A New Global CGE Database

Scott McDonald and Karen Thierfelder[1]

Addresses for correspondence:

Scott McDonald,
International Agricultural Trade & Development Gp,
Humboldt-University of Berlin,
Unter den Linden 6,
10099 Berlin.
E: scott.mcdonald@uni-hohenheim.de

Karen Thierfelder
Department of Economics
US Naval Academy
Annapolis,
Maryland, USA
E: thier@usna.edu

Abstract

This paper provides details about a new global database designed primarily for use by Computable General Equilibrium (CGE) modelers. While there are other global databases for CGE analysis and analysis using fixed price models, it is argued that these databases lack important features that are desirable in CGE models, e.g., current account transactions other than trade balances. The distinctive features this database are: its construction in accordance with the principles of the SNA; it presentation in the form of a Social Accounting Matrix (SAM); the domestic transactions data are presented as Supply and Use Tables (SUT); transactions between domestic and international institutions are explicitly identified; and the database is designed to make the process of extending transactions data and augmenting the database with satellite accounts straightforward. The database relies on the work of others, principally WIOD, and builds on research that extends and augments the GTAP database to include current account transactions. The primary purpose of the database is to demonstrate a "better mousetrap" in the hope that it will encourage the international input-output (IO) and CGE communities to develop 'complete and consistent' databases that are robust.

[1] This study makes extensive use of the methods used to develop the R23 database. The authors acknowledge the large contribution made by Terrie Walmsley to the R23 database.

© S.McDonald & K Thierfelder, Apr-19

Table of Contents

DRAFT

© S.McDonald & K Thierfelder, Apr-19

1. Introduction

> *"The engine of economic theory has reached, in the last twenty years, a high degree of internal perfection and has been turning over with much sound and fury. If the advance of economics as an empirical science is still rather slow and uncertain the lack of sustained contact between the wheels of theory and the hard facts of reality is mainly to blame." (Leontief, 1953, p 4).*

This paper provides details about a new global database designed for use by Computable General Equilibrium (CGE) modelers, i.e., flexible price models. It is reasonable to ask why the world needs yet another global database: the world already has, *inter alia*, the GTAP (www.gtap.agecon.purdue.edu), R23 (http://www.cgemod.org.uk/r23.html), WIOD (http://www.wiod.org/), OECD (http://www.oecd.org/sti/ind/measuring-trade-in-value-added.htm), ADB (www.adb.org/data/icp/input-output-tables), IDE Jetro (www.ide.go.jp/English/Data/Io), Eora (http://www.worldmrio.com/) and EXIOBASE(http://www.exiobase.eu/) databases. All these databases are a credit to those researchers who have sought "the advance of economics as an empirical science" (Leontief, 1953); despite the profession according them limited, if any, respect or honours.

However, this seeming plethora of databases masks some arguably important issues; first, only the GTAP and R23 databases are designed explicitly to support flexible-price modelling, second, all these databases are, ultimately, presented in Input-Output Table (IOT) format, third, most of these databases are presented as Multi-Region Input-Output (MRIO) tables orientated towards value-chain analyses, fourth, with the exception of GTAP and R23, the databases contain limited policy instruments, and fifth, only the GTAP and R23 databases can be rendered into 'full' SAM (McDonald and Thierfelder, 2004) although the GTAP SAM is deficient and the R23 SAM has no sectoral detail.

It is important to recognise explicitly that the production of any globally complete and consistent database, i.e., a database where all income transactions are paired with identical expenditure transactions by another agent, requires that inconsistent transactions data are reconciled. In theory, but not always in practice, national accounts will be fully reconciled, i.e., internally consistent. But, international transactions reported by one nation for transactions with another nation are not only not reconciled, they are almost invariably

© S.McDonald & K Thierfelder, Apr-19

different.[1] The processes of reconciling these international transactions almost inevitably requires re-estimation of elements within 'reconciled' national accounts.[2] The processes of reconciling an internationally consistent database are likely therefore to entail 'adjusting' estimates from national accounts agencies: in an ideal world, national accounts agencies will engage constructively with the processes, but if they do not there is the potential for disagreement. However, if the compilers of global databases wish national accounts agencies to engage with and/or cooperate, it is argued that the compilers of the global databases need to use conventions and definitions for the valuation of transactions that are consistent with the System of National Accounts (SNA).

This database addresses several issues presented by the currently available global databases for CGE models. The distinctive features of this database are

1. the database is constructed in accordance with the principles of the SNA;
2. the database is presented in the form of a Social Accounting Matrix (SAM);
3. the domestic transactions data are presented as Supply and Use Tables (SUT);
4. transactions between domestic institutions are explicitly identified;
5. international transactions between institutions are explicitly identified, which requires the identification of international current account transactions, e.g., remittances, foreign factor earnings, aid transfers, etc.;
6. the database is designed to make the process of extending transactions data and augmenting the database with satellite accounts straightforward.

The database relies on the work of others, principally WIOD, and builds on research that extends and augments the GTAP database (McDonald and Sonmez, 2004; McDonald and Thierfelder, 2004 and 2019; Sonmez *et al*., 2011; Sonmez and McDonald, 2016; McDonald *et al*., 2016).[3] The primary purpose of the database is to demonstrate a "better mousetrap"; not with the intention that "the world will beat a path to [our] door" but in the hope that it will

[1] The problems associated with reconciling international trade data are detailed in Gehlhar et al., (2008).

[2] In information theoretic terms this is not an issue. If national accounts data are confronted with additional information, e.g., external estimates of the values of transactions, then re-estimation of the national accounts data is logical consistent.

[3] The approach and processes are informed by the contribution of 27 input-output tables to GTAP (Lakatos, et al., 2008), that included providing GTAP with, and implementing (with assistance from Csilla Lakatos), the code for converting 27 EU SUT to GTAP 7's database then specification for IOT (Lakatos and McDonald, 2008). The 27 EU SUT were developed by Marc Mueller, see Mueller et al., (2009)).

© S.McDonald & K Thierfelder, Apr-19

encourage the international input-output (IO) and CGE communities to develop 'complete and consistent' [1]databases that are robust.

The decision to develop this database by using data from WIOD was easy: the WIOD method accords most closely with the guiding principles for this database.[2] The use of WIOD's published data is not always straightforward due to appropriate decisions taken by WIOD considering WIOD's guiding research agenda. It is hoped in future to reduce the scale of the issues, but at this stage the project could be presented as a proof of concept.

The rest of this paper is organised as follows. Section two reviews the guiding principles adopted; these principles provide the basis for the argument about why a different database is justified: it is suggested that the SAM with SUT that derives from these principles provides a basis on which consistent world IOT and SAM/CGE databases can be derived. Section 3 provides a critique of the GTAP database: this is a necessary, if uncomfortable, critique since in conjunction with the guidelines, in section 2, it explains why a new global database for CGE modelling is desirable.[3] An overview of the processes adopted to compile the database are reviewed in section 4, with section 5 devoted to considerations of the limitations of the current version and suggestions for further developments. Section 6 provides some concluding comments. There are also 3 appendices; the first identifies the data sources, which complements section 4, the second, provides brief notes on the estimation techniques used to produce the database, and third, provides details about an open source CGE model that can use this database (it also works with versions of the GTAP database).

[1] A database, in the context of a SAM, is 'complete and consistent' if the data are complete, i.e., all transactions are represented, and consistent if all incomes have corresponding expenditures and vice versa (see section 2 for further details).

[2] "[WIOD] rely on national supply and use tables (SUTs) rather than input output tables as our basic building blocks. Second, to ensure meaningful analysis over time, we start from output and final consumption series given in the national accounts and benchmark national SUTs to these time-consistent series. SUTs are a more natural starting point for this type of analysis as they provide information on both products and (using and producing) industries. ….. . The linking with international trade data …. and socio-economic and environmental data ….. can be naturally made in a SUT framework. In contrast, an input-output table is exclusively of the product or industry type. Often it is constructed on the basis of an underlying SUT, requiring additional assumptions." (Timmer, 2014)

[3] This critique is based on GTAP database v9. Although an illustrative aggregate variant of GTAP 10 is available it remains unclear when a final version will appear and if the structures will be unchanged.

© S.McDonald & K Thierfelder, Apr-19

2. Guiding Principles and Desirable Characteristics

The core guiding principles are that the database should conform with the SNA and that the data used should remain as close as reasonably possible to the national accounts data reported by national statistical agencies; all other considerations derive from these two core principles. The reasons for these core principles are simple. Overwhelmingly national accounts are derived based on the SNA and departure from the terms and meanings in the SNA can only serve to undermine the credibility and acceptability of the database. Moreover, if the data are transformed it is highly likely that the concordances with other data, e.g., labour quantities, emissions, etc., will be obfuscated

'Theory with Numbers'

All economy-wide multi-sector models can be categorised as members of a class of model that can be defined as combining theory with numbers (data). These models are deductive systems, that are built on the premise that the combination of theory and numbers can provide additional insights into the operation of economic systems.

It is argued that the 'accuracy' of the numbers is a crucially important. If the 'accuracy' of the numbers is not deemed important then the premise that the combination of theory and numbers can provide additional insights into the operation of economic systems, is illogical. Similarly, if it is argued that the theory is not deemed important there are few economic model reasons to compile the numbers.

System of National Accounts

The System of National Accounts was conceived, by Sir Richard Stone and others, with the implicit intent of giving empirical content to economic models; the fact that the key relationships developed by Stone and associates have proven so robust is a tribute to the insights of these pioneers.[1]

Why follow the SNA?

The reasons for following the SNA are simple. The, almost, global acceptance and use of the standards defined by the SNA means that nearly all national accounts data are collated using a

[1] The price systems built into all known CGE models are derived from accounting identities that Stone and others had worked out in the 1950s.

© S.McDonald & K Thierfelder, Apr-19

common set of standards. Although, of course, it cannot be guaranteed that all national accounts have been compiled in accordance with the SNA standards, it is beholden on those choosing to depart from the SNA's standards to justify their decisions. Moreover, if the database follows the SNA standards, the likelihood that the database will be acceptable to national accounts agencies and that the agencies will provide help will be greatly increased.

One very important feature of the SNA are the methods for collecting data implied by the 'definitions' of institutions, sectors, establishments, enterprises[1] and industries (see Chapter 4 of the SNA, 2008). Data for the production accounts of the national accounts are, overwhelmingly, collected by asking establishments/enterprises what commodities/products they **supply** and what inputs, intermediate and final, they **use**.[2] National account statisticians then allocate establishments/enterprises to activities/industries and commodities/products[3] to commodity categories, using these data. Similarly, most satellite data are collected in the same manner, which ensures that these data are directly related to the activities/industries and commodity categories used in the production accounts, e.g., labour force data record quantities of employees by activities, as do energy use and emissions data, while commodity tax data, e.g., VAT, GST, import duties, etc., are linked to commodity categories.

The reasons for following the SNA are therefore overwhelmingly pragmatic and practical. This does not mean that the SNA's methods and conventions have gone unchallenged, e.g., Pyatt, 1994a and b, or do not raise issues, e.g., the definition of the production boundary and the economic contribution of social reproduction and women's work.

Price System

The SNA contains an explicit system of prices, design by its architects, especially Stone, to support economic models. 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.

[1] It is important not to confuse 'enterprises' and 'incorporated business enterprises'.
[2] The terms make and supply and the terms absorption and use are interchangeable. In this paper the terms use and supply are used.
[3] The terms commodity/ies and term product/s, and the terms activity/ies and industries are interchangeable. In this paper the terms commodities and activities are used.

© S.McDonald & K Thierfelder, Apr-19

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 can be summarised in a figure (Figure 1). These prices form the basis for the valuation of all transactions recorded in the national accounts, which makes understanding the definitions of these prices and how they are related in a nation's data critical for ALL flexible-price models. The key prices for a database for CGE models are the basic and the purchaser prices.

Figure 1 SNA Price Relationships

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

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

DRAFT

Supply and Use Tables

The SNA (2008) accord important roles to Supply and Use tables (SUT): *inter alia* the benchmarking of national accounts and reconciliation of the goods and services accounts, i.e.,"supply and use tables, a special non-symmetric format of I/O tables first developed in the 1968 SNA, which is used to check the internal consistency of data on production and

© S.McDonald & K Thierfelder, Apr-19

products" (UN, 1999).[1] SUT tables contain the information to relate basic, producer and purchaser prices and provide the basis for the practical derivation of IOT.[2]

SUT are asymmetric tables; the Use tables define the use of commodities by agents – activities, domestic institutions and the rest of the world – while the Supply tables define the supply of commodities by agents. Typically, activities are categorised using the International Standard Industrial Classification (ISIC) scheme by reference to the principal commodity produced by an activity. In this scheme the labels for activities and commodities are identical, but each commodity can be produced by many activities and each activity can produce many commodities. Consequently, the domestic production component of the Supply table is a square matrix (activity by commodity)[3] but not a diagonal matrix, i.e., it is square but asymmetric. Similarly, the dimensions of the intermediate input component of the Use matric is commodity by activity.

Since IOT are, in practice, tables derived from SUT, it is argued that all single country and global databases for all economy-wide multi-sectoral analyses (fixed or flexible price) should be first derived and reconciled in SUT format[4]. Moreover, the term database should encompass both the transactions and satellite account data to ensure consistent definitions of commodities and activities. Ideally the reconciliation of the databases should be in the format of a SAM to ensure the databases are 'complete' and 'consistent' (see below for an explanation of the importance of this argument).

DRAFT

1 "Supply and use tables are a powerful tool with which to compare and contrast data from various sources and improve the coherence of the economic information system. They permit an analysis of markets and industries and allow productivity to be studied at this level of disaggregation. When, as is usually the case, supply and use tables are built from establishment data, they provide a link to detailed economic statistics outside the scope of the SNA." (paragraph 14.3, SNA, 2008).

2 "Supply and use tables are a necessary first step in preparing input-output tables as described in chapter 28 but have important uses on their own, both analytically and as quality control tools. When supply and use tables are first prepared, they are unlikely to balance and until they are brought into balance, GDP measured from the production approach will differ from the expenditure measure of GDP. Only supply and use tables provide a sufficiently rigorous framework to eliminate discrepancies in the measured flows of goods and services throughout the economy to ensure the alternative measures of GDP converge to the same value." (paragraph 14.15, SNA 2008).

3 All references to the dimensions of matrices are by row then column. When referring to three dimensional matrices: the first two dimensions refer to transactions of a region and the third identifies the region.

4 With or without the segmentation of the aggregate Use matrix between commodities sourced domestically and imported.

© S.McDonald & K Thierfelder, Apr-19

From Supply and Use Tables to Input-Output Tables

Once a set of reconciled SUT are produced the transition to an IOT format as envisaged in the SNA is a mechanical process. The following is a simplified description of the process defined in the wider SNA literature (see UN, 1999, and Miller and Blair (Chapter 5), 2009).

1. Segment the Use matrix into domestic and imports Use matrices (if not done prior to reconciliation).
2. Revalue the domestic and imports Use matrices from purchaser prices into basic prices; this requires the development of *inter alia* the following matrices
    a. domestic and imports matrices of taxes and subsidies on all commodity transactions, ideally there should be one set of tax matrices for each tax and subsidy instrument, e.g., VAT, G(eneral) S(ales) T(axes), excise taxes, etc.,
    b. domestic and imports matrices of trade and transport services on all commodity transactions, ideally there should separate matrices for trade and transport services.
3. Determine the account definitions required for the IOT will be commodities (a commodity by commodity IOT) or activities (an activity by activity IOT); this will determine whether the account totals derive from the commodities or the activities.
4. Determine the technology assumptions that will be used to define the technologies used to produce non-principal commodities: this process results in the definitions of commodities or activities no longer being fully consistent with those used in the SUT and wider national accounts. The technology assumptions include, *inter alia*,
    a. 'pure' commodity technology assumption (CTA) – all commodities produced with a common technology irrespective of the activity;
    b. 'pure' activity technology assumption (ITA)[1] – all commodities produced by activity with a common technology irrespective of the commodity;
    c. (multiple) hybrid technology assumptions (HTA) – uses a partitioning od the domestic production matrix so that some commodities produced with a CTA and others with an ITA;
    d. by-product technology assumption – non-principal commodities produced by an activity are treated as by-products.
5. Manual adjustments to the tables; these are typically necessary because the technology assumptions can result in negative transactions and/or the resultant IOT contain

[1] The ITA is a problematic assumption because the resulting tables are not price invariant.

© S.McDonald & K Thierfelder, Apr-19

cgemod

transactions deemed anomalous (with HTA this often relates to adjusting the partitioning of the domestic production matrix).

6. The application of the same adjustment process to all satellite account data to maintain consistent account definitions: simplistically, if the matrices are commodity by commodity, then satellite accounts define by commodity are straight forward but those by activity are problematic because the definitions of the activities have been 'redefined', and vice versa. Commodity by commodity are the most common tables[1], while satellite accounts most commonly related to activities, e.g., employment by activity, energy use and emissions by activities (and institutions), etc.

This (simplified) description indicates the massive data requirements to transition systematically and robustly from the, primary, SUT data to IOT. It is notable that some national accounts agencies produce annual SUT, with aggregated Use tables, and only produce domestic and imports Use matrices infrequently – typically at 5 or 10-year intervals when national agencies produce IOT. Only one national agency is known to make matrices of domestic and imports matrices of taxes and subsidies and trade and transport on commodity transactions, even aggregated, publicly available.[2] The sources of data used by all known global databases, except WIOD, to achieve this transition are opaque.

If the intended use of the global database is fixed price modeling then reporting taxes paid and subsidies by agents as aggregates in single row, i.e., an expenditure by each agent, is arguably a less substantial issue. For flexible-price models the information about taxes and subsidies and trade and transport margins by commodity and agents are essential, IF an IOT format is to be used in the SAM.

<u>Social Accounting Matrices</u>

> *"Since every economic model has its corresponding accounting framework, and since every such framework can be set out as a SAM, it follows that every economic model has a corresponding SAM." (Pyatt, 1987, p 330)*

There are many benefits to using a SAM as the organisational basis for the transaction and satellite data used for whole economy model.[3] A SAM is an extremely flexible data system

[1] The USA is a notable exception.

[2] The exception is Australia (http://www.abs.gov.au/AUSSTATS/abs@.nsf/DetailsPage/5209.0.55.0012015-16?OpenDocument)

[3] The potential benefits of Social Accounting Matrices are recognised in the SNA, although they are still regarded as not essential.

© S.McDonald & K Thierfelder, Apr-19

that can equally accommodate commodity and activity account data presented in SUT and IOT formats, including data on taxes and subsidies and trade and transport margins that are commodity and agent and taxes/subsidies/trade margin/transport margin specific.[1] A SAM format also has the major advantage of making the interpretation of the transactions, and associated satellite accounts, transparent. Similarly, the price relationships are explicit and can be deduced by the (simple) expedient of deriving column coefficients, in terms of all entries in a column or a subset of entries.

But from the perspective of the compilation of a national or global SAM the most important benefit from using a SAM is that the resultant SAM must be 'complete' **and** 'consistent'. 'Complete' in the sense that ALL transactions are correctly reported in terms of the transacting agents and values. 'Consistent' in the sense that the correctly valued income and expenditure transactions by all agents are allocated to other agents such that an equality of row and column totals is achieved.[2] Note that both conditions need to be fulfilled.

A database may be complete in the sense that there are value estimates for all transactions correctly allocated to the transacting agents but not consistent, i.e., the cell estimates are not correctly valued. This is an inevitable occurrence when compiling a SAM[3], and has led to a large literature on SAM estimation techniques[4], e.g., RAS (Stone et al., 1963, Bacharach, 1970), least squares (Stone, 1977)[5],-and Maximum Entropy (Robinson et al., 1998), that use different algorithms and constraints to determine estimates of the cell values that produce a 'consistent' SAM, with advocates for, and users of, many of these techniques. This a well understood issue, which, with a professional approach to data collection and assembling, should give little cause for concern.

A far more important set of concerns arise for SAM that are consistent but not complete. Identifying cells with wrong transactions values is difficult: the necessary estimation processes will mean that final values with often differ from reported values in

1 See, for example, the consolidation of the piecemeal GTAP database into a simple global SAM (McDonald and Thierfelder, 2004 and 2019).

2 It is common to hear the equality of row and column totals in a SAM being described as an objective. It is not an objective; rather it is an outcome of the allocation of transactions to the correct cells in the SAM. The equality of row and column totals does not ensure the allocation of transactions to the cells in the SAM is correct, it only indicates that the allocation of transactions is consistent.

3 See Stone (???).

4 The expression of 'estimating' a SAM is preferred to 'balancing' a SAM. The process of estimating a SAM involves estimating correct transacting values subject to satisfying certain constraints one of which is the equality of row and column totals.

5 Known as the Stone-Byron method due to the algorithm for inverting large matrices due to Byron (1976).

© S.McDonald & K Thierfelder, Apr-19

national accounts. Identifying missing transactions values is relatively easy. There are two main groups; transactions reported in national accounts but not in the SAM and transactions that mandated by observation. If a SAM is consistent but not complete, because cells values wrong or missing, the final SAM transactions will be distorted. This can cause major problems for all multi-sector models, because it distorts the price relationships. In flexible price models, the prices formation processes will be distorted: the magnitudes of these distortions cannot be quantified without access to improved data, but they will exist.

Structure of the Global Social Accounting Matrix

The databases for all whole-economy multi-sector model can be presented in a SAM format. While there are no universal roles for the structure of SAM there are certain basic requirement and 'standard' conventions. A standard arrangement for a SAM in terms of the groups of accounts is presented in Figure 2. When developing a single country SAM the choices on the actual accounts within each group can be made solely on the basis of criteria that are country specific but for a global SAM the commodity and activity accounts within each group will be the same across all regions/countries; for households, factors and tax instruments the compiler need not follow a common set of accounts, although in practice this is likely to be case, i.e., it will be necessary to determine a compromise. In addition, when developing a global SAM, it may be necessary to make other compromises: fully accounting for all the tax instruments in different countries may not be practical and information on incorporated business enterprises may not be available. Inevitably therefore a global SAM will require more compromises than a single county SAM.

Independence of Data and Model

A final, but vitally important, consideration is the need for the structure and the content of the database to be independent from any model. SAM databases only record transactions within economies for some period at some point in time: they are constructed following some set of accounting conventions, e.g., the SNA, and contain no behavioural relationships. The resultant SAM databases are therefore capable of supporting multiple models, i.e., different sets of behavioural relationships: this was one of the guiding principles that underpinned to SNA and development of SAMs.

© S.McDonald & K Thierfelder, Apr-19

DRAFT

The independence of data from model does not constrain the development of the models, but if the data are made conditional on a model the data will constrain the development of models.

© S.McDonald & K Thierfelder, Apr-19

DRAFT

Figure 2 **A Standard SUT Based Social Accounting Matrix**

	Com'dities	Margins	Activities	Factors	Private H'hold	Ent'prises	Taxes	Govt	Investment	Intl Margins	RoW	Total
Commodities		Domestic Margins	Inter'diate Inputs		Private demand			Govt Demand	Invest demand	Intl Margins	Exports (fob)	Total Demand
Domestic Margins	Domestic Margins											Domestic Margins
Activities	Supply											Total Supply
Factors			Factor payments								Factor Incomes	Factor Incomes
Private Household				Factor incomes		Dividend incomes		Transfers			Remittances	Household incomes
Enterprises				Factor incomes				Transfers			Dividend incomes	Enterprise Incomes
Taxes	Product taxes		Activity Taxes	Factor taxes	Income Taxes	Corporation Taxes						Tax Revenues
Govt							Tax revenue				Transfers	Government Incomes
Saving				Depr'tion	Savings	Savings		Savings		Margins balance	Trade balance	Total Savings
International Margins	Intl Margins											Total Int Margins
RoW	Imports (fob)			Factor Payments	Remittances	Dividend Expenditures		Transfers				Total Foreign Expenditure
Total	Total Supply	Domestic Margins	Total Output	Factor Payments	Household Expenditure	Enterprise Expnenditures	Total Taxes	Govt Exp'diture	Total Investment	Total Intl Margins	Total Foreign Income	

© S.McDonald & K Thierfelder, Apr-19

Ideally the databases would be sufficiently comprehensive to allow different users to generate reduced forms of the databases to support different models. In practice, however, there will be limits on how comprehensive databases can be for practical purposes, especially when developing a new database, and for financial reasons.

Desirable Characteristics of a Global Database

Commodity and Activity Transactions

Domestic Inter Institutional Accounts

Current Account Transactions

3. Issues Raised by the GTAP Database

> *"The issue of whether the SAM is deterministic or stochastic is crucial as the SAM provides the underlying data set upon which simple SAM-multiplier analyses and more complex Computable General Equilibrium Models (CGEs) are calibrated. Increasingly, these models are used to explore and simulate the impact of policies and exogenous shocks on the whole socio-economic system. An erroneous or inaccurate SAM invalidates the results obtained from these models" (Thorbecke, 2003, p 186)*

If this study, and database, was compiled for purposes of fixed price models then the discussion in this section would relate to one, or more, of the IO orientated databases. But this study is primarily concerned with a database for global CGE analyses, which means reference must be made to issues associated with the GTAP database.

© S.McDonald & K Thierfelder, Apr-19

Inter-industry accounts

There are several reasons to be concerned about the inter-industry accounts used by the GTAP database. In part, this reflects the fact that the origins of the GTAP database, and model, were, and are, overwhelmingly based on the SALTER database, and model, (see Jornini et al., 1994): this is still reflected in the inter-industry accounts and structure of the contributed source data (Huff, et al., 2000)[1]. This is arguably a burden of history that GTAP seems to be struggling to shake off.

Control Totals

GTAP uses 5 key macroeconomic aggregates as control total private consumption (*C*), gross fixed capital formation (*I*), government consumption (*G*), exports (*X*) and imports (*M*) (see Aguiar, 2015). *C*, *I* and *G* are valued in national accounts at purchaser prices[2], *X* is valued *fob* and *M* is valued *cif*.

This is a very limited range of aggregate control totals. It is reasonable to ask why GTAP does not use a far wider range of control totals, e.g., current account balance, trade balance, internal balance (government savings/borrowings), total factor payments, etc. If a wider range of control total was used there are potential issues to address with respect to their global reconciliation, but it has been demonstrated that this is an achievable objective (see McDonald et al., 2015).

Source Data

The guidelines for providing 'Input-output tables to the GTAP database' express a preference for "commodity by commodity tables". Note that this is a preference, not a requirement, and that there is no specification about the 'preferred' process for converting SUT into IOT[3]. The guidelines, for the "new unified format", request contributors to report transactions data as 4 matrices "Usage of input *i* in use *u*, commodity tax excluded" (UF) and " …., commodity tax

---

1 For example, "Some of you will be aware of the distinctions between commodity by commodity, commodity by industry, and industry by industry tables. Those who are, are liable to be misled by some of the terminology in table 1 and elsewhere in this document. ……. Our terminology (inherited from the Australian Industry Commission's SALTER Project) reflects usage in CGE modeling rather than input-output accounting." (Huff et al., p 2, 2000). It should be noted that the phrase "usage in CGE modelling" is questionable; GTAP follows conventions associated with the 'Australian school' and the ORANI and MONASH models, while other CGE modelers use terms derived from the SNA.

2 There is no reference to the valuation basis in the GTAP documentation (Aguiar, 2015).

3 Sometime after 2008, the GTAP technical paper on 'contributing IOT to the GTAP data base' was amended. The version in 2008 expressed a preference for commodity by commodity table generated using a pure ITA.

© S.McDonald & K Thierfelder, Apr-19

inclusive" (UP), "Output of sector *i*, non-commodity indirect tax included" (OP) and "Output of sector *i*, non-commodity indirect tax included" (MF), together with some information about sector names and concordances. This requires the contributors to have access to large amounts of typically unpublished data, i.e., data used when transforming SUT to IOT, or to deduce for themselves an appropriate allocation of tax transaction values.[1]

The guidelines for the treatment of imports (p 7) is of interest. The first option relies on the existence of "total use matrix and an imports matrix"; the second, is based on a column of imports by commodities for which GTAP requests contributors to "create an imports matrix by pro-rating the totals across uses by applying the structure implied by the total use matrix. ….. Then fill in the import matrix by multiplying each commodity total by the appropriate percentage for each sector. Finally, subtract the new imports matrix from the total use matrix to obtain the domestic use matrix" (p 7), and the third, relates to a "row vector reporting total import costs by sector, but not by commodity. If this is how imports are treated in your IO table, you need further information" (p 7).

The principle cause for concern is quality control. While GTAP check the IO data they receive, they cannot verify the account definitions used[2], the processes used to convert SUT to IOT[3], or any of the underlying data to quantify the tax revenues. Consequently, quality control over the IOT data is franchised out. Reading the documentation on the GTAP website illustrates the vast range of methods that contributors have taken when trying to provide data to GTAP in the requested format; such readings may convince the reader about the integrity of many of the contributors while convincing the reader that the request data structure is 'odd'.

But, far more important than these concerns is the problem of the basis upon which transactions are valued.

Prices

GTAP through version 9[4] has used the terms Agent, Market and World for the prices at which transactions are valued. When the GTAP database is rendered in SAM format the issues

---

1 An examination of the tax rates on commodities in the final GTAP databases indicate that for many databases it may be reasonable to conclude that contributors have applied common tax rates by commodities purchased by activities with different rates for final demand. This is not always so for agricultural and energy commodities, but these are commodity groups for which GTAP make many in-house adjustments.

2 The almost universal use of the principal commodity system of classification means that the labels for activities and commodities are indistinguishable.

3 Different countries use different technology assumptions for converting SUT to IOT.

4 The available documentation for GTAP 10 asserts that GTAP is transitioning from agent, market and world prices to basic and purchaser prices, but the documentation indicates that the process is only a relabelling of market to basic prices and agent to purchaser prices. This cannot be verified prior to the publication of GTAP 10, but if this is what is proposed, then the concerns raised also apply to GTAP 10.

© S.McDonald & K Thierfelder, Apr-19

associated with the valuation basis are transparent. In an IOT framework, as advocated by GTAP, the expenditures by agents for commodities should include the commodity valued at basic prices PLUS domestic trade and transport services and taxes less subsidies on commodities. But in the GTAP database domestic trade and transport services are not explicitly identified.[1] This has implications for the application of the reconciliation of the database with the macroeconomic control totals and input-output data.

The problem can be illustrated for the case of private consumption (C). The macroeconomic control total is in purchaser price terms, which means that the expenditure on commodities net of the commodity taxes and subsidies, that are separately identified, must be valued at basic prices PLUS trade and transport services, i.e., at producer prices. The difficulties arise because this implies that the matrix UF, "Usage of input *i* in use *u*, commodity tax excluded", request by GTAP as a source IO data (UF) should be valued at producer prices; this is not the valuation basis used for standard IOT. Alternatively, the requested matrix UP, "Usage of input *i* in use *u*, commodity tax inclusive", might be valued at purchaser prices, but again this is not the valuation basis used for standard IOT.

Assume that contributors have supplied matrices UF valued at basic prices, the most likely case since this is the price base in IOT, and then correctly deduced the values for the associated commodity taxes. In this case, an implication is that 'balancing' process also involves the implicit revaluation of the expenditures by agents on imported and domestic commodities at producer prices, but while the adjustment of transaction values may be systematic, due the FIT mechanism, there are no reasons to believe the adjustment are correct.

The disconnect between the price system used in GTAP and that used in the SNA is a major cause for concern. Changing the labels will not remove the problem.

1 The issue has been explicitly known since 2006 (see Peterson, 2006). The results from GTAP-M, which includes accounting for domestic margins, are concerning: "tying the domestic marketing activities to specific commodities changes the degree of price transmission from producers to users, compared to a model that does not include margin activities explicitly. For example, the reduction in the crop price due to global technical change in the crop sector is completely passed through to consumers in the standard GTAP Model while only about fifty to eighty percent of the reduction in the crop price is passed through to consumers in the GTAP-M Model." (Peterson, 2006, abstract).

© S.McDonald & K Thierfelder, Apr-19

Commodity accounts

The GTAP database has many accounts for agricultural commodities[1]. But this “creates a problem in data collection since not all countries’ input-output (I-O) statistics provide such fine sectoral detail” (Peterson, 2015, p1). This is arguably an understatement. Only 15 contributed tables, out of 86, contained a ‘full’ set of agricultural and food commodity accounts (2 more had full sets of agricultural commodities (Peterson, 2015); and even these numbers are flattering. Botswana is identified as having a ‘full’ set of accounts; but the base data did not have a full set and the base Botswana IO data, from 1993-4, has not been updated since GTAP 5!

However heroic the efforts of the GTAP staff to generate country specific estimates of the technical coefficients for agricultural commodity production, the fact that over 80 percent of the regions in the database require the use of third-party data must be a cause for concern. Even more so when agriculture accounts for such a small share of GTAP in many of the world’s economies. It is curious to speculate on why GTAP has so many agricultural commodities and why the number increased so much between GTAP 3 (1996) and GTAP 4 (1998).

The number of food commodities is more understandable because food commodities, unlike agricultural commodities, account for relatively large shares of private consumption. Disturbingly, it seems that national accounts agencies are reducing the amount of detail on food commodities in SUT and IOT.

GTAP also adjust the information of energy commodities and activities: “For energy, as with agriculture, we prepare a special data set not just to supplement data from sector generic sources …. but to override them” (McDougal and Lee, 2007). As with the agriculture and food data this may give rise to concerns, especially as the energy data result in vastly different commodity taxes and subsidies across different purchasing agents in the regions of the data base.

[1] “One of the main features of the GTAP Data Base is a detailed sectoral classification of agricultural and food products. Twelve sectors within agriculture and eight sectors within the area of food, beverages, and tobacco are identified.” (Peterson, 2015, p 1).

© S.McDonald & K Thierfelder, Apr-19

<u>Consistent BUT Incomplete</u>

The GTAP database is ‘consistent’ but not ‘complete’, which raises non-trivial concerns about the distortions that must have been imposed on the database as part of the estimation/balancing procedures.[1] If transactions are known to exist, then if the database is ‘consistent’ it can only have been rendered so by distortions elsewhere in the database. Moreover, if the missing transaction are critical to the determination of total incomes and/or expenditures by any accounts in a SAM representation of the database, and the database is ‘consistent’, it can only have been rendered so by distortions elsewhere in the database and the distortion are likely to be serious.

There are missing transactions in the GTAP database. Estimates of these transactions are available in national accounts data and from international databases (for details on international source see McDonald, et al., 2015). The missing transactions can be considered in terms of domestic and international inter-institution transactions, but to do so clearly it is important to mention the GTAP’s concept of a ‘regional household’.

In the GTAP database the regional household’s income is the sum of factor incomes, tax revenues, depreciation and the trade balance, and its expenditures are private and government consumption and investment. Thus, in the GTAP database there are no income taxes paid by the private household and no savings/borrowing by the private household and government. This means that the internal balance is not record, and hence a very important policy target is missing. This caused a problem when GTAP sort to introduce household income and corporation tax revenues; the model did not include the requisite instruments to accommodate income taxes and household and government savings/borrowing.[2] This was resolved by assigning household income taxes to the labour factors, over and above social security contributions, and corporation tax revenues to the other three primary factors – capital, land and natural resources. This had the advantage of not requiring changes to the model code, but at the expense of adjusting the data to fit the model[3], and at the expense of not appropriately representing inter-institution transactions.

1 GTAP uses a procedure called FIT for ‘updating’ IOT (see James and McDougall, 1993).

2 It would be better to also include transfers from the government to households but treating income taxes as net of these transfers is adequate.

3 It is acknowledged that the choices may not be ideal (see McDougall and Hagemejer, 2009).

© S.McDonald & K Thierfelder, Apr-19

International Inter-institution transactions

The international transactions components of the GTAP database only record trade, imports and exports, of goods and services and the trade balance, i.e., all other transactions on the current account, e.g., migrant remittances, factor payment abroad, aid transfers etc., are excluded. If GTAP's regional household concept is followed, then these international transactions could be assigned as aggregate income and expenditure transactions by the regional household. The empirical evidence indicates net income and expenditures on the current accounts of all countries are non-zero (see IMF, 2015a, and McDonald et al., 2015). Accounting for net international transfers on the income side, means that the income of the regional household will be incorrect, but this income must be equal to total expenditure ($C + G + I$), which is a control total. If the control totals are treated as binding, then one or more of factor incomes and/or tax revenues and/or depreciation must have been adjusted to accommodate the missing transactions data: this is a distortion.

This distortion could be important because it will impact on the parameterization of CGE models, and thereby influence the results.[1]

Simply extending the control totals to include the current account balance would do much to reduce this distortion.

But the regional household concept is a reduced form, which modelers may find unsatisfactory. In national accounts inter-institution international transfers are assigned to specific agents, e.g., remittances to/from households, aid transfer to/from governments, etc. The exclusion of international transfers mean that the incomes or expenditures of domestic institutions are distorted, with consequences for parameterization of CGE models, that may influence the results.

In some instances, the consequences can be serious. In several developing countries aid transfers are a large share of recipient government incomes, which implies that in database where the government account is explicit, the adjustment is likely to take place through adjusting tax revenues, i.e., changing important policy instruments.[2]

---

1 This issue cannot be circumvented by arguing that the simulations with the model assumed that net current account transfers were fixed and therefore 'fall' out of the model.

2 In model simulations using tax replacement instruments the exclusion of a large share of government revenue might be expected to be substantial. The currency units in which transfers are valued is also important, but this is a model, not a data, issue.

© S.McDonald & K Thierfelder, Apr-19

Domestic Inter-institution transactions

The exclusion of domestic inter-institution transfers is a concern not only because it distorts the data but because it misrepresents the flow of funds through the system. The case for including households, government and investment-savings as institution has been made. The SAN also identifies another important institution; (incorporated business) enterprises. In moat economies a large share of savings is provided by enterprises, who also pay a large amount of corporation taxes and pay dividends to the households (and governments) who are the beneficial owners of the enterprises. Thus, without explicit recognition of enterprises the flow of funds will be underrepresented.

One option is to combine the household and enterprise accounts. This is tempting because it reduces the amount of data that should be collected, but it avoids a non-trivial problem with national accounts data. All to often national accounts data record savings by households as a residual/balancing item, whereas estimates of savings by enterprises are likely to be more robust, i.e., excluding the enterprise account reduces the information content.

Implicit to this issue is what should be done 'unknown knowns', i.e., known transactions whose values are not known. Is it better to include in a database a poor estimate of a known transaction or assume a zero transaction? In information theoretic terms, and when using mathematical estimation methods, the answer is clear: a poor estimate is better. In terms of behavioural relationships in economic models the answer is also clear: a poor estimate is better.

DRAFT

4. Compiling the Database

This section is a short summary of a longer technical paper (McDonald and Thierfelder, 2019). Full details plus the source data used, in GDX format, will be available from www.cgemod.org.uk.

The major source databases used to compile the database were

1. WIOD Input for National Supply and Use Tables (for 2014) (www.wiod.org/database/sut_input16)
2. WIOD National Supply and Use Tables (for 2014) (www.wiod.org/database/nat_suts16)

© S.McDonald & K Thierfelder, Apr-19

3. WIOD International Supply and Use Tables (for 2014) (www.wiod.org/database/int_suts6)
4. WIOD World Input-Output Tables (for 2014) (www.wiod.org/database/wiots16)
5. World bank Development Indicators (WDI) (World Bank url)
6. International Monetary Fund (IMF): Balance of Payments Statistics (BOPS) and Government Financial Statistics (GFS) databases (www.imf.org/en/Data#data).
7. OECD: DAC International Development Statistics, (www.oecd.org/dac/stats/oecdstat-faq.htm).

An appendix provides details about the data series from the World Bank and the IMF.

The 'adjustment' and reconciliation processes required the use of mathematical/estimation methods. The method chosen for this study is a variant of the maximum entropy method developed by Robinson and associates (specifically an adaptation of Robinson and McDonald, 2004); in this method is that control totals can be declared initially with error and then the estimation process can be applied iteratively to develop reconciled estimates for discrete components of the database.[1] All the data compilation and reconciliation was carried out using GAMS and GDX.

The first stage was to transform all data from WIOD databases using a consistent set of labels: these data were converted from Excel to GDX format (the GDX versions of the WIOD data will be made openly available pending discussions and agreement with WIOD). The second step was to download the required data from the IMF and World Bank; the required data were identified from the construction of the R23 database (see McDonald et al., 2015a).

In all cases the details below identify the components of the WIOD databases used to develop estimates of the various sub matrices of the SAM (see Table X for a summary). The following abbreviation are used

- DCOM – domestic commodities
- MCOM – imported commodities
- MAR- domestic margins
- ACT – activities
- FAC – factors

[1] A similar process could be used by many estimation methods, e.g., generalised or modified variants of RAS (see Allen and Lecomber, 1975).

© S.McDonald & K Thierfelder, Apr-19

- INS – institutions
- TAX – taxes
- ITT – international trade and transport margins
- WOR – trade partners

Interindustry Data

The data for the columns and rows of the commodity and activity accounts were derived from the WIOD databases. In all cases the details below identify the components of the WIOD databases used to develop estimates of the various sub matrices of the SAM (see Table X for a summary). An issue that derives from the use of the WIOD database is that the final WIOT are industry/activity by industry/activity tables, so the majority of the WIOD data used are taken from the International Supply and Use tables.

Supply Tables

The International Supply tables from WIOD provide the data for domestic production valued at basic prices (the ACT:COM submatrix); imports valued *cif* (a column control total for imports valued inclusive of international trade and transport margins); international trade and transport margins (a column control total for ITT:MCOM). Data on taxes and subsidies paid by commodities are not included in the International Supply tables; however some estimates of overall tax rates by commodity can be derived from the input files for national SUTs.

An important component of the International Supply tables are the gross outputs at basic prices for activities, i.e., summations over commodities: these provide the control totals for the activity accounts of the SAM.

Use Tables

The International Use tables from WIOD provide the data for demand for domestically produced and imported commodities valued at basic prices, which requires conversion into purchaser prices. The process proceeds in nine steps

1. aggregate the imports by source region to create an aggregate imports Use matrix (MCOM:ACT, MCOM:INS),
2. define the domestic Use matrix (DCOM:ACT, DCOM:INS; DCOM:WOR),

© S.McDonald & K Thierfelder, Apr-19

3. verify the international trade and transport margins (a column control total for ITT:MCOM) and derive matrices of prior estimates of international trade and transport margins using the input files for national SUTs that match the imports Use matrix,[1]
4. use entropy estimation to derive a consistent sub-matrix of international trade and transport margins (ITT:MCOM),
5. extract data on taxes and subsidies by commodities paid by activities and institutions (control total for TAX:ACT, TAX:INS) and derive matrices of estimates of taxes the input files for national SUTs ;
6. segment taxes and subsidies by commodities paid by activities and institutions between import duties – levied only on imports – and domestic taxes – levied on domestic and imports (this uses aggregate data from the IMF Financial Statistics to derive control totals),
7. use entropy estimation to derive consistent matrices of import[2] and domestic commodity taxes (TAX:ACT; TAX:INS),
8. revalue import and domestic Use matrices at purchaser prices (DCOM:ACT, DCOM:INS, MCOM:ACT, MCOM:INS),
9. extract value added totals for activities (a column control total for FAC:ACT sub matrix.

Factor demand transactions require use of the WIOD socio-economic accounts (SEA). The SEA for 2014 identify payments to capital and labour, so in the preliminary table the factor demand accounts were limited to two factors – labour and capital – to reduce the estimation problem and remain close to the WIOD data. The also allowed the ready formation of satellite accounts for factor use quantities. Subsequent estimates have disaggregated the FAC:ACT sub matrix to derive estimates of that split payment to capital between ‘land’ and capital (using aggregated data from GTAP to inform the prior estimates) and the payment to labour between high, medium and low skilled labour (using data from WIOD SEA 2012 to inform the prior estimates). These estimates are preliminary and less ‘reliable’ than the aggregate estimates.

[1] Various alternative methods for deriving the priors have been explored. The general issue of deriving priors is returned to in section 5 below.

[2] MAcMAPS HS6 2007 was used to derive the estimated prior shares with control totals from the IMF databases and WIOD international SUT.

© S.McDonald & K Thierfelder, Apr-19

DRAFT

Trade Data

The trade data were primarily sourced from the Intranational Use tables. These tables report the demand by domestic agents/accounts for imports by commodity and source region valued *fob*.

The imports by source region, valued *fob*, are aggregated over commodities to create aggregate imports from source regions by commodity and destination region to created an imports (WOR:MCOM) sub-matrix valued *fob*, which is the required valuation for the SAM. This sub-matrix is complemented by two other sub-matrices, one of international trade and transport margins and the other of import duty revenues. Given that imports from source region *s* by commodity *c* and destination region *d* valued fob are identical to exports of commodities from source region d to destination region s, this sub-matrix also provides the transaction data for the sub-matrix (DCOM:WOR).

The import and export sub-matrices were checked against the control totals for imports and exports.

The matrix of demand by importers for International Trade Margins were derived from the sub-matrix of international trade and transport margins (ITT:MCOM), and verified against control totals. Exports of margin services were estimated using prior estimates of international trade and transport margins based on international shares in total exports of the associated commodity; the estimation was completed using the entropy estimation method.

Inter-Institutional Transfers

The data for inter-institutional transfers are derived using the method developed for the R23 database, and discussion here is brief (for more details see McDonald et al., 2015a).

Domestic Inter-Institutional Transfers

At this stage the household and enterprise accounts were aggregated. The prior estimates for income and corporation taxes and social security contributions were taken from the IMS's Government Financial Statistics, as were estimates of internal balances (government saving/borrowing) (see Appendix 1). Household savings were estimated as a residual.

Separating the enterprise account is on-going, but getting the information apparently requires getting country specific enterprise data.

© S.McDonald & K Thierfelder, Apr-19

International Inter-Institutional Transfers

Aid and Other Grants

The bilateral DAC database developed by the OECD must first be adjusted to remove any negatives and ensure that aid from multilateral organizations is supplied by donor countries. The negatives in the 2007 OECD DAC database were removed and replaced with a 3-year averages and an adjustment made to ensure balance. Supplies by donors to the multilateral organizations are assumed to come from the DAC donor countries in the same shares with which they supply aid to DAC countries.

Once adjusted and balanced, DAC aid data are subtracted from total grants and included separately. The bilateral DAC database is used to provide both the source and destination of the DAC aid, with multilateral aid flowing through the globe account. The remaining non-DAC grants are also assumed to go through Globe and are therefore not bilateral.

Remittances

The initial estimates for determining the source and destination of remittance data comes from data developed as part of the GMig2 database (Walmsley et al. 2005) which estimates the numbers of workers, wages and remittances by home and host country of migrant labour.

Factor Payments

The GMig2 bilateral remittance data (Walmsley et al 2005) is also used as the initial estimates for worker's compensation by source and destination. Finally, the bilateral FDI database developed by Guimbard, Gouel and Laborde from CEPII and adapted and documented in Lakatos and Walmsley (2010) was used to provide the initial estimates of factor payments to capital by source and destination.

5. Limitations and Further Development

Although this database has been derived with a heavy reliance on the WIOD this prototype was, deliberately, compiled without discussion with the WIOD team. It was deemed important to (a) provide independent proof of concept, (b) demonstrate that the transition to a SAM from SUT tables was practical, i.e., the necessary additional data could be compiled and reconciled, and (c) ensure that discussions with the WIOD team were focused. At this time of writing (January 2019) discussions with the WIOD team have not taken place. These

© S.McDonald & K Thierfelder, Apr-19

comments on limitations and further developments are not intended to prejudge the content or outcome of any discussions; rather they are intended to identify limitations of the current exercise and some possible future options.

Mathematical estimation techniques should really only be used after all the available data have been confronted; the data supplied to all known mathematical estimation techniques require control totals (row and column totals and aggregates for sub-matrices), initial estimates of the transactions in the matrices – these are known as prior estimates in the entropy method, and estimates relating to the error distributions associate with the priors and control totals. All known analyses of the performance of mathematical estimation techniques demonstrate the importance of the prior estimates, e.g., the early and easily replicated analyses by Lynch (1979) and Allen and Lecomber (1975). Many of the limitations relate to concerns about the derivation of prior estimates and control totals.

Supply and Use Tables

The major limitation encountered in this exercise was the revaluation of the WIOD international Use matrices from basic prices to purchaser process. The process of generating the SAM database would be simpler if the starting point (international?) domestic and aggregate import Use tables were valued in purchaser prices. From the SAM perspective this would be preferable because the asymmetric format that is preferred would be immediately available AND users would be free to generate their preferred IOT variants.

Domestic Institutions

The behaviour/role of Incorporated Business Enterprises is a critical component in the functional distribution of income, direct (corporation) tax revenues and savings-investment. Currently available data only relate to corporation tax revenues and capital (factor) incomes from abroad: this has required strong assumptions to be made for incomes from domestic factor sales (capital) and savings. The collection of additional data from country specific national accounts is being undertaken to improve these prior estimates.

International Inter-Institutional Transfers

The source data used are now dated. The database will be revised when better data have been sourced.

© S.McDonald & K Thierfelder, Apr-19

Tax Revenue data

Extensive tax and subsidy data are necessary for meaningful flexible-price models, since to a very large extent the tax instruments are critical policy instruments.

Trade tax and subsidy data

In this version it was necessary to use the MAcMAPS series from 2007; these data are somewhat out of date. The trade tax and subsidy data will, shortly, be developed direct from the International Trade Centre's open source Trade Map utility.

Domestic tax data

The indirect commodity taxes are the least satisfactory transaction data in the current database. Certain control totals are available from the WIOD, e.g., taxes and subsidies paid by activities, and estimates of aggregate revenues for different tax instruments, e.g., VAT, import duties, GST, Excise taxes, etc., are available from the IMF financial statistics. This has required some strong/heroic assumptions to generate the prior estimates and the estimates are sensitivity to these assumptions. The collection of additional data from country specific national accounts is being undertaken to cross check the IMF data and improve these prior estimates.

Exports of international margins

The export of international margins has followed the approach pioneered by GTAP; regions export margins to a global pool from which regions source margin services for imports.[1] The priors used are derived, crudely, as residuals. A greater use of primary data may improve the database.

Industry by industry choice in WIOD

The choice of an industry by industry format is appropriate given the objectives of the research programme. For a database geared to CGE analyses there are arguments for a commodity by commodity format. In an ideal world the core database would allow different users a choice.

---

1 In this SAM structure a dummy region – globe – is used for the pool. Implicit in this approach is an assumption of market clearing in the globe region, i.e., the volume of total exports is exactly equal to the volume of total imports.

© S.McDonald & K Thierfelder, Apr-19

6. Concluding Comments

Why undertake this effort? Some will argue that the world already has too many databases and insist that the current databases are adequate, while the proponents of the existing databases will not express fondness for this effort. The case for this database is, arguably, not helped by the fact that it has fewer regions, accounts and factors.

The reasons do not include pecuniary gain. In fact, the reasons stem from concerns about the existing databases for global CGE models; inter alia, concerns about data quality control, inconsistency with national accounting norms, data biases and missing data.

The arguments presented do not need the development of an alternative database; it is argued that the reasons for the concerns are sufficiently evident without the database. The purpose of the database is to demonstrate the practicality of developing a database that follows SNA conventions. By so doing demonstrate that a database which provides the basis for flexible- and fixed-price models with or without MRIO is a practical proposition, and at the same time allow the transactions in the database to be trace back directly to published national accounts. By being able to link the databases back to national accounts data the credibility of the policy analyses derived using models calibrated to the database will be enhanced, and disarm the, easy, criticism that the values in the analyses do not reflect national accounts data.

If the data are to be robust and reflect national accounts data, then it is argued that they should be structured so they can be easily related back to the published data. This makes an argument for the data to remain as close as possible to the source data and to be taken directly from published national accounts. The conclusion here is that SUT tables, in the context of a SAM, is the best option. In this, the argument follows that made by WIOD, who derived internationally consistent SUT as a stepping stone to the database requirements for their analyses. By following this approach WIOD took responsibility for quality control over the database, by generating the IOT themselves rather than contracting this out to third-parties. An implication of this approach is that it ensures the domestic transactions data are not a support set.

An additional, and important, advantage of the SUT framework is the greater ease with which the accounts can be augmented with satellite account data; factor quantities, energy use

© S.McDonald & K Thierfelder, Apr-19

and emissions, etc. Given the growing demand for analyses of environmental and socio-economic policies this is useful.

Moreover, using a SUT structure means that the regions in the database can be more readily extended, when suitably reliable published SUT are available. Extending the number of commodities and activities is far more difficult. It will also be possible for user to extend the range of factors, institutions (especially households) and taxes for a region, without having the do the same for all regions.

The discussion has also addressed the issue raised by global whole-economy databases being ‘consistent’ but ‘incomplete’. While IO models are typically not concerned inter-institution transactions, they are critical in CGE models where the method requires the articulation of the full circular flow. Leaving out important transactions, that inevitably distort the database is not desirable, and, arguably, inexcusable when the data necessary to derive estimates exist.

There remains much that can be improved in this database, and efforts are underway to establish a collaborative environment to support these developments. In the meantime, the database will be published and made available without charge by the summer of 2019.

© S.McDonald & K Thierfelder, Apr-19

References

Angel H. Aguiar, A.H., (2015). GTAP 9 Database Documentation: Chapter 6: Macroeconomic Data'. (www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=4823)

Allen, R.I.G. and Gossling, W.F., (eds) (1975). *Estimating and Projecting Input-Output Coefficients*. London: Input-Output Publishing Company.

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

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

Byron, R.P., (1976). 'The Estimation of Large Social Account Matrices', *Journal of the Royal Statistical Society*, A, Vol 141, pp 359-367.

Byron, R.P., (1996). 'Diagnostic Testing and Sensitivity Analysis in the Construction of Social Accounting Matrices', *Journal of the Royal Statistical Society*, A, Vol 159, pp 133-148.

Gehlhar, M., et al. (2008). Reconciling Merchandise Trade Data. Global Trade, Assistance, and Production: *The GTAP 7 Data Base*. B. Narayanan and T. L. Walmsley. West Lafayette, IN, Center for Global Trade Analysis, Purdue University.

Golan, A., Judge, G. and Robinson, S., (1994). 'Recovering Information from Incomplete or Partial Multisectoral Economic Data', *Review of Economics and Statistics*, Vol 76, pp 541-549.

Hussein, Z., and A. Aguiar (2012). **"**GTAP 8 Data Base Documentation - Chapter 6: Macroeconomic Data.**"** In Narayanan, G. B., et al. (2012). Global Trade, Assistance, and Production: The GTAP 8 Data Base. West Lafayette, Indiana, Center for Global Trade Analysis, Purdue University

International Monetary Fund (IMF) (2015a): *Balance of Payments Statistics* (BOPS) database (http://www.imf.org/en/Data#data). Washington, D.C., USA.

International Monetary Fund (IMF) (2015b): *Government Financial Statistics* (GFS) database (http://www.imf.org/en/Data#data). Washington, D.C., USA. #

Jornini, P., McDougall, R., Watts, G. and Dee, P.S. 1994, The SALTER model of the world economy: model structure, database and parameters, Industry commission, Canberra.

Marianne James, M. and McDougall, R., (1993). 'FIT: An Input-Output Data Update Facility for SALTER', SALTER Working Paper No. 17 (www.pc.gov.au/research/supporting/salter-model/workingpaper17/workingpaper17.pdf)

© S.McDonald & K Thierfelder, Apr-19

Lakatos, C., McDonald, S. and Mueller, M., (2008). 'GTAP 7 Database Documentation: Chapter 7Z: Europe'. (www.gtap.agecon.purdue.edu/resources/download/4613.pdf)

Lakatos, C. and McDonald, S. (2008). Code to convert SUT data to GTAP IOT format. *Mimeo*.

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

Lynch, R.G., (1979). 'An Assessment of the RAS Method for Updating Input-Output Tables', in Sohn, I., (ed) *Readings in Input-Output Analysis*. Oxford: Oxford University Press.

McDonald, S. and Sonmez, Y., (2004). 'Augmenting the GTAP Database with Data on Inter-Regional Transactions', Sheffield Economics Research Paper 2004:009. The University of Sheffield

McDonald, S. and Thierfelder, K. (2015). "R23 Model: A SAM based Global Macro CGE Model," available at http://www.cgemod.org.uk/r23.html.

McDonald, S., and Thierfelder, K., (2004a). 'Deriving a Global Social Accounting Matrix from GTAP version 5 Data', GTAP Technical Paper 23. Global Trade Analysis Project: Purdue University.

McDonald, S. Thierfelder, K. and Walmsley, T.(2015). "R23 Database: Sources and Methods," available at http://www.cgemod.org.uk/r23.html.

McDougall, R. and Lee, H., (2007). "GTAP 6 Data Base Documentation: Chapter 17: An Energy Data Base for GTAP" (www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=1951)

Mueller, M., Pérez Domínguez, I., & Gay, S. H. (2009). Construction of Social Accounting Matrices for EU27 with a Disaggregated Agricultural Sector (AgroSAM). Available at http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id=2679: JRC Scientific and Technical Reports.

Narayanan, G. B., et al. (2012). Global Trade, Assistance, and Production: The GTAP 8 Data Base. West Lafayette, Indiana, Center for Global Trade Analysis, Purdue University.

OECD (2015): DAC International Development Statistics, OECD.Stat (http://www.oecd.org/dac/stats/oecdstat-faq.htm). Paris, France.

Robinson, S. and McDonald, S. (2004). 'A General Information-Theoretic Approach to Estimating Disaggregated National Accounts, Including Input-output and SAM Accounts', International Conference on Input-Output Techniques, Brussels, Sept. http://cgemod.org.uk/samest.html

Robinson, S., Cattaneo, A., and El-Said, M., (1998). 'Estimating a Social Accounting Matrix Using Cross Entropy Methods', *mimeo*.

© S.McDonald & K Thierfelder, Apr-19

Round, J.J., (2003). 'Constructing SAMs for Development Policy Analysis: Lessons Learned and Challenges Ahead', *Economic Systems Research*, 15(2), pp 161-183.

Stone, R. (1962a) *A Computable Model of Economic Growth* (with Alan Brown). No. 1 in A Programme for Growth, Chapman and Hall, London, 1962.

Stone, R. (1962b) *A Social Accounting Matrix for 1960* (with Alan Brown and others). No. 2 in A Programme for Growth, Chapman and Hall, London, 1962.

Stone, R., (1977). 'Forward' in Pyatt, G., Roe, A.R. and Associates, *Social Accounting Matrices for Development Planning with Special Reference to Sri Lanka*. Cambridge: Cambridge University Press.

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

Timmer, M.P., (2014). 'The World Input-Output Database (WIOD): Contents, Sources and Methods', WIOD Working Paper Number: 10

UN (1968). *A System of National Accounts*. Studies in Methods, Series F, Rev 4. New York: United Nations.

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

United Nations (2009). *System of National Accounts 2008*. http://unstats.un.org/unsd/nationalaccount/docs/SNA2008.pdf (with World Bank, IMF, European Commission and OECD).

World Bank (2004-2011); *World bank Development Indicators* (WDI). http://data.worldbank.org/data-catalog/world-development-indicators. Washington, D.C., USA.

© S.McDonald & K Thierfelder, Apr-19

Appendix 1 World Bank and IMF Data Sources

Table A1-1 World Bank Development Indicator codes

BM_GSR_FCTY_CD	Primary income payments (BoP, current US$)
BM_GSR_GNFS_CD	Imports of goods and services (BoP, current US$)
BM_GSR_TOTL_CD	Imports of goods, services and primary income (BoP, current US$)
BM_TRF_PRVT_CD	Secondary income, other sectors, payments (BoP, current US$)
BM_TRF_PWKR_CD_DT	Personal remittances, paid (current US$)
BN_CAB_XOKA_CD	Current account balance (BoP, current US$)
BX_GSR_FCTY_CD	Primary income receipts (BoP, current US$)
BX_GSR_GNFS_CD	Exports of goods and services (BoP, current US$)
BX_TRF_CURR_CD	Secondary income receipts (BoP, current US$)
BX_TRF_PWKR_CD	Personal transfers, receipts (BoP, current US$)
BX_TRF_PWKR_CD_DT	Personal remittances, received (current US$)
GC_REV_GOTR_CN	Grants and other revenue (current LCU)
GC_REV_SOCL_CN	Social contributions (current LCU)
GC_REV_XGRT_CN	Revenue, excluding grants (current LCU)
GC_TAX_EXPT_CN	Taxes on exports (current LCU)
GC_TAX_GSRV_CN	Taxes on goods and services (current LCU)
GC_TAX_IMPT_CN	Customs and other import duties (current LCU)
GC_TAX_INTT_CN	Taxes on international trade (current LCU)
GC_TAX_OTHR_CN	Other taxes (current LCU)
GC_TAX_TOTL_CN	Tax revenue (current LCU)
GC_TAX_YPKG_CN	Taxes on income, profits and capital gains (current LCU)
GC_XPN_COMP_CN	Compensation of employees (current LCU)
GC_XPN_GSRV_CN	Goods and services expense (current LCU)
GC_XPN_INTP_CN	Interest payments (current LCU)
GC_XPN_OTHR_CN	Other expense (current LCU)
GC_XPN_TOTL_CN	Expense (current LCU)
GC_XPN_TRFT_CN	Subsidies and other transfers (current LCU)
NE_CON_GOVT_CD/CN	General government final consumption expenditure (current US$ or current LCU)
NE_CON_PETC_CD/CN	Household final consumption expenditure, etc. (current US$ or current LCU)
NE_CON_PRVT_CD/CN	Household final consumption expenditure (current US$ or current LCU)
NE_DAB_TOTL_CD/CN	Gross national expenditure (current US$ or current LCU)
NE_EXP_GNFS_CD/CN	Exports of goods and services (current US$ or current LCU)
NE_GDI_FTOT_CD/CN	Gross fixed capital formation (current US$ or current LCU)
NE_GDI_STKB_CD/CN	Changes in inventories (current US$ or current LCU)
NE_GDI_TOTL_CD/CN	Gross capital formation (current US$ or current LCU)
NE_IMP_GNFS_CD/CN	Imports of goods and services (current US$ or current LCU)
NY_GDP_DISC_CN	Discrepancy in expenditure estimate of GDP (current LCU)
NY_GDP_FCST_CD/CN	Gross value added at factor cost (current US$ or current LCU)
NY_GDP_MKTP_CD/CN	GDP (current US$ or current LCU)
NY_GDS_TOTL_CD/CN	Gross domestic savings (current US$ or current LCU)

Source: http://data.worldbank.org/indicator

© S.McDonald & K Thierfelder, Apr-19

Table A1-2 IMF Balance of Payments Statistics (BOPS) descriptions

Current Account, Net (BPM5)
Imports/Repayments of income/liabilities/outflows
Goods and Services, Debit (BPM5)
Income, Compensation of Employees, Debit (BPM5)
Income, Investment Income, Debit (BPM5)
Current Transfers, Debit (BPM5)
Income, Current Transfers, General Government, Debit (BPM5)
Income, Current Transfers, Other Sectors, Debit (BPM5)
Income, Current Transfers, Other Sectors, Workers' Remittances, Debit (BPM5)
Income, Current Transfers, Other Sectors, Other Transfers, Debit (BPM5)
Exports/Drawings of income/liabilities/inflows
Goods and Services, Credit (BPM5)
Income, Compensation of Employees, Credit (BPM5)
Income, Investment Income, Credit (BPM5)
Current Transfers, Credit (BPM5)
Income, Current Transfers, General Government, Credit (BPM5)
Income, Current Transfers, Other Sectors, Credit (BPM5)
Income, Current Transfers, Other Sectors, Workers' Remittances, Credit (BPM5)
Income, Current Transfers, Other Sectors, Other Transfers, Credit (BPM5)

Source: IMF BOPS, 1993

© S.McDonald & K Thierfelder, Apr-19

Table A1-3 IMF Government Financial Statistics (GFS) descriptions

	IMF taxes	Related to World Bank code
	Tax incomes	
Totals taxes	11 – total tax revenue	GC_TAX_TOTL_CN
Direct taxes	111 – taxes on income and profits (income tax)	GC_TAX_YPKG_CN
Other taxes (factor use and other)	112 – taxes on payroll (factor use tax) 113 – taxes on property (factor use tax) 116 – other taxes	GC_TAX_OTHR_CN
Goods and services taxes	114 – goods and services sales taxes (Includes VAT as a sub-category 11411)	GC_TAX_GSRV_CN
International Trade	115 – taxes on international trade (Import duties and export sub). Includes 1151 – imports duties 1152 – export taxes 1153 – rents on international trade	GC_TAX_INTT_CN includes: GC_TAX_IMPT_CN GC_TAX_EXPT_CN difference
	Other government Income	
Social security contributions	12 - social contributions total	GC_REV_SOCL_CN
Other government reventue	13 - Grants income 131 - Grants income (from foreign govts) 132 - Grants income (from Int Orgs) 133 - Grants income (other govt units) 14 - Other revenue 141 - Property income 142 - sales of G and S 143 - fines, penalties and forfeits 144 - Transfers nec 145 - Premiums, fees and claims etc	GC_REV_GOTR_CN
	Government expenses	
Compensation of employees (current LCU)	21 - compensation of employees	GC_XPN_COMP_CN
Goods and services expense (current LCU)	22 - use of G&S	GC_XPN_GSRV_CN

© S.McDonald & K Thierfelder, Apr-19

Consumption of fixed capital	23 - consumption of fixed capital	
Interest payments (current LCU)	24 - Interest paid	GC_XPN_INTP_CN
Other transfers	25 - subsidies 26 - grants expense 261 - grants expense (to foreign govts) 262 - grants expense (to int orgs) 263 - grants expense (to govt units) 27 - social benefits	GC_XPN_TRFT_CN
Other expense	28 - other exp 281 - property exp 282 - Transfers nec (exp) 283 - Premiums, fees and claims etc (exp)	GC_XPN_OTHR_CN

Source: adapted from IMF GFS (2014). World bank codes added.

Appendix 2 Estimation Method

The estimation method used for this study is 'a general information-theoretic approach to estimating disaggregated national accounts, including input-output and SAM accounts' (Robinson and McDonald, 2004). It is a development of the 'cross entropy methods' (Robinson, 1998). It is major improvements include allowing control totals to be uncertain and providing for multiplicative and additive error terms. The code is available at http://cgemod.org.uk/samest.html

Appendix 3 A Global CGE Model

The GLOBE model adapted to run using this database is a variant of GLOBE 2 (McDonald and Thierfelder, 2009) (see http://cgemod.org.uk/globe2.html). The only change required was to include the multi-product specification for activity outputs that is part of the STAGE model (2015) (see http://cgemod.org.uk/stage2.html). The GAMS codes for GLOBE 2 and STAGE 2 will be open source in the spring of 2019. The GAMS codes for the variant of GLOBE that can be calibrated with this database or the GTAP database will be available by mid summer 2019.

© S.McDonald & K Thierfelder, Apr-19