



The GTAP Data Base Construction Procedure

By

Philip Harslett¹

GTAP Working Paper No. 76

2013

¹ The views expressed in this paper are those of the author and do not necessarily reflect the views of the Productivity Commission. The author is grateful to GTAP Center staff for their valuable input. Any remaining errors remain the responsibility of the author.

THE GTAP DATA BASE CONSTRUCTION PROCEDURE

Abstract

The GTAP Data Base provides a consistent snapshot of the global economy. It consists of a set of product-by-industry input-output tables that represent the structures of more than 100 economies and are linked by bilateral merchandise and cross-border services trade. It is used to support a variety of models that are built to analyse policy changes.

The database is assembled using regional input-output tables and data on trade, energy, protection and macroeconomic aggregates from a variety of international sources. The incompatibilities inherent in the data mean that many compromises are required to produce a fully consistent representation of the global economy.

This paper provides a summary of the GTAP Data Base construction process. The process has improved continuously over the different versions. This paper refers to the process used to produce version 8.1. The detailed documentation is available from https://www.gtap.agecon.purdue.edu/databases/v8/v8_doco.asp. The purpose of this paper is to provide a link between that detailed documentation and the higher-level summary in the 'Introduction to the Global Trade Analysis Project and the GTAP Data Base' paper by Walmsley, Aguiar and Narayanan (2012).

This paper contains three sections. Section 1 provides an overview of the GTAP Data Base construction procedure. Section 2 details the data sources used and the manipulations applied to data obtained from international organizations to ensure that they are globally consistent. Section 3 details the FIT process, which is the procedure used to adjust regional I-O tables so that they are consistent with international data sources.

1. An overview of the GTAP database construction procedure

The Global Trade Analysis Project (GTAP) Data Base is used by GTAP and other global models to simulate the effects of different policy scenarios. The most recent GTAP Data Base is version 8.1. It includes 57 sectors and 134 regions and represents the world economy in 2007.² This paper summarises the processes used by GTAP researchers and external volunteers to create a consistent database. The GTAP-E Data Base, which also includes carbon emissions to enable analysis of carbon abatement policies, is based on the GTAP Data Base.

Creating such a database is not a simple task. The constructed database needs to be fully consistent to be functional as a CGE database. To achieve this, data are collected from a range of sources, and for a variety of reasons, these data are necessarily inconsistent with each other (box 1). As a result, many compromises are required.

The database construction procedure resolves inconsistencies by prioritising certain data. This occurs at three levels.

First, data from individual sources (for example, detailed bilateral trade data) are often inconsistent, and processes are used to resolve these inconsistencies. For example, the ‘Gehlhar method’ (box 2) is used when two differing values are reported for the same bilateral trade flow.

Second, data from different international organizations are made consistent with each other (these data sources are listed in box 1). In most cases, this involves prioritising the detailed trade data over other data. For example, the elements of domestic absorption (private consumption, investment and government expenditure) are scaled so that GDP is consistent with merchandise and services export and import values aggregated from the detailed trade data.³

Finally, the FIT procedure (box 3) is used to update and adjust the regional I-O tables so that they are consistent with the data sourced from international organizations.

² The GTAP 8.1 Data Base also includes a database representing the world economy in 2004. This is the same reference year as the GTAP 7 Data Base, but uses the same construction process as the 2007 Data Base.

³ A substantial proportion of the services trade data present in the GTAP Data Base is estimated due to missing observations. This estimation process is summarized in box 5.

Box 1 Data sources used in the GTAP construction procedure

Two types of data are used to create the GTAP Data Base.

- regional input-output (I-O) tables, usually based on nationally published input-output tables, submitted by researchers from around the world
- data collected from international organizations. These include:
 - merchandise trade data from the United Nations Comtrade database
 - services trade data from the United Nations Trade website and EUROSTAT
 - macroeconomic data from the World Bank Development Indicators
 - tariff data from the International Trade Center's MAcMap system
 - income and factor taxes data from the International Monetary Fund's Government Finance Statistics
 - energy data from the International Energy Agency.

There are many reasons why these data are inconsistent with each other.

- Data often refer to different reference years.
- Countries report different values for the same trade flow because partners are misidentified or commodities are misclassified or classified differently in different countries.
- National statistical agencies produce national accounts with trade figures that are not consistent globally.
- All these data are sourced at different times and are subject to frequent revisions (including after they were accessed, which means that the finalised GTAP database is inconsistent with the revised data).

Source: Narayanan, Dimaranan and McDougall (2012).

Box 2 The Gehlhar method

There are often inconsistencies in bilateral trade data reported by two countries. The Gehlhar method is the approach that is used to reconcile inconsistent merchandise and services trade data.

This approach involves determining the reliability of each country as a reporter. The reliability of each country is measured by constructing indices that calculate the proportion of a country's bilateral trade flows (with its trade partners), where the reported values of the trade flow deviate less than 20 per cent between the two reporting countries. Separate indices are calculated for a country's reliability in reporting exports and imports, for each commodity.

The reliability indices are then used to choose between the reported import flow and the reported export flow. This process is conducted at the HS6-digit level. The resultant consistent trade data are then aggregated to the GTAP sectors.

A more detailed explanation of the Gehlhar method is presented in section 2.2.

Sources: Gehlhar (1996), Gehlhar, Wang and Yao (2010).

Box 3 The FIT process

Regional I-O tables can be inconsistent with data collected from international organizations for many reasons, for example:

- they can have a different reference year
- they can be based on different underlying data collections

The FIT process adjusts regional I-O data so that they are consistent with constraints derived from the data collected from international organizations.

Constraints related to the total value of imports by commodity for each country are satisfied using minimum cross entropy methods (see box 8). Simply put, minimum cross entropy alters the structure of regional I-O tables as little as possible to satisfy the constraints.

Other constraints (such as macroeconomic aggregates) are met by applying input-output methods to the regional I-O tables (see box 9). Simply put, input-output methods involve applying aggregate changes across the aggregate's components as closely to proportionally as possible, given the information derived from the minimum cross entropy process. While input-output methods are similar to the RAS procedure (they both apply aggregate changes across the aggregate's components as closely to proportionally as possible), the implementation of the two procedures is different. Input-output methods are solved as part of a system of equations, whereas the RAS procedure is solved iteratively.

The international data are set as constraints, rather than the regional I-O tables, because they are usually more up-to-date and trade data are already subject to international balance conditions.

The FIT process is described in more detail in section 3.

Sources: James and McDougall (1993), McDougall (2006).

Table 1 Reasons why database values are not recognizable in source

<i>Data item</i>	<i>Source</i>	<i>Reason why database values are not recognizable in source</i>
GDP	World Bank's World Development Indicators (WBWDI) database	Some data are recognizable, others are not, due to subsequent revisions to WBWDI database.
Consumption, Investment, Government expenditure	WBWDI database	Data adjusted so GDP from WBWDI database are consistent with aggregate export and import, merchandise and services trade data.
Exports and imports (merchandise trade)	UN Comtrade (HS6)	Adjusted using the Gehlhar method to create consistent bilateral trade flows.
Exports and imports (services trade)	UN Service Trade website, EUROSTAT	Adjusted using the and CPB ⁴ methods to create consistent bilateral trade flows.
Regional I-O tables	Submitted by researchers (usually derived from I-O tables developed by statistical agencies)	Regional I-O tables updated to match data collected from international organizations using FIT procedure.

The amendments and compromises made to construct the GTAP Data Base are extensive. Indeed, few elements of the original data sources can be recognised in the constructed database (table 1). For example, the value of Australian aggregate consumption expenditure in 2007 in the World Bank's World Development Indicators database is US\$ 476 billion (World Bank 2013) compared with US\$ 478 billion in the GTAP Data Base. This is compounded by the fact that many of the original data sources have been revised since the GTAP Data Base was constructed, which makes it even more difficult to reconcile with published sources.

The procedure for constructing the GTAP Data Base is the product of years of development. Improvements to the procedure are due to improved data availability, feedback from users of the GTAP database, and the efforts of specialists in various areas. Some of the main modifications that have occurred include:

- modifying the FIT procedure to target agricultural production in the EU. This was implemented in response to concerns that the GTAP 5 Data Base differed considerable from EUROSTAT production data
- sourcing energy data from the International Energy Agency (IEA) rather than relying on submitted I-O tables. This was implemented in response to user concerns regarding apparent divergences between earlier GTAP releases and IEA data
- placing more weight on services trade databases as they become more reliable. In the GTAP 6 Data Base construction, aggregate data on services trade were sourced from International Monetary Fund balance of payments data, and then converted to detailed bilateral data using data on bilateral merchandise trade flows to determine their scale

⁴ Details available on this is available from van Leeuwen (2012), accessible from: <https://www.gtap.agecon.purdue.edu/resources/download/6036.pdf>

and a RAS to make it all consistent. Since version 7, and in version 8, detailed bilateral services trade data were sourced directly from the UN Service Trade website and EUROSTAT, and made consistent using a combination of methods including the Gehlhar method.

2. Data sources, initial checks and manipulations

This section presents the data sources used to construct the GTAP Data Base. It also details the checks and manipulations applied to these data prior to the FIT procedure. The manipulations applied to data from international organizations are designed to make them consistent with each other, so they can be used as targets in the FIT procedure.

2.1 Regional I-O tables⁵

The primary data source for the GTAP Data Base is a large collection of single-county I-O tables contributed by researchers from around the world. Most tables submitted to GTAP are based on I-O tables created by national statistical agencies. These tables are usually submitted at the GTAP level of aggregation.

Checks are made to I-O tables that are submitted to GTAP to ensure that:

- each table is balanced (total sales equal total costs for each sector)
- original data have been mapped according to the GTAP concordances
- tax rates implied by the data are not implausible
- values that should be positive, are positive⁶
- there are no unreasonable or unexplainable entries.

The following modifications are then undertaken:

- Production targets from the OECD are applied for 9 agricultural commodities in 37 regions.⁷ The I-O tables are adjusted as part of the FIT procedure. This is in addition to the subsequent FIT procedure that is applied to all commodities and regions (detailed in section 3). To ensure that the results of this initial targeting do not change in the subsequent adjustments, data on GDP, exports and production taxes are also targeted in this initial procedure. In some cases, the agricultural production targets are too small

⁵ Based on Chapter 7 of the GTAP 8 Data Base Documentation.

⁶ Subsidies are negative in this framework as they reduce the cost of output. Normal returns to capital should be positive in a CGE database, but are negative in an I-O table if an industry has recorded a loss for that year. If the losses are considered unusual, submitters are advised to estimate capital returns using the ratio of capital earnings to other costs in other years and to adjust 'changes in stocks' upward to maintain sectoral balance. If capital losses are usual (for example, because of a subsidy), submitters are advised to increase the returns to capital and to adjust indirect taxes downwards to maintain balance in the sector (Huff et al 2000).

⁷ Agricultural production targeting was originally employed in the GTAP Data Base construction process in response to concerns that the data for European Union (EU) member countries in the GTAP 5 Data Base, were considerably different from EUROSTAT production data (van Leeuwen 2002). These discrepancies created problems for researchers who analyzed EU agricultural reform.

to be consistent with the corresponding export target. Production targets are then increased so that export targets (consistent with UN Comtrade data) can be satisfied. More detail on the FIT procedure is included in section 3.

- Modifications are made to I-O tables that contain an incompatible treatment of government services. In many I-O tables, the value for government consumption is very low compared to the World Bank Development Indicators data. Similarly, in some tables, the share of government consumption used on government services is very low compared to other I-O tables. The I-O tables for these regions are adjusted using a representative I-O table constructed from regions with plausible values for government consumption.
- I-O tables that do not contain all 57 sectors are disaggregated using an I-O structure using data from the Food and Agricultural Organization (FAO) of the United Nations (for agricultural sectors) and a representative I-O structure derived from regions with the full set of 57 sectors (for non-agricultural sectors).
- Composite regions are created to include countries without a contributed table. These regions are primarily based on geography (for example, Rest of East Asia and Rest of South America). I-O tables for composite regions are created using a linear combination of I-O tables for selected regions that match as closely as possible each composite region's per capita income level and production pattern.

The reference years for the I-O tables vary. This is because I-O tables for most of the countries are available at five-year or longer intervals and they are often published several years after the data have been collected. These I-O tables are updated to reflect more up-to-date macroeconomic aggregates, trade, energy, and protection targets using the FIT procedure (see section 3).

2.2 Merchandise trade⁸

The primary source for bilateral merchandise trade is the United Nations Comtrade database at the HS6 level. Two types of alterations are made to these data. The first is the process for choosing between two conflicting observations for the same trade flow. The second is to attribute re-exports to their original source.

The Gehlhar method

Reported bilateral trade statistics consist of 'mirror' observations (where both importer and exporter provide a record for the same transaction) and 'single-sided' observations (where only one reporter provides a record for a given transaction). While these 'mirror' observations consist of only 42 per cent of bilateral trade flow observations (58 per cent are single-sided), they cover about 92 per cent of the total value of merchandise trade.

If there is only one observation for a transaction then that value is used. When there are two observations (the two reported values will invariably be at least slightly different, even after transport margins are included) the Gehlhar method is used to choose between two observations

⁸ Based on Chapter 9.A and 9.B of the GTAP 7 Data Base Documentation.

for the same transaction.⁹ The Gehlhar method chooses the value reported by the more reliable reporter.

The reliability of each reporter is measured by constructing a reliability index based on the proportion of a country's trade flows (with all their trade partners) where the reported values of the trade flows deviate less than 20 per cent between the two reporting countries.¹⁰ The calculated indices are commodity specific (at the HS6-level), and distinguish between a country's reliability in reporting exports and imports. All import flows reported on a *cif* basis (including cost, insurance and freight) are converted to *fob* (free on board) values using transport margins that are developed from detailed bilateral commodity data from the United States (box 4).

The reliability indices are then used to choose between the reported import flow and the reported export flow. Averages of the two reported values are not used because it is assumed that discrepancies arise in trade data because one of the reporters makes an error. The resultant consistent trade data are then aggregated from the HS6-level to the GTAP industry categories.

Box 4 International transport margins

Transport margins for the GTAP Data Base are estimated as follows.

1. Margins by commodity and transport mode (ground, vessel, air) are calculated using average bilateral transport margins (*cif/fob*) data for US trade only from the U.S. Census Bureau's Foreign Trade Statistics.
2. Trade routes between countries are classified as either contiguous, intra-continental or inter-continental.
3. Modal shares are formulated for each of these route categories.
 - Trades between USA and its immediate neighbours (Canada and Mexico) are used to calculate the modal shares for trade between contiguous countries.
 - To obtain the modal shares for intra-continental routes, the ratio of the ground to non-ground (air and vessel) modal shares for contiguous countries is reduced by a factor of ten.
 - To obtain the modal shares for inter-continental routes, the same ratio is reduced by a factor of 1000.
4. Margins by commodity and transport mode calculated in step 1 are applied to each route category (and thus bilateral trade flows for all countries) using the modal shares from step 3.

Source: Gehlhar and McDougall (2002).

⁹ Chapter 9.A of the GTAP 7 Data Base documentation states that 'single-sided observations for trade between reporters are also reconciled based on the reporters reliability indicators' (Gehlhar, Wang and Yao 2010, p. 6). This refers to instances where a reporter reports a zero value, as distinct from when a value is missing.

¹⁰ For each importer and exporter, and for a given commodity, the trading partner which produces the largest discrepancy is dropped from their set of transactions before computing their reliability index.

Re-exports

Several countries engage in re-export activity (where it exports a good which it imported in the same state). However, re-exports are not consistent with the trade theory in the GTAP model, and are therefore reallocated to their original source. Re-exports are not typically identified from other exports in trade data, which makes it difficult to attribute re-exports to their original source.

Adjustments are made to the trade data of two major re-exporting countries (the Netherlands and Singapore). For the Netherlands, re-exports are measured indirectly using supply and use tables from Dutch national accounts. Singapore's trade data are adjusted for re-exports using partner data and direct estimates of re-exports produced from Singapore's trade statistics.

Hong Kong is another major re-exporter. Because most of Hong Kong's exports originate in China, many statistical agencies attempt to correct for 'the re-export problem' by attributing all of Hong Kong's merchandise exports to China. Unfortunately, Hong Kong does provide some value-added to these exports (thus they are not pure re-exports), and Hong Kong exports some goods that originate from countries other than China. As a result, the value of Chinese goods exported to its trading partners tends to be greater in the official data of the partner than it is in China's official statistics (which better account for re-exports).

Because information is available on the cause of the discrepancies between reported values, and Chinese official statistics are considered reliable in this case, Chinese reports are used and the Gehlhar method is not applied to Chinese trade. However, using Chinese statistics exclusively does not account for the value of Hong Kong's re-export markup is not included in Chinese statistics.

Instead, a mathematical programming model is used to adjust bilateral trade balances for China and all its partners simultaneously and to estimate Hong Kong's re-export markup. The program tries to preserve the combined trade total for China and Hong Kong while adjusting re-export markups and partner flows. This method:

- accounts for the relative reliability of initial estimates
- produces markups and trade flows that are consistent with each other
- provides more flexibility compared with scaling methods (for example, additional constraints can be imposed, such as upper and lower bounds to avoid implausible values).

Adjustments are not made for other re-exporting countries (e.g. United Arab Emirates) due to a lack of adequate information.

2.3 Services trade¹¹

The following method for constructing services trade data was developed by Lejour, van Leeuwen and McDougall (2008). It was originally used in the construction of the GTAP 7 Data Base, with

¹¹ Based on Chapter 9.D of the GTAP 8 Data Base Documentation. This is only designed to deal with 'mode 1' cross-border services trade and with some income transfers that relate to commercial presence.

some minor changes made for the construction of the GTAP 8 Data Base. This method replaced the method that was used previously, which was developed by McDougall and Hagemejer (2006).

Bilateral services trade data are sourced from the United Nations Trade website (for non-EU countries) and EUROSTAT (for EU countries), for 2007. Data are available for around 54 GTAP regions and 16 data series of which 11 correspond directly to GTAP industries (table 2).

Table 2 Services trade data series^a

<i>Data series</i>	<i>Notes about concordance with GTAP sectors</i>
Total services	Sum of Sea Transport, Air Transport, Other transport Travel, Communication services, Construction services, Insurance services Financial services, Computer and information services Royalties and license fees, Other business services Personal, cultural and recreational services, Government services
Transportation	
Sea transport	Subsection of Transportation
Air transport	Subsection of Transportation
Other transport	Subsection of Transportation
Travel	
Communication services	
Construction services	
Insurance services	
Financial services	
Computer and information services	
Royalties and license fees	
Other business services	
Personal, cultural and recreational services	
Government services	
Other commercial services	Total services minus 205 (transport) and 236 (travel)

^a These data series do not exactly match with GTAP sectors. The sectors are classified according to the 'Extended Balance of Payments Services Classification'.

Source: van Leeuwen (2012).

Reporters are removed if:

- more than 50 per cent of flows for the 'Total services' trade series are missing. Five countries (Iran, Morocco, Nigeria, Chile and Uruguay) are removed under this criterion.
- more than 40 per cent of flows for the 'Total services' trade series are missing and more than 50 per cent of flows for all other data series are missing. Six countries (Indonesia, Thailand, Malaysia, Philippines, Egypt and South Africa) are removed under this criterion.

Box 5 **Estimating services trade flows**

The process for estimating missing services trade flows is as follows:

- For missing trade flows between EU countries:
 - Trade shares for Germany are used. Specifically, for each sector (excluding recreational and other services), German bilateral export shares as a proportion of total German exports to EU countries are applied to the total value of services trade of the exporting country in the missing trade flow.
 - For recreational and other services, the same procedure is used, but French export shares are used instead of German shares.
- For missing trade flows that involve non-EU countries:
 - Trade shares for Germany are again used. Specifically, for each sector (excluding recreational and other services), German bilateral export shares as a proportion of total German exports to all countries are applied to the total value of services trade of the exporting country in the missing trade flow.
 - For recreational and other services, the same procedure is used, but French export shares are used instead of German shares.
 - For some countries the total value of services trade to all countries needs to be estimated. This is either estimated using neighbouring country totals or by taking the difference of total exports and the sum of exports for sectors for which the data are available.

Source: van Leeuwen (2012).

Reliability indices are then calculated employing the Gehlhar method (presented in the previous section) using data for 2004-2008.¹² Reliability indices are then used to choose between the reported export and reported import value for the same flow. If only one flow is reported, this value is chosen. If neither the exporting or importing country has reported a value for the trade flow, the value of the flow is estimated as described in box 5.

2.4 **Macroeconomic data**¹³

The following macroeconomic aggregates are collected from the WBWDI database:

- Gross Domestic Product (GDP)
- GDP per capita
- private consumption (C)
- gross fixed capital formation (I)
- government consumption (G)

¹² Data for 2004-2008 are used to calculate the reliability indices, rather than only 2007 data, to increase the size of the sample.

¹³ Based on Chapter 6 of the GTAP 8 Data Base Documentation

Macroeconomic aggregates (C, I, G) are used as targets in the FIT procedure (discussed in section 3). However, the aggregate import and export values contained in the WBWDI database differ from those implied by the reconciled trade data. Therefore, if the macroeconomic aggregates were left unchanged, the implied value of GDP would not match the WBWDI value of GDP. The macroeconomic aggregates are scaled before they are used in the FIT procedure, to maintain the WBWDI value of GDP and the reconciled trade data.

Input-output tables for composite regions (a collection of regions in the same geographical area that do not have a contributed table) are constructed from other I-O tables. GDP per capita is used to match countries in a composite region to similar countries for which data are available. GDP data is then used to scale data for each individual country in a composite region to obtain the I-O table for the composite region.

2.5 Energy data¹⁴

A special dataset relating to energy use and production is collected from the International Energy Agency's (IEA) Extended Energy Balances (EEB) database (box 6). In most cases this dataset overrides the original data from the I-O tables, the protection dataset and the trade dataset.¹⁵ Also, a volume dataset based on these data is included in the GTAP database.¹⁶

Box 6 The IEA's Extended Energy Balances database

The Extended Energy Balances (EEB) database consists of energy flows, measured in kilotonnes of oil equivalent, indexed by year, country, flow, product, transformation process and final consumer:

- flows include production, imports, exports
- products include different types of coal and peat, natural gas, oil, biofuels and waste, electricity and heat (such as nuclear or hydro)
- transformation processes include traditional power plants, and autoproducers (plants that produce electricity or heat for their own use as an input into the plant's primary activity)
- final consumers are classified in a range of industries that consume energy products as intermediate inputs and households.

The industry classification relies on the International Standard Industry Classification (ISIC), the basis for the GTAP sectoral classification, which facilitates matching to GTAP sectors.

Source: IEA (2013).

¹⁴ Based on Chapter 17 of the GTAP 6 Data Base Documentation

¹⁵ This special treatment arises from users' concern about apparent divergences of energy data between earlier GTAP releases and International Energy Agency data (Babiker and Rutherford, 1997).

¹⁶ This facilitates the analysis of energy policies.

The following steps are taken to prepare the energy volume data for GTAP use:

- The 137 EEB countries are mapped to the 129 GTAP regions.
- The EEB industries and commodities (termed flows and products respectively in the EEB database) are mapped to the energy industries and commodities included in GTAP. The EEB classification of energy industries and commodities is much more detailed than GTAP's, but its classification of non-energy sectors (which are only final users of energy) is less detailed. First, GTAP sectors are aggregated into 22 sectors (table 3). Then for the most part, the IEA's EEB sectoral classifications are treated as disaggregations of the special-purpose GTAP Energy Dataset (EDS).¹⁷ The following concordances are used for commodities (table 4) and industries (table 5). Commodities in the EEB database that are not accounted for in the EDS are discarded (see bottom of tables 4 and 5).
- Energy flows are balanced for global consistency through rescaling. While, the IEA's EEB dataset is balanced within countries (total supply and total consumption of each product are equal for each country), it is not balanced globally (world exports and world imports of each commodity differ).

¹⁷ The EDS is an aggregation created specifically for this procedure, to facilitate the matching of industries in the GTAP and EEB datasets.

Table 3 Concordance between EDS Industries and GTAP Sectors

<i>EDS (aggregated GTAP sectors)</i>		<i>GTAP (original)</i>	
<i>Code</i>	<i>Description</i>	<i>Code</i>	<i>Description</i>
agr	Agriculture, forestry, fishing	pdr	Paddy rice
		wht	Wheat
		gro	Cereal grains n.e.c.
		v_f	Vegetables, fruit, nuts
		osd	Oil seeds
		c_b	Sugar cane, sugar beet
		pfb	Plant-based fibres
		ocr	Crops n.e.c.
		ctl	Bovine cattle, sheep and goats, horses
		oap	Animal products n.e.c.
		rmk	Raw milk
		wol	Wool, silk-worm cocoons
		frs	Forestry
		fsh	Fishing
coa	Coal	coa	Coal
oil	Oil	oil	Oil
gas	Gas	gas	Primary gas production
		gdt	Gas manufacture, distribution
omn	Minerals n.e.c.	omn	Minerals n.e.c.
fpr	Food products, beverages	cmt	Bovine cattle, sheep and goat, horse
		omt	Meat products n.e.c.
		vol	Vegetable oils and fats
		mil	Dairy products
		pcr	Processed rice
		sgr	Sugar
		ofd	Food products n.e.c.
		b_t	Beverages and tobacco products
twl	Textiles, wearing apparel, leather	tex	Textiles
		wap	Wearing apparel
		lea	Leather products
lum	Wood products	lum	Wood products
ppp	Paper products, publishing	ppp	Paper products, publishing
p_c	Petroleum, coal products	p_c	Petroleum, coal products
crp	Chemical, rubber, plastic products	crp	Chemical, rubber, plastic products
nmm	Mineral products n.e.c.	nmm	Mineral products n.e.c.
i_s	Ferrous metals	i_s	Ferrous metals
nfm	Non-ferrous metals	nfm	Non-ferrous metals
teq	Transport equipment	mvh	Motor vehicles and parts
		otn	Transport equipment n.e.c.
ome	Machinery and equipment n.e.c.	fmp	Ferrous metal products
		ele	Electronic equipment
		ome	Machinery and equipment n.e.c.
omf	Manufactures n.e.c.	omf	Manufactures n.e.c.

Table 3 continued

<i>EDS (aggregated GTAP sectors)</i>		<i>GTAP (original)</i>	
<i>Code</i>	<i>Description</i>	<i>Code</i>	<i>Description</i>
ely	Electricity	ely	Electricity
cns	Construction	cns	Construction
tpt	Transport	otp	Transport n.e.c.
		wtp	Water transport
		atp	Air transport
ser	Services n.e.c.	wtr	Water
		trd	Trade
		cmn	Communication
		ofi	Financial services n.e.c.
		isr	Insurance
		obs	Business services n.e.c.
		ros	Recreational and other services
		osg	Public administration and defence,
		dwe	Dwellings

Source: McDougall and Lee (2006).

Table 4 **Concordance between EEB and EDS Commodities^a**

<i>EDS Commodities</i>		
<i>Code</i>	<i>Description</i>	<i>EEB Commodities</i>
coa	Coal	Coking coal, Other bituminous coal and anthracite, Sub-bituminous coal, Lignite/brown coal, Peat, Patent fuel, BKB/peat briquettes
oil	Oil	Crude oil
gas	Gas	Gas works gas, Natural gas, Natural gas liquids
p_c	Petroleum, coal products	Coke oven coke and lignite coke, Gas coke, Coke oven gas, Refinery feedstocks, Refinery gas, Ethane, Liquefied petroleum gases (LPG), Motor gasoline, Aviation gasoline, Gasoline type jet fuel, Kerosene type jet fuel, Other kerosene, Gas/diesel oil, Heavy fuel oil, Naphtha, White spirit and SBP, Lubricants, Bitumen, Paraffin waxes, Petroleum coke, Other petroleum products
ely	Electricity	Electricity, Heat
...	Discarded	Blast furnace gas, Oxygen steel furnace gas, Combustible renewables and waste, Industrial wastes, Municipal wastes renewables, Municipal wastes non-renewables, Primary solid biomass, Biogas, Liquid biomass, Non-specified combust. renewables and waste, Charcoal, Additives/blending compounds, Inputs other than crude or NGL, Nuclear, Hydro, Geothermal, Solar photovoltaics, Solar thermal, Tide, wave and ocean, Wind, Other fuel sources of electricity

^a The justifications for this concordance are presented in McDougall and Lee (2006).

Source: McDougall and Lee (2006).

Table 5 **Concordance between EEB and EDS Industries^a**

<i>EDS industry</i>		
<i>Code</i>	<i>Description</i>	<i>EEB Industry^b</i>
agr	Agriculture, forestry, and fishing	Agriculture
coa	Coal	Coal mines
oil	Oil	Oil and gas extraction (part)
gas	Gas	Oil and gas extraction (part), Gas works
omn	Minerals n.e.c.	Mining and quarrying
fpr	Food products, beverages	Food and tobacco
twl	Textiles, wearing apparel, leather	Textile and leather
lum	Wood products	Wood and wood products
ppp	Paper products, publishing	Paper, pulp, and printing
p_c	Petroleum, coal products	Patent fuel plants, Coke ovens; Blast furnaces, Petrochemical industry, BKB* plants, Petroleum refineries, Liquefaction plants, LNG plants
crp	Chemical, rubber, plastic products	Chemical and petrochemical, Non-specified industry (1/16)
nmm	Mineral products n.e.c.	Non-metallic minerals
i_s	Ferrous metals	Iron and steel
nfm	Metals n.e.c.	Non-ferrous metals
teq	Transport equipment	Transport equipment
ome	Machinery and equipment n.e.c.	Machinery, Non-specified industry (1/16)
omf	Manufactures n.e.c.	Non-specified industry (1/16)
ely	Electricity	Public electricity plant, Autoproducer electricity plant, Public CHP** Plant, Autoproducer CHP plant, Public heat plant, Autoproducer heat plant, Heat pumps, electric boilers, Own use in electricity, CHP and heat plants, Pumped storage (electricity)
cns	Construction	Construction
tpt	Transport	International civil aviation, Domestic air transport, Road, Rail, Pipeline transport, Internal navigation, Non-specified transport, Commercial and public services (1/24)
ser	Services n.e.c.	Non-specified industry (1/16), Commercial and public services (23/24), Non-specified other
...	Production	Production, From other sources—primary energy
...	Private consumption	Residential
...	Exports	Exports, International marine bunkers
...	Imports	Imports
...	Special treatment	Transfers, Non-specified energy sector, Distribution losses, Non-specified industry (3/4), Non-energy use in other sectors
...	Discarded Stock changes	Statistical differences, Gasification, plants for biogas, Nuclear industry, Charcoal production plants

^a The justifications for this concordance are presented in McDougall and Lee (2006).

^b Numbers in brackets refer to the proportion of that classification that is allocated to the EDS classification. The reasoning behind these proportions are described in McDougall and Lee (2006)

Source: McDougall and Lee (2006).

- Re-exports are removed by assuming that imports and domestically produced commodities have the same propensity to be exported, and multiplying each export flow (commodity and partner-country specific) by the share of domestic production in the sum of domestic production and imports.
- Some energy industries are reported to use energy according to the EEB but are not reported to have any output in the GTAP Data Base. In these cases, the industry is eliminated, following the information in the GTAP data. For instance, in the EEB dataset, the Belgian gas sector uses gas despite not having any output in the GTAP EDS database.
- Price and tax data for the EDS commodities are constructed from a variety of sources. These data are not available at the EDS commodity level (it is too aggregated) or the EEB commodity level (it is too disaggregated), so data are obtained for nine energy commodities. These data are primarily sourced from the IEA's 'Energy Prices and Taxes' publication, however a range of other sources are used to fill critical gaps, such as the United States Department of Energy's 'Petroleum Marketing Monthly' (DOE 1997) and the Asian Development Bank's 'Energy Indicators of Developing Member Countries of ADB' (ADB 1994). The price data, which are expressed in terms of various units of measurement, are converted to a tons of oil equivalent basis and in US dollars. The data tend to be available in purchasers' prices. Basic prices are obtained either:
 - directly from source data (only available in a few cases)
 - by adding import duty rates to *cif* import prices.
 - by removing taxes (obtained from IEA data) and margins from purchasers' prices.
- Energy quantity and price data are used to construct values to replace values in the original I-O tables and trade data.
 - The RAS method is used to adjust the trade dataset so that the values for energy commodities in the trade dataset are consistent with the totals for trade in the energy dataset.

2.6 Protection and support data¹⁸

The protection dataset covers import-side, export-side, and domestic support instruments. The import-side instruments are ordinary import tariff rates and anti-dumping duties. The export-side instruments are ordinary export subsidy rates and export tax equivalent rates of quotas, voluntary price undertakings, and voluntary export restraints. Domestic support instruments only apply to agricultural industries and include the rates of total domestic support and the percentage shares of output subsidies, intermediate input subsidies, land-based payments, and capital-based payments to total domestic support.

Protection and support data are obtained from the following sources:

- Tariffs — the International Trade Center's MACMap-HS6 database. These data require little initial processing as the MACMap-HS6 database is already exhaustive and internally consistent.

¹⁸ Based on Chapter 8 of the GTAP 7 Data Base Documentation.

- Agricultural domestic support — data on budgetary transfers are sourced from the OECD Producer Support Estimate (PSE). These data are allocated to four categories; output payments, intermediate-input payments, land-based payments, capital-based payments. Most of the support is not commodity specific, reflecting the fact that domestic support is increasingly provided to farmers regardless of the commodity produced. This support is allocated across commodities using production weights.
- Agricultural export subsidies — for non-EU countries, agricultural export taxes and subsidies are based on country submissions to the WTO on the value of export subsidy expenditures. For EU countries, data are sourced from the ‘Financial report on the European Agricultural Guidance and Guarantee Fund’. The latest version includes information on these and a few non-agricultural sectors such as textiles, at bilateral-level, for over 30 exporting and 200 partner countries.

The protection data are converted to ad-valorem equivalent rates before they are applied to produce the equivalent values in the input-output tables.

2.7 Income and factor taxes¹⁹

Values for income and factor taxes are sourced from International Monetary Fund’s Government Finance Statistics. Four types of taxes are extracted:

- taxes on income and profits earned by companies
- taxes on income and profits earned by individuals
- social security contributions
- taxes on payroll, manpower and property (termed other factor employment taxes).

Any missing observations are imputed with earlier data, converted to ratios of GDP, which are applied to the reference year. If no earlier data are available, then the average tax-to-GDP ratios of countries for which data are available is applied.

Taxes on individuals are applied to skilled and unskilled labour. Taxes on companies are applied to non-labour factors (capital, land and natural resources). Other factor employment taxes (payroll taxes and taxes on property) are applied across all factors. All these allocations are conducted on a pro rata basis — in proportion to the value of each factor.

2.8 Value added²⁰

Natural resources and agricultural land

Regional I-O tables do not provide sufficient information on some of the factors included in the GTAP database, specifically natural resources (including land), skilled and unskilled labour. It is important to include a ‘natural resources’ factor separately from other returns to capital, as the

¹⁹ Based on Chapter 13 of the GTAP 7 Data Base Documentation.

²⁰ Based on Chapter 12 of the GTAP 7 Data Base Documentation.

supply response of liberalisation simulations in resource-abundant economies can otherwise be implausibly large.

Values for the ‘natural resources’ factor are calculated using cost shares. For non-agricultural natural resources, these cost shares are reverse engineered from industry-specific supply elasticities. For agricultural land, these cost shares are sourced from the economic literature on primary factor shares in agriculture, and are then altered so that they are consistent with estimates of the global supply elasticity for agricultural output as a whole.²¹ The natural resources factor cost shares are value for natural resource use.²²

Skilled and unskilled labour

Regional I-O tables do not differentiate between skilled and unskilled labour. Furthermore, there is no global dataset available to disaggregate employment by worker type in each industry. The shares of skilled and unskilled labour were estimated with data from labour force surveys and national censuses of 13 countries, based on the work of Vo and Tyers (1996). For the other countries, a statistical model was developed to explain labour payment shares by sector in the regions for which data were available. This model was used to predict labour payment shares in the remaining countries (box 7).

Box 7 Statistical model used to predict labour shares

The labour share data used by Vo and Tyers (1996) do not cover all GTAP regions. Therefore a statistical model was developed to estimate labour payment shares in the sample regions. This model is used to predict skilled and unskilled labour payment shares in 30 aggregated sectors in the unobserved regions.

A relationship between skilled labour payments and region-specific factors, including stage of development and educational attainment, was postulated. GDP per capita and the average number of years of tertiary education in the population were used to represent stage of development and educational attainment, respectively. Regression revealed a systematic relationship between skilled labour payment shares and these variables.

The model was then used to predict labour splits, by sector, in the remaining GTAP regions. The values of the explanatory variables for all regions were plugged into the regression model to obtain the skilled labour payments for 20 aggregated sectors in each GTAP region. The predicted values of skilled labour payment shares were only used where survey data unavailable.

Source: Dimaranan and Narayanan (2010).

²¹ In the GTAP Data Base, natural resources in agricultural industries are considered to be a separate factor to natural resources in non-agricultural industries. The former is termed ‘land’ and the latter is termed ‘natural resources’.

²² Cost shares are not calculated in a way that allows them to be applied to total capital value (even though theoretically natural resources rents would be a component of capital returns) because total capital value is a more volatile measure than value-added for a given year.

2.9 Assessing the adjustments to data from international agencies

The adjustments made to data collected from international organizations can sometimes be substantial. This section assesses the magnitude of these adjustments for (merchandise and services) trade data and macroeconomic data. A comparison of export-to-GDP and absorption-to-GDP ratios between the GTAP database and the World Bank's World Development Indicators (WBWDI) shows that most adjustments are in fact small and that the bigger ones affect data from less reliable sources or relate to re-exports. That said, these aggregate diagnostics do not show the adjustments that occur at more disaggregated levels.

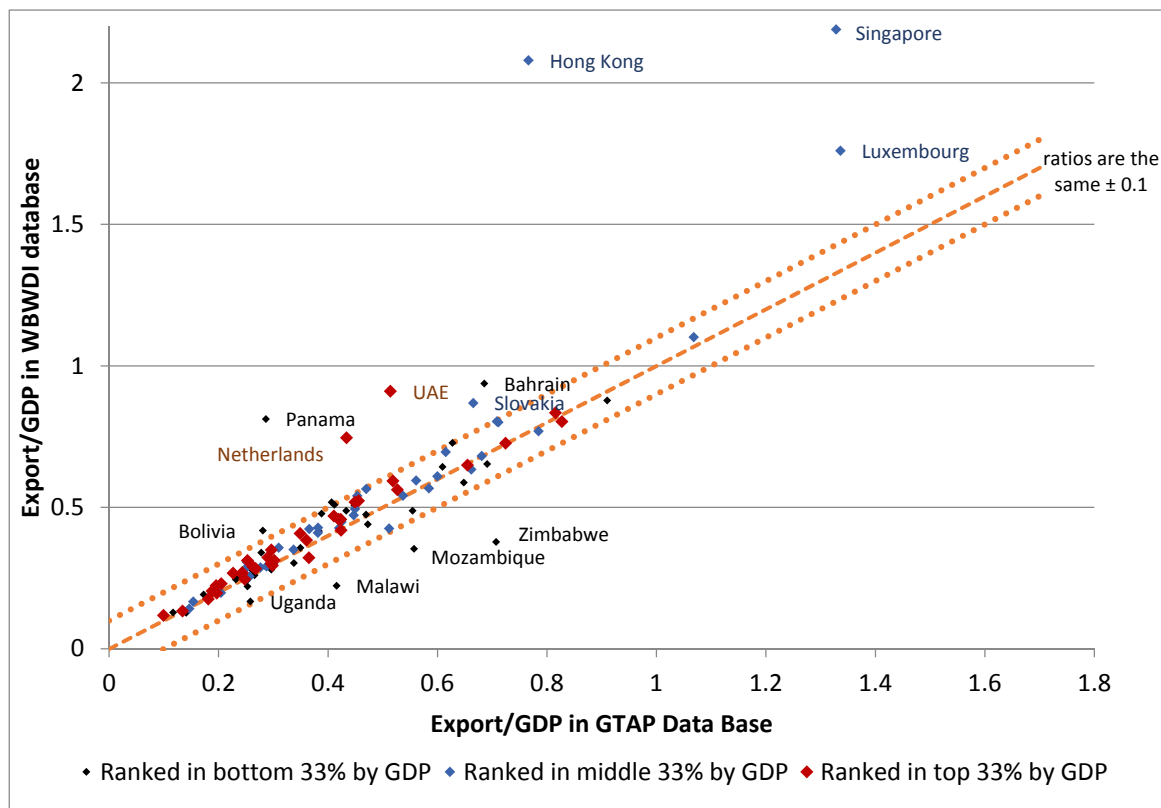
Merchandise and services trade data

Comparing the export to GDP ratios in the GTAP Data Base with the WBWDI provides an indication of the changes that occur to merchandise and services trade data as part of the GTAP Data Base construction procedure (figure 1). Because data are revised periodically by statistical agencies, it is important that any analysis of the changes made to data as part of the GTAP database construction procedure uses data that were released at approximately the same time. Most of the data sourced from the WBWDI database for this comparison were released on June 2012, one month after the GTAP Data Base construction process accessed trade data, which maximises the likelihood that they match the data sourced in May by the GTAP Center.²³ That said, this comparison methodology is still relatively crude. For instance:

- only exports (not imports) are considered
- for each country, export data in the GTAP Data Base are aggregated over goods and services and over trade partners. This means that large changes in individual flows will not be discernible.

²³ The data were sourced independently from Econdata's DX archives. For a few countries, the value of GDP in the GTAP Data Base and the June 2012 release of the WBWDI database were significantly different. In these cases, values from an earlier revision (January 2011) of the WBWDI database were used as they matched more closely the GTAP values.

Figure 1 **Comparison of export to GDP ratio in GTAP and WBWDI^a**



^a Exports include services and merchandise exports. Most of the data sourced from the WBWDI database for this figure were released on June 2012, one month after the database construction process accessed trade data. These data have since been revised. For some countries, data sourced from the WBWDI database for this figure were released on January 2011.

For most countries, there is little difference in the export to GDP ratio between the two data sources, indicating that the adjustments made to trade data are small (figure 1). However, there are some exceptions:

- The largest differences in the export to GDP ratio occur for major re-exporting countries (such as Hong Kong and the Netherlands). These changes are due to the aforementioned re-export adjustments.
- The difference in the export to GDP ratio for Luxembourg is likely to be due to the fact that:
 - financial services represent a significant proportion of Luxembourg's exports
 - financial services are difficult to consistently value (different accounting standards are partly responsible), which means that large adjustments in financial services trade data are required to obtain a consistent database
- The differences in the export to GDP ratio for some countries (such as Singapore, the United Arab Emirates and Zimbabwe) is likely to be due to revisions in the WBWDI data. It is not possible to observe this directly. That said, the large differences between

the GDP values for these countries in the GTAP and WBWDI datasets (irrespective of which WBWDI revision is used), imply that significant revisions occurred both before and after the data were collected for the GTAP construction process.

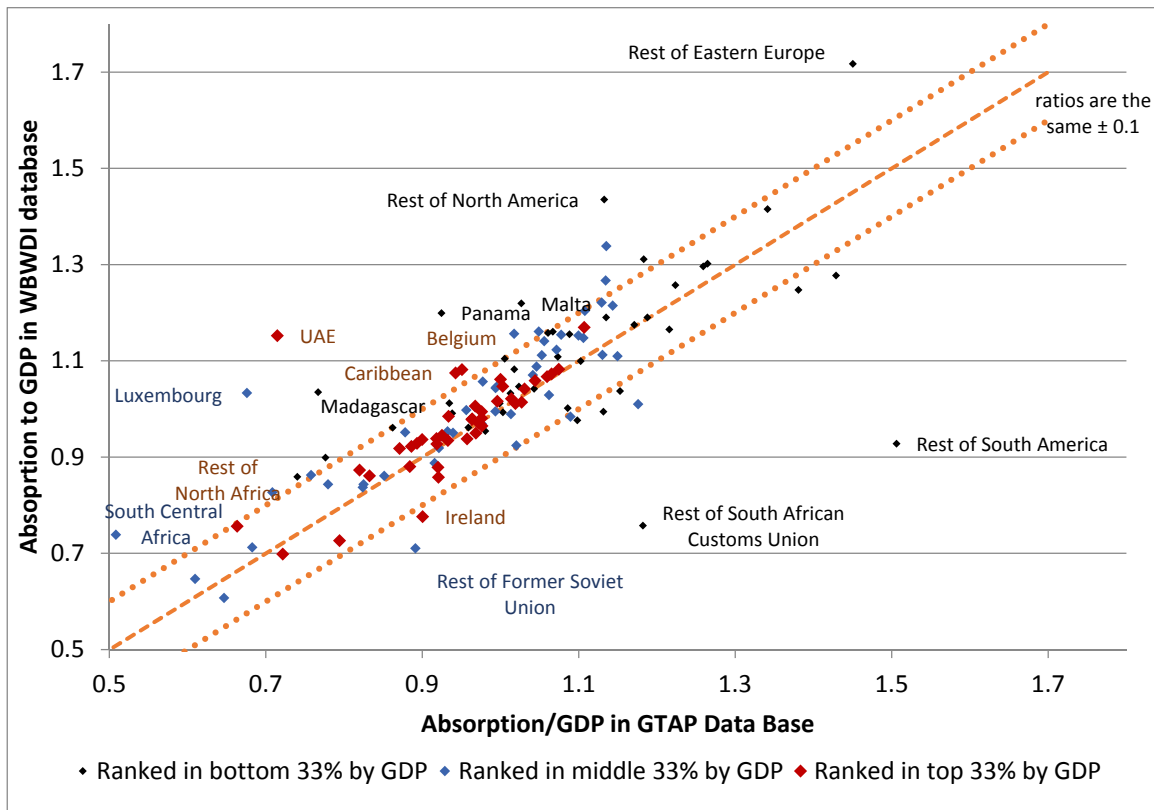
- The remaining countries with notable differences in the export to GDP ratio are small economies that are consistently considered unreliable reporters according to the Gehlhar method. The trade values for these countries need to be substantially adjusted to match data from more reliable trade partners.

Macroeconomic data

Comparing the absorption (C+I+G) to GDP ratio in the GTAP Data Base with the WBWDI provides an indication of the scaling applied to absorption in the GTAP Data Base construction procedure (figure 2). Most of the WBWDI data used in figure 1 were published in June 2012, whereas the WBWDI data used in figure 2 were accessed by GTAP researchers in May 2012. Although chronologically very close, there are still some differences between the revisions.

For most of the larger countries, the required scaling does not significantly alter the absorption to GDP ratio (figure 2, table 6). For some of the smaller countries (and for composite regions), scaling can produce relatively large changes in the absorption to GDP ratio (figure 2).

Figure 2 **Comparison of absorption to GDP ratio in GTAP and WBWDI ^a**



^a Data from the WBWDI database were accessed by GTAP researchers in May 2012. The WBWDI data have since been revised.

Data sources: Narayanan, Aguiar and Robert McDougall (2012), Hussein and Aguiar (2012).

Table 6 **Comparison of absorption to GDP ratio between the GTAP and WBWDI databases — G20 countries^a**

<i>G20 member countries</i>	<i>Absorption/GDP (GTAP)</i>	<i>Absorption/GDP (WBWDI)</i>	<i>Difference between GTAP and WBWDI</i>	<i>GDP^b (US\$b)</i>
South Africa	1.02	1.01	0.01	286
United States of America	1.04	1.06	-0.02	14 062
Canada	0.98	0.99	-0.02	1 424
Mexico	0.97	0.97	0.00	1 026
Brazil	0.96	0.98	-0.02	1 366
Argentina	0.96	0.94	0.02	261
China	0.89	0.92	-0.04	3 494
Japan	0.98	0.98	-0.01	4 378
Korea Republic of	0.98	0.97	0.01	1 049
India	1.00	1.05	-0.04	1 233
Indonesia	0.97	0.95	0.02	432
Russian Federation	0.87	0.92	-0.05	1 300
Turkey	1.06	1.07	-0.01	647
Germany	0.93	0.93	0.00	3 329
France	1.01	1.02	-0.01	2 638
United Kingdom	1.03	1.04	-0.01	2 799
Italy	1.00	1.02	-0.02	2 116
Saudi Arabia	0.72	0.70	0.02	384
Australia	0.97	1.01	-0.04	857

^a Excludes the European Union. ^b The value for GDP are the same in both databases. Data from the WBWDI database were accessed by GTAP researchers in May 2012. These data have since been revised.

Sources: Narayanan, Aguiar and Robert McDougall (2012), Hussein and Aguiar (2012).

In summary, the adjustments that are required to construct a globally consistent database from a large range of inconsistent data sources necessarily involve choosing to preference certain data over other. Despite the need for a large amount of data processing, the resulting database is broadly consistent with the aggregate data for the base year. Many adjustments are required at the more disaggregate levels to accommodate this and other information about the relative magnitudes of different sectors or economic indicators. The continuous improvements in the production of the database have increased its quality over the years, especially in terms of the representation of certain features of the global economy, such as protection and the energy sectors.

3. Adjusting regional I-O tables – the FIT procedure²⁴

This section describes the process for adjusting regional I-O tables so that they are consistent with international data sources (which are made globally consistent as detailed in section 2). The FIT procedure is a data processing system, which modifies raw input-output tables by:

²⁴ Based on Chapter 19 of the GTAP 6 Data Base Documentation, and James and McDougall (1993).

- updating regional I-O tables to the base reference year (I-O tables tend to be less recent than international data)
- removing any remaining discrepancies between regional I-O tables and international data (for example, the value of trade in a regional I-O table may not be consistent with the trade dataset, possibly because of the adjustments made to the trade dataset).
- eliminating ‘changes in stocks’ from the database.

The FIT procedure achieves these objectives by setting the data collected from international organizations (international datasets) as constraints, and using entropy-theoretic methods and input-output methods to adjust the I-O tables. These methods are explained below.

3.1 International datasets are set as constraints

The following datasets sourced from various international organisations are applied as constraints. These data are consistent due to the adjustments summarised in section 2. While all of these datasets have been manipulated to some degree, some have been altered more than others. The constraints are set by:

- The macroeconomic dataset — values at purchaser’s prices for aggregate private consumption, government consumption and investment.
- The trade dataset — border values of exports and imports by commodity.
- The protection and support dataset which includes:
 - import duty rates, by commodity
 - export subsidy rates, by commodity
 - non-commodity indirect tax rates, by industry
 - commodity tax rates on intermediate usage, by industry and commodity
 - rates of tax on private consumption of energy, by commodity
- The energy dataset
 - basic values for intermediate usage of energy, by industry (including non-energy industries) and energy commodity
 - basic values for consumption of energy, by energy commodity, and final user.

2. Updating the regional I-O tables

For each element of the I-O table, percentage changes (in values) are calculated, that transform the original I-O table to the updated I-O table. These changes are linked to equations that determine the percentage changes in prices and quantities. Price and quantity equations are required because some of the international database constraints (such as tax and tariff rates) relate to prices. If these

constraints did not exist, a single set of value equations would be sufficient (as the remaining constraints are in values).

Elements in regional I-O tables need to adjust for three reasons:

- Some elements need to match international datasets. For example, the value of exports for a given commodity in a country's I-O table, needs to match trade data.
- Some elements are a component of a larger aggregated variable that needs to match international datasets. For example, the value of household and government consumption on individual commodities must sum to the aggregate values of household and government consumption in the macroeconomic dataset.
- Some elements of the regional I-O tables will need to adjust to ensure that database constraints hold (such as the requirement that total costs equal total sales for each sector). For example, setting the quantity of 'changes in stocks' to zero or adjusting exports to match international datasets requires that the other use categories adjust to ensure that total costs equal total sales for each sector.

These types of adjustments are undertaken using the price and quantity equations. Two methods are used simultaneously — input-output methods (which are used for both the quantity and price equations) and entropy-theoretic methods (only used for quantity equations). Value equations are used to link changes in prices and changes in quantities to changes in values.

Input-output methods

Input-output methods are used to modify tables by essentially assuming (near) fixed proportions between quantities of inputs and outputs. If there were only value constraints to be met, then input-output methods could assume constant prices and would apply aggregate quantity changes to the output of a commodity across all the users of that commodity as closely to proportionally as possible, while keeping the new database balanced. However, there are both value and price constraints in the FIT process. For input-output methods to solve for the changes in prices and quantities of each commodity:

- the price change for factors must be exogenous
- it must be assumed that there is one basic price for each commodity

Box 8 presents a brief explanation of input-output methods and their relationship to the price and quantity equations (these equations are characterised as different models in the FIT process). A worked example of using input-output methods to update an I-O table is presented in Section 3.3.

I-O quantity model

In the FIT process, the I-O quantity model adjusts values for aggregate consumption, investment and government expenditure in the I-O tables to match international datasets. The I-O quantity model is also used to remove ‘changes in stocks’. Because the international datasets are in values, the quantity model depends on the price model to make these adjustments by:

- feeding the final demand changes (differences between the macroeconomic aggregates implied by the I-O table and the corresponding values in the macroeconomic dataset) through the demand and market-clearing equations to determine new levels for:
 - commodity-specific levels of consumption, investment and government expenditure
 - commodity-specific levels of intermediate usage
 - primary factor employment by industry
- setting the change in the ‘changes in stocks’ use category such that the value of the ‘changes in stocks’ use category becomes zero. The values of this category are then moved into other use categories. This reallocation is as proportional as is possible given the other constraints.

Box 8 An explanation of input-output methods

Input-output methods — essentially the assumption of fixed proportions between quantities of inputs and outputs — can be used to scale an existing I-O table so that any given final demands are allocated across industries in a consistent way. For each commodity i , total revenue is equal to inter-industry sales plus sales to final users, i.e.

$$(1) \quad q_i p_i = \sum_j (q_{ij} p_i) + q_{iF} p_i$$

when prices are uniform across all sales and there is a single final buyer (F). Using input-output coefficients ($a_{ij} = q_{ij}/q_j$) this can be written as

$$(2) \quad q_i = \sum_j (a_{ij} q_j) + q_{iF}$$

For any given set of final demands (q_{iF}) the n equations in (2) can be solved simultaneously for n industry outputs (q_i) assuming fixed input-output coefficients (a_{ij}).

Similarly, input-output methods can be used to calculate output prices for given primary factor prices. For each commodity i , total costs are equal to total intermediate and primary factor costs, i.e.

$$(3) \quad q_i p_i = \sum_j (q_{ji} p_j) + q_{Vi} p_{Vi}$$

when prices are uniform and there is a single primary factor (V). Using input-output coefficients for intermediate and primary inputs this can be written as

$$(4) \quad p_i = \sum_j (a_{ji} p_j) + a_{Vi} p_{Vi}$$

Then for any given set of primary factor prices (p_{Vi}), the n equations in (4) can be solved for n commodity prices (p_i) assuming fixed input-output coefficients.

Instead of exogenising the *quantity* of final demands q_{iF} and factor prices p_{Vi} , it is possible to exogenise the *value* of final demands x_{iF} and factor prices by adding n additional equations

$$(5) \quad x_{iF} = q_{iF} p_i$$

and solving the $3n$ equations in (1), (3) and (5) simultaneously for q_i , p_i and q_{iF} for given p_{Vi} and x_{iF} . This system of equations is no longer linear but for small changes can be solved by linearising in percentage changes as

$$(1') \quad \hat{q}_i + \hat{p}_i = \sum_j (\lambda_{ij} (\hat{q}_j + \hat{p}_j)) + \lambda_{iF} (\hat{q}_{iF} + \hat{p}_i)$$

$$(3') \quad \hat{q}_i + \hat{p}_i = \sum_j (\theta_{ji} (\hat{p}_j + \hat{q}_j)) + \theta_{Vi} (\hat{p}_{Vi} + \hat{q}_i)$$

$$(5') \quad \hat{x}_{iF} = \hat{q}_{iF} + \hat{p}_i$$

where:

$\lambda_{ij} = q_{ij}/q_i$ = distribution shares across an input-output row

$\theta_{ji} = q_{ji} p_j / q_i p_i$ = cost shares down an input-output column

Since the distribution shares sum to unity, (1') can be written — as FIT does — by subtracting $\hat{p}_i$ from both sides, leaving only percentage changes in quantities. Similarly since the cost shares sum to unity, $\hat{q}_i$ in (3') can be removed.

I-O price model

The I-O price model adjusts the basic and post-tax prices for intermediate usage, final demands and primary factors so that they are consistent with the data on assistance and taxes.

The model contains three groups of equations:

- Zero pure profit equations — these equations make the value of total sales and total cost equal for each industry.
- Tax equations — these equations define purchaser prices as the sum of basic prices and taxes.
- Cost of capital equations — these equations link the price of capital services (the user cost of capital) for each industry to the rate of return of that industry and the price of investment.

Minimum cross entropy methods

Minimum cross-entropy (MCE) methods are employed to adjust the regional I-O tables to meet international data on imports (box 9). Input-output methods cannot be used for this task because the number of endogenous variables (the value of imports, by commodity and by use category) outnumber the number of exogenous variables (the total value of imports by commodity).

As part of the MCE process, the input-output structure is divided into two levels. The first level includes the commodity shares — the share of individual commodities in each use category as a proportion of total expenditure on that commodity or the share of a primary factor in total expenditure on primary factors. The second level includes the source shares — the share of domestic products or imports in total expenditure for each commodity in each use category.

These two levels of shares are allowed to vary to meet the disaggregated import targets. As minimum cross entropy is employed, the difference between the original shares and the new shares is minimised. The objective function is formulated so that the commodity shares are altered only if altering the domestic/import shares cannot produce a solution that satisfies the constraints with a reasonable adjustment.

Box 9 Minimum cross entropy methods

Minimum cross entropy (MCE) estimation involves dispersing the values associated with unknowns so that they reflect priors specified by the researcher as closely as possible, given a set of constraints. This technique is a generalisation of the maximum entropy approach. Maximum entropy involves dispersing the values associated with unknowns as widely as possible in the absence of a prior.

In the context of the FIT process, the MCE estimation objective function can be represented by the following equation:

$$\min_{\mathbf{p}} \mathbf{p}' \ln(\mathbf{p}/\mathbf{q})$$

$$\text{s. t. } \mathbf{p}' \mathbf{1} = 1 \text{ and } \mathbf{y} = \mathbf{X}\mathbf{p}$$

Where:

- $\mathbf{p}$ represents a vector of positive shares to be estimated.
- $\mathbf{q}$ represents a vector of priors about the values for the shares (the original regional I-O tables shares)
- $\mathbf{y} = \mathbf{X}\mathbf{p}$ represents any other constraints placed upon the unknowns (the constraints from the internationally sourced data).

This objective function produces values for $\mathbf{p}$ that are as close to $\mathbf{q}$ as possible, given the constraints.

Sources: James and McDougall (1993), Golan, Judge and Miller (1996).

3.3 An example of input-output methods

This section presents two examples using a simple input-output table to illustrate how the input-output methods are used to modify an input-output table:

- In the simple example, final demands are required to increase by 5 per cent
- In the more complex example, the prices of value added for industry 1 and 2 are required to increase 2 per cent and 3 per cent, respectively

Setting up the problem

This section sets up the database and framework before the solutions to the updates are shown. Table 7 presents an I-O table, which abstracts from imports.

Table 7 **A simple I-O table**

	<i>Industry 1</i>	<i>Industry 2</i>	<i>Final demands</i>	<i>Total Sales</i>
Commodity 1	4	2	1	8
Commodity 2	2	6	2	12
Value added	4	4		
Total Production	8	12		

The input-output equations in the following sections are all expressed in percentage changes.

Value equations

Equations (1) through to (4) link the changes in values to changes in quantities and changes in prices. (Price and quantity equations are required because some of the international database constraints, such as tax and tariff rates, relate to prices.)

$$(1) \quad \hat{x}_{ij} = \hat{p}_{ij}^F + \hat{q}_{ij}$$

Equation (1) links the percentage change in intermediate input use of commodity i in value ($\hat{x}_{ij}$) for industry j to the percentage change in the final price of the intermediate input ($\hat{p}_{ij}^F$) and the percentage change in the quantity of intermediate input use ($\hat{q}_{ij}$).

$$(2) \quad \hat{x}_{iF} = \hat{p}_{iF}^F + \hat{q}_{iF}$$

Equation (2) links the percentage change in final demands in value ($\hat{x}_{iF}$) of commodity i to the percentage change in the price of household consumption ($\hat{p}_{iF}^F$) and the percentage change in the quantity of final demands ($\hat{q}_{iF}$).

$$(3) \quad \hat{x}_{Vj} = \hat{p}_{Vj} + \hat{q}_{Vj}$$

Equation (3) links the percentage change in value added (labour and capital) in value ($\hat{x}_{Vj}$) for industry j to the percentage change in the price of value added ($\hat{p}_{Vj}$) and the percentage change in the quantity of value added ($\hat{q}_{Vj}$).

Quantity equations

Equations (5) through to (8) are used to derive quantity changes. The shifter coefficients would usually be derived from the entropy component of the FIT process.

$$(5) \quad \hat{q}_{ij} = \hat{a}_{ij} + \hat{q}_j, \text{ where } \hat{a}_{ij} = \frac{\hat{q}_{ij}}{\hat{q}_j}$$

Equation (5) links the percentage change in intermediate input use ($\hat{q}_{ij}$) for industry j to the percentage change in an input-output coefficient ($\hat{a}_{ij}$) and the percentage change in the output of industry j ($\hat{q}_j$).

$$(6) \quad \hat{q}_{iF} = \hat{a}_{iF} + \hat{q}_F, \text{ where } \hat{a}_{iF} = \frac{\hat{q}_{iF}}{\hat{q}_F}$$

Equation (6) links the percentage change in final demands ($\hat{q}_{iF}$) of commodity i to the percentage change in an input-output coefficient ($\hat{a}_{iF}$) and the percentage change in total final demand ($\hat{q}_F$).

$$(7) \quad \hat{q}_{Vj} = \hat{a}_{Vj} + \hat{q}_j, \text{ where } \hat{a}_{Vj} = \frac{\hat{q}_{Vj}}{\hat{q}_j}$$

Equation (7) links the percentage change in the use of value added ($\hat{q}_{Vj}$) for industry j to the percentage change in an input-output coefficient ($\hat{a}_{Vj}$) and the percentage change in the output of industry j ($\hat{q}_j$).

$$(8) \quad \hat{q}_i = \sum_j (\lambda_{ij} \hat{q}_{ij}) + \lambda_{iF} \hat{q}_{iF}$$

Equation (8) is a market-clearing equation that equates total output of commodity i to total usage of commodity i . The variables λ_{ij} and λ_{iF} denote the share of total sales of commodity i that is attributable to usage by intermediate inputs and final demands, respectively.

Price equations

Equations (9) through to (12) are used to derive price changes. There is assumed to be a uniform price for commodity i across users.

$$(9) \quad \hat{p}_{ji}^F = \hat{p}_i^B + \hat{t}_{ji}$$

Equation (9) links the percentage change in the post-tax price of intermediate inputs ($\hat{p}_{ji}^F$) for industry j to the percentage change in the before-tax price of intermediate inputs ($\hat{p}_i^B$) (it is assumed that there is one basic price across users of commodity i) and the percentage change in power of the tax on intermediate input use ($\hat{t}_{ji}$).

$$(10) \quad \hat{p}_{Fi}^F = \hat{p}_i^B + \hat{t}_{Fi}$$

Equation (10) links the percentage change in the post-tax price of final demand ($\hat{p}_{Fi}^F$) on commodity i to the percentage change in the before-tax price of commodity i ($\hat{p}_i^B$) and the percentage change in the power of the tax on final demand ($\hat{t}_{Fi}$).

$$(11) \quad \hat{p}_j^B + \hat{q}_j = \hat{t}_j + \sum_i \{ \theta_{ij} (\hat{p}_{ji}^F + \hat{q}_{ij}) \} + \theta_{Vj} (\hat{p}_{jV} + \hat{q}_{Vj})$$

Equation (11) is a market-clearing equation that equates the before-tax value of output in industry j to the total costs of that industry. The variables $\hat{p}_j^B$, $\hat{q}_j$ and $\hat{t}_j$ denote the percentage changes in the basic price of the output of industry j , the quantity of output of industry j , and the power of the tax on industry j . The variables $\hat{p}_{Vj}$ denotes the price of value added for industry j . The variables θ_{ij} and θ_{Vj} denote the share of total costs of industry j that is attributable to intermediate inputs and value added.

A simple example

In this example, it is assumed that:

- all the input-output coefficient changes are equal to zero — these terms are determined by entropy equations in the FIT process
- there are no changes to prices. This means that changes in values are equal to changes in quantities.
- the total change in final demands is equal to 5 per cent

These assumptions imply:

$$\hat{q}_F = 5, \hat{q}_{ij} = \hat{q}_j, \hat{q}_{Vj} = \hat{q}_j,$$

Shares from the initial database produce the market clearing equations:

$$(8a) \quad \hat{q}_1 = 0.5(\hat{q}_1) + 0.25(\hat{q}_2) + 0.25(5)$$

$$(8b) \quad \hat{q}_2 = 0.167(\hat{q}_1) + 0.5(\hat{q}_2) + 0.33(5)$$

This system of linear equations produces the following solution:

$$\hat{q}_1 = 5, \hat{q}_2 = 5$$

Combined with the assumptions above, this solution produces all the percentage changes to apply to the original database (table 8). Consistent with the assumptions, an increase in final demand simply scales up the entire table; this illustrates how input-output methods results in percentage changes being apportioned as equally as possible in very simple cases.

Table 8 **Calculated percentage changes**

	<i>Industry 1</i>	<i>Industry 2</i>	<i>Final Demand</i>
Commodity 1	5	5	5
Commodity 2	5	5	5
Value added	5	5	

Table 9 presents the updated database (the original database multiplied by the calculated percentage changes).

Table 9 **Updated I-O table**

	<i>Industry 1</i>	<i>Industry 2</i>	<i>Final Demand</i>	<i>Total Sales</i>
Commodity 1	4.2	2.1	2.1	8.4
Commodity 2	2.1	6.3	4.2	12.6
Value added	2.1	4.2		
Total Production	8.4	12.6		

A more complicated example

In this next example, the same assumptions are applied, except:

- the price of value added for industry 1 and 2 is assumed to be 2 per cent and 3 per cent larger than in the initial data, respectively.
- all other prices changes are endogenous

These assumptions are equivalent to the following:

$$\hat{p}_{V1} = 2, \hat{p}_{V2} = 3, \hat{x}_F = 5$$

$$\text{Where } \hat{x}_F = \sum_i (\theta_{iF} \hat{x}_{iF})$$

These assumptions imply the following input-output equations:

$$(2) \quad 5 = 0.33\hat{x}_{1F} + 0.67\hat{x}_{2F}, (3) \quad \hat{x}_{1V} = 2 + \hat{q}_{V1}, \hat{x}_{2V} = 3 + \hat{q}_{V2}$$

$$(8) \quad \hat{q}_i = \sum_j (\lambda_{ij} \hat{q}_j) + \lambda_{iF} \hat{q}_F$$

$$(11a) \quad \hat{p}_1^B + \hat{q}_1 = 0 + \sum_i \{\theta_{i1} (\hat{p}_i^B + \hat{q}_{i1})\} + \theta_{V1} (2 + \hat{q}_{V1})$$

$$(11b) \quad \hat{p}_2^B + \hat{q}_2 = 0 + \sum_i \{\theta_{i2} (\hat{p}_i^B + \hat{q}_{i2})\} + \theta_{V2} (3 + \hat{q}_{V2})$$

Substituting equations (2) and (3) into equations (8), (11a) and (11b) results in equations (8c), (8d), (11c) and (11d). These equations are sufficient to solve for $\hat{x}_1$, $\hat{x}_2$, $\hat{p}_1$, $\hat{p}_2$.

$$(11c) \quad \hat{p}_1^B + \hat{q}_1 = 0.5(\hat{p}_1^B + \hat{q}_1) + 0.25(\hat{p}_2^B + \hat{q}_1) + 0.25(2 + \hat{q}_1)$$

$$(11d) \quad \hat{p}_2^B + \hat{q}_2 = 0.167(\hat{p}_1^B + \hat{q}_2) + 0.5(\hat{p}_2^B + \hat{q}_2) + 0.33(3 + \hat{q}_2)$$

$$(8c) \quad \hat{q}_1 = 0.5(\hat{q}_1) + 0.25(\hat{q}_2) + 0.25(5 - 0.33\hat{p}_1^B - 0.67\hat{p}_2^B)$$

$$(8d) \quad \hat{q}_2 = 0.167(\hat{q}_1) + 0.5(\hat{q}_2) + 0.33(5 - 0.33\hat{p}_1^B - 0.67\hat{p}_2^B)$$

These equations are sufficient to solve for $\hat{q}_1$, $\hat{q}_2$, $\hat{p}_1^B$, $\hat{p}_2^B$.

$$\hat{q}_1 = 2.333, \hat{q}_2 = 2.333, \hat{p}_1^B = 2.4, \hat{p}_2^B = 2.8$$

Table 10 contains all the calculated percentage changes that will be applied to the original database. With the addition of price changes, the percentage changes in consumption differ by commodity. Table 11 contains the updated database.

Table 10 Calculated percentage changes

	<i>Industry 1</i>	<i>Industry 2</i>	<i>Final Demand</i>
Commodity 1	4.789	4.789	4.789
Commodity 2	5.199	5.199	5.199
Value added	4.380	5.403	

Table 11 Updated I-O table

	<i>Industry 1</i>	<i>Industry 2</i>	<i>Final Demand</i>	<i>Total Sales</i>
Commodity 1	4.192	2.096	2.096	8.383
Commodity 2	2.104	6.312	4.208	12.624
Value added	2.088	4.216		
Total Production	8.383	12.624		

3.4 Some exceptions

As noted previously, the general rule in the fitting procedure is to adjust the regional I-O tables to the international datasets, rather than the other way around. However, there are some exceptions.

- Target tax rates on intermediate inputs used in agriculture are based on the share of intermediate input subsidies in total domestic support, as given in the protection dataset. However, the domestic support shares in the fitted tables may differ from the target shares, as the cost structures of the agricultural industries change during the FIT procedure.

- In a few cases, the target shares for domestic support cannot be attained. This tends to occur when the share of subsidies to intermediate inputs in total domestic support is large, but the share of intermediate inputs in total costs is small. To avoid this problem, a ceiling on the subsidy rate of 90 per cent is imposed for all countries.
- In a few cases, the energy targets are infeasible. This tends to happen where there is a relatively large energy usage target for a relatively small industry. Since the size of each industry after fitting is not known in advance, the energy usage targets cannot be adjusted ex ante. Instead, the FIT procedure is revised so that it adjusts not only the energy use values in the regional I-O table, but also the energy use target from the energy dataset. This is fashioned so that where possible, most of the adjustment is made to the regional I-O table rather than to the energy use target.

3.5 Comparison of original and fitted I-O tables

A comprehensive comparison of the original and fitted I-O tables is impractical given the number of cells contained in the I-O tables. McDougall (2006) uses an ‘entropy-theoretic’ measure to estimate changes between the original and fitted I-O tables (box 10).²⁵ The entropy-theoretic measure produces a single value for each region, which is interpreted as a summary measure of the distance between the fitted and unfitted data.

Table 12 is reproduced from McDougall (2006); its production requires access to the original I-O tables. It contains results for the GTAP 6 Data Base. This type of diagnostic has not been published since. As one might hope, the largest changes occur for regions whose I-O tables are subject to greater uncertainty (table 12). That said, and as might be expected, the order of magnitude of the changes in the better quality results (the first part of the table) are significantly larger than those for the example in box 8.

²⁵ The use categories include the total costs for each industry and the total expenditure in each final demand category. The input categories include the market value of sales for each domestic commodity, the market value of sales for each imported commodity and the market value of each primary factor.

Box 10 Using an 'entropy-theoretic' measure to compare original and fitted I-O tables

Totals for various input and use categories are used to calculate an 'entropy-theoretic' measure that is used to compare original and fitted I-O tables. The use categories include the total costs for each industry and the total expenditure in each final demand category. The input categories include the market value of sales for each commodity and the market value of each primary factor.

For both the unfitted and fitted I-O tables:

- the input and use categories are combined into a hybrid input/use category
- each input and use category is expressed as a share of the total of the input/use category.

The entropy-theoretic measure is:

$$(1) \quad D = \frac{1}{2} \sum_i (F_i - U_i) (\log_e F_i - \log_e U_i)$$

Where U_i and F_i denote *unfitted* and *fitted* shares of various use and input categories from the table.

The percentage change in U_i can be approximated by

$$(2) \quad \log_e F_i - \log_e U_i \approx \frac{(F_i - U_i)}{U_i}$$

Therefore (1) can be approximated by:

$$(3) \quad D \approx \frac{1}{2} \sum_i \frac{(F_i - U_i)^2}{U_i}$$

Thus the entropy-theoretic measure is approximately equal to a weighted sum of the squared changes in the input and use categories, divided by two.

A worked example using the unfitted and fitted I-O tables from box 8

Table A presents the totals and shares for each input and use category for the unfitted and fitted I-O tables derived from the second example in section 3.3. Inserting the values for U_i and F_i into equation (1) results in $D = 0.00005$.

Table A Calculating shares for hybrid input/use category

	Unfitted I-O table		Fitted I-O table	
	Values	Shares (U_i)	Values	Shares (F_i)
Total costs for industry 1	8	0.1653	8.384	0.1670
Total costs for industry 2	12	0.2481	12.624	0.2506
Total expenditure on final demand	4	0.0865	6.304	0.0874
Total costs for commodity 1	8	0.1653	8.384	0.1637
Total costs for commodity 2	12	0.2481	12.624	0.2456
Total costs of value added	4	0.0865	6.304	0.0857
Total of hybrid input/use category	52	1.0000	54.624	1.0000

Source: McDougall (2006).

Table 12 Changes in I-O structure from FIT procedure

	<i>Entropy-theoretic index</i>
Regions with smallest changes to regional I-O table	
Canada	0.01
Japan	0.01
Australia	0.01
United States	0.02
Colombia	0.02
Regions with largest changes to regional I-O table	
Cyprus	0.71
Malta	0.51
Botswana	0.47
Madagascar	0.44
Rest of South African Customs Union	0.38

a The changes in I-O structure presented in this table are for the construction of the GTAP 6 database. An explanation of the measure used to make this comparison is presented in box 10.

Source: McDougall (2006).

References

- Babiker, M. and Rutherford, T. 1997. 'Input-output and general equilibrium estimates of embodied carbon: A data set and static framework for assessment' University of Colorado, Working Paper No. 97-2.
- Dimaranan, B. and Narayanan B. 2010, 'GTAP 7 Data Base Documentation - Chapter 12.B: Skilled and Unskilled Labor Data', Center for Global Trade Analysis, Purdue University.
- Gehlhar, M. 1996, 'Reconciling Bilateral Trade Data for Use in GTAP', GTAP Technical Paper No. 10, Center for Global Trade Analysis, Purdue University.
- Gehlhar, M. and McDougall, R. 2002, 'GTAP 5 Data Base Documentation - Chapter 15.D: Transport Margins and Modes', Center for Global Trade Analysis, Purdue University.
- Gehlhar, M., Wang, Z. and Yao, S. 2010, 'GTAP 7 Data Base Documentation - Chapter 9.A: Reconciling Merchandise Trade Data', Center for Global Trade Analysis, Purdue University.
- Golan, A., Judge, G. and Miller, D. 1996, *Maximum Entropy Econometrics: Robust Estimation with Limited Data*. John Wiley and Sons, Chichester.
- Huff, K. McDougall, R. and Walmsley, T. 2000, Contributing Input-Output Tables to the GTAP Data Base. GTAP Technical Paper No. 1.
- Hussein, Z. and Aguiar, A. 2012, 'GTAP 8 Data Base Documentation –Chapter 6: Macroeconomic Data', Center for Global Trade Analysis, Purdue University.
- IEA (International Energy Agency) 2013, 'Energy Balances of OECD Countries: Documentation for Beyond 2020 Files'.
- James, M. and McDougall, R. 1993, 'FIT: An Input-Output Data Update Facility for SALTER', Salter Working Paper Number 17, Industry Commission, Canberra.
- Lejour, A., van Leeuwen, N. and McDougall, R. 2008, 'GTAP 7 Data Base Documentation - Chapter 9.D: Services Trade Data', Center for Global Trade Analysis, Purdue University.
- McDougall, R. 2006, 'GTAP 6 Data Base Documentation - Chapter 19: Updating and Adjusting the Regional Input-Output Tables', Center for Global Trade Analysis, Purdue University.
- McDougall, R. and Hagemejer, J. 2006, 'GTAP 6 Data Base Documentation - Chapter 15.B: Services Trade Data', Center for Global Trade Analysis, Purdue University.
- McDougall, R. and Lee, H. 2006, 'GTAP 6 Data Base Documentation - Chapter 17: An Energy Data Base for GTAP', Center for Global Trade Analysis, Purdue University.
- Narayanan B., Aguiar, A. and McDougall, R. Eds. 2012, 'Global Trade, Assistance, and Production: The GTAP 8 Data Base', Center for Global Trade Analysis, Purdue University.

- van Leeuwen, N. 2002. GTAP 5 Data Base Documentation - Chapter 11.M: European Union (Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Luxembourg, Portugal, Spain, Sweden)', Center for Global Trade Analysis, Purdue University.
- van Leeuwen, N. 2012, 'GTAP 8 Data Base Documentation - Chapter 9.D: Services Trade Data (DRAFT)', Center for Global Trade Analysis, Purdue University.
- Vo, T. and Tyers, R. 1996. 'Splitting Labor by Occupation in GTAP: Source and Assumptions', Australian National University.
- Walmsley, T., Aguiar, A. and Narayanan, B. 2012, 'Introduction to the Global Trade Analysis Project and the GTAP Data Base', GTAP Working Paper No. 67, Center for Global Trade Analysis, Purdue University.
- World Bank 2013, World Development Indicators database, Washington DC.