	t2_m_con_cif	Total imports cif		t2_m_con_cif			
15	t2_vat	Total VAT tax revenues		t2_vat			
16	t2_excise	Total excise tax revenues		t2_excise			
17	t2_tariff	Total tariff tax revenues		t2_tariff			
18	t2_export	Total export tax revenues		t2_export			
19	t2_sales	Total other commodity tax revenues		t2_sales			
20	t2_c_sub	Total commodity subsidy expenditures		t2_c_sub			
21	t2_comtax	Total commmodity taxes		t2_comtax			
22	t2_prod	Total taxes on production		t2_prod			
23	t2_a_sub	Total activity subsidy expenditure		t2_a_sub			
24	t2_acttax	Total activity taxes and subsdies		t2_acttax			
25	t2_ent_dtax	Total direct taxes on enterprises		t2_ent_dtax			
26	t2_hh_dtax	Total direct taxes on households		t2_hh_dtax			
27	t2_yf	Total factor income		t2_yf			
28	t2_yf_cap	Total factor income to capital		t2_yf_cap			
29	t2_yf_lab	Total factor income to labour		t2_yf_lab			
30	t2_yf_employ	Total factor income to employees		t2_yf_employ			
31	t2_yf_land	Total factor income to land		t2_yf_land			
32	t2_yh	Total household income		t2_yh			
33	t2_ye	Total enterprise income		t2_ye			
34	t2_yh_fact	Total household income from factors		t2_yh_fact			

| Layout | Sets | err_values | controls | Prior-SAM | Row-Col | Macro-SAM | **Agg_totals** | SAM-elements | SAM-blocks |

SAM-blocks

This worksheet identifies blocks of the SAM that should be accorded special treatment and the associated standard errors. Blocks are defined by reference to the macro-SAM structure.

Table 5.8   **SAM-blocks Worksheet**

SAM-elements

This worksheet allows for setting specific constraints on individual cells in the SAM. The cells are identified by the row:column references; respectively by columns B and C. Columns D, E and F of the worksheet are used to define the nature of the error, `code_tv`, in terms of value/coefficient and additive/multiplicative error, the target value, `value_tv`, and the standard error, `stderr_tv`.

The option to provide transaction specific controls is probably under used. It should be used wherever there are transactions that are of major 'importance' and/or about which there is good information. In addition, it should be used if the other settings result in changes to prior point estimates that are unreasonably large (or small).

Table 5.9   **SAM-elements Worksheet**

	A	B	C	D	E	F
1	SAMOPT TV(sac,sacp,nsam3)					

2	code_tv: 0 - value and additive error
	            1 - coefficient and multiplicative error;
	            2 - value and multiplicative error
	value_tv: 0 - proto-sam value is used
	            ne 0 - value is used. Unscaled, so units must match
	             the proto-sam.
	stderr_tv: 0 - default standard error is used
	            gt 0 - value is used for standard error
	            lt 0 - zero standard error is specified					

3						
4		row	column			
5	Optional description of cell	sac	sacp	code_tv	value_tv	stderr_tv
6						
7						
8						
9						
10						

6. Control Totals

Control totals are provided as 4 parameters from Excel via GDX in 4 worksheets. They could be provided directly from GDX, but these options are not coded: it is judged that this is not appropriate because all control totals should be subject to scrutiny. All control totals are assumed to be estimated with error unless the user determines they are not measured with error.

The programme assumes that the prior-SAM and ALL control totals are recorded in the same units, e.g., thousands or millions or billions, and the same currency units. The same scaling factor, samscal, is applied to all data in the programme.

Row and Column Controls

The row and column control totals are assigned in two columns in the worksheet 'Row-Col': the code tests for large differences between the row and column from the P_SAM and the worksheet controls and reports an error if the differences are excessive – see `rowcolchk`. The third column allows the user to assign specific standard errors to row and column controls: non-zero values overwrite the default standard errors while a minus number sets the associated standard error to zero.

If row and column control totals for commodities and activities are derived from the SUT divergence of the controls from those calculated from an extended `P_SAM`, i.e., a `P_SAM` with multiple representative household groups (RHG), factors and indirect taxes, may be substantial.

Macro-SAM Controls

The macro-SAM control totals (`MSAM(ss,ss)`) are assigned in the worksheet 'Macro-SAM'. The dimensions in the example are based on the typical account structure found in SUT augmented with aggregates often reported in national accounts to produce a minimal macro-SAM consistent with standard SUT. Note the following.

1. Commodity/Product (cmdty) and Activity/industry (actvty) accounts – as defined in the SUT

2. Margins (mrgn) are assigned explicitly, i.e., not as a mix of negative and positive entries in the commodity:commodity submatrix.

3. Factor accounts distinguish between

    a. Labour (lbr) – compensation of employees

    b. Mixed income (fmix) – returns to land. capital and self-employed labour that have not been separately identified.

    c. Capital (kptl) - returns to capital that have been separately identified.

    d. Land (land) - returns to land that have been separately identified.

4. One aggregate household account (hhold)

5. One aggregate incorporated business enterprise account (entrp)

6. Tax accounts – aggregate indirect taxes (itax) and aggregate direct taxes (dtax)

7. One aggregate investment account (invst)

8. One aggregate rest of the world account (wrld)

Given full reconciliation between the SUT and the national accounts the row and column totals for the macro-SAM will equate. However, the recording of inter-institutional transactions varies. Since inter-institutional transactions are the norm, these should be included in the macro-SAM or as an aggregate total (see below) if control totals are required for inter-institutional transactions.

The macro-SAM does not need to be complete (in which case it should be inconsistent) if certain entries cannot be estimated: BUT if transactions are missing from the macro-SAM estimates need to be included in the prior-SAM and the row and column control totals. If NOT, the estimated SAM will not be complete.

Users can modify the structure of the macro-SAM. However, it should be noted that any additional control totals that may be added to a macro-SAM could be easily included as additional aggregate control totals (see below), so oft times extended the macro-SAM will not improve the estimates. Moreover, the macro-SAM has the itax account that can be useful in ensuring that all indirect tax revenues are accounted for in the estimated SAM.

Below the macro-SAM control totals in the worksheet 'Macro-SAM', there is a matrix with the same dimensions for imposing case specific standard errors (`stderrs_msam(ss,ss)`): Non-zero values in this matrix overwrite the default standard errors while a minus number sets the associated standard error to zero.

Aggregate Total Controls

The (default) arrangements and structures for row and column and macro-SAM control totals and associated standard errors rarely justify changes. But the aggregate control totals, from the worksheet 'Agg_totals', will often reward customisation.

The worksheet 'Agg_totals' includes a series of 'standard' aggregates identified by the set aggr (in column A with descriptions in column B), which defines the third dimension of the parameter `aggagg(sac,sac,aggr)`. This parameter is used to define the elements of the SAM that are aggregated to equal the value of the aggregate (`value_agg` that is assigned in column F), that is only applied for the subset `aggr2(aggr)` whose elements are identified in column E. Special standard errors can be associated with each aggregate control total applied, `stderr_agg`, are assigned in column G; non-zero values in this column overwrite the default standard errors while a minus number sets the associated standard error to zero.

Extending the elements in the set `aggr` requires the user making changes in two files. First, the worksheet 'Agg_totals' needs changing to extend the sets `aggr` and `aggr2` and associated control totals and standard errors need to be assigned. And second, the file `samest_aggtotals.inc` needs extending to define the parameter `aggagg`. The pre-coded examples illustrate how simple the definitions for elements of `aggagg` if the aggregate is defined over a contiguous matrix or vector, e.g., `aggagg ("vattax",c,"t2_vat") = 1.0`, but involves multiple lines when the aggregates are defined over discontinuous vectors and matrices, e.g., `aggagg(c,h,"t2_gdp_exp") = 1.0`, etc. The expenditure measure of GDP definition also illustrates how a subtraction can be defined.

SAM Element Controls

The worksheet 'SAM-elements' provides a facility of defining control totals and standard errors for individual cells in the SAM. This is a cell by cell operation and therefore most often

applied when there are features about a transaction that mean a specific treatment is justified. The features that may justify specific treatment may be data related, e.g., a transaction value that is relative to others in a group is large/small and/or has wider/narrower error bounds, or solution value related, e.g., a transaction value in the solution that involves very large changes relative to the prior.

The cells need to be identified by row (column B) and column (column C), each cells needs to be assigned three parameters: a code that identifies whether it is estimated as a value with additive or multiplicative error or a coefficient with multiplicative error (column D with the options listed in cell A1), a transaction value (column E) and standard error (column F using the 'standard' code for standard errors).

These options are typically used if there are known characteristics about cells or, more usually, in response to the issues that arise in the solution and/or the performance of the programme.

SAM Blocks

The 'SAM-blocks' worksheet follows the structure of the macro-SAM and uses that to defined special treatment for blocks of cells in the prior SAM that map to the macro-SAM. Again, these options are typically used if there are known characteristics about the blocks or, more usually, in response to the issues that arise in the solution and/or the performance of the programme.

The matrix `CMAC(ss,ss)` identifies whether 'blocks', sub matrices, are estimated in value, the default, or coefficient form. The value default can be a 'strong' default in that the coefficient form can/should be avoided if practical. The matrix `SIGMAC(ss,ss)` is used to assign specific standard errors to the blocks that diverge from the general standard errors; non-zero values in this column overwrite the default standard errors while a minus number sets the associated standard error to zero.

7.   Standard Errors and Support Sets

The theory relating to standard errors and support sets in SAMEST is detailed in the technical document for the programme (SAM_Estimation CE.pdf); this section is concerned with the practical implementation of standard errors and support sets in SAMEST. The assignment of error support sets is in the worksheet `err_values`, as is the assignment of 'general' standard errors, for which there is a default set, `alt0`, and two option sets, `alt1` and `alt2`.

The assumed default for the error support sets is uninformative priors, i.e., 7, for the R(ow)C(olumn) controls, the Micro-SAM (`MSAM`) controls, the transaction value (*TV*) controls and the aggregate (`AGG`) controls. The values for error support sets must be 7, 5 or 3 and must be assigned in the Excel worksheet: if they are not assigned, or not one of the accepted values, the programme will abort with an error message. There are no hard coded default values in the programme.

The 'general' standard errors are also for each of the controls; `RC`, `MSAM`, `TV` and `AGG`. It is common to presume that the standard errors are smaller for the aggregate control totals (`MSAM` and `AGG`) and larger for the cell specific control totals (`TV` and `RC`). One estimation approach is to start with 'larger' standard errors, i.e., less constraining errors, and then progressively reduce the standard errors, i.e., tighten the 'general' constraints. This approach may/will throw up errors that either induce improvements in the prior SAM and/or the control totals or induce assigning specific, looser, standard errors to components of the SAM.

Table 7.1   err_values in Excel

	A	B	C	D	E	F	G	H	I	J	K	L
	Default Values for Errors											
	These parameters are used to set prior standard errors - sigmay_RC, sigmay_MAC, and sigmay_TV											
	Default prior standard errors for various SAM components. Choice is determined by value of consign.											
	stderrs	Alt0 values are default and cannot be changed. Alt1 and alt2 values are set here.										
		alt0	alt1	alt2								
	sig_RC	0.25	0.15	0.1		sac	row/col sums					
	sig_AGG	0.05	0.05	0.03			macro aggregates					
	sig_TV	0.25	0.15	0.1		sac, sacp	cell values or coefficients					
	sig_MSAM	0.05	0.05	0.03		ss	macro SAM aggregates					
	Use these error support set numbers. DEFAULT values are 7. If NOT assigned 3 or 5 or 7 the programme will ABORT with an error message.											
	Esupp	No. elements		Description								
	esupp_RC	7		Errors on column/row sums								
	esupp_AGG	7		Errors on aggregate control totals								
	esupp_TV	7		Errors on cell values/coefficients								
	esupp_MSAM	7		Errors on standard macro-SAM aggregates								

8.   **Programme Files**

This section reviews the various INCLUDE files called from the programme file, `samest_5.gms`, that are that are in two directories, `1_data` and `2_include`. The results are written out in a third directory, `3_ouput`.

It is assumed that users are using GAMS Studio as the editor, have set up Studio to produce a reference file for each Run, that each Run (is) with GDX Creation (`F10`), and makes use of the `View>Pin` option to view the GMS or INC file simultaneously with the reference file. When first exploring the programme and/or when loading new data, users are advised to make frequent use of the `$stop` option.

samest_5.gms

The main programme file contains all the set, parameter, variable, and equation statements, in the process of which several ***.inc files are called before the MODEL and SOLVE statements. The first is a file that controls the loading of data and associated assignments, while the other files contain all the assignment and testing instructions prior to the solve statement.

Nearly ALL the configuration of the programme is implemented through the data and settings contained in an Excel data file; this file has a standard configuration but is project specific. The Excel and GDX data files are identified through a $SETGLOBAL command: the output files are labeled by a $SETGLOBAL command, see below.

```
*~~~~~~ Substitutions to control Dataload
*~~ Identify the Excel workbook
$SETGLOBAL samest_data    country_BTN_sm
*~~ Identify the location of the Prior SAM (data_in if the prior is
in Excel) else a GDX file
$SETGLOBAL p_sam_in      data_in
*~~ Identify the location of the Prior SAM (data_in if the prior is
in Excel) else a GDX file
$SETGLOBAL macsam_in     data_in
*~~~~~~ Substitution to Assign labels for project specific outputs
$SETGLOBAL dstats        country_BTN_sm
```

In the examples the `prior SAM` is available from the Excel workbook; the $SETGLOBAL `p_sam_in` command allows for the possibility the prior SAM is 'supplied' from a GDX file. Similarly, the `macro-SAM` in the examples is available from the Excel workbook; the $SETGLOBAL `macsam_in` command allows for the possibility the macro-SAM is 'supplied' from a GDX file.

Unless changes are made to the programme users should not need to edit the master programme and only edit TWO of the various include files.

samest_dataload.inc

This file is found in the 1_data directory and controls the loading of data to the programme. Data from Excel are converted to the file `data_in.gdx` and from there read into the programme.

There are eight major components in this file.

1. Converts sets and data in Excel to GDX (Figure 8.1)

2. Load sets and data (not prior SAM) from `data_in.gdx`

3. Load Prior SAM from `data_in.gdx` or `'gdxname.gdx'`

4. Define sets that exclude total and other elements.

5. Set DEFAULTS, in Excel, to control the model (Figure 8.2).

6. Assign parameters for Aggregate Targets

7. Select ERROR Support Sets

8. 'Manual' or 'Automatic' SAM scaling (Figure 8.3)

Figure 8.1   Extract from samest_dataload.inc

```
*~~~~~~ 1. Convert EXCEL file data to GDX
~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
$CALL "GDXXRW i=1_data\%samest_data%.xlsx o=1_data\data_in.gdx
INDEX=LAYOUT!A4
$if errorLevel 1 $abort problems with GDXXRW
*~~~ 2. Load sets and data (not prior SAM) from data_in.gdx ~~
```

```
$GDXIN 1_data\data_in.gdx

* Load Micro and Macro Sets from data_in.gdx

$LOADdc sac c m a f l k lnd h g e i w emrgn etax

$LOADdc gtind gtdir

$LOADdc ss

* Control Sets

$LOADdc macro macro2 d_cons m_cons

$LOADdc sig alt mt esuppset esupp

* Macro SAM Data

$LOADdc macscale

* Values for controlling parameters

$LOADdc d_controls m_controls stderrs rowcol_tot stderrs_MSAM

*$LOADdc mactotal0  stderrs2

$LOADdc macromat

$LOADdc sigmac cmac

*Load arbitrary SAM aggregator

$LOADdc nsam SAMOPT_TV

$GDXIN

*~~~~~~ 3. Load Prior SAM and MACRO SAM ~~~~~~~~~~~~~~~~~~~~~~~

*~~~~~ Prior SAM

$onText

The prior SAM can be accessed from data_in.gdx if the SAM is in the

Excel workbook or an alternative GDX file. SETGLOBAL 'p_sam_in'

identifies the source

$offtext

$GDXIN 1_data\%p_sam_in%.gdx

*~~~ Load Prior SAM

$LOADdc P_SAM  = P_SAM

$GDXIN

*~~~~~ Exogenous Macro SAM

$onText

  The Macro SAM can be accessed from data_in.gdx if the Macro SAM is

in the Excel workbook or an alternative GDX file. SETGLOBAL

'macsam_in' identifies the source

$offtext
```

```
$GDXIN 1_data\%macsam_in%.gdx

*~~~ Load Prior macro-SAM

$LOADdc MACSAM_EXOG

$GDXIN
```

Figure 8.2 Options that control the Programme

```
92   *~~~~~~~ 5. Set DEFAULTS to control the model ~~~~~~~~~~~~~~~~~~~~~~~~~~~~~
93
94   * Set options to manage settings. default_SAM is set in code.
95    opt_blocks        = m_controls("con_opt_blocks","base") ;
96    opt_macro         = m_controls("con_opt_macro","base") ;
97    opt_RC_exog       = m_controls("con_opt_RC_exog","base") ;
98    opt_tv            = m_controls("con_opt_tv","base") ;
99    opt_msam          = m_controls("con_opt_msam","base") ;
100
101   opt_blocks$((opt_blocks ne 0.0) and (opt_blocks ne 1.0)}    = 0.0 ;
102   opt_macro$((opt_macro ne 0.0) and (opt_macro ne 1.0)}       = 0.0 ;
103   opt_RC_exog$((opt_RC_exog ne 0.0) and (opt_RC_exog ne 1.0)} = 0.0 ;
104   opt_tv$((opt_tv ne 0.0) and (opt_tv ne 1.0)}                = 0.0 ;
105   opt_msam$((opt_msam ne 0.0) and (opt_msam ne 1.0)}          = 0.0 ;
106
107   * If opt_macro = 0, set sigmac and cmac to zero
108    SIGMAC(ss,ssp)$(opt_blocks ne 1)     = 0 ;
109    CMAC(ss,ssp)$(opt_blocks ne 1)       = 0 ;
110
```

SAMEST: A User Guide

Figure 8.3   Automatic Scaling in samest_dataload.inc

```
183  *~~~~~~ Assign parameters for Autoscaling of the SAM
184
185   samscal           = d_controls("samscal","base") ;
186
187  * Assigning values to programme control parameters
188  * This assigns samscal for autoscaling scaling.
189  * It is overwritten is if manual scaling is chosen,
190
191   samautoscal       = d_controls("samautoscal","base") ;
192   cutoff0           = d_controls("cutoff0","base") ;
193
194  *~~~~~~ SAM Scaling
195
196   ABORT $(d_controls("scaltarg","base") LE 0.0)
197   "samautoscaltarg is less than or equal to ZERO - set to positive value" ;
198
199 ▼ WHILE{
200 ▼  ( (SUM[(sacn,sacnp)$(P_SAM(sacn,sacnp)
201 ▼          AND (ABS[P_SAM(sacn,sacnp)/samscal ]
202              < d_controls("scaltarg","base") } ), 1.0 ] }
203         /SUM[(sacn,sacnp)$P_SAM(sacn,sacnp), 1.0 ]) < d_controls("scalprop","base") ,
204
205  *    samscal is increased in multiples of 10 until the condition is satisfied
206       samscal = samscal * 10;
207   } ;
208
209  *~~ Select manual scaling
210 ▼ $onText
211  If samautoscal eq 0, then samscal is set by autoscaling, else manual
212  using d_controls("samscalman","base") as the scaling factor
213  $offtext
214
215   samscal$(samautoscal eq 0.0 ) = d_controls("samscalman","base") ;
```

NB: the same scaling factor, `samscal`, is applied to the exogenous macro-SAM, other aggregate totals (`MACTOTAL0`), the row and column totals from Excel (`rowcol_tot`) and the cutoff parameter (`cutoff` – see below).

The programme will abort if there are egregious errors in the scaling of `P_SAM` and the `macro-SAM` and between `P_SAM` and the row and column totals. The tests do NOT cover all possible errors, so they are not a substitute for care in the selection of units.

It is recommended that the same units of measurement are used for all data; this is the default assumption in SAMEST.

samest_single_set.inc

This file is embedded in the `samest_download.inc` file. By using singleton sets that are defined in this file the subsequent code can be customised to allow for the use of different account name by eliminating most quoted labels from the programme.

It will need editing if the accounts names for the macro-SAM and selected aggregate control totals have been changed.

<u>samest_stats.inc</u>

These descriptive statistics refer to the PRIOR SAM used in the estimation procedure, i.e., after all scaling and adjustments. These statistics are a useful summary of the prior SAM. They can also help if the programme is having difficulties; they provide information for setting the scaling and cutoff parameters.

The same statistics are calculated and reported for the estimated SAM - SAMCE.

<u>samest_prior.inc</u>

There are 3 procedures applied in this file. These procedures may result in `SAM0 = P_SAM`, **IFF** all adjustments have been made to `P_SAM` before it is loaded into the programme.

i.   These procedures may be redundant.

ii. BUT these procedures can be used to remove 'small' TVs from P_SAM as part of a process used to successively estimate a final SAM, e.g., increase the cutoff to get a first estimate, further condition the model and then reset the cutoff.

The procedures

1. Remove 'tiny' transaction values (TVs)

    a. Very small TVs may cause solution problems and/or be irrelevant.

    b. The definition of 'tiny' is defined by the parameter `'cutoff0'` that is assigned in Excel (`d_controls` worksheet)

2. Identify Rows (`irow`) and Columns (`icol`) for inclusion in the estimation and use `irow` and `icol` to assign `sacbal`.

    a. This MAY be a subset of `P_SAM`

3. Assign the SAM (SAM) used as the prior in the estimation.

    a. using `irow` and `icol` and `sacbal`

    b. compute differences between row and column sums

c. compute some information on coefficients and non zeros.

samest_msam_emdog.inc

This file produces an aggregated macro-SAM, `MACSAM_ENDOG`, from the prior SAM, `P_SAM`. This macro-SAM may be used to assign a macro-SAM used to provide control totals for the estimation process, IFF an 'ideal' exogenously defined macro-SAM has NOT been provided in the Excel database.

The aggregation is controlled by the set `map_ss_sac`, which is defined from subsets in the Excel file. Sets are defined in the main program and loaded from excel. This mapping set is assigned in the file `samest_msam_endog.inc`. The specification of `map_ss_sac` must be changed if the aggregates in the Excel file are changed.

This file ONLY provides the Transaction Values for the macro-SAM. The ERROR BOUNDS AND SUPPORT SETS are defined in `samest_err_supp.inc`.

Assign Prior Macro SAM (in samest_5.gms)

The macro-SAM is defined as either the endogenously estimated macro-SAM, in `samest_msam_endog.inc` or an exogenously defined macro-SAM. The latter option is preferred.

IDEAL:

1. An exogenous macro-SAM that is

    a. Is complete and consistent.

    b. Distinguishes between the factors

        i. labour

        ii. capital

        iii. land

    c. distinguishes between indirect and direct taxes.

    d. includes domestic margins.

2. If the exogenous macro-SAM is INcomplete and/or INconsistent

      a. limit the cells in the macro-SAM used as `controls`

      b. place wider bounds on controls used.

ALTERNATIVES:

1. Do not impose macro-SAM `controls`

2. Derive an implicit macro-SAM, `macsam_endog`, from the prior SAM, P_SAM

      a. The endogenous macro-SAM will be Incomplete, and/or

      b. The endogenous macro-SAM will be INconsistent

            i. THEREFFORE

                 1. Limit the cells in the macro-SAM used as control totals

                 2. place wide bounds on controls used.

After the file `samest_msam_endog.inc` is run the macro-SAM used in the model, `MACSAM0`, is assigned. The choice of using an exogenous macro-SAM, loaded from Excel, or an endogenous macro-SAM, derived from the prior SAM is controlled by the parameter `m_controls("con_macsam","base")` that is set in the Controls worksheet.

samest_aggtotal.inc

Use aggregate control totals as constraints if `opt_agg=` `0.0 or 1.0.`

This process allows for the creation of selected control total that are not provided for by a standard macro-SAM, which in this programme is constrained to operate across contiguous accounts in the P_SAM. Examples of `aggtotals` include

1. GDP = C + I + G + X - M

2. Components of INDIRECT TAXES, e.g.,

      a. Import duties

      b. VAT

      c. Excise taxes

      d. Taxes on emissions

    e. etc

3. Components of DIRECT TAXES, e.g.,

    a. Household income taxes

    b. Corporation taxes

    c. Factor income taxes

    d. Etc.

Some of these Control totals could be captured using an extended macro-SAM, so the decisions are user and case specific, however it is often easier to implement additional aggregate controls using the `aggtotals` options.

These control totals are additional to the macro-SAM control totals that are defined in a standard macro-SAM, `MACROSAM0`. Any aggregates of the elements in the micro-SAM can be specified as a aggregation constraints, with their own prior standard error. This is done in the Excel file in the worksheet 'Agg_totals'.

The EXOGENOUS target aggregate totals in `AGGTOTAL0` are assigned in the Excel worksheet, 'Agg_totals', using the parameter `aggagg(sac,sacp,aggr)` that is assigned in the file `samest_aggtotal.inc`. A series of standard assignments for `aggagg(sac,sacp,aggr)` are pre coded in the file `samest_aggtotal.inc`; these assignments can be extended to meet case specific requirements. The aggregate controls actually used by `samest_5.gms` are determined by membership of the set `aggr2(aggr)`, which is assigned in column E of the Excel worksheet.

The ENDOGENOUS target aggregate totals, `AGGTOTAL2`, are assigned from the micro-SAM, `P_SAM`, using the parameter `aggagg(sac,sacp,aggr)` to control the aggregation. This is NOT a recommended option.

The EXOGENOUS target aggregate totals are applied if `m_controls("con_opt_agg") = 0.0`, while the ENDOGENOUS target aggregate totals are applied if `m_controls("con_opt_agg") = 1.0`. In both cases the aggregate control totals used ACTUALLY imposed are identified by the subset - `aggr2(aggr)`.

If the range of aggregates is increased, the user needs to update the worksheet Macro and revise the file `samest_mactotal.inc`.

samest_cells.inc

This file identifies cells to be

1. included in the estimates

2. cells to be estimated as values

3. sets cells estimated as values to have additive error

4. cells to be estimated as coefficients by references to macro SAM accounts

The default/standard assumptions are: IF `opt_blocks` and `opt_macro` and `opt_RC_exog = 0`, then:

1. Cell constraints are all on values, `IVAL(irow,icol)`.

2. Constraints are set on all valid column and row sums, `icol2` and `irow2`.

3. Prior errors are assumed to be additive.

4. Prior standard errors are set by SIGM setting.

Define column sums that can be constrained. Users may wish to exclude columns with only one entry, or, alternatively, leave them in and define an entry constraint in terms of coefficients. Any other columns for which column constraints do not make a lot of sense.

Accounts which sum to zero should probably be constrained exactly, with no error. This can be done by simply setting their prior standard error, `sigmay_RC`, to zero, which will happen automatically if the target column sum is zero. See section on defining `sigmay_RC`.

For cells that are estimated as coefficients, by default, the cells will have multiplicative errors. It is natural to specify errors as multiplicative, since then the coefficients cannot become negative.

The `IF opt_blocks` statement specifies values and multiplicative errors if `cmac(ss,acmt) = 2`. However, this limits the adjustment domain in values, which may

lead to odd results if the discrepancies in row and column targets are large. Additive errors widen the domain of adjustment but may also lead to odd results.

The file `samest_coef_block.inc` allows hard coding for coefficients; the code in this file is illustrative.

OPT_BLOCKS alternative assumptions

1. Choices determined for blocks of the SAM as specified in CMAC and SIGMAC

2. Set accounts to be in terms of coefficients by turning off ivalue. ICOEFF is set as NOT ivalue.

3. Coefficients are treated "equally" and the estimation procedure will fail if there are large discrepancies in the accounts. Spreading large adjustments evenly across coefficients leads to odd results. In that case, it is better to work with values, which results in more asymmetric coefficient adjustment and large cells adjusting more.

4. Set coefficients for cells in blocks of the SAM, defined in the macro-sam, cmac(ss,ssp) using map_ss_sac(ss,sac). If cmac(ssn, ssnp) = 1, then the SAM cells in the corresponding block (defined by map_ss_sac) are all specified as coefficients.

5. If cmac(ss,ssp) = 2, treat cells as values and errors as multiplicative.

6. If standard errors are present in sigmac, use them.

LARGE CELL VALUE CHANGES

If cells have to change a lot, multiplicative errors may be infeasible so try additive errors, e.g.,.

acell(irow,a)$(not izero(irow,a))   = yes ;

acell(c,h)$(NOT izero(c,h))   = yes ;

samest_RC_targets.inc

The row and column targets are defined in this file. The code offers 4 options: 3 are based on the prior SAM while the fourth requires data from the Excel workbook.

Using 'exogenous' data from the Excel workbook is the preferred option: the reason is that the row and column targets are important control totals that should be explicitly assigned, rather than implicitly assigned from the prior SAM.

IF BETTER PRIOR INFORMATION ABOUT COLUMN and ROW SUMS IS KNOWN, USE IT.

There are 5 steps.

1. Assign targets for row and column totals

    a. If `opt_RC_exog = 0,` assign row and column total exogenously from worksheet Row-Col: this will overwrite the values for `colsum0(sac)` and `rowsum0(sac)`

    b. If `opt_RC_exog = 1,` assign row and column total endogenously from P_SAM

2. Check the differences between the totals of the rows and columns are not excessive; if they are large a judgements are required about the reliability of the row and/or column targets.

3. Assign targets based on Row and column controls (`rowsum0` and `colsum0`). The choice is controlled by the parameter target.

    - Target = 1 - target totals are mean of prior SAM row & column totals

    - Target = 2 - target totals are prior column totals

    - Target = 3 - target totals are prior row totals

4. Assign Targets for blocks specified as coefficients. If row-column totals are flagged in `CMAC(ssn,ssnp)`, then specify the targets for the corresponding blocks of cells defined by `map_ss_sac(ssn,sacn)`. This option requires 'opt_blocks' is set to 1.

5. Assign Targets for blocks specified as coefficients. IF `opt_tv = 1,` set `ivalue` and `acell` for prior SAM values depending on SAMOPT values.

<u>samest_consist.inc</u>

This file checks for consistency in the specifications for cells. The checks are designed to ensure that cells to be estimated as values are not also estimated as coefficients, and vice versa, that cells not estimated are not included in the estimation, that cells to be estimated are either estimated in levels or coefficients but not both, etc.

There are 2 parts each with several steps

1. Ensure cells NOT specified as values (`ivalue`) are specified as coefficients.

    a. CHK `ivalue` has a NON-zero column total (`icolnz`)

    b. CHK `ivalue` is NEGATIVE (`ineg`)

    c. REMOVE `ivalue` if all entries are zero (`izero`)

    d. Set `icoeff` if NOT `ivalue` and NOT ZERO

    e. Ensure all `ivalue` and `icoeff` are estimated

    f. Assign `ifixv` as cells not estimated

    g. Set additive errors for cells that are NEGATIVE and ESTIMATED

    h. Set multiplicative error for cells ESTIMATED without ADDITIVE

2. Check for errors in setting of constraints. These errors cause the programme to ABORT with error message

    a. Cells with both `ivalue` and `icoeff`

    b. Cells to be estimated but neither `acell` nor `lcell`

    c. Cells to be estimated but both `acell` and Courier New

    d. Cells to be estimated but initially zero

    e. Accounts in `irow` or `icol` but not in `sacn`

<u>samest_err_supp.inc</u>

Assume a prior on standard deviation of the errors, sigma, and then assume some prior knowledge of the standard error (perhaps due to measurement error) of the column sums. It is

assumed that all column sums have a standard error of 5%, or, alternatively, a user-defined value. This is a Bayesian prior, not a maintained hypothesis.

1. The estimated error is weighted sum of elements in an error support set, e.g.,

    `ERR_**(**,**) =E= SUM(jwt_**, W_**(**,**,jwt_**)*vbar_**(**,**,jwt_**V));`

    where the `W_**` are estimated in the CE procedure.

2. The prior variance of these errors is given by: `(sigmay(sacn))**2 = SUM(jwt, WBAR(sacn,jwt)*(VBAR(sacn,jwt))**2 )` where the `WBARs` are the prior on the weights.

3. The VBARs are chosen to define a domain for the support set. The prior on the weights, `WBAR`, are then calculated to yield the specified prior on the standard error, `sigmay`.

Weights

We define four sets of errors with separate weights, `W_RC`, `W_MAC`, `W_TV` and `W_MSAM`. The first is for specifying errors on column sums, the second for errors on macro aggregates (defined in the set macro), and the third for errors in cell values and the fourth for errors on the macro-SAM TVs.

If `sigmay` is set to zero, the corresponding constraint is assumed to be exact, with no error.

Define standard errors for errors on column sums, `sigmay_**`. In this version, if the initial column sum is zero, then `sigmay_**` will also be zero and the account will be constrained to sum to zero. This is appropriate for a "balancing" account. However, a nonzero sum with or without error can also be specified.

7 weights:

This is an 'uninformative' prior that incorporates only information about the outer bounds between which the errors must fall. In Bayesian analysis, the continuous uninformative prior is the uniform distribution between the bounds. Assuming the bounds are specified as + or - 3*s*, where *s* is a constant, then the prior mean is zero and the variance of the continuous uniform distribution can be calculated.

The decision is here is to specify a finite distribution with an evenly spaced, seven-element support set, with uniform prior weights all equal to 1/7, for which the variance can be calculated.

The seven-element finite prior distribution provides a conservative specification of an uninformative prior. Adding more elements would more closely approximate the continuous uniform distribution. Using fewer elements yields a higher prior variance. The seven-element specification permits the estimated posterior distribution to be essentially unconstrained. In this case, with limited data, it will not be possible to recover much information about the posterior distribution.

3 weights:

This code assumes a prior mean of zero and a two-parameter distribution with specified prior standard error.

There are three weights, `W(sacn,jwt)`, to be estimated. The actual moments are estimated as part of the estimation procedure.

5 weights:

This code assumes a prior mean of zero, no skewness (symmetric distribution), and a prior value of kurtosis consistent with a prior normal distribution with mean zero, variance `sigmay**2`, and kurtosis equal to `3*sigmay**4`. The addition of a prior on skewness and kurtosis requires estimation of 5 weights (`jwt = 5`). The prior weights `wbar` are specified so that: `SUM(jwt, wbar(sacn,jwt)*vbar(sacn,jwt)**4) = 3*sigmay(sacn,jwt)**4` as well as defining the variance as above.

The prior weights and support set are also symmetric, so the prior on all odd moments is zero (no skewness). The choice of +/- 1 standard error for `vbar(sacn,"2")` and `vbar(sacn,"4")` is arbitrary. The actual moments are estimated as part of the estimation procedure.

9. Concluding Comments

SAM estimation programmes can be categorised by how quickly they can achieve a consistent SAM. But, as Thorbecke observed, "[A]n erroneous or inaccurate SAM invalidates the results obtained from these models" (Thorbecke, 2003, p 186), which carries the implication that SAM estimation programmes should be judged on the 'accuracy' of the resultant SAM. That however is a difficult criterion to quantify since ultimately a SAM that 'perfectly' reflects the transactions in an economy is unknown. The SAMEST programme does provide some evidence through the optimand and the marginal values for TVs.

It is suggested that the speed of SAM estimation programmes is a distraction. Rather more emphasis should be placed on deriving the prior SAM and control totals and the theoretical properties of estimation programmes in the presence of partial and incomplete data.

References

Allen, R. G. and Lecomber, J. R. C. (1975) Some tests on generalised versions of RAS. In Estimating and Projecting Input-Output Coefficients (eds R. G. Allen and W. F. Gossling), pp. 43-56. London: Input-Output Publishing Co.

Allen, R.I.G. and Lecomber, J.R.C., (1975). 'Some Tests on a Generalised Version of RAS' in Allen, R.I.G. and Gossling, W.F., (eds) Estimating and Projecting Input-Output Coefficients. London: Input-Output Publishing Company.

Allen, R.I.G., (1974). 'Some Experiments with the RAS Method of Updating Input-Output Coefficients', Oxford Bulletin of Economics and Statistics, Vol 36, pp 215-228.

Bacharach, M. (1965) Biproportional matrices and input-output change. PhD Thesis. Department of Economics, Cambridge University.

Bacharach, M., (1970). Biproportional Matrices and Input-Output Change. Cambridge: Cambridge University Press.

Barker, T., van der Ploeg, F. and Weale M. (1984) A balanced system of national accounts for the United Kingdom. Rev. Inc. Wlth, 30, 461-485.

Byron, R. P. (1978) The estimation of large social account matrices. J. R. Statist. Soc. A, 141, 359-367.

Deming, W. E. (1943) Statistical Adjustment of Data. New York: Wiley.

Friedlander, D. (1961) A technique for estimating a contingency table, given the marginal totals and some supplementary data. J. R. Statist. Soc. A, 124, 412-420.

Friedlander, D., (1961). 'A Technique for Estimating a Contingency Table, given Marginal Totals and some Supplementary Data', Journal of the Royal Statistical Society, Series A, Vol 124.

Giinliik-Senesen, G. (1984) Sosyal Hesaplar Matrisi ve Turkiye icin bir Uygulama (The social accounting matrix and an application for Turkey). PhD Thesis. Istanbul Teknik Universitesi.

Giinliik-Senesen, G. and Bates J.M, (1988). 'Some Experiments with Methods of Adjusting Unbalanced Data Matrices', Journal of the Royal Statistical Society. Series A, Vol. 151, No. 3, pp. 473-490

Günlük-Senesen. G. and Bates, J.M., (1988). 'Some Experiments with Methods of Adjusting Unbalanced Data Matrices', Journal of the Royal Statistical Society, Vol 151, pp 473-490.

Harrigan, F. and Buchanan, I. (1984) A quadratic programming approach to input-output estimation and simulation. J. Reg. Sci., 24, 339-358.

Lahr, M. and L. De Mesnard. 2004. "Biproportional techniques in input-output analysis: table updating and structural analysis." Economic Systems Research, 16(2): 115-134.

Lecomber, J. R. C. (1975) A critique of methods of adjusting, updating and projecting matrices. In Estimating and Projecting Input-Output Coefficients (eds R. G. Allen and W. F. Gossling), pp. 1-25. London: Input-Output Publishing Co.

Lecomber, J.R.C., (1975). 'A Critique of Methods of Adjusting, Updating and Projecting Matrices' in Allen, R.I.G and Gossling, W.F., (eds) Estimating and Projecting Input-Output Coefficients. London: Input-Output Publishing Company.

Lemelin, A., I. Fofana, and J. Cockburn. 2013. "Balancing a Social Accounting Matrix: Theory and application (revised edition)". https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID2439868_code458963.p df?abstractid=2439868&mirid=1.

Lenzen, M., B. Gallego, and R. Wood. 2009. "Matrix balancing under conflicting information." Economic Systems Research, 21(1): 23-44.

Lynch, R.G., (1990). 'The Agriculture Sector in the Compilation of Input-Output Tables in the UK', paper given at Conference on Input-Output in Agriculture, Wye College, April 1990.

Matuszewski, T., Pitts, P.R., and Sawyer, J.A., (1964). 'Linear Programming Estimates of Changes in Input-Output Coefficients', Canadian Journal of Economics and Political Science, Vol 30.

McDougall, R.A. 1999. "Entropy theory and RAS are friends". GTAP Working Paper 6. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=1006&context=gtapwp.

Omar, F. H. (1967) The projection of input-output coefficients with application to United Kingdom. PhD Thesis. Department of Economics, University of Nottingham.

Peters, J.C., and T.W. Hertel. 2016. "Matrix Balancing with unknown total costs: preserving economic relationships in the electric power sector." Economic Systems Research, 28(1): 1-20

Robinson, S., A. Cattaneo, and M. El-Said. 2001. "Updating and estimating a social accounting matrix using cross entropy methods." Economic Systems Research, 13(1): 47-64

Round, J. 2003. "Constructing SAMs for development policy analysis: lessons learned and challenges ahead." Economic Systems Research, 15(2): 161-183

Round, J. I. (1982) Social accounting matrices and development planning. In Development Planning Models (ed. S. B. Dahiya), vol. 2, pp. 212-225. New Delhi: Interindia Publications.

Schneider, M.H., and S.A. Zenios 1990. 'A comparative study of algorithms for matrix balancing', Operations research, 38(3), 439-455

Stone, R. (1977) The development of economic data systems. In Social Accounting for Development Planning (G. Pyatt and A. R. Roe), Foreword. New York: Cambridge University Press.

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

Stone, R., Bates, J. M. and Bacharach, M. (1963) A Programme for Growth, vol. 3, Input-Output Relationships 1954-1966. London: Chapman and Hall.

Stone, R., Bates, J.M. & Bacharach, M., (1963). 'Input-Output Relationships 1954-66', Vol 3 in Stone, R, (ed) A Programme for Growth. London: Chapman & Hall.

Stone, R., Champernowne, D. G. and Meade, J. E. (1942) The precision of national income estimates. Review of Economic Studies., 9, 111-125.

Theil, H. (1967) Economics and Information Theory. Amsterdam: North-Holland.

Weale, M. (1985) Testing linear hypotheses on national accounting data. Rev. Econom. Statist., 67, 685-689.

A1. History

This programme uses a cross entropy (CE) technique to estimate the cells of a consistent SAM assuming that the initial data are inconsistent and measured with error. The basic references are:

Golan, Amos, George Judge, and Sherman Robinson. 1994. "Recovering Information from Incomplete or Partial Multisectoral Economic Data." The Review of Economics and Statistics. Vol. LXXVI, Number 3, pp. 541-549.

Robinson, Sherman, Andrea Cattaneo, and Moataz El-Said. 2001. "Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods." Economic Systems Research, Vol. 13, No. 1, pp. 47-64.

See also A. Golan, G. Judge, and D. Miller, Maximum Entropy Econometrics, John Wiley & Sons, 1996. The Golan, Judge, Robinson (1994) article is also presented there.

One approach described in the Golan, Judge, and Robinson article is to treat every cell in the SAM as being specified with an error support set whose weights are estimated. There is a prior specified for each error distribution. This treatment differs from the version of CE-SAM in Robinson, Cattaneo, and El Said in which the column coefficients were included directly in the cross entropy minimand.

In this program, the CE minimand only includes probability weights for the various error support sets. The SAM coefficients are not treated as analogous to probabilities.

Negative entries and accounts with zero sums do not require any special treatment in this approach.

The approach allows specification of a prior estimate of the standard error of: (1) cell entries, expressed either as values or column coefficients, (2) column sums, (3) various aggregates, and (4) macro aggregates from a standard macro-SAM aggregated from the micro-SAM, or from a user-defined aggregate SAM. The errors on cell entries can be specified as additive or multiplicative.

Fixed constraints are achieved by setting standard errors to zero.

Stochastic

- Version 1 was coded by Sherman Robinson and Scott McDonald Jan-April 2006.

- Version 2 was coded in May-June 2006 by Sherman Robinson.

- Version 3 was coded in August 2006 by Sherman Robinson.

- Version 4 was coded in August 2016 by Scott McDonald.

Various earlier versions were done by Scott McDonald and Sherman Robinson, with input from Melt van Schoor, 2004 to 2006.

The earlier versions were based on an initial version programmed by Sherman Robinson, August 2003, which was a major extension in approach to that of an earlier version by Sherman Robinson and Moataz El-Said, March, 2002.

Previous Versions:

The earlier method is described in S. Robinson, A. Cattaneo and, M. El Said, "Updating and Estimating a Social Accounting Matrix Using Cross Entropy Methods." Economic Systems Research, Vol. 13, No. 1, 2001, pp. 47-64.

Also available as TMD Discussion Paper No. 58, IFPRI, August 2000. The Discussion paper is downloadable in PDF format from the IFPRI web page "www.cgiar.org/IFPRI/" (requires Acrobat reader software available free on the web "www.adobe.com/products/acrobat/readstep2.html"

The GAMS code for this method is documented in: Robinson, Sherman, and Moataz El-Said. 2000. "GAMS Code for Estimating a Social Accounting Matrix (SAM) Using Cross Entropy (CE) Methods." TMD Discussion Paper No. 64. Washington, D.C.: International Food Policy Research Institute.

A2. Extensions to SAMEST

The discussion of SAMEST in the User Guide assumes that the matrix being estimated is a Social Accounting Matrix (SAM), partly for ease of exposition. In fact, the principles can be applied to any square matrix and with some care the rectangular matrices. This appendix touches on two topics that were set aside in the main body of text despite potentially important; estimating Supply and Use Table (SUT) and estimating a SAM that extends beyond the SNA's production boundary towards a general production boundary. The comments on estimating a SUT are not controversial, but the comments about a SAM with a General Production Boundary are questionable in national accounting terms.

Estimating Supply and Use Tables

If an SUT is unavailable, then an SUT must be estimated, which is difficult without access to primary data. The primary data required are censuses and/or survey of manufacturing, services, agriculture, etc., to provide estimates of inter-industry transactions, plus estimates of trade flows and domestic final demand. In addition, the national accounts should be able to provide control totals adequate to generate a macro-SAM. If the primary data are not available, it is questionable whether compiling a SAM is a worthwhile exercise due to the probability that the final SAM will be unreliable.

tbc

SAMs with a General Production Boundary

tbc