

# Constructing a CGE Database Using GEMPACK for an African Country

E. L. Roos<sup>1,\*</sup>, P. D. Adams<sup>1</sup>, J. H. van Heerden<sup>2</sup>

<sup>1</sup>Centre of Policy Studies, Victoria University, Melbourne, Australia

\*e-mail: louise.roos@vu.edu.au

<sup>2</sup>Department of Economics, Pretoria University, Pretoria, South Africa

**Abstract** This paper describes how we transformed the 2002 Ugandan Supply Use Table (SUT) into the required structure of a database for the static Ugandan CGE model Dixon et al. (ORANI: A multisectoral model of the Australian economy, 1982). We describe the unique features captured in the Ugandan SUT as well as that of the CGE database. We highlight the structural differences of the published data and that of the CGE database. In describing the SUT we identify data issues, such as negative capital rentals and omitted data entries that had to be addressed before the database could be constructed. The ideas put forward in this paper describe, in a pragmatic manner, not only how to transform published data into a CGE database, but also how to create an additional sector in the CGE database. For the Ugandan CGE database, we created an additional Oil sector.

**Keywords** Computable general equilibrium (CGE), Database, Africa, Supply Use Tables

**JEL Classification** C81 · C68 · O55

## 1 Introduction

Papers reporting results from applications of a computable general equilibrium (CGE) model often focus on describing the economic theory of the model and explaining the results of a policy simulation. Very few articles include a description of the database

that underlies the application, not because it is unimportant, but because there is simply not enough space in an article to discuss the database creation process in detail and to note implausible flows and anomalies in the original data sources that need correction. It is common for authors to merely state that the model is calibrated to a social accounting matrix (SAM) or Supply Use Table (SUT), without providing any detail about the process of deriving the data from official statistics. For example, in a recent article by [Roos and Giesecke \(2014\)](#) the development of a dynamic CGE model is described, including a detailed description of a labour-market module, capturing the prevalence and incidence of HIV/AIDS for South Africa. The authors state that “two large databases form the initial solution to the SAGE-H model. The core CGE database is calibrated from the 2002 Supply-Use Tables. In parameterising the labour supply module we begin with working age population data on labour market activities by gender and race.” ([Roos and Giesecke 2014](#)). The authors then proceed to shortly describe the sources of HIV data and the use of these data, without going into great detail of the steps taken to transform the initial data into the required databases.

The initial database of a CGE model is of importance because: (1) it contains information regarding the structure of the economy to be modelled in the base year; (2) it is a critical aid in the interpretation of results and (3) in a Johansen-style CGE model, it is the initial solution to the CGE model ([Horridge et al. 1993](#)). Accordingly, there is a need for journal-based descriptions of database construction, such as this paper. In this paper we describe numerous aspects of constructing a database for a comparative-static CGE model of Uganda. This work was carried out for the Ugandan Ministry of Finance, which employed the authors to construct a multi-purpose CGE model for use by various government departments. The Ministry of Finance was particularly interested in the prospect of the economic impact of the discovery of new oil resources. The new oil revenue, if managed effectively, has the potential to significantly enhance real income in Uganda. Currently, Uganda does not produce oil and therefore no allowance for this sector is made in the country’s official statistics.<sup>1</sup>

The rest of this paper is organised as follows: Sect. 2 describes the construction process of a multiple single region database and highlights the difficulties faced in standardising and updating this database. The features of the Ugandan SUT are described in Sect. 3. Section 4 describes the structure of the Ugandan CGE database. Section 5 describes the steps taken to transform the SUT into the required database format, while Sect. 6 describes the creation of an additional crude oil sector in the database. The paper ends with a conclusion.

## 2 Examples of Other Work: Salter and GTAP Databases

In this section we provide a brief overview of the type of work associated with constructing a model database from official statistics. This is not a literature review. The examples cited have been chosen for the interesting aspects of the data construction,

---

<sup>1</sup> At the time of writing this paper, Uganda’s most recent SUT and SAM are for the year 2006. In these data only one mining commodity is specified, namely “Mining and Quarrying”, which contributes less than 1 % of total domestic output.

rather than for its chronological importance. The first example is the Salter database. The second example is the GTAP database which is better known but based on the Salter work.

Brown et al. (1993) describes the steps involved in constructing a 16 region, 37 sector multi-region database, created for the SALTER (Sectoral Analysis of Liberalisation of Trade in East Asian Region) model.<sup>2</sup> This database is based on a number of single region input–output tables, international trade data, industry-specific data and macroeconomic data (Brown et al. 1993, p. 1). The data originated from different sources, had unique sector classifications and were created for different base years. Thus, the objective was to homogenise the format of the databases, update them to a common base year (1988), and reconcile the databases with international trade and macroeconomic data (Brown et al. 1993, p. 1). This large database was created in three major steps.

The first major step was to standardise each single region database to the Salter format via a sequence of specific and generic adjustments. Three types of adjustment were undertaken before a single region database was considered to be in the Salter format. They are individual, commonplace and compulsory adjustments (Brown et al. 1993, p. 3). Individual region adjustments are unique to each region and convert each region's input–output data to the standard Salter format. The Salter format for example identifies the use<sup>3</sup> of commodities from domestic and imported sources, indirect taxes by commodity, use and source, labour, capital and land rental payments and industry taxes. The single region input–output table is seldom presented in the Salter format and therefore requires individual adjustments. For example, in the Korean database private and government fixed capital formation was added to yield investments (Brown et al. 1993, p. 35). Commonplace adjustments correct for variances between individual input–output tables and the Salter format (Brown et al. 1993, p. 4). Examples of commonplace adjustments include calculating imports at basic values; removing negative values from for example, gross operating surplus (GOS) entries; deriving imports by commodity and user; balancing of databases and disaggregation of commodities as to fit the commodities identified in the Salter format. The final adjustment imposed to create a database in concordance with the Salter format refers to compulsory adjustments. They include adjustments to intermediate usage of commodities and the splitting of gross operating surplus over capital and land. Intermediate use adjustments refer to the extraction of Processed Rice and Gas from the Other Food Products and Oil and Gas products, respectively. In extracting these sectors, it is possible to generate undesirable sales structures. For example, a country with relatively little Processed Rice may find it sells most of the Paddy Rice to the Other Food industry rather than the Processed Rice industry. With the intermediate use adjustment, intermediate usage of Paddy Rice was allocated from the Other Food Products to the Processed Rice industry (Brown et al. 1993, pp. 8–9). In terms of splitting gross operating surplus, the task was difficult

---

<sup>2</sup> The SALTER model was developed in Australia to provide insights to the effects of trade policy between Australia and its near trading partners through the Asia-Pacific Economic Cooperation (APEC) Group (Zeitsch et al. 1991; Jomini et al. 1994). It is the basis for the well-known GTAP model and database (Hertel 1997).

<sup>3</sup> The use of commodities includes intermediate, investment, consumption, government, stocks and exports.

due to the lack of data. Therefore, every single region database included in the Salter database the GOS was split equally over capital and land (Brown, 1993, p. 11).

After all the adjustments, the single region databases were in the Salter format, but valued at different currency and for different years. All databases were then scaled to a level of gross domestic Product (GDP) for each region for 1988 and denominated in \$US million (Brown et al. 1993, p. 12). Finally, in each database, values that were targeted in the update procedure that were equal to zero in the database were corrected. For example, zero values could exist for commodity-specific exports whereas international trade data may report values for this commodity. This is problematic because the update procedure imposes a percentage change between two values. A zero denominator makes it impossible to update the specific value. Brown et al. uses an averaging procedure to ensure that each value targeted in the update process has a very small base value in the base year (Brown et al. 1993, p. 12).

The second major step in creating the Salter database was to update the homogenous single region databases with an input–output and trade structure that was consistent with 1988 (Brown et al. 1993, p. 13). In doing so, simulations were conducted with FIT, a program for fitting national databases to international targets (Powell 2007, p. 3). Values in the single region databases were then adjusted to be consistent with 1988 data collected from external sources. Values targeted in each single region database included the main aggregate components of gross national expenditure (GNE) and the trade structure for 1988. Fiscal data were incorporated and tax rates in the databases were updated for 1988. Via a facility in the FIT model, changes in stocks were set to zero. Finally, to improve the update of the Oil commodity, a relative oil price index was shocked to account for changes in oil prices between the year of the initial input–output table and 1988 (Brown et al. 1993, p. 13). The shocks imposed to the FIT model were calculated using a TABLO generated program. This program read in the standardised single region model and calculates the current values that were to be updated. The program also read in target data for macroeconomic, fiscal and trade data. The shocks were then calculated as the percentage change difference between the target value and the current database values. The FIT simulations used the standard single country database and the shocks to update the database (Brown et al. 1993, p. 20).

The final major step in creating the Salter database was to combine the updated single region databases and parameters into a large multi-region database (Brown et al. 1993, p. 23). Again a TABLO generated program was used to facilitate this step. The TABLO program also calculated the marginal budget shares for each region, disaggregated export taxes and exports over destination, disaggregated import taxes and imports over source countries and derived capital mobility data (Brown et al. 1993, p. 23).

The Salter database and model is the basis for the better known GTAP system developed by Tom Hertel and others in the 1990s.<sup>4</sup> GTAP (Global Trade Analysis Project) provides an analytical framework capturing trade and capital flows between countries on a global scale. The aim was to include as many countries as possible to enable detailed analysis of trade policy. This analytical framework required a comprehensive database depicting: (1) bilateral trade flows between each pair of countries;

---

<sup>4</sup> A description of the creation of GTAP and its links to the Australian Salter project is described in Powell (2007).

(2) protection levels imposed by each country on these trade flows; (3) each countries production technology; (4) the resource endowments of each country; (5) the consumption and savings behaviour in each country and (6) information on the tax structure in each country (Powell 2007, p. 5). Creating such a large database is problematic and requires many adjustments as explained in the creation of the Salter database. To create this large multi-country database (hereafter the GTAP database), the core structure of the input–output tables of the Salter database were used and combined with a new approach to reconcile trade data (Powell 2007, p. 7). To date, 8 versions of the GTAP database have been released.<sup>5</sup> The current release, GTAP 8, includes two reference years (2004 and 2007) as well as 129 regions for 57 commodities (Narayanan et al. 2012). GTAP users are encouraged to submit country-specific databases and improve on contributions to the GTAP database. Each database is accompanied by a short description noting the source data, strength and limitations of the published input–output data, steps taken to process anomalies in the raw data and a mapping between the input–output sectors and the GTAP sectors.

### 3 Data Source for the Ugandan Database

The building blocks for creating a database for a CGE model are official statistics from an input output (IO) table, or from a SUT, or from a SAM.<sup>6</sup> Before creating the CGE database, it is important to review and understand the valuation, structure and unique features of the published data, as these dictate what data manipulating steps are required (see Sect. 5).

The 2008 System of National Accounting recommends three ways to value the production of goods and services: in basic prices, in producers’ prices and in purchasers’ prices. The basic price is the “amount receivable by the producer from the purchaser for a unit of a good or service produced as output, minus any tax payable, and plus any subsidy receivable on that unit as a consequence of its production or sale. Basic prices exclude any transport charges involved separately by the producer.” The producers’ price is the “amount receivable by the producer from the purchaser for a unit of a good or service produced as output, minus VAT, or similar deductible tax, invoiced to the producer. It excludes any transport charges invoiced separately by the producer.” Finally, the purchasers’ price is the “amount paid by the producer, excluding any deductible VAT or similar deductible tax, in order to take delivery of a unit of good or service at the time and place required by the purchaser. The purchasers’ price includes any transport charges paid separately by the purchaser to take delivery at the required time and place” (United Nations 2009, p. 275).

In this section we describe the basic structure of the Ugandan SUT where the supply table is valued at basic price and the Use table valued at purchasers’ price (Ugandan Supply Use Table 2002). This will be our starting point in creating a CGE database.

---

<sup>5</sup> For details on the version of GTAP databases see: <https://www.gtap.agecon.purdue.edu/databases/default.asp>.

<sup>6</sup> For a review of SUT, we recommend the System of National Accounts (United Nations 2009) Chap. 14. For the structure and compilation of the Ugandan SUT, we refer to reader to the report by Alarcon et al. (2006).

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Size	1	1	1	1	1	1	IND 1..i
C ↓	Total supply at purchasers price	Imports of goods and service	c.i.f/f.o.b Adjustments	Taxes and subsidies	Trade and transport margins	Total domestic supply	MAKE

C = 142 commodities

**Fig. 1** The format of the supply table

Figure 1 is a simplified illustration of the Ugandan supply table. The first column lists the 142 commodities as they appear in the supply table. Column 2 lists the total supply of all commodities at purchasers' prices. In theory, the values in this column should be equal to the commodity-specific use, valued at purchasers' prices. Column 3 lists the value of commodity-specific imports c.i.f and column 4 lists the commodity-specific c.i.f/f.o.b adjustments. Column 5 lists the commodity-specific taxes, which include taxes and duties on imports and value added type taxes. Column 6 lists the commodity-specific trade and transport margins. Column 8, the multi-production matrix (MAKE), shows the domestic production of 142 commodities by 241 domestic industries valued at basic prices. Column 7 shows the value of the domestic supply of commodities valued at basic price, that is, the MAKE matrix summed over industries.

Total supply valued at basic price is calculated by adding the domestically produced commodities (Column 7) to the imported commodities after taking into account the c.i.f/f.o.b adjustment (Column 3 and 4). Total supply valued at basic price is transformed into producers' price by adding the taxes on commodities (Columns 5). Finally, total supply at purchasers' price is calculated by adding the trade and transport margins (Column 6). In principle the total supply of commodities at purchasers' prices (Column 2) is equal to the total use at purchasers' prices.

The use table, valued at purchasers' prices is illustrated in Fig. 2. This table contains information on the sales of commodities to final users as well as information on factor payments. Commodities may be used as intermediate inputs by producers or final demanders such as investors, households, exporters and government. For example, in the Ugandan use table, the intermediate use matrix, V1PUR\_S, is a 142\*241 matrix and records the flow of each commodity  $c$  used by industry  $i$  in current production.<sup>7</sup> The final demand vectors consist of investors (V2PUR\_SI), a representative private household (V3PUR\_S), exporters (V4PUR) and government (V5PUR\_S) and changes in inventories (V6PUR\_S). Column 8 is the value of total use at purchasers' price and is the sum over all users of commodity  $c$ . In theory the commodity-specific values in column 8 is equal to the commodity-specific total supply at purchasers' price (Fig. 1, Column 2).

<sup>7</sup> In keeping with the GEMPACK naming convention, underscore S (\_S) implies that the data is summed over source, that is, the published data does not distinguish between domestic and imported sources in the use table.

Users		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		Size	Producers	Investors	Household	Exports	Government	Inventories	Total use at purchasers price
			← 1 →	1	1	1	1	1	1
(1)	Flows	C ↓	V1PUR_S	V2PUR_SI	V3PUR_S	V4PUR	V5PUR_S	V6PUR_S	Sum over columns
(2)	Compensation of employees	1	V1LAB_O	C = 142 commodities I = 241 industries					
(3)	Operating surplus	1	OPSUR						
(4)	Depreciation	1	DEPR						
(5)	Net taxes on production	1	V1PTX						

**Fig. 2** The format of the use table

The components of value added include compensation of employees (V1LAB\_O), operating surplus (OPSUR) and depreciation (DEPR). Compensation of employees records the value of each industry's total payment to labour and is presented by industry only. The operating surplus (OPSUR) and depreciation (DEPR), records industry-specific capital payments and depreciation respectively. The vector V1PTX records the value of all production taxes paid by each industry.

The process of converting published data to a CGE database is largely determined by the differences between the published data and structure of the model database, as well as the perceived implausible data items and misprints in the SUT. We converted the published SUT tables to the required form in a number of data manipulating steps. In doing so, additional data from reputable sources could be incorporated. For example, the use table (Fig. 2, row 2) provides data on labour payments by industry. The CGE database, however, requires labour payments by industry and occupation. The occupational distribution by industry could for example be adopted from the Ugandan SAM (Ugandan Social Accounting Matrix 2002).<sup>8</sup> However, it is likely that for resource scarce countries, additional data may not be readily available or data may

<sup>8</sup> Note that the number of industries and commodities may differ between a SUT and SAM. Depending on the dimensions in the model, the modeller will disaggregate the appropriate wage bill matrix. Traditionally the SUT includes more commodities and industries than the SAM and therefore includes more detail on the supply and use of commodities.

be out of date. In the next section we describe the structure of the database for the Ugandan CGE model.

#### 4 Basic Structure of the CGE Database

The database for our comparative-static CGE model for Uganda consists of a set of matrices capturing the unique economic structure and a set of parameters describing the Ugandan economy. The matrix structure of the CGE database is illustrated in Fig. 3. Matrices are defined by a number of sets. These sets include elements related to commodities, industries, source and margins included in the model.<sup>9</sup> The final CGE database for Uganda includes: (i) 83 commodities and industries; (ii) 16 occupational types; (iii) 2 margin commodities; and (iv) 2 sources. For example, the matrix called V3BAS (Fig. 3, row 1 column 3) has the dimensions of commodity  $c$  from source  $s$ . This implies that V3BAS contains values for all 83 commodities from 2 sources.

The first three numbered rows form the absorption matrix, rows 4 to 7 the production matrix and the two satellite matrices are the multi-production matrix and the tariff matrix. In the absorption matrix, users are identified in the column headings. Users are denoted by the column numbers:

1. Domestic producers divided into  $i$  industries;
2. Investors divided into  $i$  industries;
3. A single representative household;
4. An aggregate foreign purchaser of exports;
5. Government demand; and
6. Changes in inventories.

The matrices in the first row, i.e. V1BAS to V6BAS, represent direct flows of commodities, from both domestic and imported sources, to users valued at basic prices. The first matrix, V1BAS, shows the direct flow of commodity  $c$ , from source  $s$ , used by industry  $i$  as an input into current production. V2BAS shows the direct flow of commodity  $c$ , from source  $s$ , used by industry  $i$  as an input to capital formation. V3BAS shows the flow of commodity  $c$  from source  $s$  that is consumed by a representative household. V4BAS is a column vector and shows the flow of commodity  $c$  to exports. V5BAS and V6BAS show the flow of commodity  $c$  from source  $s$  to the government and changes in inventories respectively. In the database, no imported commodity is exported without being processed in a domestic industry. Hence, V4BAS has no import dimension. These matrices contain only direct flows valued at basic prices. The basic price of a domestic commodity is the price the producer receives and excludes margin costs and sales taxes. The basic price of an imported commodity is the duty-paid price, i.e. the price at the port of entry just after the commodity has cleared customs. It excludes all sales taxes and margin costs but includes tariffs. We assume that the basic price of a commodity is the same for all users. The row sums are the total direct usage of a commodity from all sources. It should be noted that all the values, with the exception of V6BAS, are positive. V6BAS records the *changes in inventories*, and can therefore be positive or negative.

---

<sup>9</sup> In Fig. 3, the sets defined for each matrix are listed under or next to the heading called "Size".



			Absorption Matrix					
			1	2	3	4	5	6
			Producers	Investors	Household	Export	Government	Change in Inventories
Size			← I →	← I →	← 1 →	← 1 →	← 1 →	← 1 →
1	Basic Flows	C×S ↓	V1BAS	V2BAS	V3BAS	V4BAS	V5BAS	V6BAS
2	Margins	C×S×M ↓	V1MAR	V2MAR	V3MAR	V4MAR	V5MAR	n/a
3	Taxes	C×S ↓	V1TAX	V2TAX	V3TAX	V4TAX	V5TAX	n/a
4	Labour	OCC ↓	V1LAB	C = 83 commodities I = 83 industries S = 2 sources (domestic, imported) OCC = 16 occupation types M = 2 commodities used as margins				
5	Capital	1	V1CAP					
6	Land	1	V1LND					
7	Production Tax	1	V1PTX					

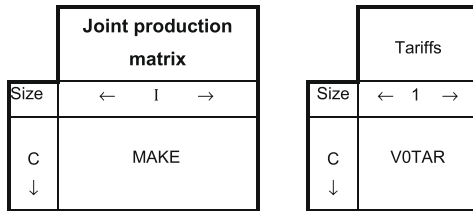


Fig. 3 The CGE database

A special feature of ORANIG-style models is the detailed treatment of margins (Dixon et al. 1982, p. 3). Margin activities include wholesale trade, retail trade and various forms of transports such as air and road transport.<sup>10</sup> For example, to determine

<sup>10</sup> The Ugandan Supply table shows margin commodities as goods road transport, railway transport, retail trade, wholesale trade, retail sale of automotive fuel and the sale of motor vehicles.

the total demand for Ugandan road transport, the model determines the demand for road transport that will be used to facilitate the many commodity flows recognised in the model. This is an important feature especially when the output of margin industries constitutes a large share of GDP. The Ugandan SUT shows that approximately 18 % of GDP is accounted for by the activities produced by the margins industries.<sup>11</sup> The second row in Fig. 3, V1MAR to V5MAR, represents the value of commodities used as margins to facilitate the basic flows listed in row 1. In ORANIG-style models, we assume that all margins are produced domestically. V1MAR is a 4 dimensional matrix and shows the cost of margin service  $m$  used to facilitate the flow of commodity  $c$ , from source  $s$  to industry  $i$  for use in current production. V2MAR can be interpreted in a similar way. V3MAR and V5MAR are 3 dimensional and show the cost of margin service  $m$  that facilitates the flow of commodity  $c$  from source  $s$  to the representative household and the government respectively. V4MAR is a 2 dimensional matrix and shows the cost of margin service  $m$  that facilitates commodities flows to exporters. There are flows that do not require any margins and therefore the values in these matrices are zero. This is mainly for service commodities and inventories.

The third row in Fig. 3, V1TAX to V5TAX, represents the tax matrices. These matrices show the taxes paid in the delivery of domestic and imported commodities to the different users. Taxes will have a positive value and subsidies will be negative. For example, a positive element in V1TAX and V2TAX can be interpreted as the tax associated with the delivery of commodity  $c$  from source  $s$  used by industry  $i$  as an input into current production and capital formation respectively. A negative value is interpreted as a subsidy paid on commodity  $c$ , from source  $s$ , used by industry  $i$ . V3TAX and V5TAX are interpreted as the taxes associated with the delivery of commodity  $c$  from source  $s$  used by households and government respectively. V4TAX is associated with the taxes paid for the delivery of commodities to exporters. Taxes are not paid on inventories and therefore there is no V6TAX matrix. It is possible that commodity-specific tax rates may differ between users and sources.

Rows 4 to 6 contain matrices that provide a breakdown of the cost of primary factors used by industry in current production. These matrices include the inputs of three factors of production: occupation-specific labour (V1LAB), capital (V1CAP) and agricultural land (V1LND). The inclusion of these factors allows the model to determine the percentage change in the industry-specific demand for factors due to a change in relative price. V1LAB (Fig. 3, row 4) shows the purchase of labour of class (or occupation)  $o$  by industry  $i$  that is used as an input into current production. V1CAP (Fig. 3, row 5) contains the rental value of each industry's capital and V1LND (Fig. 3, row 6) shows the rental value of agricultural land used by each industry. Industry also pays production taxes such as business licenses, payroll taxes and stamp duties (System of National Accounts 2009, pp. 147–148). These taxes are contained in V1PTX in row 7. The database shows that labour, capital, land and production costs are only used in current production and therefore these matrices are absent from entries in the capital formation, household consumption, exports, government and change in inventories columns.

---

<sup>11</sup> Based on the 2002 SUT, the value of margins is approximately 2,087 billion shillings and GDP at market price 11,811 billion shillings. Thus,  $(2,087/11,811)*100 = 17.67\%$ .

The satellite matrices illustrate the multi-production matrix (MAKE) and tariff matrix. Each element in the MAKE matrix refers to the basic value of commodity  $c$  produced by industry  $i$ . In principal there are two different types of MAKE matrices. The first is where the entries in the matrix are diagonal, i.e. an industry can only produce one commodity and a commodity can only be produced by one industry. All non-diagonal values are zero. The second type of matrix is a joint production matrix where an industry can produce more than one commodity and a commodity can be produced by more than one industry. Therefore, a number of the off-diagonal values are non-zero. The implication of a joint production matrix is that a producer will choose to produce a combination of output commodities that will maximise their revenue. For example, as the market price of commodity 1 increases relative to the price of commodity 2, producers will shift their resources towards the production of commodity 1 and away from commodity 2.

The final matrix, tariffs (VOTAR), contains tariff revenue by imported commodity. The tariff matrix is separate from the absorption matrix because the values of tariff revenues are already included in the basic price of imports, i.e. they are included in the basic flows in row 1. It enables calculation of ad valorem rates as the ratio between tax revenues and the relevant basic flows of commodities on which the taxes are levied.

Four basic requirements must be satisfied by the database. The first requirement is that the values in most matrices should be positive. Only matrices relating to taxes and change in inventories, may include negative values. In some years some industries might experience negative profit, which would be revealed in the official statistics as a negative entry in the cost-of-capital vector (VICAP). Negative profit, however, is not allowed in the model. The second requirement is market clearing, namely that the value of output of domestically produced commodities must equal the total value of the demand for these commodities. That is, for all non-margin commodities the sum of all inputs used in an industry should equal the sum of all basic values of the direct use of the corresponding commodities. For margin commodities the sum of all inputs into the production of margin  $m$  should equal the sum of all direct usage of  $m$  plus the sum of all usage of  $m$  as a margin. This reflects two features of the database: firstly, the valuation of the MAKE matrix and the absorption matrix are the same, i.e. basic price; and secondly, the columns of the absorption matrix identify all possible uses of domestically produced commodities. The third requirement is that of zero-pure profit, which implies that for each industry the value of output must equal the total production cost. That is, the column sums of the MAKE matrix should equal the sums corresponding to industry demand for inputs into current production. This is because the columns of the absorption matrix recognise all input costs that form part of production costs at basic price, including the profits earned by owners of the fixed factors employed in each industry. The final requirement of the database is that GDP from the income side is equal to GDP from the expenditure side.

Although the Ugandan SUT contains detailed supply and use data of 142 commodities and 241 industries, a number of differences with the CGE database are noted. Firstly, the use table contains matrices representing the flow of commodity  $c$  to user  $u$  valued at purchasers' prices (Fig. 2, row1). The model theory requires these flows to be disaggregated into: (i) flows of commodity  $c$ , from source  $s$  to user  $u$  valued at basic prices (Fig. 3, row 1); (ii) commodity  $c$ , source  $s$ , user  $u$  and margin-specific

matrices (Fig. 3, row 2) and (iii) commodity  $c$ , source  $s$  and user-specific tax matrices (Fig. 3, row 3).

Secondly, the tax column in the supply table records commodity-specific taxes paid on domestic and imported goods (Fig. 1, column 5). Similarly, the margin column (Fig. 1, column 6) records commodity-specific margins on commodities. However, there is no information on who pays these taxes or the user-specific margins on commodities. Instead, the taxes paid by all users and the margins are included in the purchasers' values of the commodity flows in the use table. This tax column (Fig. 1, column 5) will need to be split into user-specific tax matrices (Fig. 3, row 3) and the margin column (Fig. 1, column 6) split into margin matrices reflecting the flow of commodity  $c$ , from source  $s$  to final user  $u$  (Fig. 3, row 2).

Thirdly, the import column in the supply table (Fig. 1, columns 3) records imports of commodity  $c$ . We use this column to disaggregate the use of commodity  $c$ , valued at purchasers' prices, into two source dimensions namely, domestic and imported sources.

Fourthly, the payments to labour recorded in the use table are not distinguished by occupation (Fig. 2, row 2). The model theory requires labour payments by industry and occupation (Fig. 3, row 4).

Fifthly, the use table records industry-specific operating surplus and depreciation in two separate rows (Fig. 2, rows 3 and 4). The model theory, however, requires gross capital payments by industry.

Sixth, factor payments are represented by labour payments and gross operating surplus. There is no data on land rentals in the SUT. For land-using industries, factor payments will need to be reallocated to identify payments to labour, capital and land.

Seventh, the SUT provides no data on the industry-specific structure of capital creation for each of the investors. That is, the SUT only includes one column for investment by commodity (Fig. 2, column 3). This single column needs to be disaggregated into  $i$  number of columns, representing the investment activity of each of the industries in the CGE model.

Eighth, all values, except for taxes and changes in inventories, have to be positive. However, there are some negative values present in the published SUT data. For example, some industries report negative value added and/or negative capital rentals. For modelling purposes these negative values have to be adjusted. This adjustment is described in the next section. Finally, the change in inventories column (Fig. 3, column 6) may include values for service commodities. There are no stocks for service commodities and therefore these values will be removed.

Apart from the obvious gaps between the format of the available data in Figs. 1 and 2 and the required data in Fig. 3, a number of apparently implausible data items in the published data might be identified. Firstly, at the industry level, data might be omitted. For example, in the Uganda SUT, 13 industries employ no labour. The SUT further suggests that the share of labour in total factor payment is 31 %. Our first guess would be that this share is too low compared to other African countries.<sup>12</sup> Ninety SUT

---

<sup>12</sup> Data from the GTAP database for other East Africa countries suggests a more even split between capital and labour. Examples of the labour\capital distributions are: Ethiopia 54:46, Malawi 53:47, Tanzania 50:50, Uganda 53:47 and Zambia 51:49 (Narayanan and Walmsley 2008).

industries have labour shares of less than 25 %.<sup>13</sup> Possible reasons for the low shares are that (1) the labour payments only include payments to employed workers and not those who are self-employed or unpaid family workers and (2) all of the mixed income seems to be added to capital rentals. Secondly, of the 241 industries in the Ugandan use table, 64 industries pay production tax. The data further suggests that 59 % of all production taxes are paid by one industry (Sawmilling and planing of wood). We find this implausible because this industry is relatively small and contributes approximately 1 % to total domestic output.

## 5 Creating the CGE Database

In creating the CGE database we write a sequence of data programs coded in GEM-PACK (Harrison and Pearson 1996). Each step in the data manipulation process addresses a specific data query. This automated process has a number of advantages. Firstly, recording each step promotes transparency and avoids any “black box” issues, that is, the data programs become a permanent documentation of the data manipulation process. Secondly, each data program serves as a record of the process used to manipulate the data. Thirdly, adjustments and corrections to formulas can easily be made. Finally, the automation enables fast replications of the process when needed. This is very useful when new data becomes available.<sup>14</sup> The next section describes the steps taken to convert the published data into the required CGE database.

### 5.1 Step 1: Overview of Published Data and Adjustments

Before proceeding with creating the CGE database, we spent time reviewing the published data, identifying the accounting identities within the data and the presence of implausible values. We checked that the: (1) GDP calculated from the income side is equal to GDP from the expenditure side; (2) commodity-specific aggregate supply equals aggregate demand; (3) share of the components of the GDP identities are plausible; (4) zero pure profits condition by industry is satisfied; and (5) industry-specific factor shares are plausible.

After reviewing the data we proceeded with the adjustments recorded in the SUT. The supply table required one adjustment, the c.i.f/f.o.b adjustment. Imported goods c.i.f. includes values for (i) imported goods f.o.b, (ii) transport services rendered by both resident and non-resident transporters and (iii) insurance services rendered by both resident and non-resident insurers (United Nations 1999, p. 31). If the services mentioned in (ii) and (iii) are rendered by non-residents, then their value are already included in the total value of imports of market services. If these services are rendered by residents, it is part of domestic output and should not be treated as imports. If no adjustment is made, there is double counting to the value of the transport and insurance services. The adjustment column consists of a deduction from the services items for

---

<sup>13</sup> This calculation excludes those industries with negative value added values. Negative value added occurs when capital rentals records large negative values that exceed that of labour payments.

<sup>14</sup> The Centre of Policy Studies Archive contains a number of technical notes, presentations, programs and databases that may be useful in creating country-specific databases. (<http://www.copsmodels.com/archivep.htm>).

transport and insurance equal to the c.i.f/f.o.b adjustment for the commodities (United Nations 2009, p. 508). Commodity-specific imports after the adjustment is calculated as:

$$\text{Imp\_adj}_{(c)} = \text{Imp}_{(c)} + \text{Adj}_{(c)} \quad \text{for } c \in \text{COM} \quad (1)$$

where  $\text{Imp}_{(c)}$  and  $\text{Adj}_{(c)}$  is the commodity-specific imports c.i.f and adjustment respectively. Both these column vectors appears in the supply table.

## 5.2 Step 2: Reviewing Industry-Specific Factor Payments and Production Taxes

In reviewing published data a number of industry-specific issues were identified. In the Ugandan SUT we found that (1) no labour payments were recorded for some industries, (2) there was no occupational dimension in the labour payment data recorded in the SUT, (3) some industries recorded negative capital rental and/or value added values; and (4) there were no land payments to land-using industries.<sup>15</sup>

Labour payment data from the Ugandan SUT and SAM showed that some industries recorded no labour payments or recorded only one labour class being employed. After establishing that these industries do employ labour, we adopted the labour payment distribution pattern over labour class from the average labour payments of the industries falling in the same major classification.

For example for the Millet and Sorghum industry appearing in the Ugandan SAM we adopted the average labour class shares calculated for the agricultural sector. We followed a similar strategy for industries recording labour payments for only one labour class. An alternative option would be to adopt the labour class shares from an industry with a similar production structure.

For modelling purposes all flows, except changes in stocks and subsidies, should always be positive. The Ugandan SUT data showed that for some industries, the gross operating surplus and/or value added was negative. Often statistical offices do not have additional industry-specific data to adjust the negative GOS values. A possible solution to correct for the negative GOS values is to simply set these negative values to zero and then redistribute the total payments between GOS and labour based on the industry-specific capital-labour shares of an industry with a similar production structure. A second option is to adopt the shares captured in the GTAP database for the specific country (Narayanan and Walmsley 2008). For the Ugandan CGE database we set the negative values to zero and then redistributed the total payments based on the shares adopted from the GTAP database.

ORANIG-style models distinguish between 3 types of factors of production, namely labour, capital and land. The use table includes information on compensation of employees, gross operating surplus and indirect taxes, but there are no values for land rentals. We therefore needed to allocate some part of the gross operating surplus to land-using industries. External land rental data might be adopted from the statistical office, but for resource poor countries these data might not be available. An alternative is to adopt the factor shares from the GTAP database to reallocate the total factor payments summed over labour, capital and land for all agricultural, forest and mining

<sup>15</sup> Land refers to cultivated land area, forests and natural resources such as minerals, coal, gold and oil. In the database, only agricultural and mining industries use land.

industries. In the Ugandan SUT we identified 62 agricultural and mining industries as land-using industries. We adopted the factor shares from the GTAP database to redistribute labour payments, rental and land payments over the land-using industries. For non-land using industries, none of the gross operating surplus was allocated to land.

Production taxes mainly comprise taxes on the ownership or the use of land and other assets used in the production process. For Uganda the industry-specific production tax data suggests that 64 industries paid production tax. Of the total production tax, one industry contributes 59 % of all production taxes, leaving the majority of industries to contribute the remainder of the tax. After consulting with statistical office employees, we found this implausible because based on the industry output data this industry contributes less than 1 % of total industry output. In the absence of additional industry-specific data we assumed that the production tax paid was proportional to industry-specific output.

### 5.3 Step 3: Adjustments to the MAKE Matrix

Each element in the MAKE matrix refers to the basic value of commodity  $c$  produced by domestic industry  $i$ . The initial MAKE matrix supplied in the Ugandan supply table shows a high level of multi-production in the economy, implying that an industry can produce more than one commodity and a commodity can be produced by more than one industry. For example, the MAKE matrix in the Ugandan SUT shows 77 of the 241 industries producing real estate activities. Upon further review of the data we found additional implausible results. The statistical office suggested that these features of the MAKE matrix were probably a result of misprints and classification problems.

The MAKE matrix was adjusted in three steps. First, we removed all flows that constituted less than 1 % of the commodity and/or industry flows. In the Ugandan MAKE matrix, 243 elements were identified as having a share of less or equal to 1 % of total commodity or industry output. After the adjustment most cells from the MAKE matrix kept their original values, while the negligible flows were turned to zero. The values of the negligible flows were then added to the diagonal elements. The second task at hand was to remove all implausible elements in the MAKE matrix. Based on the advice from the Ugandan Statistical Office, a number of commodity flows by industry were identified as implausible. For example, the Ugandan multi-product matrix suggests that part of the commodity called *Bee keeping and natural honey* was produced by the domestic *Fish* industry. The implausible elements were set to zero and the values of the elements were either added to the diagonal element or to a suggested element. Finally, we adjusted elements based on the clear views and opinions of the Statistical Office. For example, the Ugandan Statistical office advised that all agricultural commodities should be produced by a single unique industry. We therefore moved all the non-diagonal flows for the agricultural sector to the diagonal value.

The above mentioned adjustments preserved commodity-specific output totals (i.e. column sums of the MAKE matrix), but altered the industry-specific output totals (i.e. row sums of the MAKE matrix). We accepted the adjusted MAKE matrix as our target matrix and scaled the remaining matrices, using RAS, in order to retain commodity and industry balances across the SUT (Horridge 2011).



#### 5.4 Step 4: Removing Negative Flows

Upon revision of the Ugandan SUT we identified further negative flows in matrices capturing the use of a commodity  $c$  by industry  $i$  in the production process (V1PUR\_S in Fig. 2) and the matrix capturing commodity-specific change in inventories (V6PUR\_S in Fig. 2).

In the matrix representing the use of commodity  $c$  by industry  $i$  in the production process, we identified 29 elements containing negative flows. Examples include values for the commodity *Collection, purification and distribution of water* used by the *Dairy Production* industry ( $-16$ ), *Bakery Production* ( $-675$ ), *Tobacco* industry ( $-15$ ), *Textiles* ( $-26$ ) and the *Publishing* industry ( $-101$ ). The values of these flows are relatively small compared to the overall use of these commodities and were therefore set to zero.

The second matrix containing negative values is the changes in commodity-specific inventories (column 7 in Fig. 2). In the CGE database we do allow this matrix to include negative values for all non-service commodities. We do not allow values for service commodities because services usually cannot be stored, and hence cannot be put into inventories. The original Ugandan SUT data contains stock entries for six service commodities. After consultation with the Statistical Office we removed these flows from the inventories column. To restore the total value of inventories to the original SUT value, inventories for all non-service commodities were scaled to the original aggregate values of inventories. A consequence of the adjustments to the intermediate use and change in inventories matrices, commodity-specific aggregate demand did not equal aggregate supply and industry cost did not equal industry output. We used the RAS procedure to ensure that the balancing conditions hold (Horridge 2011). With the RAS procedure we take an original matrix say  $A(r,c)$ , where  $r$  refers to the number rows of size  $1-r$  and  $c$  is the number of columns of size  $1-c$ , and identify target vectors of row totals  $RT(r)$  and column totals of  $CT(c)$ . The RAS attempts to find a new matrix  $B(r,c)$ , so that:

$$\sum_c B(r,c) = RT(r) \quad \text{for all } r \text{ row totals} \quad (2)$$

$$\sum_r B(r,c) = CT(c) \quad \text{for all } c \text{ column totals} \quad (3)$$

The new  $B(i,j)$  matrix is related to the original  $A(r,c)$  via (E.4).

$$B(r,c) = rm(r) \times cm(c) \times A(r,c) \quad \text{for } r = 1, \dots, r \text{ and } c = 1, \dots, c \quad (4)$$

where  $rm(r)$  is a vector of row multipliers and  $cm(c)$  is a vector of column multipliers (Horridge 2011).

#### 5.5 Step 5: Splitting Flows into Domestic and Imported Sources

The previous steps described data adjustments imposed on published data to address anomalies, misprints and negative values. We now proceed with a description of how



we transformed the published data into the matrices that constitute a CGE database. Our first aim was to create the domestic and imported flows of commodity  $c$  to user  $u$ . This task was made difficult because the SUT only presents commodity-specific import values and not commodity-and-user-specific import values.

Figure 3, row 1 shows that all final users, except exporters, use both domestically produced and imported commodities. For example, V3BAS is defined by commodity  $c$  and source  $s$  and shows the flow of commodity  $c$  from domestic and imported sources consumed by households. The basic prices of imported goods are the prices received by importers, excluding transport and other margin costs involved in transferring imports to final users, that is, the basic price of imported goods is the c.i.f. price of imported goods plus the tariffs (Dixon et al. 1982, p. 112). Import duties were included in the *Taxes and subsidies* column (Fig. 1, column 5). The Statistical office provided additional data on the total value of import duties. The data enabled us to split the import duties from the taxes column and calculate the basic prices of imported commodities, which in turn were used to calculate commodity-specific import shares. Alternatively, if no external import duty data were available, the tariffs matrix in the CGE database could be set to zero and the commodity-specific import numbers in the SUT could then be used without any further adjustments.

The commodity-specific flows, as they appeared in the use table, were valued at purchasers' prices and include imports (Fig. 2, row 1). As no information was available on user-specific imports, we calculated the commodity-specific import share as:

$$\text{IMP\_SHR}_{(c)} = \frac{\text{VOIMP}_{(c)}}{\sum_{u \in \text{USERS}} \text{VPUR}_{(c,u)}} \quad \text{for } c \in \text{COM}, u \in \text{USER} \quad (5)$$

where  $\text{VOIMP}_{(c)}$  is the commodity-specific imports valued at basic price<sup>16</sup>; and  $\text{VPUR}_{(c,u)}$  is the value of commodity  $c$  used by user  $u$  at purchasers price.

The dimension  $u$  refers to the following users (1) industries, (2) investors, (3) household, (5) government and (6) stocks. Note that exporters are omitted from the  $u$  dimension. This implies that no imported good was directly re-exported. Schematically, the matrices related to exporters in Fig. 3 (V4BAS, V4MAR and V4TAX) will have an  $s$  dimension, but for the imported source the values will be zero. In the absence of detailed import data by user, we further assume that the commodity-specific import share was the same over all users. This implies that if imported meat made up 10 % of total sales of meat from both domestic and imported sources, then all users of meat would use 10 % of imported meat in their meat purchases. The value of imports by commodity and user can now be calculated as:

$$\text{VPUR}_{(c, \text{"imp"}, u)} = \text{IMP\_SHR}_{(c)} \times \text{VPUR\_S}_{(c,u)} \quad \text{for } c \in \text{COM}, u \in \text{USER} \quad (6)$$

where  $\text{IMP\_SHR}$  is the commodity-specific share calculated in (E.5) and  $\text{VPUR\_S}_{(c,u)}$  is the total value of commodity  $c$  by user  $u$  valued at purchasers' price. To determine the domestic source, the commodity-and-user-specific imports were then deducted from

<sup>16</sup> The basic values of imports in the Uganda database are the commodity-specific c.i.f valuations inclusive of tariffs.

the total use valued at purchasers' price. The end result was matrices showing the use of commodities valued at purchasers' prices by domestic and imported sources. An alternative way to determine end use of imports is to prorate imports across end users using shares calculated from the end usage of domestically produced commodities. This method assumes that if most of the domestically produced commodities are consumed by say households, then most of the imports will be consumed by households (Brown et al. 1993, p. 7).

## 5.6 Step 6: Creating Margin Matrices

Trade and transport margins are the difference between the purchasers' price and the producers' price of a product. It is therefore possible that a product can be sold at different purchasers' prices due to differences in commodity-specific margins and net taxes (United Nations 2009, p. 275). We include the detailed treatment of margin for several reasons. Firstly, about 18 % of Uganda's GDP is accounted for by the activities of the margin sectors (wholesale trade, retail trade and transport). Therefore, to determine the total demand for road transport, the model determines the demand for transport used in facilitating the flow of commodity  $c$ , from source  $s$  to the final user  $u$ . Secondly, it allows for the detailed modelling of variations in purchasers' price of commodity  $c$ , from source  $s$  across user  $u$ . Thirdly, we can simulate the effects of technical change involving particular use of margins (Dixon et al. 1982, p. 3). For example, we may want to simulate the impact of constructing the Kampala-Entebbe Highway. Apart from the obvious impacts of the highway, such as an increase in capital spending, additional benefits of the highway includes a decrease in congestion and shorter travel time for both travellers and goods. These additional impacts are simulated via changes in technical change in the use of road margins.

For all non-service commodities listed in the Ugandan SUT, margins are given as the sum of trade and transport margins. Four features are noted. Firstly, there are no margins for services provided because services are delivered directly from producers to consumers, and hence do not require margins. Secondly, there are no margins on inventories because they comprise of unfinished commodities and materials. Thirdly, trade margins are usually produced domestically but apply to goods produced domestically and imported. Transport margins may be provided by residents and non-residents and used by residents and non-residents (United Nations 2009, p. 276). Finally, the margin column in the supply table shows a positive value for each commodity where margins apply, with an offsetting negative value for each of the margin commodities. Note that the values for trade and transport services, reflected in the use table, show only those that are consumed directly and do not include any margins. Instead, margins are included in the value of the goods at purchasers' prices shown in the rest of the use table. Consequently, in the supply table, the trade and transport margins should be deducted from the total supply of margin services (United Nations 2009, p. 276).<sup>17</sup>

---

<sup>17</sup> For example, direct use of transport is when a person buys a bus ticket and use the bus as a form of transport. When transport is used as a margin, it facilitates the movement of commodities to the final users. In the first case, the bus ticket is billed separately whereas in the second case, the margin is included in the purchasers' price of the commodity.

The margins included in the supply table shows the total value added to each commodity to which margins apply. In the use table, valued at purchasers' price, each commodity-specific value includes the basic price, net indirect tax and total margins associated with that flow. The data does not distinguish between the value and types of margins added to each commodity used by the different users.

We created the margin matrices by calculating the commodity-specific margin share in the total use of each commodity. The margin share is calculated as:

$$\text{MAR\_SHR}_{(c)} = \frac{\text{MARGIN}_{(c)}}{\sum_{u \in \text{USER}} \sum_{s \in \text{SRC}} \text{VPUR}_{(c,s,u)}} \quad \text{for } c \in \text{COM}, s \in \text{SRC}, u \in \text{USER} \quad (7)$$

where  $\text{MARGIN}_{(c)}$  is the commodity-specific margins; and  $\text{VPUR}_{(c,s,u)}$  is the value at purchasers price of commodity  $c$  from source  $s$  used by user  $u$ . The  $u$  dimension refers to the following users, (1) industries, (2) investors, (3) household, (4) exporters and (5) government. Because stocks are defined as unsold commodities, there are no margins associated with stocks. These shares were then multiplied with the user specific flows valued at purchasers' price to create the commodity-and-user-specific aggregate margins.

$$\text{VMAR\_M}_{(c,s,u)} = \text{MAR\_SHR}_{(c)} \times \text{VPUR}_{(c,s,u)} \quad \text{for } c \in \text{COM}, s \in \text{SRC}, u \in \text{USER} \quad (8)$$

$\text{VMAR\_M}_{(c,s,u)}$  is the total value of margins facilitating the flow of commodity  $c$  from source  $s$  to user  $u$ . Next we defined these margin matrices in terms of transport and trade margins. This split was based on the share of each margin in the total value of margins, that is:

$$\text{MARCOM\_SHR}_{(m)} = \frac{\text{MARCOM}_{(m)}}{\text{TOT\_MAR}} \quad \text{form } m \in \text{MAR} \quad (9)$$

where  $\text{MARCOM\_SHR}$  is the share of margin commodity  $m$ ,  $\text{MARCOM}$  is the value of margin commodity  $m$  and  $\text{TOT\_MAR}$  is the total value of margins.

Margin-specific flows are then calculated as:

$$\text{VMAR}_{(c,s,u,m)} = \text{MARCOM\_SHR}_{(m)} \times \text{VMAR\_M}_{(c,s,u)} \quad \text{for } c \in \text{COM}, s \in \text{SRC}, u \in \text{USER}, \text{ and } m \in \text{MAR} \quad (10)$$

Several assumptions were made to create the user-specific margins. Firstly the commodity-specific margin share calculated in (E.7) was the same for all users, that is, if the margin share for commodity  $c$  is 8 %, it is 8 % for all users of commodity  $c$  from source  $s$ . Secondly, we assumed that all users use the same proportion of trade and transport margins as calculated in (E.9).

### 5.7 Step 7: Creating Indirect Tax Matrices for All Users

The supply table shows the total tax levied on each commodity (Fig. 1, column 5). Taxes on products are payable on goods and services when they are produced, delivered, sold, transferred or otherwise disposed of by their producers (United Nations 2009, p. 146). The modellers' task is to allocate these taxes across users. ORANIG-styled models exclude taxes on inventories (Fig. 3, column 6). Tax matrices are determined via eq. (11) and (12).

$$VTAX_{(c,s,u)} = IND TAX_{(c,s)} * TAX FAC_{(u)} * VPUR_{(c,s,u)} / WTOT_{(c,s)} \quad (11)$$

$$WTOT_{(c,s)} = \sum_{u \in USER} TAX FAC_{(u)} * VPUR_{(c,s,u)} \quad (12)$$

where VTAX is the value of indirect tax by commodity  $c$ , source  $s$  and user  $u$ ; IND TAX is the value of commodity and source-specific taxes in the supply table. TAX FAC by user is a tax factor assigned to each user, reflecting the proportion of tax paid by that user. VPUR is the value of commodities from source  $s$  used by user  $u$  at purchasers' price and WTOT is the factor weighted value of commodities  $c$  from source  $s$ . In the absence of any official data to differentiate commodity tax rates by user (as well as by commodity) we allow intuition to guide us. That is, we supposed that exporters, government and investors paid a much lower rate of VAT than households. We assigned a tax weight factor to industries and investors of 1 and to households of 3. For a highly disaggregated SUT table, some commodities are mainly sold to a single user. In this case we might assume that users of this commodity pay the same tax rate.

### 5.8 Step 8: Creating Matrices for the Basic Flows

The aim of this step was to create the domestic flow valued at basic prices of commodity  $c$  from source  $s$  to user  $u$  (Fig. 3, row 1). As mentioned before, the flows at purchases prices include domestic and imported flows valued at basic value plus margin costs plus taxes. The imported flows valued at basic prices were calculated in Step 5, margin flows in Step 6 and tax matrices in Step 7. To calculate the domestic basic flows, we subtracted the imports, margin and tax matrices from the total purchases values summed over source. The end result was a matrix with positive values reflecting the flow of commodity  $c$ , from all domestic sources to final users  $u$  valued at basic prices.

### 5.9 Step 9: Creating an Industry Dimension for the Investments Column

The matrix pertaining to investors in the use table Fig. 2, column 3 was a column vector implying that there was only one representative investor. However, we know that investors buy commodities to construct capital in each industry. We therefore had to split the investment vector into an investment matrix with  $c$  commodities and  $i$  industries. To do so, we calculated the industry-specific shares in gross capital rentals (VICAP) and multiplied it with the total investment value. We were then left with

a matrix where we knew the row and column totals. The next step was to fill in the investment matrix. In the absence of additional data, we assumed that the industry-specific shares were uniform over all commodities. An alternative method would be to fill in the matrix using guess-work reflecting common sense ideas. For example investment by the rail industry accounts for most of the use of locomotives. If the matrix does not add up correctly, the RAS procedure can be used to adjust the matrix (Horridge 2011).

#### 5.10 Step 10: Final Balancing of the Database

In the final step we checked that the data added up and that the balancing conditions still held. We checked that the commodity-specific flows valued at purchasers price added up to the commodity flows at basic price plus the taxes plus margin flows. The first balancing condition states that industry costs should equal the industry output. Considering Fig. 1, the column totals for each industry in the producer column should equal the domestic output of each industry. The domestic output per industry is the MAKE matrix summed over commodities. The second condition states that the domestic commodity output valued at basic price should equal the total domestic sales valued at basic price. After we completed all the data adjustments procedures, there was a slight discrepancy in the balancing conditions. The RAS procedure was used to balance the final CGE database (Horridge 2011).

Finally, the database with 241 industries and 142 commodities were mapped to 82 commodities and industries. The advantage of a smaller database is that it saves computational time when simulations are conducted.

#### 5.11 Parameters

The final section of the database is the parameters. It is unlikely that, for third world countries, enough time series data exists to estimate elasticities via econometric techniques. To our knowledge, no published parameter estimations exist for Uganda. Elasticities and parameters can be adopted from neighbouring countries or countries with a similar economic structure for which data are available. For the Ugandan CGE database we adopted the elasticities from the GTAP database (Narayanan and Walmsley 2008).

## 6 Creating an Oil Sector

Many African countries, such as Uganda, have recently announced the discovery of natural resources, such as oil and gas, as well as plans to extract these resources. However, published SUT's and SAM's do not explicitly include these sectors because they do not currently exist. For example, the Ugandan SUT includes 4 mining industries and one mining commodity.<sup>18</sup> The oil sector is omitted. In this section the creation of an oil sector, with an appropriate cost and sale structure, is explained. We create

---

<sup>18</sup> The mining industries are: (1) mining of non-ferrous metal ores, except uranium and thorium, (2) quarrying of stone, sand and clay, (3) extraction of salt and (4) other mining and quarrying N.E.C. There is only one mining commodity, namely Mining and quarrying.

this sector for two reasons. Firstly, the intuition behind creating any new sector is very similar to the process described below. In creating a sector, the cost and sales structure of the new sector are included in the database. Secondly, our choice of creating a new oil sector stems from the discovery of new oil reserves in Uganda and the lack of representation of this sector in official statistics. This is important because we would want to inform policy makers about the economic consequences of growth in this sector due to an increase in capital spending and continuous future production. The results would be influenced by the cost and sales structure of new oil sector.

We follow a two-step procedure. Firstly, we created an additional industry called *Crude Oil*. We did this by disaggregating the total *Mining* sector into a *Mining* sector, which mainly captures all mining activities other than oil drilling, and a *Crude Oil* sector, which represents all oil drilling activities. Secondly, we created an additional commodity in the aggregated database called *Crude Oil*. Conceptually, we created an additional industry (column) for all the matrices listed in columns 1 and 2 and the MAKE matrix in Fig. 3, and an additional commodity for all the matrices listed in rows 1 to 3 and the satellite matrices in Fig. 3.

Three specific characteristics distinguish the oil sector from other sectors. Firstly, the oil sector is a new sector for which no current data exists. Although the initial values in the database will be negligible, the values do capture a cost and sales structure. Secondly, we assumed that all crude oil would be produced domestically. Thirdly, we assumed that the total value of crude oil was used as an intermediate input into the *Petroleum* industry.<sup>19</sup> Finally, the *Crude Oil* industry used a combination of intermediate commodities, labour, capital and land as inputs in the production of the *Crude Oil* commodity.

As the oil sector is a new sector (with no data available) we imposed hypothetically small values to illustrate how this sector was created. As with all the previous data manipulation procedures described above, the creation of this sector was automated. We could therefore easily accommodate any changes to the imposed values. The following information regarding the new sector was required: (1) the value of output, (2) input structure, and (3) sales structure.

We did not have information regarding the total value of the oil sector. We therefore assumed that the value of the *Crude Oil* industry was 10 % of the mining industry. The remaining value of the *Mining* industry is 90 % of the initial value.

Recall from our discussion in Sect. 4 that commodity output is equal to the sum of the total domestic use valued at basic price. Our next step was to distribute the total value of crude oil over the users of the commodity, *Crude Oil*.

To create the sales structure of the *Crude Oil* commodity, we assumed that 100 % of the total value of the *Crude Oil* commodity is used as an intermediate input in the domestic *Petroleum* industry.<sup>20</sup> To accommodate the new commodity, we adjusted the use structure of the commodity *Mining* by deducting the use structure of the commodity *Crude Oil* from the use structure of the *Mining* commodity. To adjust

---

<sup>19</sup> The Petroleum sector is already accounted for in the SUT.

<sup>20</sup> There may be alternative sales structures. For example, some part of the crude oil commodity could directly be exported. If data on a different sales structure becomes available, we can easily change the automated data program.

the use of margins to facilitate movement of *Crude Oil* to the *Petroleum* industry, we assumed that the margin rate for the commodity *Crude Oil* was the same as for *Mining*. We then recalculated the margin matrices and scaled them to the total margin value.

Finally we created the cost structure of the *Crude Oil* industry. Recall from our discussion in Sect. 4 that industry cost equal industry output. We do know the total value of the industry cost, but we do not know how these costs are to be distributed over the various inputs. This proved to be difficult because of the lack of information regarding industry-specific input and cost structures. To create the cost structure, we assumed that the cost structure of the *Crude Oil* industry was similar to the *Mining* sector. Hence, we adopted from the *Mining* industry the following; (1) the source-specific intermediate input commodities; (2) use of labour, land and capital and (3) the share in which each of these inputs are used. Since we assumed that the value of the *Crude Oil* sector is 10 % of the total *Mining* sector, we allocated 10 % of all input costs from the *Mining* industry to the *Crude Oil* industry.

## 7 Conclusion

This paper describes the process of transforming published data, such as the Ugandan SUT, into a database for a static CGE model. We show that it is possible to use data that included anomalies, misprints and implausible flows to create a database by incorporating expert knowledge and using common sense.

The paper started with a discussion of an example of how a large, multiple country databases for Australia and her Asian trading partners were constructed. This Salter database includes 16 regions with 37 sectors. The database is based on single region input–output tables, international trade data, macroeconomic and industry-specific data (Brown et al. 1993). The input–output structure of the Salter database was used as an input to the well-known GTAP database.

The paper continues with a review of the Ugandan SUT for 2002. We present the features and structure of the SUT and highlight issues with the data that made it difficult to create a CGE database. The 2002 SUT is the primary data source for the construction of the CGE database, which is the main focus of the paper. The construction process included allowance for an additional sector, namely the *Crude Oil* sector. We created this sector because new deposits of oil were discovered after 2002 and any CGE analysis reflecting contemporary conditions must make allowance for the new discovery. For the creation of this sector we required information on the value, cost and sales structure of each of these sectors. We adopt the same cost structure as the existing mining industry. In terms of sales we assume that all crude oil is used as an intermediate input into the petroleum industry. A large share of the output of the petroleum industry is then exported.

As mentioned in the introduction, database construction is a necessary, but often downplayed, aspect of economic modelling, particularly CGE modelling. It is not a science. It is more a process based on experience, the availability of primary data, and the determination of the researcher to draw on secondary information to generate the best available model data set. The benefits of carefully thought out database construction are many. Credibility of the modelling rests in part on the quality of the database.



But a database is more than just something to facilitate a simulation. It serves as a data source for a whole range of analyses associated with input–output statistics, such as industry concentration indexes and multiplier analysis. Moreover, the process of interacting with statistical agencies frequently has spill over benefits as the agencies strive to improve delivery and quality to meet a new source of demand for its services.

## References

- Alarcon, J., Baryahirwa, S., Bahemuka, S., van Heemst, J., Kasirye, I., & Rudaherwanwa, N., et al. (2006). *Social Accounting Matrix Uganda 2002. Project Report: Output, procedures and methodology*. EPRC/ISS/UBOS.
- Brown, S., Strzelecki, A., & Watts, G. (1993). Matching input–output data to international trade data and assembling a salter database, Salter Working Paper No. 19. Industry Commission, Canberra, December.
- Dixon, P. B., Parmenter, B. R., Sutton, J., & Vincent, D. P. (1982). *ORANI: A multisectoral model of the Australian economy*. Amsterdam: North-Holland.
- Harrison, W. J., & Pearson, R. K. (1996). Computing solutions for large general equilibrium models using GEMPACK. *Computational Economics*, 9, 83–127.
- Hertel, T.W. (Ed.). (1997). *Global trade analysis: modeling and applications*. Cambridge: Cambridge University Press.
- Horridge, J. M. (2011). *Four ways to RAS in GEMPACK, Zip file TPMH0093*. Melbourne: Centre of Policy Studies, Victoria University.
- Horridge, J. M., Parmenter, B. R., & Pearson, K. R. (1993). Orani-F: A general equilibrium model of the Australian economy. *Economic and Financial Computing*, 3(2), 71–140.
- Jomini, P., McDougall, R., Watts, G., Dee, P.S. (1994). The SALTER model of the world economy: model structure, database and parameters. Salter Working Paper No. 24, Industry Commission, Canberra, Australia.
- Narayanan, G. B., Aguiar, A., & McDougall, R. (Eds.). (2012). *Global trade, assistance, and production: The GTAP 8 Data Base*. Center for Global Trade Analysis, Purdue University, West Lafayette.
- Narayanan, G. B., & Walmsley, T. L. (Eds.). (2008). *Global trade, assistance, and production: The GTAP 7 Data Base*. Center for Global Trade Analysis, Purdue University, West Lafayette.
- Powell, A.A. (2007). Why, how and when did GTAP happen? What has it achieved? Where is it heading? GTAP Working Paper No. 38. Center for Global Trade Analysis, Purdue University, West Lafayette.
- Roos, E. L., & Giesecke, J. A. (2014). The economic effects of lowering HIV incidence in South Africa: A CGE analysis. *Economic Modelling*, 39, 123–137.
- Ugandan Social Accounting Matrix (SAM). (2002). Ugandan Bureau of Statistics. Kampala.
- Ugandan Supply Use Table (SUT). (2002). Ugandan Bureau of Statistics. Kampala.
- United Nations. (2009). *System of National Accounts 2008*. [http://unstats.un.org/unsd/publication/SeriesF/SeriesF\\_74E.pdf](http://unstats.un.org/unsd/publication/SeriesF/SeriesF_74E.pdf). Accessed June 2012.
- United Nations. (1999). *Handbook of input–output table compilation and analysis*. Series F, No. 74. Department for Economic and Social Affairs, Statistics Division.
- Zeitsch, J., McDougall, R., Jomini, P., Welsh, A., Hambley, S., Brown, S., & Kelly, J. (1991). SALTER: A general equilibrium model of the world economy. Salter Working Paper No. 4, Industry Commission, Canberra, December.