Deriving a Global Social Accounting Matrix and Satellite Data from GTAP 11 Databases

CGEMOD Technical Paper 15

Scott McDonald[1]

Abstract

This paper describes how the GTAP 11 databases can be converted into global Social Accounting Matrices (SAM) with linked satellite accounts; the GDX version of GTAP 11 does not include a SAM representation. The focus of attention is on the generation of a SAM representation of the GTAP database that is consistent with the GTAP model. The resultant SAM can be used to calibrate global CGE models that have been coded in GAMS or GEMPACK. Since GTAP version 7, the GTAPAgg programme has produced the transaction values in the GTAP database as a SAM using the method described in this paper. As of mid-2019, the GAMS codes used to generate the SAM representation of the GTAP database are available for GTAP versions 5 to 10 inclusive.

Correspondence Address:

e: jrs.mcdonald@gmail.co.uk

---

[1] This paper is a development of McDonald and Thierfelder (), and substantial parts of the text draw on the previous papers. The author acknowledges the explanations of the GTAP model and GEMPACK code provided by Terrie Walmsley and the identification of an error in the code by Vitaly Kharitonov.

Table of Contents

1.   Introduction

The GTAP database is the major source of data used by global Computable General Equilibrium (CGE) models. Version 11 of the GTAP database is supplied in the Header Array (HAR) file format, which is used by GEMPACK, and GAMS Data Exchange (GDX) file format, which is used by GAMS. This paper is concerned only with rendering the all the GTAP 11 in GDX formatted data into a global SAM with associated satellite accounts as a single GDX file.[2]

Since the databases for ALL whole economy models can always be represented in a Social Accounting Matrix (SAM) format, and many modelers choose to present their model databases in SAM format, a SAM format has attractions. The main reasons for developing a Social Accounting Matrix (SAM) version of the GTAP database are

1. the greater accessibility of the information in the database to modelers and policy makers;

2. the explicit identification of the price system and processes that generate the price definitions;

3. the increased use by economists of the SAM format, especially now that it is formally part of the System of National Accounts (SNA) (UN, 1993 and 2008); and

4. the greater ease with which the data for a single region can be assessed and related to national account aggregates.

The GTAP 11 database in GDX does not include a representation of the data in SAM format. Moreover, the GDX data uses multiple GDX files for which the account labels are not consistent with those that would be used in a SAM format. The GAMS code provided addresses both these issues and reports all the data as a single GDX file in SAM and satellite account format.

The transformation of the GTAP database to a SAM format does involve some complications. These arise because the GTAP database does not appear, at first sight, to observe the 'law of one price', i.e., 'buyer' prices are not common across all agents, and the

---

[2]   It is believed that the GTAPAgg routine produces a global SAM based on a revised version of the GEMPACK code that was developed from McDonald and Thierfelder (2004) by Walmsley. This has not been checked.

fact that GTAP does not adopt the price definitions set out in the SNA.[3] Thus when developing a SAM representation, it is necessary to identify the price system used in the GTAP database.

SAMs that are consistent with the SNA use three key price definitions - purchaser, producer and basic prices. Historically, the GTAP system used three different price definitions - agent, market and world prices. In the GTAP literature for an imported commodity, the values of transactions at 'market' prices were defined as the values at 'world' prices plus trade and transport margins plus import duties, and the values at 'agents' prices were then defined as the values at 'market' prices plus any sales/purchase taxes. For domestic commodities, the values of transactions at 'agents' prices for domestic consumption were the values at 'market' prices plus any sales/purchase taxes, while for exports the values at 'world' prices were the values at 'market' prices less any export taxes. With the release of version 11 GTAP has changed the language with respect to price definitions but the 'meanings' remain consistent with those used historically.[4] The relationship between the SNA and GTAP price systems is not immediately obvious.

The guiding principle followed in converting the GTAP database to a SAM is to ensure that the 'law of one price' holds in the resulting SAM, i.e., that all transactions in each row are valued at a common price. Thus, for example, purchasers of commodities for consumption/demand by domestic agents, make separate payments to the supplying agents (domestic or foreign) and the government with respect to indirect taxes on commodities. This approach ensures that the national accounting identities for the SAM are maintained, while transforming the GTAP database into a global SAM format that is FULLY consistent with the price and tax systems in the GTAP model.[5]

A SAM representation has the further advantage of making it easier to augment the transactions data and relate the information in satellite accounts to specific transactions. Examples of data augmentation include separately identifying various inter regional current account transactions, e.g., migrant remittances, aid transfers, etc., (see McDonald and

---

[3]   The issue of prices is explored further below.

[4]   The change in language may herald a more substantive change for later variants, but for now the change in language makes no difference to the price system in GTAP.

[5]   The GTAP database has always been constructed to be consistent with the GTAP model. This arguably unfortunate. The decision to make the global SAM consistent with the GTAP model is NOT an endorsement of the SAM structure.

Sonmez, 2004; Walmsley *et al*., 2007; McDonald *et al*., 2016) and decomposing the transfers between domestic institutions, e.g., savings and income taxes paid by households and government borrowing/savings (see McDonald *et al*., 2004; Walmsley & Minor, 2013).

The programmes used to produce the SAM representation for GTAP version 11 were developed in GAMS (using Studio as the editor) and are freely available in a zipped archive. The programmes for previous versions are available in the form of a GAMS User Library;[6] these programmes will have limited future use given the GAMS's decision to cease supporting HAR2GDX by GAMS. The programmes include checks to ensure data consistency, but users are encouraged to carry out checks for themselves.[7] The SAM representation detailed in this paper was first developed for GTAP version 5 and, with minor adjustments that mostly reflect changes in the labels used for parameters/coefficients, is essentially unchanged. The rather extensive changes to the parameters used for GTAP 11 are actually little more than changes to names. The GAMS programmes convert all GTAP data, transactions and other data. in the standard GTAP database into a global SAM with satellite accounts that are saved in a single GDX file.

The rest of this paper is organised as follows. The next section provides a description of the Social Accounting Matrix. The pertinent aspects of the data provided by the GTAP database are discussed in section 3, while section 4 provides a description of how the matrices of data from the GTAP database are used to construct the SAM. The SAM produced by the programme does not balance exactly; since this is a necessary condition for a 'true' SAM section 5 reports the magnitudes and relative importance of differences in the account totals for both SAMs, which are very minor, and then reports the steps taken to ensure balanced SAMs are produced. The maximum and minimum transactions values for each version of the database by region are reported in section 6, while section 7 offers some concluding comments.

---

[6]   The GAMS programmes can all be run using demo/student GAMS that is available from the GAMS website (www.gams.com). Early versions of the code, up to an including version 7, used HAR2GAMS and GAMS2HAR, to convert between HAR and GDX formats; these programmes were written by Mark Horridge and Tom Rutherford and versions were subsequently made standard parts of the GAMS language. Earlier code for use with version 5.0 of the GTAP database used SEEHAR, written by Ken Pearson for converting the HAR file into a GAMS readable input file. These programmes were converted to GEMPACK code by Terrie Walmsley and incorporated into the GTAPAgg programme by Mark Horridge to produce a SAM.
For GTAP versions 9 and 10 the disaggregated global SAM from GTAPAgg were used.

[7]   If errors are uncovered, please let the authors know; preferably with a copy of the code used to generate and detect the errors.

2.   Price Definitions: GTAP & System of National Accounts

An important feature of all SAMs is the explicit identification of the system of prices. An understanding of the system of prices in a SAM is important if the data are to be used appropriately in any economic model and is critical to the development of the price system in any CGE model. The price system apparently used in the GTAP database is not consistent with the price system used in the System of National Accounts (SNA). Although the price systems have a degree of commonality, it is important to appreciate the differences between the two systems when interpreting the GTAP data and relating the reported values to (most) published national accounts. It is convenient to start with a brief description of the system of prices in the SNA and then review the system used in the GTAP database; this order allows an easier explanation of the differences.

SNA Price System

This section explains the price system in the System of National Accounts and its implications for the interpretation of a SAM consistent with the price system in the SNA. As part of this explanation, it is important to understand the distinction between prices in input-out tables (IOT) and Supply and Use tables (SUT).

Prices in the SNA

The SNA defines three key prices (see SNA, 2008, 6.49 to 6.69):

1. Purchaser Prices - the prices paid by purchasers, which include transport and distribution margins and any VAT payable.

2. Producer Prices - the price paid to the producer by the purchaser less any VAT or other deductible tax and any transport or distribution charges invoiced separately.

3. Basic Prices - the price paid to the producer by the purchaser less any tax payable plus any subsidy receivable but excluding any transport or distribution charges invoiced separately.

The relationships between these prices are summarised in Figure 2.1.

Figure 2.1   **SNA Price Relationships**

<div align="center">

Basic prices

+

Taxes on products excluding invoiced VAT

-

Subsidies on products

=

Producers' prices

+

VAT not deductible by the purchaser

+

Separately invoiced transport charges

+

Wholesalers' and retailers' margins

=

Purchasers' prices

</div>

Source: SNA (2008), Figure 6.1, p 103.

The key prices for a CGE model are the basic and the purchaser prices. Typically, in a SAM consistent with the SNA, commodities supplied to an economy by domestic activities and the rest of the world will be valued in basic prices, while commodities used by an economy will be valued in purchaser prices. It is useful to see how these prices are recorded in a SAM, but to do so requires determining the prices paid by agents when purchasing commodities.

If a SAM is constructed around Supply and Use tables (SUT), i.e., the inter-industry Use matrices can be asymmetric[8], then prices for each commodity will be **purchaser** prices and, implicitly, identical for all agents. Domestic production and imports, *cif* paid plus import duties/subsidies, are valued at basic prices.

If a SAM is constructed around input-output tables (IOT), i.e., the inter-industry Use matrices are symmetric, and the SAM follows SNA principles, then, in one possible arrangement, prices in the rows for each commodity will be **basic** prices and identical for all agents. The derivation of purchaser prices for each agent then requires the existence of additional information that identifies the various taxes and trade and transport margins

---

[8]   Commodities and activities are separately identified, even if a principal product classification scheme is used for activities, and the Supply matrix is/may be non-square and/or will have entries off the diagonal.

associated with purchases. This could be conceived of as two matrices in which each row corresponds to a commodity and each column to an agent that purchases commodities.

1. The columns of the first set of matrices should contain all the agent specific taxes and subsidies paid by the agent for purchases of each commodity;[9] and

2. the columns of the second set of matrices should contain all the agent specific transport charges and wholesaler and retailer margins paid by the agent for purchases of each commodity.[10]

This type IOT structure is arguably an intermediate stage in the development of the format in which IOT are typically published.[11] If, and only if, these data are available the **purchaser** prices are agent and commodity specific because the data define agent specific taxes and subsidies paid and agent specific transport charges and wholesaler and retailer margins for purchases of each commodity.

In the SUT based method for deriving IOT, the account totals for commodities/products or activities/industries determine the account totals for the IOT. Since the SUT commodity and activity account totals are valued in basic prices, the IOT account totals are in basic prices. In these formulations trade and transport margins are services produced by a trade and transport services industry and purchased by domestic agents, including industries and final demand (see UN, 1999, chapter 4, and Miller and Blair, 2009, pp146-9).

GTAP Price System

With GTAP 11 the implied price system in the GTAP database has the appearance of having changed from the (implicit) definitions for three prices in previous version: Agents prices, Market prices and World prices. These definitions can be defined as

1. Agents prices – the agent specific prices paid by agents for commodities;

2. Market prices – the prices of imported commodities delivered to a region inclusive of international trade and transport costs and any import duties/subsidies, **and** the prices

---

[9]   Each matrix in the set would relate to a specific tax instrument, e.g., General Sales Taxes (TST), VAT, excise taxes, etc.

[10]   Each matrix in the set would relate to a specific trade and transport margin, e.g. transport charges and wholesaler margins, etc.

[11]   See Chapter 4, and Tables 4.4 to 4.6, in the 'Handbook of Input-Output Table Compilation and Analysis' (UN, 1999)

received by agents when selling commodities on the domestic market or as exports; and

3. World prices – the prices paid for exports after any export taxes/subsidies, and the prices received for imports after the payment of international trade and transport costs but before the payment of import duties/subsidies.

But the relabelling does not actually change the price definitions, as is made evident when transforming the data into SAM format.

Basically, market prices have been relabelled as Basic prices, world prices are now labelled *fob* and *cif*, and agents prices are now Purchaser or Supply prices. An illustration of GTAP's new prices system will be added as Figure 2.2.

Figure 2.2   GTAP Price Relationships

Reconciling the SNA and GTAP Price Systems

The data requested by GTAP in the matrices UF, "usage of input i in use u, commodity tax excluded", are in fact standard IOT data, which should be valued in basic prices, while the matrices UP, "usage of input i in use u, commodity tax inclusive", are defined as the matrices UP plus matrices of total taxes paid by purchasers. If the matrices UF are inclusive of trade and transport margin services, which is the expected format for official IOT, then transactions in UP should be valued in purchaser prices.

There are some complications. First, while GTAP states that "our preference is for commodity by commodity tables" (Huff, *et. al*., 2022) it is evident that GTAP are uncertain whether the IOT supplied are commodity-by-commodity or industry-by-industry tables. But the values at basic prices for the commodity and industry accounts differ, which means that the vectors MF, "imports of commodity i, import duties excluded", that are defined by commodity definitions may be inconsistent. And second, GTAP records import duties as tariff **equivalent** revenues and not as tariff revenues. This means that the total values of the supplies of imports of commodity c to an economy are only equal the basic price valuation if all rents (difference between tariff **equivalent** revenues and tariff revenues) accrue to the source region. But GTAP's formulation presumes the rents accrue to the regional household

account. This means that the prices paid for imported intermediate inputs by domestic agents are inflated by the rents.

GTAP state that "Our terminology (inherited from the Australian Industry Commission's SALTER Project) reflects usage in CGE modeling rather than input-output accounting" (Huff, 2022, p 2).[12] The terminology chosen by GTAP is misleading. Setting aside the complications mentioned above, it transpires that the price system for the data requested by GTAP is roughly consistent with the SNA's basic and purchaser prices. The complications arise from how GTAP present the price system.

---

[12]   It is appropriate to note that most of the single country CGE models do not use this terminology.

3.   Social Accounting Matrix

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 (UN, 1993 and 2009), with adjustments for the limited data on intra-institutional accounts and GTAP's price system. The basic structure is identified in Table 3.1, 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 'basic' prices (GTAP market prices). Associated with each purchase by an agent there is an additional payment to the government that represents sales taxes, i.e., the implicit prices paid are 'purchaser' prices. 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), i.e., 'basic' prices. For domestically produced commodities market prices are the prices received by domestic activities, i.e., 'basic' prices, 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, i.e., 'basic' prices.

Table 3.1   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																

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 taxes[13] 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.

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:

---

[13] 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).

i)   underlying the data for each region is a symmetric input-output table that means the Supply Matrix – sub matrix 3:2 in Table 3.1 – 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 rates[14] 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.

Table 3.2   **SAM Dimensions**

Description	Version 5.4	Version 6.0	Version 7.1	Version 8.1	Version 9	Version 10	Version 11
Commodities	57	57	57	57	57	65	65
Activities	57	57	57	57	57	65	65
Factors	5	5	5	5	8	8	8
Regions	78	87	112	134	140	141	160
Margin Commodities	3	3	3	3	3	3	3
Domestic Institutions	4	4	4	4	4	4	4

For trade relationship, each region **can** import from and export to all other regions, hence for each region there needs to be r accounts to record trade transactions. In addition, there needs to be $2*r$ accounts for each region to capture trade taxes – import duties and export taxes – up to 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.[15] And since there are three transport margins associated with each trade transaction there needs to be $m*r$ 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.

---

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

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

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). Table 3.3 provides a plan of the dimensions of the resulting SAM matrices.

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 consequence of the number of regions, r, in the database (r trade accounts plus 2 or $3r$ for trade taxes) and the fact that m types of transport margins on imports are identified with imports from each region ($m*r$ 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 $(4c + a + 2f + 7 + mr + 3 \text{ or } 4r)* (4c + a + 2f + 7 + mr + 3 \text{ or } 4r)*(r)$, where c and a are the number of sectors in the database, f the number of factors and r the number of regions. The dimensions of the SAM matrices for different versions of the GTAP database are

version 5.4 – 770*770*78 (46,246,200 cells),

version 6.0 – 824*824*87 (59,070,912 cells),

version 7.1 – 974*974*112 (106,251,712 cells),

version 8.1 – 1,240*1,240*134 (206,038,400 cells),

version 9.1 – 1,288*1,288*140 (232,252,160 cells).

Version 10 - 1,335*1,335*141 (251,293,725 cells).

Version 11 – 1,468*1,468*160 (344,803,840 cells)

.

Table 3.3 **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	$1*c$	$1*c$	$1*a$	$1*f$	$1*1$	$1*1$	$1*2k$	$1*c$	$1*c$	$1*f$	$1*1$	$1*1$	$1*1$	$1*1$	$1*m*k$	$1*m$	$1*k$	

Table 3 also highlights several important features of the GTAP database.

- Information is concentrated in the trade accounts.
- The SAM matrices are sparse, i.e., there are only 7,619,480 non-zero entries in the GTAP SAM for version 11.[16]
- 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.

---

[16]   With GTAP 11 the number of non-zero zeros recorded in the database has been greatly reduced where the non-zero zeros are not required when log differencing.

4. The GTAP Data

With version 11 the GTAP database was made available in GDX format with five GDX files for each year; one outcome from this programme is to condense the data into one GDX file.

GTAP and GEMPACK

The solution algorithms used by GEMPACK solve for rates of change; this requires that all cells in any sub matrix of the SAM, for which the model has non linear 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 represent real world transactions.

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 consistent[17]; 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.

Rates of Change

One feature of GEMPACK that might cause discomfort to GAMS modelers is the solution of models in terms of rates of changes – in essence the models are written in terms of log differentials. One area in which this can cause some confusion is the treatment of taxes. Typically, 'levels modelers' might be inclined to talk about the *tax rates*, whereas 'rates of changes modelers' might talk about the *powers of taxes*. The explanation is straightforward. Consider the level relationship for an *ad valorem* tax

$$P_1 = P_2 * (1+T)$$

where *T* is the *ad valorem* tax rate and (1 + *T*) is defined as the **power of the tax**. In terms of rates of change this can be written as

---

[17] The equality of row and column totals in a SAM can be used to identify consistency, the reported transactions are consistent, but NOT that they are complete.

$$d \ln \left(P_1 \right) = d \ln \left(P_2 \right) + d \ln \left(1 + T \right)$$
$$p_1 = p_2 + t$$

such that t is the proportionate change in the power of the tax. It is useful to note this when reading the GTAP code for the tax system.

SETS and Control of SETS

It is important, for GAMS users, to note that when reading the extracts of the GTAP model that the model is coded in GEMPACK where the sets are local – hence the inclusion of local definitions for variables and parameters. To understand the local control of sets consider the following extract of code from a GTAP tablo[18] 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 (see the schematic above), 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.

Regional Household

The GTAP database and model use the concept of a Regional Household. Since this is crucial to an understanding of the relationships between agents it is useful to provide a brief description of the Regional Household concept. In the GTAP database all incomes from taxes (direct and indirect), all factor incomes, depreciation and the trade balance accrue to the Regional Household. These incomes are then distributed between the Private Household[19], the Government and the Capital (saving-investment) accounts, which are the only domestic institutions responsible for the final demand of commodities.[20] Hence, the Regional

---

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

[19]   Implicitly the Private household account encompasses households, incorporated business enterprises and NPISH.

[20]   This distribution of regional household expenditure between private household, government and investment is, in fact, a macroeconomic closure rule embedded in the model. It approximates a balanced

Household is the sole source of income to the private household, the government and the capital accounts. Consequently, the private household does not record transactions directly with either the government or the capital accounts and the government account does not record transactions with the capital or private household accounts; hence in the database the private household neither saves nor pays income taxes and the government does not run a surplus of deficit on its current account.

The Regional Household provides a method by which the limitations imposed by the absence of data on intra-institutional transactions can be partially circumvented.[21] This method does however have a cost in modeling terms; it involves the imposition of a specific distribution of income between domestic institutions.

Trade and Transport Margins

The GTAP database records the transactions with respect to three types of international trade and transport margin for each commodity imported by each destination region from each source region. The sum of the value of these margins for each commodity imported from each source are defined as the value of the difference between the *fob* valuation of the commodity exported from source region *s* to destination region *y* and the *cif* valuation of each commodity imported by destination region *d* from source region *s*. But, the database does not record data on the regions that supply each trade and transport margins transaction, i.e., there is not full information on bilateral trade in trade and transport margins. Instead for the export of trade and transport margins by each region the database only records the total value, whereas for margin imports the database records, separately, the value of the trade and transport cost associated with every import transaction valued at *fob* prices. The global sum of the values of the exports of each trade and transport margin must equal the global sum of the values of the respective imports of trade and transport margins. But it is not necessary that the values of the exports of trade and transport margins by each region must equal the values of the imports of trade and transport margins by each region.

---

macroeconomic closure, although the shares change slightly due to the CDE demand function in the GTAP model.

[21]   McDonald and Sonmez (2004) demonstrate that it is relatively straightforward to compile adequate additional data to at least partially address the absence of data on intra-institutional transactions. McDonald *et al*., (2016) extend this process and identify sources of data.

In the GTAP model this is handled by the concept of a global pool for trade and transport margins. This pool supplies all the demands for (the import of) trade and transport margins and then balances the global market for trade and transport margins by purchasing the margins from the regions that export trade and transport margins. Hence, each region runs a trade account with the global pool for trade and transport margins; *ex post* the sum of the balances on these trade accounts must be zero. Since the total values of margin exports and imports for each region and each margin goods are known, the balance in trade and transport margins trade can be readily calculated.

The Global Bank and Trade Balance

A related issue exists with respect to global savings. The values of total savings and investment within each region must be equal for all regions. But total savings by domestic institutions within each region need not equal the total value of investment within each region because of inter-regional capital account transactions. Furthermore, the *ex post* the balance on the capital accounts must be equal and opposite to the balances on the current account for each region and the global sum of the balances of trade in goods and services – current account - must equal zero, as must the global sum of inter-regional capital account transactions.

This presents economic logic problems for experiments since the logic of a price driven model suggests that differences in investment expenditures across regions would be influenced by differences in returns to capital in different regions. Nevertheless, this is not a substantive issue for the reporting of transactions in the database since it simply records transactions in the base period and these must satisfy the requisite accounting conditions. Hence, for database purposes it is sufficient to record the calculated balance of trade in goods and services for each region with every other region, although issues may remain about the appropriate modeling approach.

5.   Constructing the Social Accounting Matrix

This section describes how the various sub matrices of the SAM are assigned. Table 5.1 presents a map of the assignments where the notation is derived from the GTAP data labels used the GDX file that contains the data. The major issues all relate to the treatment of taxes and hence to the systems by which prices are formed.

The description of the assignment statements starts with the components of domestic demand, and domestic production, then the factor accounts. A large part of the code relates to tax revenues and thence government income, which involves specifying the tax transactions. The discussion then moves onto the purchases of inputs by activities and to the components of regional income. Finally, the discussion turns to the treatment of trade and the associated trade and transport margins and thence the income to the capital account, i.e., the balances on the capital account with respect to the rest of the world.

GAMS Sets

The following extract from the GAMS code identifies the sets used in the programme.

```
ss               GTAP SAMG Accounts
mcc(ss)          Imported Commodities
dcc(ss)          Domestic Commodities
cc(ss)           Aggregate commodities
aa(ss)           Activities

ff(ss)           Factors
ll(ff)           Labour

ttm(ss)          Import duties and Ad valorem equivalent Import duties
ttms(ss)         Specific Import duties
tte(ss)          Export taxes
tts(ss)
ttsm(ss)         Sales Taxes on imported goods and services
ttsd(ss)         Sales Taxes on domestic goods and services
ttff(ss)         Taxes on Factors

ootpm(ss)        Other trade and transport costs
wwtpm(ss)        Water trade and transport costs
aatpm(ss)        Air trade and transport costs
vvst(ss)         Exports of trade and transport services

ii(ss)           Investment

ww(ss)           Rest of the World
rr               Regions

* Energy commodities and taxes
```

```
cc_e(cc)              Aggregate energy commodities
mcc_e(mcc)            Imported energy Commodities
dcc_e(dcc)            Domestic energy Commodities
tts_e(ss)             Sales Taxes on aggregate ENERGY goods and services
ttsd_e(ttsd)          Sales Taxes on imported ENERGY goods and services
ttsm_e(ttsm)          Sales Taxes on domestic ENERGY goods and services
```

The outer set *SS* contains all the within region accounts with a series of subsets, e.g., *MCC*(*SS*) indicates that *MCC* is a subset of *SS*.

There is also a need for large number of mapping sets (although I am sure the number could be greatly reduced)

```
map_mcc_comm(mcc,comm)   Map to import commodities from comm
map_dcc_comm(dcc,comm)   Map to domestic commodities from comm
map_cc_comm(cc,comm)     Map to aggregate commodities from comm
map_cc_mcc(cc,mcc)       Map to aggregate from import commodities
map_cc_dcc(cc,dcc)       Map to aggregate from domestic commodities
map_aa_acts(aa,acts)     Map to activities from at
map_ff_endw(ff,endw)     Map to factors from endw

map_ttm_reg(ttm,reg)     Map to ad val import duties from regions
map_ttms_reg(ttms,reg)   Map to specific import duties from regions
map_tte_reg(tte,reg)     Map to export taxes from regions

map_ww_reg(ww,reg)       Map to trade partners from regions
map_reg_reg(reg,reg)     Map regions to regions

map_rr_reg(rr,reg)       Map to region rr from region reg

map_ttsm_comm(ttsm,comm) Map to import sales taxes from comm
map_ttsd_comm(ttsd,comm) Map to domestic sales taxes from comm
map_ttff_endw(ttff,endw) Map to factor use taxes from endw

map_ootpm_reg(ootpm,reg) Map to other import margins from regions
map_wwtpm_reg(wwtpm,reg) Map to water import margins from regions
map_aatpm_reg(aatpm,reg) Map to air import margins from regions

map_ww_wwtpm(ww,wwtpm)   Map to trade partners from water import margins
map_ww_aatpm(ww,aatpm)   Map to trade partners from air import margins
map_ww_ootpm(ww,ootpm)   Map to trade partners from other import margins
```

Table 5.1 **Mapping Transactions: Social Accounting Matrix for a Representative Country**

		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			VMFB			VMPB							VMGB	VMIB				1:18
2	Domestic Commodities			VDFB			VDPB							VDGB	VDIB		VST	VFOB	2:18
3	Activities		MAKB																3:18
4	Factors			EVFB															4:18
5	Regional Household				EVOS - VDEP			TMTAX + TETAX	ISTAX	DSTAX	FTAX	PTAX	DTAX						5:18
6	Private Household					YH													6:18
7	Trade Taxes	VMTS	VFOB - VXSB																TMTAX TETAX
8	Import Sales Taxes			VMFP-VMFB			VMPP-VMPB							VMGP-VMGB	VMIP-VMIB				ISTAX
9	Domestic Sales Tax			VDFP-VDFB			VDPP-VDPB							VDGP-VDGB	VDIP-VDIB				DSTAX
10	Factor Taxes			EVFP-EVFB															FTAX
11	Production Taxes			MAKB-MAKS											*PTAXINV*				PTAX
12	Direct Taxes				EVFB-EVOS														DTAX
13	Government					YG													13:18
14	Capital				VDEP	SAVE											VTWR-VST	VIWS-VTWR-VXWD	14:18
15	Import Margins	VTWR																	15:18
16	Export Margins														VTWR				16:18
17	Rest of World	VIWS-VTWR																	17:18
18	Totals	18:1	18:2	18:3	18:4	18:5	18:6	18:7	18:8	18:9	18:10	18:11	18:12	18:13	18:14	18:15	18:16	18:17	

A naming convention is adopted in the GAMS code for each GTAP parameter/coefficient used in the programme. All the GTAP parameters are read in using the names and sets assigned to them in the GDX file. But the local control of sets within GEMPACK means that it is possible for a set to be used with multiple definitions, which goes against the logic of the SAM account structure and the global control of sets in GAMS. Hence as part of the transition process it is necessary to re label the accounts for each parameter. Overwhelmingly, this is carried out by the application of mapping sets that allow the relabelling of set elements through the use of conditional statements. For instance in the following block of GAMS code

```
SAM_gtp(mcc,aa,reg)        = SUM((comm,acts)
                             $(map_mcc_comm(mcc,comm)
                             $map_aa_acts(aa,acts)),
                             VMFB(comm,acts,reg)) ;
```

the elements *comm* and *acts* in the parameter *VMFB* are relabeled as elements *mcc* and *aa* in the parameter *SAM_gtp* via the mapping sets *map_mcc_com* and map_aa_acts.

Commodity Demand by Domestic Agents

Commodity demand by domestic agents consists of final demand by domestic institutions – private household, government and investment – and intermediate demand by activities. Domestic demand is further disaggregated by distinguishing between imported and domestically produced commodities. If standard national accounting conventions for a SAM are followed, purchases by domestic agents would be valued at purchaser prices, but for GTAP 11 these are recorded at GTAP's 'basic' price valuations.

Imported Commodity Demand

There are 4 equations – one for each GTAP parameter.

Import Intermediate Use matrices (1:3) - *VMFB*

```
SAM_gtp(mcc,aa,reg)        = SUM{(comm,acts)$[map_mcc_comm(mcc,comm)

                                 $map_aa_acts(aa,acts)],

                                 VMFB(comm,acts,reg)} ;
```

Household import demand vectors (1:6) - *VPGB*

```
SAM_gtp(mcc,h_h,reg)   = SUM{comm$map_mcc_comm(mcc,comm),
```

```
                                  VMPB(comm,reg)} ;
```

Government import demand vectors (1:13) - *VMGB*

```
SAM_gtp(mcc,g_t,reg)    = SUM{comm$map_mcc_comm(mcc,comm),

                             VMGB(comm,reg)} ;
```

Investment import demand vectors (1:14) - *VMIB*

```
SAM_gtp(mcc,i_is,reg)   = SUM{comm$map_mcc_comm(mcc,comm),

                              VMIB(comm,reg)} ;
```

Domestic Commodity Demand

There are 6 equations – one for each GTAP parameter.

Domestic Intermediate Use matrices (2:3) - *VDFB*

```
SAM_gtp(dcc,aa,reg)        = SUM{(comm,acts)$[map_dcc_comm(dcc,comm)

                                $map_aa_acts(aa,acts)],

                                 VDFB(comm,acts,reg)} ;
```

Household Domestic demand vectors (2:6) - *VDPB*

```
SAM_gtp(dcc,h_h,reg)     = SUM{comm$map_dcc_comm(dcc,comm),

                              VDPB(comm,reg)} ;
```

Government Domestic demand vectors (2:13) - *VDGB*

```
SAM_gtp(dcc,g_t,reg)     = SUM{comm$map_dcc_comm(dcc,comm),

                              VDGB(comm,reg)} ;
```

Investment Domestic demand vectors (2:14) - *VDIB*

```
SAM_gtp(dcc,i_is,reg)    = SUM{comm$map_dcc_comm(dcc,comm),

                               VDIB(comm,reg)} ;
```

Export of transport services (2:16) - *VST*

```
SAM_gtp(d_otm,vst_w,reg)     = VST(v_wtm,reg) ;

SAM_gtp(d_wtm,vst_o,reg)     = VST(v_otm,reg) ;

SAM_gtp(d_atm,vst_a,reg)     = VST(v_atm,reg) ;
```

Exports of Goods and Services (2:17) - *VFOB*

```
SAM_gtp(dcc,ww,reg)        = SUM{[comm,dst]$[map_dcc_comm(dcc,comm)

                                 $map_ww_reg(ww,dst)],

                                  VFOB(comm,reg,dst)} ;
```

Domestic Commodity Output (3:2)

Domestic output (*MAKB*) is defined as the production of commodities by domestic activities
valued at 'basic' prices – a MAKE matrix

Make Matrix (3:2)

```
SAM_gtp(aa,dcc,reg)      = SUM{[comm,acts]$[map_dcc_comm(dcc,comm)

                                $map_aa_acts(aa,acts)],

                                  MAKB(comm,acts,reg)} ;
```

Production Costs by Activities

Production costs by activities fall under three headings: purchases of intermediate inputs,
primary inputs and payment of indirect taxes. Purchases of imported and domestic
intermediate commodity demand, i.e.,

```
SAM_gtp(mcc,aa,reg)       = SUM{(comm,acts)$[map_mcc_comm(mcc,comm)

                                   $map_aa_acts(aa,acts)],

                                    VMFB(comm,acts,reg)} ;

SAM_gtp(dcc,aa,reg)       = SUM{(comm,acts)$[map_dcc_comm(dcc,comm)

                                   $map_aa_acts(aa,acts)],

                                    VDFB(comm,acts,reg)} ;
```

To these costs must be added factor payments by activities (4:3)

```
SAM_gtp(ff,aa,reg)         = SUM{[endw,acts]$[map_ff_endw(ff,endw)

                                    $map_aa_acts(aa,acts)],

                                     EVFB(endw,acts,reg)} ;
```

and all factor use taxes paid by activities, i.e., factor taxes paid by activities (10:3)

```
SAM_gtp(ttff,aa,reg)  = SUM{[endw,acts]$[map_ttff_endw(ttff,endw)

                                 $map_aa_acts(aa,acts)],

                              [EVFP(endw,acts,reg) - EVFB(endw,acts,reg)]} ;
```

and production taxes paid by activities (:3)

```
SAM_gtp(a_tax,aa,reg)  = SUM{acts$map_aa_acts(aa,acts),
                                [SUM(comm,(MAKB(comm,acts,reg) -
                                    MAKS(comm,acts,reg)))]} ;
```

Sales of factor services to domestic activities are the sole source of factor incomes in the database.

Factor Expenditures

Factor incomes are distributed to domestic agents in three sub matrices. The first is a charge for economic depreciation (14:4)

```
SAM_gtp(i_is,f_k,reg) = VDEP(reg) ;
```

The second is the payment of 'income' taxes levied on factors

```
SAM_gtp(d_tax,ff,reg)  = SUM[aa,SAM_gtp(ff,aa,reg)]
                            - SUM{acts,SUM[endw$map_ff_endw(ff,endw),
                                    EVOS(endw,acts,reg)]} ;
```

What remains after payment of income taxes and the depreciation charge is paid to the regional household. For convenience this is done as two equations; the first assigns the after-tax factor incomes to the regional household, and the second removes the depreciation charge that is only levied on the capital factor (*f_k*), i.e.,

```
SAM_gtp(reg_h,ff,reg)    = SUM{acts,SUM[endw$map_ff_endw(ff,endw),
                                    EVOS(endw,acts,reg)]} ;

SAM_gtp(reg_h,f_k,reg)   = SAM_gtp(reg_h,f_k,reg) - VDEP(reg) ;
```

Indirect Tax Expenditures

There are 11 parameters reporting indirect tax expenditures. Import Duty Income (7:1) is reported for *ad valorem* import duties

```
SAM_gtp(ttm,mcc,reg)   = SUM{[comm,src]$[map_mcc_comm(mcc,comm)
                                $map_ttm_reg(ttm,src)],
                                VMTS(ad_val,comm,src,reg)} ;
```

And specific import duties

```
SAM_gtp(ttms,mcc,reg)   = SUM{[comm,src]$[map_mcc_comm(mcc,comm)
```

```
                              $map_ttms_reg(ttms,src)],

                               VMTS(specif,comm,src,reg)} ;
```

There are 8 sub matrices that report sales tax expednitures. Sales taxes on imports are computed for Intermediate Inputs (8:3)

```
    SAM_gtp(ttsm,aa,reg) = SUM{[comm,acts]$[map_ttsm_comm(ttsm,comm)

                              $map_aa_acts(aa,acts)],

                              [VMFP(comm,acts,reg) - VMFB(comm,acts,reg)]} ;
```

Household Demand (8:6)

```
    SAM_gtp(ttsm,h_h,reg)  = SUM{comm$map_ttsm_comm(ttsm,comm),

                                 [VMPP(comm,reg) - VMPB(comm,reg)]} ;
```

Government Demand (8:13)

```
    SAM_gtp(ttsm,g_t,reg)  = SUM{comm$map_ttsm_comm(ttsm,comm),

                                 [VMGP(comm,reg) - VMGB(comm,reg)]} ;
```

Investment Demand (8:14)

```
    SAM_gtp(ttsm,i_is,reg)  = SUM{comm$map_ttsm_comm(ttsm,comm),

                                  [VMIP(comm,reg) - VMIB(comm,reg)]} ;
```

Sales taxes on domestic produced commodities are computed for Intermediate Inputs (9:3)

```
    SAM_gtp(ttsd,aa,reg) = SUM{[comm,acts]$[map_ttsd_comm(ttsd,comm)

                              $map_aa_acts(aa,acts)],

                              [VDFP(comm,acts,reg) - VDFB(comm,acts,reg)]} ;
```

Household Demand (9:6)

```
      SAM_gtp(ttsd,h_h,reg)  = SUM{comm$map_ttsd_comm(ttsd,comm),

                                   [VDPP(comm,reg) - VDPB(comm,reg)]} ;
```

Government Demand (9:13)

```
    SAM_gtp(ttsd,g_t,reg)  = SUM{comm$map_ttsd_comm(ttsd,comm),

                                 [VDGP(comm,reg) - VDGB(comm,reg)]} ;
```

Investment Demand (9:14)

```
    SAM_gtp(ttsd,i_is,reg)  = SUM{comm$map_ttsd_comm(ttsd,comm),
```

```
                                  [VDIP(comm,reg) - VDIB(comm,reg)]} ;
```

Indirect tax expenditures are also incurred by activities. Factor taxes paid by activities (10:3)

```
    SAM_gtp(ttff,aa,reg) = SUM{[endw,acts]$[map_ttff_endw(ttff,endw)

                            $map_aa_acts(aa,acts)],

                            [EVFP(endw,acts,reg) - EVFB(endw,acts,reg)]} ;
```

and production taxes paid by activities (11:3)

```
    SAM_gtp(a_tax,aa,reg)  = SUM{acts$map_aa_acts(aa,acts),

                              [SUM(comm,(MAKB(comm,acts,reg) -

MAKS(comm,acts,reg)))]} ;
```

The income taxes paid by households and incorporated business enterprises are assumed, in the GTAP database, to be paid by factor accounts. Income taxes paid by factors (4:10) are

```
    SAM_gtp(d_tax,ff,reg)  = SUM[aa,SAM_gtp(ff,aa,reg)]

                                - SUM{acts,SUM[endw$map_ff_endw(ff,endw),

                                    EVOS(endw,acts,reg)]} ;
```

Finally, there are export tax expenditures (7:2)

```
SAM_gtp(tte,dcc,reg)  = SUM{[comm,dst]$[map_dcc_comm(dcc,comm)

                          $map_tte_reg(tte,dst)],

                          [VFOB(comm,reg,dst) - VXSB(comm,reg,dst)]} ;
```

Regional Household Income

Regional household income consists of distributed factor incomes, which were explained above, and total tax revenues, which are defined as indirect tax income plus direct tax income. Because of the Regional Household system all after tax and depreciation factor incomes and tax revenues accrue to the regional household, hence private households pay no direct/income taxes and the private household and government do not save.

Factor Incomes AFTER depreciation and income taxes (5:4) are, for convenience, done as two equations; the first assigns the after-tax factor incomes to the regional household, and the second removes the depreciation charge that is only levied on the capital factor (f_k), i.e.,

31

```
SAM_gtp(reg_h,ff,reg)      = SUM{acts,SUM[endw$map_ff_endw(ff,endw),

                                  EVOS(endw,acts,reg)]} ;

SAM_gtp(reg_h,f_k,reg)     = SAM_gtp(reg_h,f_k,reg) - VDEP(reg) ;
```

Then a series of equations assign indirect tax revenues to the regional household. Import duties and export taxes (5:7)

```
SAM_gtp(reg_h,ttm,reg)     = SUM(mcc,SAM_gtp(ttm,mcc,reg))   ;

SAM_gtp(reg_h,ttms,reg)    = SUM(mcc,SAM_gtp(ttms,mcc,reg))   ;

 SAM_gtp(reg_h,tte,reg)     = SUM(dcc,SAM_gtp(tte,dcc,reg)) ;
```

followed by domestic sales tax revenues (5:8)

```
SAM_gtp(reg_h,ttsd,reg)  = SUM(aa, SAM_gtp(ttsd,aa,reg))

                              + SAM_gtp(ttsd,h_h,reg)

                              + SAM_gtp(ttsd,g_t,reg)

                              + SAM_gtp(ttsd,i_is,reg) ;
```

and imported sales tax revenues (5:9)

```
SAM_gtp(reg_h,ttsm,reg)  = SUM(aa,SAM_gtp(ttsm,aa,reg))

                              + SAM_gtp(ttsm,h_h,reg)

                              + SAM_gtp(ttsm,g_t,reg)

                              + SAM_gtp(ttsm,i_is,reg) ;
```

and production tax revenues (5:11)

```
SAM_gtp(reg_h,a_tax,reg)  = SUM(aa,SAM_gtp(a_tax,aa,reg)) ;
```

and factor use tax revenues (5:10)

```
SAM_gtp(reg_h,ttff,reg)   = SUM(aa,SAM_gtp(ttff,aa,reg)) ;
```

Finally, direct/income tax revenues on factors (5:12) are assigned

```
SAM_gtp(reg_h,d_tax,reg)   =  SUM[ff,SAM_gtp(d_tax,ff,reg)] ;
```

Trade Accounts

Trade transactions are at the heart of the GTAP database. A distinctive feature of the GTAP database is the recording of trade margins; in essence this allows the identification of origin specific trade and transport costs associated with imports of commodities by each region with each region contributing to a global trade and transport 'pool' from which trade and transport

services are drawn. Consequently, while the use of trade and transport services associates the use of services with specific imports, the specific suppliers of these services are not identified.

Exports of Goods and Services

Exports of goods and services are reported (2:17)as valued 'free on board' (*fob*), which is a basic price where all export taxes have been paid, hence

```
SAM_gtp(dcc,ww,reg)      = SUM{[comm,dst]$[map_dcc_comm(dcc,comm)

                                 $map_ww_reg(ww,dst)],

                                  VFOB(comm,reg,dst)} ;
```

Noting that the source and destination regions are identified and how the set ordering in the SAM differs from that in the parameter *VFOB*.

Exports of Margin Services

Similarly, the export of margin/transport services (2:16) is valued fob for each of the three margin services

```
SAM_gtp(d_wtm,vst_o,reg)      = VST(v_otm,reg) ;

SAM_gtp(d_otm,vst_w,reg)      = VST(v_wtm,reg) ;

SAM_gtp(d_atm,vst_a,reg)      = VST(v_atm,reg) ;
```

but note how margins exports are only recorded by the source region and type of margin service.

Imports of Goods and Services

Imports of goods and services (1:17) are valued *fob*, and in this code are derived the values of exports by transposing destination (*dst*) into *VFOB* into the regions and source (*reg*) in *VFOB* into the source (*ww*), i.e.,

```
SAM_gtp(ww,mcc,reg)    = SUM{[comm,src,dst]$[map_mcc_comm(mcc,comm)

                                  $map_ww_reg(ww,src)

                                  $map_reg_reg(reg,dst)],

                                   VFOB(comm,src,dst)} ;
```

The transposition of labels is something that can be error prone and something users can find confusing. Hence additional calculations and options are calculated. First, a separate parameter for imports fob is calculated, i.e.,

```
Imports_fob(ww,mcc,reg) = SUM{[comm,src,dst]$[map_mcc_comm(mcc,comm)

                                  $map_ww_reg(ww,src)

                                  $map_reg_reg(reg,dst)],

                                  VFOB(comm,src,dst)} ;
```

which can make checking easier. Second, the value of imports valued 'carriage, insurance and freight' paid (*cif*) is calculated, i.e.,

```
Imports_cif_calc(ww,mcc,reg) = SAM_gtp(ww,mcc,reg) +

        SUM[ootpm$map_ww_ootpm(ww,ootpm),SAM_gtp(ootpm,mcc,reg)] +

        SUM[wwtpm$map_ww_wwtpm(ww,wwtpm),SAM_gtp(wwtpm,mcc,reg)] +

        SUM[aatpm$map_ww_aatpm(ww,aatpm),SAM_gtp(aatpm,mcc,reg)] ;
```

that can be compared with the parameter *VCIF*(*comm,reg,dst*) where the set ordering for *VCIF* differs from that in the SAM. So, code is provided for converting *VCIF* data into the SAM format for both *fob* and *cif* valuations, i.e.,

```
Imports_cif(ww,mcc,reg)  = SUM{[comm,dst]

                              $[map_mcc_comm(mcc,comm)

                              $map_ww_reg(ww,dst)],

                              VCIF(comm,reg,dst)} ;

Imports_fob(ww,mcc,reg)  = SUM{[comm,dst]

                              $[map_mcc_comm(mcc,comm)

                              $map_ww_reg(ww,dst)],

                              [VCIF(comm,reg,dst) -

                              SUM(commp,VTWR(commp,comm,reg,dst))]};
```

These calculations are switched off in the distributed code.

Imports of Margin Services

The import margin matrices (15:1) are derived as three equations from the 4 dimensions parameter *VTWR*(*comm,comm,src,dst*) for consistency with the 3-dminensioned SAM, i.e.,

```
SAM_gtp(ootpm,mcc,reg)    = SUM{[comm,src,dst]$[map_mcc_comm(mcc,comm)

                                  $map_ootpm_reg(ootpm,src)

                                  $map_reg_reg(reg,dst)],
```

```
                                                    VTWR(v_otm,comm,src,dst)} ;

        SAM_gtp(wwtpm,mcc,reg)      = SUM{[comm,src,dst]$[map_mcc_comm(mcc,comm)

                                        $map_wwtpm_reg(wwtpm,src)

                                        $map_reg_reg(reg,dst)],

                                         VTWR(v_wtm,comm,src,dst)} ;

        SAM_gtp(aatpm,mcc,reg)      = SUM{[comm,src,dst]$[map_mcc_comm(mcc,comm)

                                        $map_aatpm_reg(aatpm,src)

                                        $map_reg_reg(reg,dst)],

                                         VTWR(v_atm,comm,src,dst)} ;
```

Capital Account (14:4, 14:5, 14:16, 14:17)

Income to the capital account consists of four elements, two of which are related to the domestic accounts and two to the trade accounts. Domestic savings consist of the allowance for depreciation and net domestic saving. This leaves the balance on the trade account. Strictly this could be defined as the differences in the values of exports *fob* and import *cif*, but the treatment of trade margins through the device of a global transport pool, i.e., while specific transport transactions are associated with imports from specific origins, they are not mapped onto the region that supplies the transport services. Rather transport services are exported to a global transport pool. Hence the balance on the trade account with the rest of the world consists of elements relating to both commodities and transport services. The current account balance with each trading partner is defined as the value of imports valued *cif* less the value of transport services exported to the global transport pool less the value of export. The trade account balance for each region with the global transport pool is defined as the values of transport services imported by region less the value of transport services exported to the global transport pool.

Depreciation charges (14:4) are defined as

```
SAM_gtp(i_is,f_k,reg) = VDEP(reg) ;
```

The trade account balances for trade in transport services are calculated for each margin service

```
SAM_gtp(i_is,vst_o,reg)  = SUM((mcc,ootpm),SAM_gtp(ootpm,mcc,reg))
```

```
                                    - SUM(dcc,SAM_gtp(dcc,vst_o,reg)) ;

SAM_gtp(i_is,vst_w,reg)  = SUM((mcc,wwtpm),SAM_gtp(wwtpm,mcc,reg))

                                    - SUM(dcc,SAM_gtp(dcc,vst_w,reg)) ;

SAM_gtp(i_is,vst_a,reg)  = SUM((mcc,aatpm),SAM_gtp(aatpm,mcc,reg))

                                    - SUM(dcc,SAM_gtp(dcc,vst_a,reg)) ;
```

And the trade account balances by trade partner on commodity trade are calculated as

```
SAM_gtp(i_is,ww,reg)       = SUM(mcc,SAM_gtp(ww,mcc,reg))

                              - SUM[dcc,SAM_gtp(dcc,ww,reg)] ;
```

The domestic savings (14:5) are computed from known values for other savings to ensure investment is fully funded. Hence the value of savings by domestic institutions is defined as the value of investment NET of trade balance plus balance on margin trade and depreciation.

```
invest(reg)      = {SUM[mcc,SAM_gtp(mcc,i_is,reg)]

                      + SUM[ttsm,SAM_gtp(ttsm,i_is,reg)]

                      + SUM[dcc,SAM_gtp(dcc,i_is,reg)]

                      + SUM[ttsd,SAM_gtp(ttsd,i_is,reg)] } ;

trade_bal(reg)     = SUM[ww,SAM_gtp(i_is,ww,reg)] ;

marg_bal(reg)      =  SAM_gtp(i_is,vst_o,reg)

                      + SAM_gtp(i_is,vst_w,reg)

                      + SAM_gtp(i_is,vst_a,reg) ;

deprec(reg)        = VDEP(reg) ;

SAM_gtp(i_is,reg_h,reg)  = invest(reg)

                          - deprec(reg)

                          - trade_bal(reg)

                          - marg_bal(reg) ;
```

Checking the SAM

The checking of the SAM starts with calculating the row and column total and then computing the differences between the row and column total for each region (DIFFTOT). Further calculations are implemented to derive results for

36

```
BIGDIFFTOT(*,reg)       BIG (>0.01) difference in column & row totals

SUMDIFFTOT(reg)         Sum of differences in row and column totals

DIFFPER(*,reg)          Percentage differences in column and row totals

BIGDIFFPER(*,reg)       BIG (>0.01) Percent diffs in column and row totals

DIFTOTPERCNT(step,case) Counts of big total and percentage differences
```

The row and column totals for the SAM will not exactly balance because the underlying GTAP data are not exactly balanced. Hence, it is a judgment call as to whether the SAM is 'correct'. The parameter `DIFTOTPERCNT(step,case)` is the easiest guide where the set *case* identifies differences in totals, percentage and totals and percentages and the set *step* flags differences from > 5 to greater than 0.000001.

```
step     steps for counts of big total and percentage differences   /

stgt5        > 5 total OR percent diffs according to case

..

stgt0000001  > 0.000001 total OR percent diffs according to case /

case    number of cases in each step  /

case1        Differences in Account totals

case2        Differences in Account percentages

case3        Differences in Account totals AND percentages   /
```

In the run for GTAP 11 for 2017 the counts for DIFTOTPERCNT were

	case1	case2	case3
stgt01	6		
stgt001	217		
stgt0001	1435	14	4
stgt00001	4893	215	153
stgt000001	9583	4862	3847
stgt0000001	12651	11566	10825

This seemed to be acceptable.

6. Satellite Accounts

The conversion of other GTAP data into satellite accounts that match with specific sub matrices of the GTAP database is relatively straightforward. No judgment is made about the data; they are simply reported as satellite account data.

GTAP Database Elasticities Data

The GTAP elasticity data are mapped to following parameters where the satellite account versions as distinguished by using an underscore.

```
E_SUBD(mcc,rr)      Armington CES for domestic_imported allocation

E_SUBM(mcc,rr)      Armington CES for regional allocation of imports

E_SUBVA(aa,rr)      CES between primary factors in production

E_SUBT(aa,rr)       Elasticity of intermediate input substitution

E_SUBC(aa,rr)       CES elasticity across intermediate inputs

E_TRAE(ff,rr)       CET between sectors for sluggish primary factors

E_SUBQ(dcc,rr)      CES demand elasticity for 'making' commodities

E_SUBCDE(dcc,rr)    CDE substitution parameter

E_INCCDE(dcc,rr)    CDE expansion parameter

E_SUBG(rr)          Government expenditure CES elasticity

E_SUBI(rr)          Investment expenditure CES elasticity

RFLX(rr)            Expected rate of return flexibility parameter
```

GTAP Energy Volume Data

The GTAP energy volume data are mapped to following parameters where the satellite account versions are

```
E_VolD(dcc,ss,rr)   volume of input domestic purchases by agents (toe)

E_VolM(mcc,ss,rr)   volume of input import purchases by agents (toe)

E_VolT_A(mcc,src,dst)   volume of bilateral trade (toe)

E_VolT_B(mcc,rr,dst)    volume of bilateral trade (toe)

E_VolT(mcc,rr,rr)       volume of bilateral trade (toe)
```

A series of calculations are performed to ensure the process is completed accurately.

GTAP Emissions Data

The GTAP emissions data are mapped to following parameters where the satellite account versions are

> *GHG Emissions*
>
> `EMIT_CO2(ss,ss,rr)`   CO2 emissions
>
> *NON GHG Emissions*
>
> ` EMIT_NGHG(ghgn,ss,ss,rr)`   NON GHG emissions on commodity accounts
>
> *Emissions*
>
> `EMIT(em,ss,ss,rr)`   ALL emissions on commodity accounts
>
> *Land Use Emissions*
>
> `EMIT_LU(em,lu_cat,lu_subcat,rr)`   LAND USE emissions

Some of the GTAP emissions data are not transparent and documentation has not been found; hence some GTAP emission data are omitted.

7   A Brief User Guide

The programme requires data from two sources: the distributed GTAP data in GDX format and an Excel file. Users must obtain a GTAP license to access the GTAP data; hence we do not and cannot provide GTAP data. The Excel file is packaged with the distributed code; the version is explicitly for GTAP 11 for 2017 but we believe it can be used for all years of the GTAP 11 release.

1. No GTAP data are provided. The user must have a licensed version of the GTAP database.

2. All the sets and set mapping required to use the GAMS programmes for version 11 of the GTAP database are provided from a single Excel workbook.

The programme is set up to work off a directory structure shown in Figure 7.1.

Figure 7.1   Directory Structure
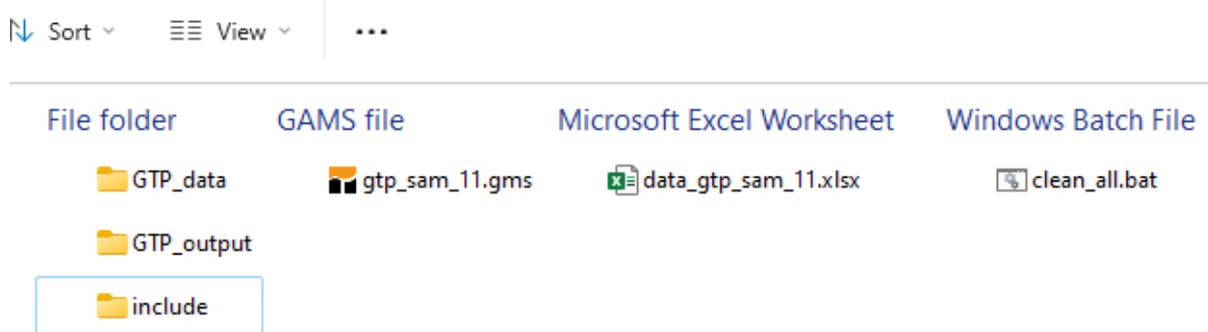
Each of the subdirectories contains a small text file that holds the subdirectory open when zipped. The include directory contains three files, only one of which, `single_set.inc`, users may choose to change; this is only likely to be the case where users what to change the information in the Excel file - `data_gtp_sam_11.xlsx` – to set account labels different from those used in the distributed code. The `GTAP_data` subdirectory contains the GTAP data in a year specific subdirectory that comes from the distributed GTAP data. The subdirectory `GTP_output` will be empty except for the `*.txt` file. The other two files are the programme file `gtp_sam_11.gms` and a batch file – `clean_all.bat` - used to clean the directory (note the distributed version has the suffix `dat` to avoid problems with distribution.

40

Customising the programme to use data for different years is controlled by three $SETGLOBAL assignments, see Figure 7.2. The first identifies the Excel file to be used, the second the subdirectory containing the GTAP data and the third the name for the resultant GDX file.

In many cases these are the only parts of the programme that users should need to adjust.

Figure 7.2 $SETGLOBAL Assignments

```
19
20   *~~~~~ SETGLOBAL assignments
21   * Assign the Excel workbook (this should work for all years i
22
23   $SETGLOBAL    xl_data    data_gtp_sam_11
24
25   * Assign the year for the directory with the GTAP data
26
27   $SETGLOBAL    year       GDX17
28
29   * Assign the name for the output GDX file
30
31   $SETGLOBAL    data_out  gtpdatav11_17
32
33   *############# 2. NOTES ON USING GTAP DATA ################
34 ▼ $onfold
```

The Excel file is where most users will encounter issues. The `Layout` Worksheet (Figure 7.3) is a standard index worksheet as used by GDXXRW; there are non-trivial number of sets and mapping sets, but these should only need changing if users want to use different labels. Problems encountered with this worksheet mostly arise when users change the content of the `name` and/or `location` of the data.

If users want to change labels for the sets, they will need ensure the changes are consistent if every case where a set is referenced. The `Sets_gtp` worksheet, Figure 7.4, contains one, albeit crude, method for deriving the members of the set ss. It relies on adding prefixes to the set labels used by GTAP; these are set in column AG of the worksheet. A more sophisticated system using, say embedded Python, could no doubt be produced; so, feel free to do so and share the method with me. But for now, I have better things to do.

USERS NEED TO RECOGNISE THAT CHANGING THE LABELS CAN PRODUCE PROBLEMS THAT ARE NOT ALWAYS EASY TO TRACK DOWN.

Figure 7.3   Layout Worksheet

Figure 7.4   Set_gtp Worksheet

Commodities and Activities			ENDW_COMM		Regions
comm	(description)		endw_comm	(description)	reg
pdr	Paddy rice		tech_aspros	Technical and professional workers	aus
wht	Wheat		clerks	Clerical workers	nzl
gro	Cereal grains nec		service_shop	Service shop	xoc
v_f	Vegetables fruit nuts		off_mgr_pros	Management	chn
osd	Oil seeds		ag_othlowsk	Agriculture and other low-skill workers	hkg
c_b	Sugar cane sugar beet		Capital	Capital	jpn
pfb	Plant-based fibers		Land	Land	kor
ocr	Crops nec		NatlRes	Natural resources	mng
ctl	Bovine cattle sheep and goats horses				twn
oap	Animal products nec				xea
rmk	Raw milk				brn
wol	Wool silk-worm cocoons				khm
frs	Forestry				idn
fsh	Fishing				lao
coa	Coal				mys
oil	Oil				phl
gas	Gas				sgp
oxt	Other Extraction				tha
cmt	Bovine meat products				vnm
omt	Meat products nec				xse
vol	Vegetable oils and fats				afg
mil	Dairy products				bgd
pcr	Processed rice				ind
sgr	Sugar				npl
ofd	Food products nec				pak
b_t	Beverages and tobacco products				lka
tex	Textiles				xsa
wap	Wearing apparel				can
lea	Leather products				usa
lum	Wood products				mex
ppp	Paper products publishing				xna
p_c	Petroleum coal products				arg
chm	Chemical products				bol
bph	Basic pharmaceutical products				bra

<u>Software Requirements</u>

1. This programme is known to work with versions of GAMS 250 after distribution 39. Users do NOT need a GAMS license to use the programmes, i.e., it runs with the DEMO/Student versions of GAMS.

<u>User Instructions</u>

An include file with User Instructions will be added when it has been revised.

NB: Limited Display statements are now used; we recommend running the programme in GAMS Studio with the Run with GDX Creation option (F10) and with the creation of a reference file. If you find problems, we recommend using the CERR option, set to 10 or less, and periodic $Stop statements to help track any problems.

<u>Warning</u>

Be warned that the files created when using these programmes can be large, although with development of GDX, they are smaller now (2024) than there were in 2004.

The programmes are not fast, but then it will only be run once, or a few times, with each GTAP release.

8.   Concluding Comments

These programmes convert the GTAP database into a global SAM and that, barring a few very small adjustments that have no modeling implications, the resultant SAM conforms to a standard definition of a SAM consistent with IOT accounting conventions. Such a result is hardly surprising; since the data for ALL whole economy models can always be represented in a SAM format provided the models are consistent, which is the case for the GTAP model. Nevertheless the exercise has several substantial benefits. It provides

i)   SAM databases that can be used by SAM based CGE modelers whose preferred programming language is GAMS;

ii)   a method by which individual region SAMs can be easily extracted from the GTAP database, which may make it easier for some researchers to evaluate the data for specific regions; and

iii)   an alternative representation of the GTAP database that might be more easily interpreted by some users of the database.

Since this technical paper was first written (late 2003) there have been several developments.

i)   a GEMPACK version of the code was developed (Kharitonov and Walmsley, 2004);

ii)   the SAM representation of the GTAP database was included in GTAPAgg as part of the GTAP 7 database;

iii)   The GTAP 11 database is distributed in HAR and GDX formats;

iv)   the documentation of the new version of the GTAP model includes a SAM representation of the GTAP database; and

v)   various methods for augmenting the GTAP database, especially with respect to current account transactions have been developed (see McDonald and Sonmez, 2004; Walmsley and Minor, 2013; McDonald et al., 2016).

References

GTAP (2002). 'GTAP61.TAB',
http://www.gtap.agecon.purdue.edu/products/models/current.asp

Hertel, T.W., (1997). *Global Trade Analysis: Modeling and Applications*. Cambridge: Cambridge University Press.

Huff, K., McDougall, R. and Walmsley, T. (2000). *Contributing Input-Output Tables to the GTAP Data Base*. GTAP Technical Paper No. 1 Release 4.2

Huff, K., McDougall, R., Walmsley, T. and Aguiar, A., (2022). *Contributing Input-Output Tables to the GTAP Data Base*. GTAP Technical Paper No. 1 Release 5

Kharitonov, V.V. and Walmsley, T.L., (2004). 'GEMPACK Code for Creating a Globald SAM form GTAP Data', *mimeo*.

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., (2004). 'Developing and Aggregating Reduced Form Social Accounting Matrices from the GTAP Database', *mimeo*.

McDonald, S., and Sonmez, Y., (2004). 'Extending a Global GTAP Social Accounting Matrix to incorporate Inter-institutional and Inter-Regional Transactions', paper presented at the *Seventh Annual Conference on Global Economic Analysis*, World Bank, Washington, June.

McDonald, S. and van Schoor, M., (2004). 'An Aggregation Program for a Global SAM from GTAP Data', *PROVIDE Technical Paper* (forthcoming).

McDougall, R. and Hagemjer, J., (2016). 'Chapter 13: Income and Factor Taxes', (GTAP 9 Data Base Documentation) (www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4884)

Miller, R.E. and Blair, P.D., (2009). *Input-Output Analysis: Foundations and Extensions*. CUP: Cambridge.

Pyatt, G. and Round, J.I., (1977). 'Social Accounting Matrices for Development Planning', *Review of Income and Wealth*, Vol 23, pp 339-364.

Pyatt, G., (1991). 'Fundamentals of Social Accounting', Economic Systems Research, Vol 3, pp 315-341.

Reinert, K.A. and Roland-Holst, D.W., (1977). 'Social Accounting Matrices' in Francois, J.F. and Reinert, K.A., (eds) *Applied Methods for Trade Policy Analysis: A Handbook*. Cambridge University Press: Cambridge

Rutherford, T.F., (1998). 'GTAPinGAMS: The Dataset and Static Model', *mimeo*.

UN, (1993). *System of National Accounts* 1993. UN: New York.

UN (1999). *Handbook of Input-Output Table Compilation and Analysis*. Studies in Methods Series F, No. 74, Handbook of National Accounting. UN: New York.

UN, (2009). *System of National Accounts* 2008. UN: New York.

Walmsley, T. and Minor, P., (2013). 'MyGTAP: A Global Model with Single Country Flexibilities', (http://mygtap.org/wp-content/uploads/2013/12/Overview.pdf)

Walmsley, T., Winters, A. and Ahmed, S.A., (2007). 'Measuring the Impact of the Movement of Labor Using a Model of Bilateral Migration Flows' *GTAP Technical Paper* No. 28 (www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=2529)