

GTAP 10 in a SAM: A User Guide

Scott McDonald & Karen Thierfelder

Addresses for correspondence:

Scott McDonald,
International Agricultural Trade &
Development
Humboldt-University of Berlin
Hannoversche Str. 27, Haus 12,
10099 Berlin,, UK.
Email: jrs.mcdonald@gmail.com
Tel: +44 789 979 3671

Karen Thierfelder
Department of Economics

US Naval Academy
Annapolis,
Maryland, USA
E: thier@usna.edu
T: +1 410 293 6887

Abstract

This paper is a user's guide to the programmes that convert the complete GTAP database, version 10, into a SAM format with satellite accounts in GDX format. This is convenient for any GAMS users that wish to work with the GTAP database in GDX format. It is also useful to persons trying to understand the accounting relationships that underpin the GTAP database and model. The programme also contains routines that convert the GTAP data for use by the various versions of the GLOBE model.

This is a draft that is undergoing continuing development and may contain less than perfect guidance. It is provided on that basis. Comments on the current content are encouraged.

Table of Contents

GTAP 10 in a SAM: A User Guide.....	1
Table of Contents	2
1. Introduction.....	3
2. GTAP as a Social Accounting Matrix	5
3. Some Notes on the GTAP Database and GEMPACK.....	11
4. Downloading and Installing the Programme for GTAP 10	17
5. Extracting GTAP Data Using GTAPAgg	20
6. Running the Programme	26
7. Aggregating a SAMs.....	29
8. Concluding Comments.....	40
References	41

1. Introduction

This User Guide is concerned with the production of the Social Accounting Matrix (SAM) and associated satellite accounts that can be derived from the GTAP database. The guide also notes how the SAM data can be converted for use in variants of the GLOBE model. The associated code uses the General Algebraic Modelling System (GAMS) software; ultimately the documentation of the processes is the GAMS code, which is always correct whereas this user guide is less reliable. The User Guide does not provide information about the theory of SAMs and satellite accounts or the mechanics of GAMS. The computer code is open source see www.cgemod.org.uk. No data are provided; users need to access GTAPAgg¹, a version of the GTAP database and the requisite licence.

This User Guide does not provide details about how the transactions data in the GTAP databases are converted into a SAM; this process is detailed elsewhere (McDonald and Thierfelder, 2004 and 2019) and has been implemented by GTAPAgg since version 7 of the GTAP database. A brief outline of the SAM is provided.

The computer code converts the entire GTAP database, which is in Header Array File (HAR) format², into a single GDX file. Currently, the code generates SIX versions of the database³:

1. GTAP in a SAM;
2. GTAP data for GLOBE 1;
3. GTAP data for GLOBE 1 energy;
4. GTAP data for GLOBE 2;
5. GTAP data for GLOBE 2 energy; and
6. GTAP data for GLOBE 3;

The first version – GTAP in a SAM – is a rendering of the GTAP database into a SAM format consistent with the GTAP model 6.2 (Hertel *et al.*, 2007)⁴ with all the other data in the GTAP database recorded as satellite files. It is a generic version intended for use by global CGE

¹ Alternatively, users can use FLEXAgg. This document assumes GTAPAgg is used.

² One file – default.prn – is a HAR file but for some unknown, and unfathomable, reason is saved by GTAPAgg as a prn file.

³ It will be necessary to update/modify the code when GTAP becomes available, but there are no plans to extend the number of versions.

⁴ Documentation for a GTAP model v7 (Corong *et al.*, 2017) has been published. We understand that the data requirements are the same, BUT we understood that version 10 of the database was intended to be consistent with version 7 of the GTAP model.

modelers using the GAMS software. The problems associated the large number of non-zero zeros in the GTAP database are addressed by truncating the transactions data at a user defined number of decimal places and then rebalancing the database so that the row and column totals equate at double precision.

The first version – GTAP in a SAM – is a model independent database. It is independent in the sense that it is not conditioned to be implement with any model, i.e., it is a transactions matrix with no presumptions about the behavioural relationships that determine the transactions. However, it should be noted that, as far as we can ascertain, the Center for Global Trade Analysis regard the GTAP database and model as inseparable, and hence the GTAP in a SAM data reflect the behavioural relationships embedded in the GTAP model.

The remaining five versions are derived from the GTAP in a SAM version for use by different versions of the GLOBE model. The code for GLOBE 1 and 2 is open source, but the file that generates the version for GLOBE 3 is currently only available to graduates of the various GLOBE courses.

The rest of this document is organised as follows. The next section provides a description of the GTAP database as a SAM. Section 3 contains some brief comments about GEMPACK and GTAP; these are intended to provide some information for those who do not know GEMPACK about some of the issues that arise when rendering GTAP data into a SAM format. The documentation of the programme assumes the user is working in GAMSIDE.⁵

⁵ At the time of writing GAMS Studio did not, yet, contain some GAMS features that the authors find invaluable. If users choose to work in GAMS Studio or their preferred editor, they will need to adapt the instructions to the requirements of their chosen editor.

2. GTAP as a Social Accounting Matrix⁶

The structure of the Global SAM is illustrated by means of a representative SAM for one region – the structure of the SAM for every region is identical. The description of the SAM structure is limited, since it is assumed that the reader is familiar with the basic principles of a SAM (see King, 1985; Sadoulet and de Janvry, 1995; Pyatt and Round, 1977; Pyatt, 1991; and Reinert and Roland-Holst, 1997); those interested in practical introduction to SAMs can follow the (free) online course at www.cgemod.org.uk/int_sam.html. In general terms the SAM structure adopted follows the conventions of the System of National Accounts for 1993 (UN, 1993), with adjustments for the limited data on intra-institutional accounts and GTAP's price system. The basic structure is identified in Table 1, wherein the transactions captured by each of the active sub matrices of the SAM are defined/described. The SAM layout has an implicit number code for each sub-matrix, given by the row and column numbers; this system of code numbers is used throughout the text to identify the sub-matrix currently under consideration.

SAM Transactions

The SAM reports six groups of agents that demand commodities (reading across the rows of the commodity accounts); activities, private households, government, investment, global transport services and other regions. These transactions take place at GTAP market prices. However, associated with each purchase by an agent there is an additional payment to the government that represents sales taxes. The SAM demonstrates that for imported commodities market prices are the price received by the exporting region plus the per unit transport cost plus the per unit tariff rate (reading down the columns of the commodity accounts). For domestically produced commodities market prices are the prices received by domestic activities, and hence export taxes are recorded as expenditures by the domestic commodity accounts. Domestic market prices are derived from the production costs, which are made up of the costs of intermediate inputs valued at market prices plus the sales taxes paid on intermediate inputs, plus the expenditure on factor use and production taxes and payments to primary inputs.

⁶ This section is taken verbatim from McDonald and Thierfelder (2019); it can be skipped by those who fully understand the SAM representation of the GTAP database.

Deriving a Global SAM from GTAP Data

All income initially accrues to the regional household before being distributed to the private household, government and the capital account. Regional income is made up of payments to factors, net of factor income taxes⁷ and depreciation, plus indirect taxes, which are made up of trade, sales, factor use and production taxes, and income taxes. The artifice of a regional household means that the private household does not pay income taxes nor does it save, and the government neither saves nor borrows, i.e., the implicit balance on the government budget is zero.

Trade consists of two elements; expenditures on commodities and expenditures on transport margins. Exports are valued *fob* (free on board) and after the payment of any export duties. Exports of transport services to the global transport pool are recorded separately. Imports of commodities are also valued *fob*, with transport services recorded separately. The sum of the two represents expenditure on imports inclusive of carriage, insurance and freight (*cif*). Consequently, there are two sets of trade balances: the first, represents the trade balance with respect to each of the transport services while the second are the trade balances with all other regions.

SAM Dimensions

The dimensions of the SAM are determined by accounts that are identified. Since these change over time it is useful to have a generalised representation of the dimensions of the global SAM. Specifically the dimensions of the SAM can be defined in general terms by using sets for the number of commodities, c ; activities, a ; factors, f ; and the number of regions, k . The numbers of members in each set for several versions of the GTAP database are defined in Table 2.

⁷ Since version 7 of the GTAP database, factor income taxes are defined as taxes paid directly by factors, e.g., social security payments, plus income taxes paid by households and incorporated business enterprises. Household income taxes are assigned to labour and corporation taxes are assigned to capital (see McDougall and Hagemjer, 2016).

Table 1 Transactions in the Social Accounting Matrix for a Representative Region

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
		Imported Commod	Domestic Commod	Activities	Factors	Regional Household	Private Household	Trade Taxes	Import Sales Taxes	Domestic Sales Tax	Factor Taxes	Production Taxes	Direct Taxes	Govt	Capital	Import Margins	Export Margins	Rest of World
1	Imported Commodities			Intermediate Inputs			Private demand							Government demand	Investment demand			
2	Domestic Commodities			Intermediate inputs			Private demand							Government demand	Investment demand		Exports of Transport Services	Exports of Goods and Services
3	Activities		Supply matrix															
4	Factors			Payments to Factors														
5	Regional Household				Distributed Factor Incomes			Import Duties and Export duties	Import Sales Taxes	Domestic Sales Taxes	Factor Use Taxes	Production Taxes	Income Taxes					
6	Private Household					Private Household income												
7	Trade Taxes	Import Duties	Export duties															
8	Import Sales Taxes			Sales Taxes - Imports			Sales Taxes - Imports							Sales Taxes - Imports	Sales Taxes - Imports			
9	Domestic Sales Tax			Sales Taxes - Domestic			Sales Taxes - Domestic							Sales Taxes - Domestic	Sales Taxes - Domestic			
10	Factor Taxes			Factor Use Taxes														
11	Production Taxes			Production Taxes														
12	Direct Taxes				Income Taxes													
13	Government					Government income												
14	Capital				Depreciation	Savings											Trade balance	Trade balance
15	Import Margins	Transport margins on imports																
16	Export Margins															Imports of transport services		
17	Rest of World	Imports of Goods and Services																

In the GTAP database there is no distinction between commodities and activities; rather GTAP identifies sectors, where the term sector is used interchangeably to define imported and domestic produced commodities and production activities. This is possible for two reasons:

- i) underlying the data for each region is a symmetric input-output table that means the Supply Matrix – sub matrix 3:2 in Table 1 – is strictly diagonal; and
- ii) the GTAP database is not organised in a single matrix structure.

In single matrix (SAM) structure it is necessary to define $2*c$ commodity accounts (c for imported and c for domestically produced commodities) and a activity accounts. The database allows each domestic agent to be charged different purchase, or sales, tax rates⁸ for both imported and domestic commodities; hence it is necessary to have $2*c$ accounts for sales taxes. Similarly, each production activity can be charged an activity specific indirect/production tax, which necessitates another account. Production also involves the use of primary inputs, of which f are identified in GTAP; so, there are f accounts for factor purchases and f accounts for taxes paid on factor use by activities.

For trade relationship, each region **can** import from and export to all other regions, hence for each region there needs to be k accounts to record trade transactions. In addition, there needs to be $3*k$ accounts for each region to capture trade taxes – import duties and export taxes – upto and including version 7.1 - and $3*k$ accounts for each region since version 8 to capture trade taxes – *ad valorem* and specific (quantity) import duties and export taxes.⁹ And since there are three transport margins (m) associated with each trade transaction there needs to be $m*k$ accounts for each region to capture trade and transport costs; with a further m accounts to capture exports by each region to the global trade and transport pool. Finally, there are four domestic institutional accounts: the regional household that collects all regional income, either from factor sales or taxes (indirect and direct). This income is dispersed to the other three domestic institutions; the private household, government and capital (investment/savings).

The most immediately obvious points about the SAM are the large number of accounts and the relative scarcity of entries in the SAM. The large size of the SAM is primarily a

⁸ The database does not distinguish between different types of tax instruments.

⁹ The SAM generated by the GTAPAgg programme does not separate out *ad valorem* and specific (quantity) import duties.

Deriving a Global SAM from GTAP Data

consequence of the number of regions, k , in the database (k trade accounts plus $3/2k$ for trade taxes) and the fact that m types of transport margins on imports are identified with imports from each region ($m*k$ accounts), while the second main reason is the series of commodity/user specific sales taxes on imports (c) and domestic commodities (c)

The resultant SAM is of the dimensions $(4c + a + 2f + 9 + mk + 3/4k) * (4c + a + 2f + 9 + mk + 3/4k) * (k)$, where c and a are the number of sectors in the database, f the number of factors and k the number of regions. The dimensions of the SAM matrix for version 10 is **1,357*1,357*140 (257,802,860 cells)**.

Table 2 also highlights several important features of the GTAP database.

- Information is concentrated in the trade accounts.
- The SAM matrices are sparse, e.g., there are **4,580,307 non-zero** entries in the GTAP SAM for version 10.
- The within regional information emphasises inter-industry and final demand transactions.
- The tax information relates overwhelmingly to indirect taxes on sales and purchases.
- There is very little information about domestic institutions other than for consumption, e.g., savings are not separately identified for private households and governments.

Table 2 Dimensions of the Social Accounting Matrix for a Representative Region

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
		Imported Commod	Domestic Commod	Activities	Factors	Regional Household	Private Household	Trade Taxes	Import Sales Taxes	Domestic Sales Tax	Factor Taxes	Production Taxes	Direct Taxes	Govt	Capital	Import Margins	Export Margins	Rest of World	Totals
1	Imported Commodities			$c*a$			$c*1$							$c*1$	$c*1$				$c*1$
2	Domestic Commodities			$c*a$			$c*1$	$c*c$						$c*1$	$c*1$		$c*m$	$c*k$	$c*1$
3	Activities		$a*c$																$c*1$
4	Factors			$f*a$															$f*1$
5	Regional Household				$1*f$			$1*c$	$1*c$	$1*c$	$1*f$	$1*1$	$1*1$						$1*1$
6	Private Household					$1*1$													$1*1$
7	Trade Taxes	$1/2k*c$	$k*c$																$2k*1$
8	Import Sales Taxes			$c*a$			$c*1$							$c*1$	$c*1$				$c*1$
9	Domestic Sales Tax			$c*a$			$c*1$							$c*1$	$c*1$				$c*1$
10	Factor Taxes			$f*a$															$f*1$
11	Production Taxes			$1*a$															$1*1$
12	Direct Taxes				$1*f$														$1*1$
13	Government					$1*1$													$1*1$
14	Capital				$1*f$	$1*1$											$1*m$	$1*k$	$1*1$
15	Import Margins	$m*k*c$																	$m*k*1$
16	Export Margins															$m*m*k$			$m*1$
17	Rest of World	$k*c$																	$k*1$
18	Totals	$1*c$	$1*c$	$1*a$	$1*f$	$1*1$	$1*1$	$1*2k$	$1*c$	$1*c$	$1*f$	$1*1$	$1*1$	$1*1$	$1*1$	$1*m*k$	$1*m$	$1*k$	

3. Some Notes on the GTAP Database and GEMPACK

GEMPACK

Understanding the structure and content of the GTAP database is much easier if the user can read GEMPACK code. This is because the structure of the GTAP database is integral to the relationships embedded in the GTAP model. If the user understands the basic principles of a SAM and the behavioural relationships in the GTAP model, then the source ‘coefficients’ can be more easily associated with the various submatrices of a SAM.¹⁰

Since GEMPACK influences the structure and content of the GTAP database, these brief notes on some aspects of GEMPACK are intended to help users when first trying to understand GEMPACK code. These notes fall into 5 categories; modelling in terms of rates of change, sets and the local control of sets, the regional household, trade and transport margins and the global bank. As a matter of convention, in the GTAP model all variables, i.e., the rates of change, are in lower case and the coefficients/parameters are in upper case. This raises an important terminological difference: GEMPACK uses the term coefficients to describe elements that are known as parameters in GAMS – in this paper the term parameter is adopted although in this context the terms are interchangeable. Subsequently, this section goes on to identify the data from the GTAP database that are used to define transactions.

GTAP and GEMPACK

The solution algorithms used by GEMPACK solve for rates of change (log differences); this requires that all cells in any sub matrix of the SAM, for which the model has behavioural relationships, must have non-zero values. In terms of single precision this requires that cells for which there are no real-world transactions report values of 0.000001 (ONE US Dollar) or greater; it is apparently not possible to identify which of these small non-zero transactions do not represent ‘real’ world transactions.

¹⁰ It was evident, when developing the GLOBE model, that many users of the GTAP database did not understand the accounting identities/relationships embedded in the GTAP database. Prior to 2003 it was argued, by some GTAP ‘experts’, that the GTAP database could not be rendered in a SAM format. This was self-evidently incorrect. This error indicates that some of these ‘experts’ did themselves understand the accounting identities. The rendering of GTAP in a SAM was undertaken to help users of the GTAP database understand the accounting identities.

This has implications that should be noted:

1. The GTAP database, and SAM, contains non-zero zeros, i.e., the number of ‘real-world’ transactions are less than the number of active cells in the database;
2. the row and column totals for a GTAP SAM are not identical, which limits the extent to which the accounting identities can be used to verify the data are consistent¹¹; and
3. any elimination (or truncating) of very small transactions, say less than \$(US)100, will result in differences in the row and column totals of the SAM.

SETS and Control of SETS

Sets in GEMPACK are ‘local’. This means that when coding in GEMPACK it is necessary to include local definitions for sets used by variables and parameters.

Consider the following extract of code from the GTAP tablo¹² file

```
Coefficient (all,i,DEMD_COMM) (all,j,PROD_COMM) (all,r,REG)
  VFA(i,j,r) # producer expenditure on i by j in r valued at agents' prices #;
Formula (all,i,ENDW_COMM) (all,j,PROD_COMM) (all,r,REG)
  VFA(i,j,r) = EVFA(i,j,r);
Formula (all,i,TRAD_COMM) (all,j,PROD_COMM) (all,s,REG)
  VFA(i,j,s) = VDFA(i,j,s) + VIFA(i,j,s);
```

This first defines the coefficient (parameter in GAMS) VFA over i , j and r with members of i defined as DEMD_COM, of j as PROD_COM and of r as REG. The subsequent two formulae are then constrained to operate over subsets of i , i.e., ENDW_COM and TRAD_COM, but they can reuse the character i because of the system of local control of sets. Similarly, it is perfectly acceptable to redefine i for subsequent use.

A major consequence of this feature of GEMPACK is the necessity to identify explicitly the set definitions that apply to each block of GEMPACK code¹³.

¹¹ The equality of row and column totals in a SAM can be used to verify consistency: the reported transactions may be consistent, but that does NOT mean they are complete.

¹² In GEMPACK a tablo file is the name given to a programme file – the programme for editing these files is known as TABMATE.

¹³ This may appear burdensome to some GAMS users, but it does avoid the potential problems with set ordering in GAMS. There are features of GEMPACK that are arguably simpler than the equivalent in GAMS and vice versa.

GTAP Database

Sets

Ideally, the sets definition used by GTAPAgg would be used to reduce the amount of data read in from Excel by the programme. But, GTAP sets data use the SAME labels for some regions and sectors, which causes set ORDER problems in GAMS: it is less error prone to define many of the sets and some maps in Excel and then define some maps dynamically in GAMS. It has been indicated that GTAP may change its labels as some point in the future, which may allow a review of this decision.

GTAP Data

The GTAP database is saved in HAR format and are organised to be convenient for the GTAP model and GEMPACK. The transaction values (TV) in GTAP are reported in millions of US dollars to SIX decimal places, i.e., GTAP reports TVs with an IMPLIED accuracy of a single US dollar. This is not a correct interpretation/understanding.

In the GTAP database ALL TVs in ALL 'active' sub matrices, i.e., sub matrices for which the GTAP model has behavioural relationships MUST have NON- ZERO values. This is (apparently) because

1. the GTAP CGE model is implemented in GEMPACK,
2. the database is estimated using the FIT programme that uses GEMPACK, and
3. GEMPACK is single precision and uses log differences.

This has non- trivial implications for users of the data

1. the database has many data points that are NON- ZERO, but do not represent real economic transactions, i.e., transactions that should be zero ('non- zero zeros'),
2. it is not, apparently, possible to determine which NON- ZERO transactions are in fact 'non- zero zeros' (we gave up asking, fruitlessly, after 10 years and after learning that FIT was a GEMPACK programme),
3. given the presence of 'non- zero zeros' it is reasonable to conclude that some apparent TVs should be ignored from an economic perspective, while from a software, e.g., GAMS, perspective these very small TVs may also be an 'inconvenience',

4. in the GTAP database, presented as a SAM, the row and column totals are not identical, which limits the ability to use accounting identities to check the database,
5. as the number of sectors, factors and regions in the GTAP database increases so the data problems presented by 'non- zero zeros' increase,

6. the GTAP 10 (2014) database contains a very large number of 'small' transactions.

There are 4,580,307 data points in active sub matrices of the GTAP SAM, of which

- 4,278,532 are transactions of greater than \$(USD)1,
- 3,830,984 are transactions of greater than \$(USD)10,
- 3,201,405 are transactions of greater than \$(USD)100,
- 2,444,050 are transactions of greater than \$(USD)1,000
- 1,66,791 are transactions of greater than \$(USD)10,000,
- 980,814 are transactions of greater than \$(USD)100,000;

7. the definition of a 'small' transaction is not the same for the USA as for, say, Mauritius, because the scale of transactions varies enormously, e.g., in GTAP 10 the largest transaction for the region with the smallest 'largest' transaction, 'Rest of the World' (xtw), is \$102.434 m, while the largest transaction for the region with the largest 'largest' transaction, 'USA' (usa), is \$10,887,626.000 m

Users of the GTAP database must make decisions about how they choose to address the implications of these aspects of the database if they wish to use the GTAP database. Some of those choices relate to the data, some to the model that will be calibrated with the data and some the software.

Decisions About the GTAP Database

The decisions made about the how aspects of the GTAP database will be handled in this programme reflect three considerations that were important to us:

- the output would be geared towards the need of users of the GAMS software (the dominant software used for CGE applications);
- separation of data and model, i.e., a database can support multiple models¹⁴; and

¹⁴ We define a model by reference to the behavioural relationships embedded in the model

- needs of the GLOBE family of models.

Consequently, we made the following choices.

1. It is more efficient to resolve all data issues **once**. Hence, we choose to transform the whole of the GTAP database from HAR to GDX.
2. The GTAP database is reformatted into a transactions SAM and a series of linked satellite accounts using a common set of account labels.
3. The use of accounting identities to check the database is too valuable to discard (we also use these identities in our models). Hence, it is necessary to address and resolved all inconsistencies in the GTAP database, e.g., all row and column totals must equate at double precision, all imports (*fob*) by region *d* must be exactly equal to exports (*fob*) from region *s*, etc.
4. Some truncation of the non-zero elements in the GTAP SAM is justified, because, by definition, there are large numbers of 'non-zero zeros', BUT it is NOT possible to determine which 'non-zero zeros' do not represent transactions. Therefore, the code allows the user to set her own truncation factor.
5. 'Small' TVs that create small shares are 'problematic' when using functional forms that are standard in CGE models (CES, CET, LES, translog, etc), irrespective of the preferred solution method.
6. The imposition of the requisite accounting identities is achieved manually to avoid (i) creating new 'small' transactions, and (ii) distorting the tax rates etc. Hence, we choose not to use a mathematical algorithm, e.g., RAS¹⁵ or entropy.
7. The trade margin accounts do not 'balance' and are 'tricky' to resolve using only within region data; hence we create a dummy region, 'GLOBE', that supplies and demands all margin services. This region is used in our models and is implicit in the GTAP model.

¹⁵ If we were to use a mathematical algorithm it would be RAS. The problem here is one of final balancing of a matrix, which is what the RAS method was developed to do, and not an estimation problem, for which the entropy method was developed.

8. We provide a MS Excel VBA programme to generate aggregations of the SAMs created by this programme and GAMS programme with an associated Excel workbook.

Programme Features

We also made certain decisions about feature of the programme:

1. The routines in this programme do NOT include provisions to augment the GTAP database. This means that some changes to the GTAP database are approximations; this is particularly the case with the removal of the REGIONAL HOUSEHOLD construct used in the GTAP database.
2. The routines in this programme include checks at each stage of the process, hence there are a very large number of checks. This reflects the experience that for each release of the GTAP database it has been necessary to identify release specific features of the GTAP database, e.g., changes in labels used.
3. The routines in this programme report a lot of information about the 'structure' of the GTAP database; this is important because it allows us to better understand the transaction relationships.

4. Downloading and Installing the Programme for GTAP 10

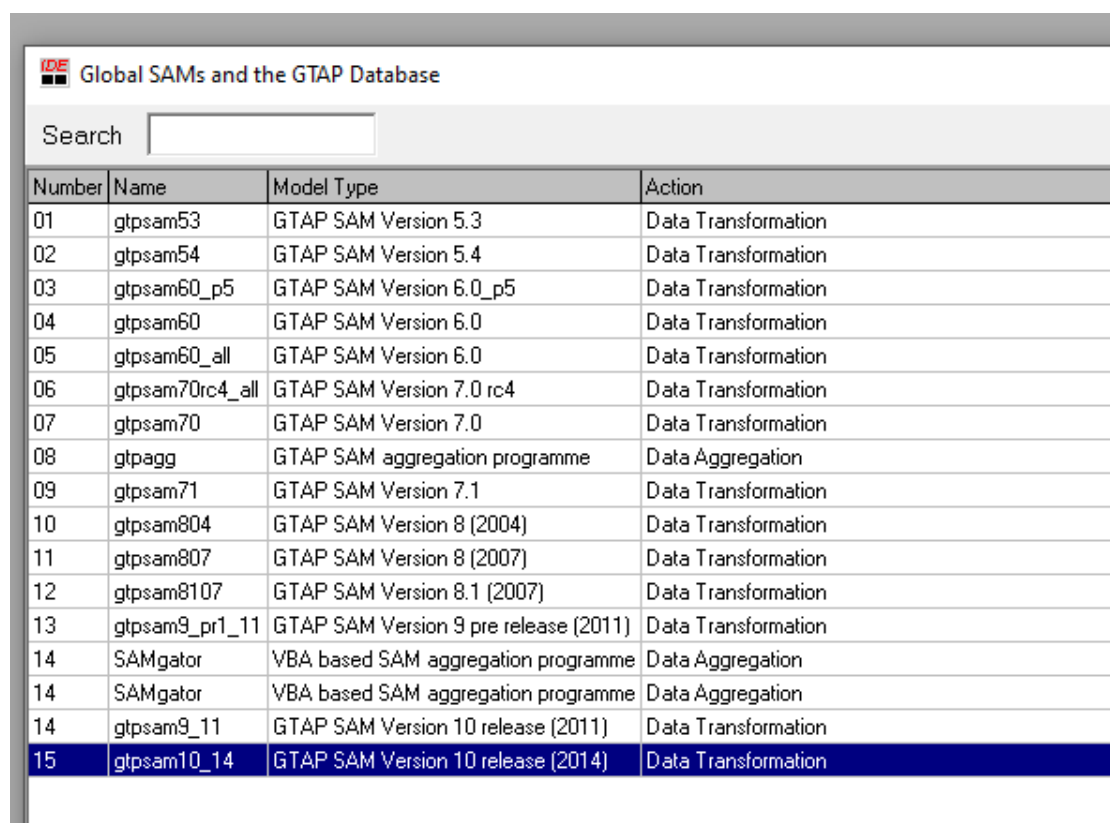
The codes for extracting the GTAP database to GAMS are supplied as a ‘User Model Library’. The ‘User Model Library’ facility in GAMSIDE provides an indexing facility whereby collections of models can be archived and then easily accessed from GAMSIDE. All the files contained within a Model Library are stored in a single directory and accessed via a simple Library manager that is accessed from GAMSIDE. The Library manager allows the user to download a file, OR selection of files, to the current working directory according to the criteria specified by the compiler of the library. (GAMS uses this facility, in the form of the GAMS Model Libraries, to make available a wide range of GAMS models that illustrate the various capabilities of the GAMS language).

An introduction to the ‘User Model Library’ facility is available at [www.cgemod.org.uk/Intro toGAMSIDE.pdf](http://www.cgemod.org.uk/Intro%20toGAMSIDE.pdf) (pp 22-26).

Steps

The GTAPinaSAM programmes are supplied as a WinZip archive that contains ALL the versions we have produced since GTAP 5. The steps in this section refer to versions 9 and 10.

1. Create a working directory for processing the data (say C:\GTAPinaSAM) and add a project file.
2. Download the GTAPinaSAM WinZip archive from www.cgemd.org.uk????
3. Unzip the archive to a subdirectory in the ‘GTAPinaSAM’ directory, say ‘gtapsamlib’.
4. In GAMSIDE access the User Model Library – gtapinasam – and select the files for GTAP version 9 (line 14) or GTAP version 10 (line 15) database that you are using. (see Figure 4.1)

Figure 4.1 User Model Library for GTAPinaSAM


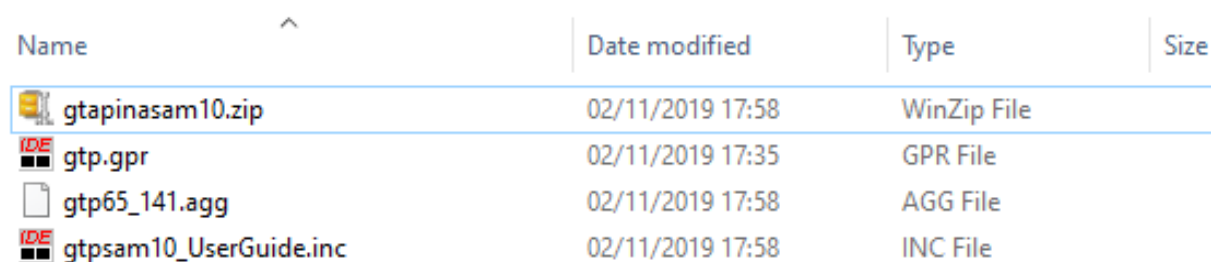
IDE Global SAMs and the GTAP Database

Search

Number	Name	Model Type	Action
01	gtpsam53	GTAP SAM Version 5.3	Data Transformation
02	gtpsam54	GTAP SAM Version 5.4	Data Transformation
03	gtpsam60_p5	GTAP SAM Version 6.0_p5	Data Transformation
04	gtpsam60	GTAP SAM Version 6.0	Data Transformation
05	gtpsam60_all	GTAP SAM Version 6.0	Data Transformation
06	gtpsam70rc4_all	GTAP SAM Version 7.0 rc4	Data Transformation
07	gtpsam70	GTAP SAM Version 7.0	Data Transformation
08	gtpagg	GTAP SAM aggregation programme	Data Aggregation
09	gtpsam71	GTAP SAM Version 7.1	Data Transformation
10	gtpsam804	GTAP SAM Version 8 (2004)	Data Transformation
11	gtpsam807	GTAP SAM Version 8 (2007)	Data Transformation
12	gtpsam8107	GTAP SAM Version 8.1 (2007)	Data Transformation
13	gtpsam9_pr1_11	GTAP SAM Version 9 pre release (2011)	Data Transformation
14	SAMgator	VBA based SAM aggregation programme	Data Aggregation
14	SAMgator	VBA based SAM aggregation programme	Data Aggregation
14	gtpsam9_11	GTAP SAM Version 10 release (2011)	Data Transformation
15	gtpsam10_14	GTAP SAM Version 10 release (2014)	Data Transformation

GTAP 10 in a SAM

When the files are downloaded GAMSIDE will automatically open the file 'gtpsam10_UserGuide.inc', but the other files will be in the WinZip archive 'gtapinasam10.zip' (see Figure 4.2). Extract the contents of the WinZip archive to the directory GTAPinaSAM; do NOT place the contents in another directory. The result should be identical to that shown in Figure 4.3.

Figure 4.2 GTAPinaSAM Directory after Download from Library






Name	Date modified	Type	Size
 gtapinasam10.zip	02/11/2019 17:58	WinZip File	
 gtp.gpr	02/11/2019 17:35	GPR File	
 gtp65_141.agg	02/11/2019 17:58	AGG File	
 gtpsam10_UserGuide.inc	02/11/2019 17:58	INC File	

Figure 4.3 GTAPinaSAM Directory

Name	Date modified	Type	Size
data	02/11/2019 17:17	File folder	
GLB_Output	02/11/2019 17:16	File folder	
GTP_Output	02/11/2019 17:16	File folder	
include	02/11/2019 18:00	File folder	
stats	02/11/2019 17:16	File folder	
clean_all.bat	02/11/2019 16:58	Windows Batch File	1 KB
clean_top.bat	24/03/2019 18:42	Windows Batch File	1 KB
data_glbsam_10_2014.xlsx	11/09/2019 12:24	Microsoft Excel W...	197 KB
glbsam10_2014.gms	02/11/2019 17:09	GMS File	15 KB
gtapinasam10.zip	02/11/2019 17:58	WinZip File	220 KB
gtp.gpr	02/11/2019 17:35	GPR File	1 KB
gtp65_141.agg	02/11/2019 17:58	AGG File	20 KB
gtpsam10_UserGuide.inc	02/11/2019 16:56	INC File	15 KB

These processes have established all the file provided by CGEMOD. It is now necessary to process the database provided by GTAP – see Section 5 below. Once that is complete, the programme can be run – see Section 6 below.

5. Extracting GTAP Data Using GTAPAgg

The GTAPAgg programme is used to extract the GTAP data as a fully disaggregated database. Prior to the release of the GTAP 9 database, the codes provided constructed the SAM from the various matrices produced by GTAPAgg. But since GTAP 7, the GTAPAgg programme has also produced a SAM based on the procedures reported by McDonald and Thierfelder (2004 and 2019); this SAM is used as a starting point for this programme.¹⁶

Installing GTAPAgg

Install GTAPAgg on the licensee's PC. We recommend adopting the default settings. Make sure the licence file purchased from GTAP is included in the directory created

Using GTAPAgg

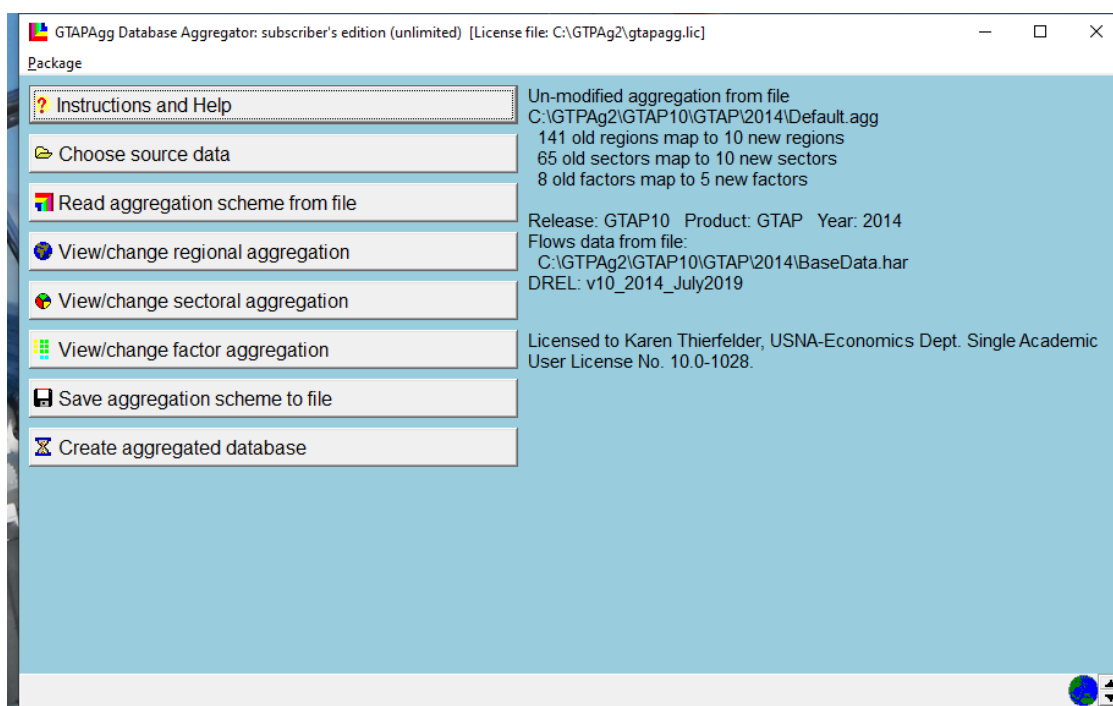
There are two ways GTAPAgg can be used to output the GTAP database for this programme. The first uses an 'aggregation scheme' and Excel workbook we have created; in this case you will be using the labels we have defined. The second requires the user create an 'aggregation scheme' and make any adjustments to the Excel workbook to render it consistent with the user's 'aggregation scheme'. We recommend the first option because it is simpler for the user and the labels are not important since model applications will, almost invariably, work aggregations of the database in which case the user can define her preferred labels.

Using CGEMOD's 'Aggregation Scheme' for GTAP 10 (2014)

Figure 5.1 shows the GTAPAgg 'front end' developed by Mark Horridge. To use the aggregation scheme supplied, select the option "Read aggregation scheme from file" and find the file 'gtp65-141.agg' that will have been placed on the top level of the directory 'GTAPinaSAM' when you extracted the files for GTAPinaSAM.

Open this file into GTAPAgg and then select "Create aggregated database". We suggest that you save the WinZip file created – the default name is (usually) 'GTAP10_GTAP_2014_65x141.zip' – into the 'data' subdirectory of the 'GTAPinaSAM' directory.

¹⁶ It is unknown whether GTAP will maintain the SAM format and, if they do, whether the format they choose will be appropriate. This decision will be reviewed for subsequent versions of the GTAP database.













Figure 5.1 GTAPAgg Front End

Then unzip the database into the ‘data’ subdirectory of the ‘GTAPinaSAM’ directory. Finally, change the three-letter file extension to the file “default.prn” to “default.har”.¹⁷ The data directory should look identical to that in Figure 5.2. **NB:** default.prn has been changed to default.har.

The programme is now ready to run.

¹⁷ The “default.prn” file is a har file; for some reason GTAP choose to label this file a ‘prn’ file.

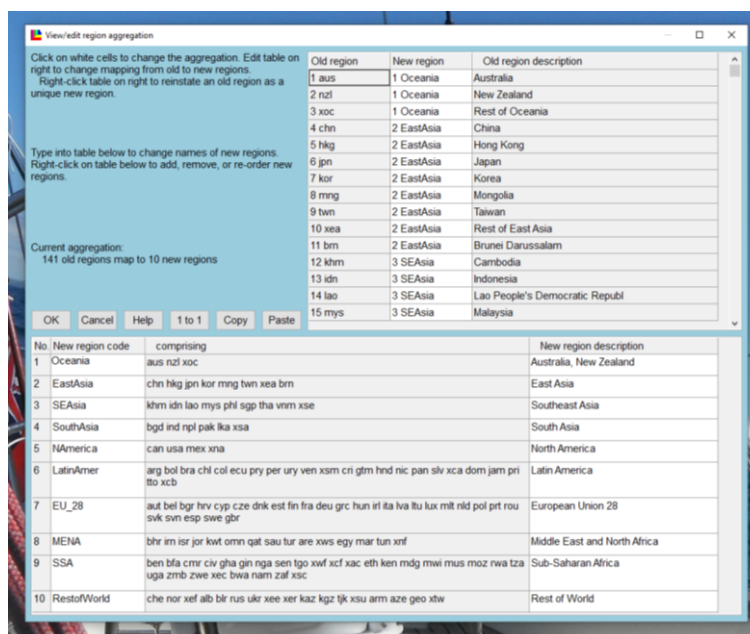
Figure 5.2 Data Subdirectory with GTAP Database

Name	Date modified	Type	Size
 basedata.har	16/08/2019 08:16	Header Array file	91,936
 baserate.har	16/08/2019 08:16	Header Array file	25,536
 baseview.har	16/08/2019 08:16	Header Array file	96,317
 co2.har	16/08/2019 08:17	Header Array file	392
 data_open.txt	02/11/2019 17:03	Text Document	1
 default.har	16/08/2019 08:16	Header Array file	244
 gsdvole.har	16/08/2019 08:17	Header Array file	936
 GTAP10_GTAP_2014_65x141.zip	16/08/2019 08:18	WinZip File	118,179
 gtapsam.har	16/08/2019 08:17	Header Array file	64,018
 gtp65_141.agg	16/08/2019 08:16	AGG File	20
 metadata.har	25/07/2019 16:52	Header Array file	16
 sets.har	16/08/2019 08:16	Header Array file	22

Using a User Defined ‘Aggregation Scheme’

The aggregation scheme we defined is designed to extract all the GTAP data in a completely disaggregated format and uses a set of **factor** labels that need to be defined by the user (see below). The description below assumes the user wants to extract all the GTAP data in a completely disaggregated format but wishes to use different labels for the factor names.

Open GTAPAgg, see Figure 5.1, and select “View/change regional aggregation”. The default shows a 10*10 aggregation. To choose a 1 to 1 aggregation, i.e., for NO aggregation of the region accounts select the **1 to 1** button, see Figure 5.3. Now select “View/change sectoral aggregation”. The default shows a 10*10 aggregation. To choose a 1 to 1 aggregation, i.e., for NO aggregation of the sectoral accounts select the **1 to 1** button.

Figure 5.3 Sector and Region Aggregation 1

Now select “View/change factor aggregation”. The default shows a mapping from 8 factors to 5 factors, where the aggregation scheme aggregates the 5 labour types to 2 labour types, see Figure 5.4. For this aggregation there is no **1 to 1** button to select. Therefore, for a 1 to 1 aggregation the user needs to make the changes manually.

Table 5.1 ‘GLOBE Default’ Factor Labels

Label	Description
tech_aspros	Technicians& Associated professionals
clerks	Clerks
service_shop	Service & Shop workers
off_mgr_pros	Officials & Managers
ag_othlowsk	Agricultural and Unskilled

This requires two steps for each labour type. Right click on the lower table and insert three more rows between the original rows 2 and 3. In the column headed ‘New factor code’ add the labels for the 5 labour types. In order to avoid making changes in the Excel workbook

“data_glbsam_10_2014.xlsx” the labels in Table 5.1 must be used. In the column headed “ETRAE value or “mobile”” type mobile for each labour type.¹⁸

Figure 5.4 Factor Aggregation GTAP Default

Click on white cells to change the aggregation. Edit table on right to change mapping from old to new factors. Right-click table on right to reinstate an old factor as a unique new factor.

Type into table below to change names of new factors. Right-click on table below to add, remove, or re-order new factors.

Current aggregation:
8 old factors map to 5 new factors

OK Cancel Help Copy

Old factor	New factor	Old factor description
1 Land	1 Land	Land
2 tech_aspros	3 SkLab	Technicians/AssocProfessional
3 clerks	2 UnSkLab	Clerks
4 service_shop	2 UnSkLab	Service/Shop workers
5 off_mgr_pros	3 SkLab	Officials and Managers
6 ag_othlowsk	3 SkLab	Agricultural and Unskilled
7 Capital	4 Capital	Capital
8 NatlRes	5 NatRes	Natural Resources

No.	New factor code	comprising	ETRAE value or "mobile"
1	Land	Land	-1.000000
2	UnSkLab	clerks service_shop	mobile
3	SkLab	tech_aspros off_mgr_pros ag_othlowsk	mobile
4	Capital	Capital	mobile
5	NatlRes	NatlRes	-0.001000

Once all the new factor labels are entered go to the ‘New factor’ column in the upper table. Click on the first labour type in the column and a drop-down menu will appear. Select the appropriate ‘New factor’ code to match the ‘Old factor’. When complete the screen be identical to that in Figure 5.4.

NOTE THAT IF THE ‘NEW FACTOR CODES’ ARE NOT IDENTICAL TO THOSE LISTED IN TABLE 5.1 THE GAMS EXTRACTION PROGRAMME WILL NOT OPERATE CORRECTLY. IT SHOULD EXIT WITH A COMPILATION ERROR OR ABORT DUE TO FAILING A TEST.

(We cannot guarantee that a user has not created an error we have not found or allowed for.)

¹⁸ The information in this column is used by the GTAP model to characterise the operation of labour markets. The information is irrelevant in many other models and has no impact of the extraction of the GTAP data to GDX.

Figure 5.4 **Factor Aggregation 2**

View/edit factor aggregation

Click on white cells to change the aggregation. Edit table on right to change mapping from old to new factors.
Right-click table on right to reinstate an old factor as a unique new factor.

Type into table below to change names of new factors.
Right-click on table below to add, remove, or re-order new factors.

Current aggregation:
8 old factors map to 8 new factors

OK Cancel Help Copy

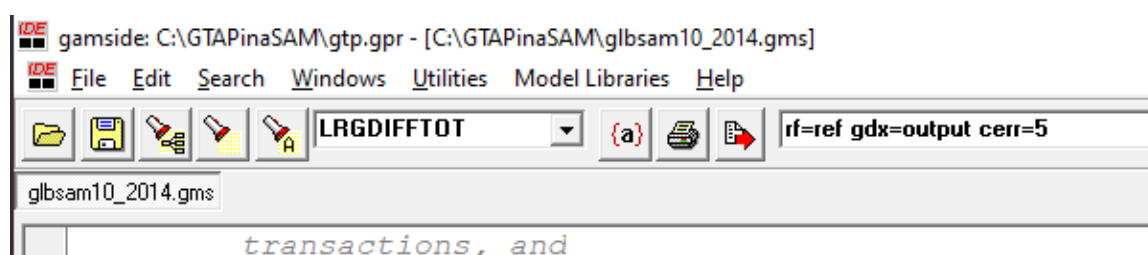
Old factor	New factor	Old factor description
1 Land	1 Land	Land
2 tech_aspros	2 tech_aspros	Technicians/AssocProfessional
3 clerks	3 clerks	Clerks
4 service_shop	4 service_shop	Service/Shop workers
5 off_mgr_pros	5 off_mgr_pros	Officials and Managers
6 ag_othlowsk	6 ag_othlowsk	Agricultural and Unskilled
7 Capital	7 Capital	Capital
8 NatlRes	8 NatRes	Natural Resources

No.	New factor code	comprising	ETRAE value or "mobile"
1	Land	Land	-1.000000
2	tech_aspros	tech_aspros	mobile
3	clerks	clerks	mobile
4	service_shop	service_shop	mobile
5	off_mgr_pros	off_mgr_pros	mobile
6	ag_othlowsk	ag_othlowsk	mobile
7	Capital	Capital	mobile
8	NatlRes	NatlRes	-0.001000

6. Running the Programme

The programme is made available as part of the User Model Library. Working in the GAMSIDE and using the 'GTAPinaSAM' directory created before, open the file `glbsam10_2014.gms`. We recommend adding the following to the command line – “`rf=ref, gdx=output, cerr=5`”; these ensure the production of a ref(ERENCE) file, pass all programme information to the file `output.gdx` and stop the programme once 5 errors are encountered.

Figure 6.1 Command Line



At this stage the most likely errors will be labels, incorrect data file from GTAP or changes made by the user to the file `data_glbsam_10_2014.xlsx`.

Conditioning the Programme

The programme uses the ASK utility twice. The first use is to let the user define the number of decimal places at which the GTAP data are truncated (see Figure 6.2). There are a substantial number of datapoints in the GTAP database that ONLY exist to ensure that models written in GEMPACK run correctly – no entry in any submatrix of the SAM that has a non-zero entry can be zero. In a database where null transactions been agents are equal to zero these non-zero zeros entries can all be set equal to 0.000001 to satisfy the needs of GEMPACK. However, GTAP uses GEMPACK to balance the submatrices and therefore no entry can be less than 0.000001, and some non-zero zeros will be greater than 0.000001.¹⁹

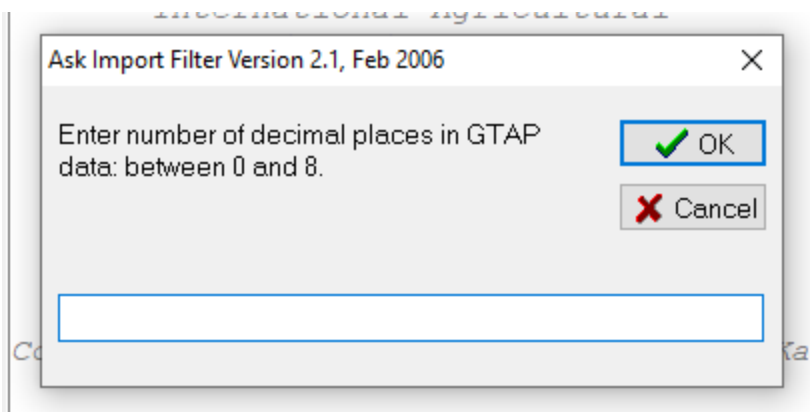
Since an entry of 0.000001 in the GTAP database is \$(USD) 1.00, many of these very small datapoints do not contain economically meaningful data.

¹⁹ GEMPACK is typically single precision and therefore these very small values do not impact on the model. This also means that the GTAP database is not balanced to produce identical row and column totals in double precision, as is standard in GAMS.

BUT, no one knows which non-zero entries have meaning and which are true zeros. We are confident that all entries at 5 and 6 decimal places (0.00001 and 0.000001) are meaningless. After that it is a judgement call; we typically truncate at 2 or 3 decimal places and could easily make an argument to truncate at 0 or 1 decimal places.

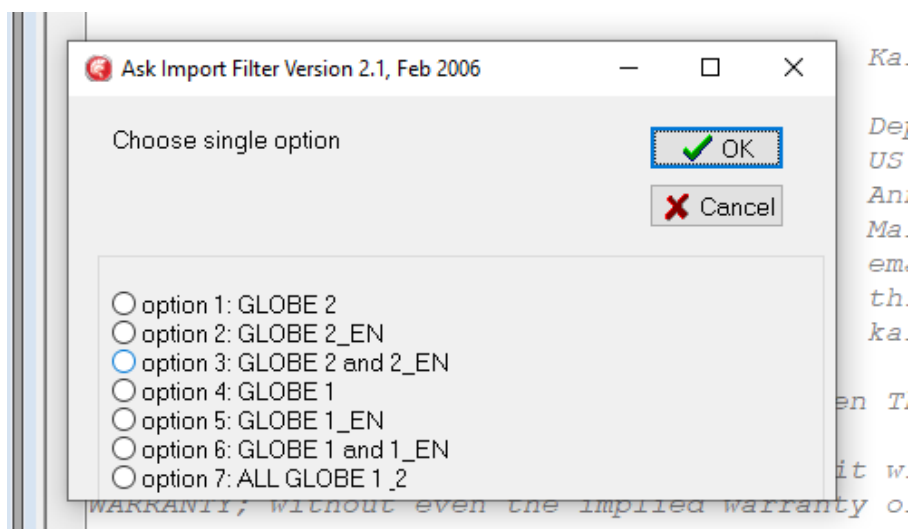
THE DECISION IS THEREFORE LEFT TO THE USER.

Figure 6.2 **Decimal Places for Truncation of Values**



The second use of the ASK utility is to control output. The truncated GTAP database in SAM and satellite account format is produced by default. The user can choose the format for various versions of the GLOBE model (see Figure 6.3).

Figure 6.3 **ASK Utility: GLOBE Model Output Specification**



Outputs

The outputs from the programme are saved in two subdirectories. The truncated GTAP database in SAM and satellite account format is written to the sub directory GTP_output. All the truncated data are reported in the file `gtp_data_10_14.gdx`; in addition, the GTAP SAM in.gdx format, as loaded from the GTAP database is written out as `SAMG_gtp.gdx`.

The sub directory GLB_ouput contains all the variants of the database conditioned for the GLOBE model.

7. Aggregating a SAMs

The dimensions of a SAM may be such as to require, for practical purposes, that the database is aggregated before being used to calibrate a CGE model. There are two main reasons for this:

1. If the database is not aggregated the resultant model would be too large to
 - a. solve within realistic limits of computer capacity,²⁰ OR
 - b. yield results that would be susceptible to practical analyses and interpretation.
2. It is argued that models should be focused upon specific issues rather than being general and that part of the process of focus depends upon the identification of an appropriate aggregation for the purposes at hand.

An aggregation of a SAM requires that the database is (simultaneously) aggregated in three dimensions, i.e., the income and expenditure accounts are simultaneously aggregated to reduce the number commodities/activities, factors, trade partners and associated transactions, and the regions are aggregated. GAMS is an extremely efficient medium for implementing the calculations required to aggregate a large SAM, but setting up the sets and (set) mappings to control the aggregation is potentially time consuming and subject to errors. One method for reducing the time costs involved in setting up a new aggregation, and simultaneously reducing the time required to configure the sets and other data needed to implement a version²¹ of GLOBE with a new aggregation, is to use the SAMgator software (PROVIDE, 2004).

SAMgator is a Visual Basic for Applications (VBA) programme implemented using Microsoft (MS) Excel; as with all VBA applications it can experience problems and hence the provision of another option. The user declares, describes and defines the new (aggregated) sectors (commodities and activities), factors and regions in an Excel template, identifies the source and destination data files and then runs GAMS remotely from within the Excel

²⁰ Developments in solver technologies have allowed solutions to be derived for ever larger models so this reason is increasingly not relevant.

²¹ The term “version of GLOBE” is used to refer to an implementation of a GLOBE model with a specific aggregation of the database. A “version of GLOBE” is used to refer to an implementation of a version of the GLOBE model that contains either ‘limited’ or no variations in the behavioural relationships of the GLOBE model. More substantial changes in the GLOBE model are identified by ‘extending’ the GLOBE name, e.g., GLOBE-AGR.

programme²². The SAMgator programme generates the set and mapping files, checks to ensure that all the mappings are ‘legal’ and then generates the GAMS programme file. In essence all that SAMGator does is convert the set information provided into the sets and maps required to run an aggregation.

Aggregation Guidelines

Some general guidelines for aggregations are given below. In the main these are common sense.

The aggregation of accounts in a SAM is overwhelmingly a matter of the context for which the aggregation is being compiled; consequently, the notes below are general and are not a replacement for economic reasoning and/or common sense.

- Aggregations across the categories (commodities, activities, factors, regions) of accounts are NOT economically appropriate even though they are mathematically possible. SAMgator allows such aggregations. Checks in the GLOBE model code will often identify inappropriate aggregations, but
- Aggregations within categories should typically avoid mappings that are difficult to rationalise, e.g., aggregating unskilled labor with capital and keeping skilled labour separate, aggregating an agricultural commodity with a heavy manufacturing commodity.

SAMgator

At the heart of SAMgator is a single (GAMS) equation that aggregates the SAM database in three dimensions (see below).

$$\begin{aligned} \text{NEWSAM}(sp, spp, k) &= \text{SUM}((ss, ssp, kk) \\ &\quad \$(\text{MAPSAMAG}(sp, ss) \\ &\quad \$\text{MAPSAMAG}(spp, ssp) \\ &\quad \$\text{MAPREGAG}(k, kk)), \\ &\quad \text{SAMG}(ss, ssp, kk)) ; \end{aligned}$$

²² The GAMS code at the heart of SAMgator is available from www.cgemod.org.uk ..

The parameter $SAMG(ss,ssp,kk)$ contains the disaggregated database where ss ²³ is the set that defines the row and columns labels for the SAM and kk the regions. The parameter $NEWSAM(sp,spp,k)$ contains the aggregated database where s is the set that defines the row and columns labels for the aggregated (NEW)SAM and k the aggregated regions. The set $MAPSAMAG(sp,ss)$ defines the members of ss that aggregated into sp by the rows of $SAM(ss,ssp,kk)$, while the set $MAPSAMAG(spp,ssp)$ defines the members of ssp that aggregated into ssp by the columns of $SAM(ss,ssp,kk)$. (Note that there is only one mapping set $MAPSAMAG$ with its implementation being solely driven by the sets ss and s and their aliases). The set $MAPREGAG(k,kk)$ defines the members of kk that aggregated into k by the ‘sheets’ of $SAM(ss,ssp,kk)$.

Using SAMgator

The user interface to SAMgator consists of three worksheets ‘Control’, ‘MappingConfig’ and ‘Sets’. A user only needs to access these three worksheets; although more experienced users may wish to use other features of SAMgator the discussion here is limited to these three worksheets.

The ‘Control’ worksheet is used to identify the location of GAMS.EXE, input and output file names, check for errors, write output and to run the aggregation programme. A screen shot of the worksheet is given in Figure 71.

In order to run GAMS from within MS Excel it is necessary for Excel to know the location of the GAMS.EXE file; this information is provided in the ‘GAMS.EXE location’ box.²⁴ In order to test that Excel can implement a GAMS programme first click on the ‘Pause after execution’ box – so it has a tick in the box as in Figure 7.1 – and then click on the ‘Test GAMS’ button. A DOS window will appear (Figure 7.2); if this indicates a normal completion this is confirmation that Excel can trigger GAMS correctly and the user can press any key to continue.

²³ The set ssp is an alias for ss .

²⁴ With some versions of GAMS spaces in the paths can lead to problems. Note how the screen shot indicates that GAMS.exe is stored in a GAMS directory on the C drive and not in the ‘Program Files’ directory that is the default destination when installing GAMS.

Figure 7.1 Control Worksheet for SAMgator

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O												
Aggregation program for global SAM from GTAP data										MODEL - GLB V2 and GLB V2_en																
This version also generates weighted averages of the GTAP elasticities ESUBD, ESUBM and ESUBVA. A correctly formulated input database is required.																										
Options:																										
GAMS EXE location:										C:\GAMS\GAMS.EXE					Test GAMS											
Pause after execution:										<input type="checkbox"/>																
<table border="1"> <thead> <tr> <th></th> <th>.GDX file name</th> <th>Parameter name</th> <th>Set Name</th> </tr> </thead> <tbody> <tr> <td>Input</td> <td>glb_samg_v2.gdx</td> <td>SAMG</td> <td>ss</td> </tr> <tr> <td>Output</td> <td>samg19_8_19.gdx</td> <td>SAMG</td> <td>sac</td> </tr> </tbody> </table>																.GDX file name	Parameter name	Set Name	Input	glb_samg_v2.gdx	SAMG	ss	Output	samg19_8_19.gdx	SAMG	sac
	.GDX file name	Parameter name	Set Name																							
Input	glb_samg_v2.gdx	SAMG	ss																							
Output	samg19_8_19.gdx	SAMG	sac																							
Aggregation program (.gms): globaggr19_8_19.gms																										
Write output file					Write to Sets sheet																					
Check for errors																										
Run aggregation																										

The user should now define the input file, which is required to be a GAMS Data Exchange (GDX) file; in this illustration the file is called 'glb_samg_v2.gdx'.²⁵ Then name the output file; in this illustration 'samg_19_8_19.gdx'.²⁶ In both instances it is necessary for the user to identify the parameter within the GDX file that is the source of the data and the destination and the set labels used in the source and destination parameters – if the user has not changed the sets and parameters in the supplied programme these labels will not need changing. And finally, the user should provide a name for the GAMS programme file; in this illustration 'globaggr_19_8_19.gms' – by saving the GAMS programme that generates the aggregation the user preserves all the information that documents the aggregation mappings and can regenerate the aggregation without accessing SAMGator.

²⁵ GAMS has changed the formatting of GDX files over recent years. The changes are backward compatible but not forward compatible. This can lead to problem if the user wishes to access a more recent version of a GDX file that is not consistent with the version of GAMS being used; this should be rare. If it does happen there are two most obvious solutions; (1) recompile the input database using the more recent version of GAMS or (2) recompile the input database using the options in GAMS to write out GDX files in the earlier formats.

²⁶ Our habit is to use the format samg_C_F_R.gdx where C, F and R are the numbers of sectors, factors and regions in the aggregated database.

Figure 7.2 **Test GAMS**

```

C:\WINDOWS\SYSTEM32\cmd.exe
*** GAMS Development Corporation
*** 2751 Prosperity Ave, Suite 210
*** Fairfax, VA 22031, USA
*** +1 202-342-0180, +1 202-342-0181 fax
*** support@gams.com, www.gams.com
***
*** GAMS Release      : 26.1.0 rf2b37b9 WEX-WEI x86 64bit/MS Windows
*** Release Date     : Released Feb 2, 2019
*** To use this release, you must have a valid license file for
*** this platform with maintenance expiration date later than
*** Feb 2, 2019
*** System Directory : C:\GAMS\
***
*** License          : C:\GAMS\gamslice.txt
*** Scott McDonald   : S181220/0001AN-GEN
*** Humboldt-University of Berlin, Intern. Agricultural Trade and Dev
*** DC14680 01COM5GEPT
*** License for teaching and research at degree granting institutions
*** mcdonald@hu-berlin.de
***
*** Licensed platform      : Generic popular platforms
*** The installed license is valid.
*** Evaluation expiration date (GAMS base module) : Dec 15, 2019
*** Note: For solvers, other expiration dates may apply.
*** Run gamslib model licememo for more details.
*** Status: Normal completion
--- Job ? Stop 11/03/19 15:38:58 elapsed 0:00:00.007
C:\GTAP 10 4 GLB\GLB_GTP SAMgator v10>pause
Press any key to continue . . .

```

Now make sure that a file called `data_C_F_R.xlsx` is in the directory `2_output` and that the file has no data, but does have multiple worksheets. A copy of this file with no data is included as `data_C_F_R - Copy.xlsx` is included in case the base version is overwritten.

MappingConfig

The role of the ‘MappingConfig’ worksheet is to generate the sets for the aggregated SAM and set up the mappings, which will control the aggregation. The process has several safeguards built in to ensure there the mappings are legitimate and that there are no errors. Although users may choose to adapt a previous aggregation the description given here assumes that the user starts with a version of SAMgator that contains no aggregation specific set or mapping data.

Figure 7.3 illustrates how the ‘MappingConfig’ worksheet might look when there is no aggregation specific set or mapping data. The user needs to create sets and mappings for the commodities/activities, factors and households.

Figure 7.3 Mapping and Configuration Worksheet 1

The steps undertaken to declare and define aggregate sets and to generate the mapping file are as follows.

1. Select a category: choose commodities/activities, factors or regions in the 'Category' box (top left).
2. Declare (and describe) an aggregate set element: type the name in the 'Name' box and below it a description in the 'Descr' box and then click the 'Add' button. This element will be a member of an output set for which each element will appear with its description in the 'Elements in output set for this category' box.
3. Select elements in the input set that are to be mapped to the selected member of the output set: select an element in the 'Elements in output set for this category' box and then select those elements that are to be mapped to that output set element from the 'Unmapped elements of input set' box (bottom right) by clicking on each in turn (NB the SHIFT and Ctrl click options do not operate). Click on the <- button to define the mapping. Note how the selected element of the input set are moved to the 'Elements in input set mapped to selected element in output set' box (bottom left).

- The <- button moves selected elements in the unmapped element box to the mapped elements box; the -> button moves selected elements in the mapped element box to the unmapped elements box; the <<< button moves all elements in the unmapped element box to the mapped elements box; >>> button moves all elements in the mapped element box to the unmapped elements box.
 - Figure 7.3 illustrates that for the category ‘c – Commodities’, which includes activities.
4. Continue adding elements to the output set in each category until all the elements in the input set are mapped to elements in the output set. Note how each element in the input set can only be mapped to one element in the output set – thereby avoiding ‘double counting’. The error checking also checks to avoid duplicates in the output set.
 5. Ordering of the output set: as each new element of an output set is declared it is added to the end of the list of members of the output set for that category and this defines the set ordering that will be used by GAMS. The order of elements within an output set can be altered by selecting elements in the ‘Elements in output set for this category’ box and using the ‘Move Up’ and/or ‘Move Down’ buttons.²⁷
 6. Deleting an output set element: select the element to delete in ‘Elements in output set for this category’ box and click on the ‘Delete element’ button.
 7. Select another category and complete steps 2 to 7 as appropriate until output sets and mapping sets for all three categories have been generated.

Sets

The worksheet ‘sets’ provides a listing of the set names and descriptions required by the GLOBE model. These can be copied and pasted into the worksheet that contains set information for GLOBE. This worksheet has been largely superseded by code that writes out sets, maps and default/dummy elasticity parameters to a separate Excel workbook.

²⁷ Note that GAMS lists set members in the order that they were declared in the programme. If the same name, say ‘abc’, is used in 2 different sets this can create ordering problems for the user: assume that the order wanted for reports is the order of members in the second set declared, but because ‘abc’ is in the first set declared it will be the first member of the second set. Use names for members of sets that are unique to the sets to which they belong.

Running the Programme

The programme can be spawned from within Excel.

There are FOUR Command buttons in the lower left-hand side of the Control sheet (roughly in cells B25:F31). These four buttons should be run in sequence (not critical but valuable) doing so for each row in turn starting each time with the left-hand button (in the order of the list below).

1. Write output file – this button generates the `**.gms` file
2. Write to Sets sheet – this button writes information to the worksheet sets
3. Check for errors – this button checks for mapping errors, e.g., same input label mapped to two different output labels, etc.
4. Run aggregations – this button spawns the GAMS programme

What to do if the programme does not generate the expected outputs. While the checks are reasonably comprehensive, they will not always capture all errors. Since the purpose of SAMGator is to generate a programme file that runs the aggregation, it must create a `*.gms` file. This will be written out to the directory in which the SAMGator Excel file is located. If the `*.gms` file is created, which it should do if the check for errors is passed, the simplest way to identify the cause of problems is to run the programme in GAMS, with `rf=ref gdx=out cerr=5` in the command line (`cerr=5` will cause the programme to stop after 5 compilation errors). Most of the problems encountered are due to typing errors with labels.

Outputs

The programme produces two outputs, in addition to the standard outputs from a GAMS programme.

GAMS Data Exchange (GDX) Database

The GDX file produced by the SAMgator programme contains an aggregated SAM. The GLOBE model assumes that the user accesses the SAM data directly from GDX. The GDX database also includes other information from the GTAP database that can/will be used by the GLOBE model.

GLOBE Model Excel file

The aggregation programme also writes out a lot of information to an Excel workbook that is used by the GLOBE model. The reserved name for this file in its empty state is data_C_F_R.xlsx. After the SAMGator programme has been run this file should be saved with a unique name that, ideally, links it to the *.gms that generated the aggregation.

Aggregation without Using SAMGator

Not only are VBA programmes somewhat ‘flaky’, there are users who object to spending a couple of hours learning to use SAMGator but would rather spend time setting up the sets and maps manually. A set of files to help users who want to work this way are provided. BUT, while SAMGator is supported we do not support the process of aggregating ‘manually’ – sorting out the errors people make with this process is far too time consuming.

Excel File for Sets and Maps

The Excel file is set up to take advantage of the regularity of the GTAP database: there are effectively three ‘dimensions’ to the GTAP database – the sectors, factor and regions – from which all the aggregated sets are derived. In addition, there are a series of ‘institutional accounts that do not vary.

Figure 7.4 New_Sets Worksheet

A	B	C	D	E	F	G	H	I	J
1	Sets								
2	Sets for the Aggregated database								
3									
4	s	(description)	k	(description)	c	a	f	t	
5	ccereal	grains wheat and paddy rice	H_Asia	high income Asia	ccereal	acereal	fland	tmH_Asia	t
6	coillseed	oilseeds	China_HK	China and Hong Kong	coillseed	aoillseed	fl_tech	tmChina_HK	t
7	cothr_ag	other agriculture	Othr_Asia	other Asia	cothr_ag	aothr_ag	fl_clerks	tmOthr_Asia	t
8	cmining	mining	India	India	cmining	amining	fl_svc	tmIndia	t
9	cfood	food beverages and tobacco	Othr_S_Asia	other south Asia	cfood	afood	fl_prof	tmOthr_S_Asia	t
10	ctextl	textiles apparel and leather	Canada	Canada	ctextl	atextl	fl_agr	tmCanada	t
11	cchem	chemicals	US	United States and Puerto Rico	cchem	achem	fcap	tmUS	t
12	cfin_mfg	final manufactured goods	Mexico	Mexico	cfin_mfg	afin_mfg	fnres	tmMexico	t
13	cint_mfg	intermediate manufactured goods	Brazil	Brazil	cint_mfg	aint_mfg		tmBrazil	t
14	cautos	autos	Othr_L_Amer	other Latin America	cautos	aautos		tmOthr_L_Amer	t
15	ctrnsp_mfg	other transport	C_Amer	Central America	ctrnsp_mfg	atrnsmp_mfg		tmC_Amer	t
16	cmach	electronics and machinery	EU_28	EU_28	cmach	amach		tmEU_28	t
17	cenergy	energy	Othr_Eur	Other Europe	cenergy	aenergy		tmOthr_Eur	t
18	cothr_svc	other services	FSU	Former Soviet Union	cothr_svc	aothr_svc		tmFSU	t
19	ctrnsp_svc	transportation services	MENA	Middle East and North Africa	ctrnsp_svc	atrnsmp_svc		tmMENA	t
20	cbus_svc	business services	Nigeria	Nigeria	cbus_svc	abus_svc		tmNigeria	t
21	acereal	grains wheat and paddy rice	SACU	South African Customs Union				tmSACU	t
22	aoillseed	oilseeds	Othr_Afr	Rest of Africa				tmOthr_Afr	t
23	aothr_ag	other agriculture	Rest_W	Rest of the World				tmRest_W	t
24	amining	mining	Glb	Globe				tmGlb	t

The user decides on the aggregation for the current study and then defines the sets in the worksheet ‘New_Sets’, see Figure 7.4. Once the aggregated sets are defined it is possible to populate this worksheet. But, while it is logical to start with the ‘New_Sets’ worksheet, it is

easier to populate this worksheet at the end of the process, and start with the ‘Map create’ worksheets.

There is one worksheet for each ‘dimension’: ‘Sector map Create’, ‘Factor map Create’ and ‘Region map Create’ and each is setup in the same way. Figure 7.5 shows the Sector map Create sheet. In column B the sector set members from GTAP are listed, with the descriptions in column A. The contents of Columns A and B only change with version of the GTAP database. Column C identifies the aggregate sectors into which the GTAP sectors are aggregated, with the associated descriptions in Column D. In the example workbook the aggregation mapping is the same as that used with SAMGator.

Figure 7.5 Sector Map Create

Creation of Sector Maps								
	A	B	C	D	E	F	G	H
1	Creation of Sector Maps							
2								
3	GTAP Sectors			Aggregate Sectors			Automatic fill to MAPS	
4	(description)	gtp	s	(description)		c	a	tss
5	Paddy rice	pdr	cereal	grains wheat and paddy rice		ccereal	acereal	tsscereal
6	Wheat	wht	cereal	grains wheat and paddy rice		ccereal	acereal	tsscereal
7	Cereal grains nec	gro	cereal	grains wheat and paddy rice		ccereal	acereal	tsscereal
8	Vegetables fruit nuts	v_f	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
9	Oil seeds	osd	oilseed	oilseeds		coilseed	aoilseed	tssoilseed
10	Sugar cane sugar beet	c_b	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
11	Plant based fibers	pfb	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
12	Crops nec	ocr	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
13	Bovine cattle sheep and goats	ctl	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
14	Animal products nec	oap	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
15	Raw milk	rmk	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
16	Wool silk worm cocoons	wol	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
17	Forestry	frs	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
18	Fishing	fsh	othr_ag	other agriculture		cothr_ag	aothr_ag	tssothr_ag
19	Coal	coa	mining	mining		cmining	amining	tssmining
20	Oil	oil	mining	mining		cmining	amining	tssmining
21	Gas	gas	mining	mining		cmining	amining	tssmining

The user works down Column C identifying the member of the aggregated database into which the original GTAP sectors are aggregated. Columns F, G and H will then be populated to define the members of the subsets *c*, *a* and *tss* of the set *s*. The same process is then followed for the worksheets ‘Factor map Create’, which populates the subsets *f* and *tff*, and the worksheet ‘Region map Create’, which populates the sets *k*, *tmr*, *tmrs*, *ter*, *w* and various permutations of the international transport margins. A complication is that that there are six alternative international margins aggregation depending on how the three international margin commodities are aggregated (3 factorial); all 6 possibilities are populated.

The user can now populate the ‘New_Sets’ worksheet using copy and paste special – values. Note that the user needs to decide which of the international margin aggregates are

required. As a precaution the non-changing members of the set s are reported with descriptions in columns X and Y.

The two mapping sets, $\text{mapsamag}(s,ss)$ and $\text{mapregag}(k,kk)$, are largely automatically populated. Only international margin elements of the set mapsamag requiring the user to manually populate the mapping; these need copy and paste from the relevant columns in the worksheet 'Region map Create' into column C; the section is highlighted in GREEN. NB the elements in columns D, F, K and M only change with changes in the GTAP database version.

The workbook is 'protected' to reduce the likelihood of errors. The password is 'Globe'.

Running the GAMS Programme

The code for the aggregation is provided as the file `globagg_GAMS.gms`. The user needs to change the text for the `$SETGLOBAL` elements to match the file names for the GDX disaggregated data file and the Excel file that provided the set and mapping information.

The structure of the output is the same as that for SAMGator.

8. Concluding Comments

This User Guide covers the conversion of the GTAP database into a Social Accounting Matrix (SAM) and satellite account format. It is argued that such a transformation of the GTAP database is a great help in making the content of the GTAP database more accessible while making it far easier to appreciate the accounting relationships that underpin ALL global models that use GTAP data.

In addition the guide also provides information on programmes that facilitate the aggregation of the database.

References

- Corong, E.L., Hertel, T.W., McDougall, R., Tsigas, M.E. and van der Mensbrugghe, D., (2017). 'The Standard GTAP Model, Version 7', *Journal of Global Economic Analysis*, Vol 2 (1).
(<https://jgea.org/resources/jgea/ojs/index.php/jgea/article/view/47>)
- King, B.B., (1985). 'What is a SAM?', in Pyatt, G. and Round, J.I. (ed), *Social Accounting Matrices: A Basis for Planning*. Washington: World Bank.
- McDonald, S. and Thierfelder, K., (2004). 'Deriving a Global Social Accounting Matrix from GTAP version 5 Data', *Global Trade Analysis Project Technical Paper 23*, 2004 Center for Global Trade Analysis: Purdue University.
- McDonald, S. and Thierfelder, K., (2019). 'Deriving a Global Social Accounting Matrix from GTAP', *CGEMOD Technical Paper 12*. Cgemod. ()
- PROVIDE (2004). 'SeeResults: A Spreadsheet Application for the Analysis of CGE Model Results', *PROVIDE Project Technical Paper 2004:1*. PROVIDE Project: Elsenburg.
- Pyatt, G., (1991). 'Fundamentals of Social Accounting', *Economic Systems Research*, Vol 3, pp 315-341.
- Sadoulet, E. and de Janvry, A., (1995). *Quantitative Development Policy Analysis*. Baltimore: John Hopkins University Press.
- UN, (1993). *System of National Accounts 1993*. New York: UN.