ABARE CONFERENCE PAPER 03.2

Data issues in general equilibrium modelling

Benjamin Buetre, Gil Rodriguez and Hom Pant
Australian Bureau of Agricultural and Resource Economics

47th Australian Agricultural and Resource Economics Society Conference
Freemantle, Western Australia, 11-14 February 2003

In recent years global general equilibrium (GE) models have been used to analyse policy issues such as trade liberalisation, regional integration and environmental policies. While a global CGE model is an excellent tool to perform a comprehensive analysis of such policies, it requires an enormous amount of data to calibrate or parameterise the model itself. Quality differences in the data sets might affect the results of the model simulation. Identifying and removing problems in the database is an important task for the effective use of a CGE model.

In this paper, we suggest an approach to detect and remove major inconsistencies from the database of a CGE model. To illustrate, we examine and modify selected interindustry transactions in the GTAP version 5 database, a general equilibrium database of the global economy. The revised database was evaluated against the original by undertaking technical change and trade liberalisation experiments using Global Trade Analysis Project (GTAP) model. Results show that estimates using the modified database are more credible than those from the original database.

ABARE project 2712
ISSN 1447 3666

abareconomics

GPO Box 1563 • Canberra • 2601 • Australia • Fax +61 2 6272 2001 • Tel +61 2 6272 2000 • www.abareconomics.com

1 Introduction

It is now well understood that different sectors, even if seemingly unrelated to each other and supposed to behave competitively, share the same input market and compete directly or indirectly for the budget of the consumer. Because of this interdependence, a policy change may deliver gains to some agents while delivering losses to others within and across other countries. The need to analyse these effects that recognize the interdependence among various agents in the economy and in the world led to an increasing use of computable global general equilibrium (CGE) models.

One particular difficulty in modelling, however, is that CGE models are very intensive in data use. A CGE model demands enormous amount of consistent and precise data on all relevant transactions of all agents, including behavioural parameters of the producers and the households. When the model is a global one, then the construction of the model database becomes even more challenging. The construction of a global database is dependent on information supplied by various national and international agencies, which gather data for different reasons, using different data collection schemes such as a census, statistical sampling and subjective estimates by experts. These national or international data may already contain sampling and non-sampling errors as well as personal biases of experts where they have been used.

In addition to that, the process of assembling a global GE database from these databases may itself introduce additional errors as this exercise involves combining and merging data from various sources, which often are inconsistent, and filling in the missing information by some alternative sources or techniques. Quite often it also requires the use of some balancing technique, such as entropy maximization or RAS procedure, to meet the basic consistency requirements of an equilibrium database. An example of these conditions is the satisfaction of zero profit conditions by all production sectors in all regions. Another example could be that the sum of global exports and imports of a commodity is always zero, and so on. While implementing such conditions and forcing a balance in the equilibrium database it is possible that the routine might allocate the balancing quantities on cells that are not supposed to have any significant positive transactions or may virtually eliminate a significant transaction from where it should be and place it elsewhere in unwanted cells. Since adjusting a cell entry in an input-output table would affect other entries, the balancing procedure could lead to a description of an industry that may not coincide with the one represented by the original data set or with the real industry structure. For example, the re-balanced input-output table might have a zero entry for the paddy rice used by the rice milling industry. It is therefore quite likely that these errors might creep into the database of the global economy, or any GE database for the matter.

While substantial progress had been made in developing the theoretical aspects of CGE models, the quality of data required to calibrate them is seldom discussed. Furthermore, the implications of errors in databases used to calibrate CGE models are rarely analysed (Hazledine and MacDonald 1992; Mercenier and Yeldan 1999).

Since a CGE model is a simultaneous equation system, any data error in one of its components may have a pervasive effect in its overall solution. For example, the presence of errors in a country specific data of a global CGE model could have effects on the results in other countries through the trade and investment linkages.

So far, researchers have focused their attention on the reliability of the estimates of the model parameters and examine the sensitivity of the model solutions to changes in its key parameter values (for example US International Trade Commission 2002). Errors in the social accounting matrix or in the basic input-output tables have remained out of sight. Leaving aside Mercenier and Yeldan (1999), which showed how incorrect data on sectoral labour shares in a CGE model could result in misleading policy prescriptions because of differing direction of welfare effects, it is difficult to find literature that looks into the problem of errors in the basic equilibrium database of CGE models.

This paper intends to contribute further to the analysis of error in the equilibrium database used to calibrate (global) CGE models. In so doing it focuses on the possible errors in the interindustry transaction matrix of a given database and suggests a practical method to mitigate the errors if there are any. This approach has been applied to a selected subset of the GTAP database version 5 (Dimaranan et al. 2002), which is the most recent and comprehensive database available for calibrating global general equilibrium models. We simulated the GTAP model version 6 using both the original and the modified data sets and the solutions are compared.

The remaining paper is organised into five sections. In section two a suggested framework to identify major database problems is described. In section three, a data adjustment approach is presented. This approach has been applied to examine and revise the GTAP database version 5 in section four. Section five reports the results of four illustrative simulations of the GTAP model using the original and the revised GTAP version five databases. Main concluding points are made in section six.

2 A framework for identifying database problems

Social accounting matrix (SAM) is the main database used to calibrate CGE models. The input-output table is the principal component of the social accounting matrix while other components are mainly related to the flow-of-funds and the international trade. The quality of data on the flow-of-funds between agents (or institutions) can be checked against the data from various components of national accounts, as they are mostly aggregates. However, checking the integrity of the cell values of the input-output table and matrix of international trade is rather cumbersome mainly because of their dimensions.

In this paper, we focus our attention only on the underlying input-output tables of the global SAM, in particular the interindustry transaction of the agricultural sectors. This does not mean that we do not think that there would not be any problems in other parts of the database; instead we believe that a similar approach may be applicable in identifying and rectifying the problems in those areas as well.

Table 1a: Original transaction matrix

	C1	C2	C3	Total
C1	5	0	12	**17**
C2	0	7	8	**15**
C3	10	15	0	**25**
Sum	**15**	**22**	**20**	

Table 1b: Modified transaction matrix

	C1	C2	C3	Total
C1	6	8	3	**17**
C2	7	7	1	**15**
C3	2	7	16	**25**
Sum	**15**	**22**	**20**	

Once a database has been obtained or assembled, the first question that comes to mind is whether the database is reasonable. How do we know whether a database is of acceptable standard or not? One of the frequently used checks is to see whether it is balanced or not. Whether the row and column sums satisfy the criteria they are supposed to satisfy. This is also one of the principal criteria used in maintaining the quality of the GTAP database (McDougall, 2002).[1] Clearly, as shown by cell values in Tables 1a and 1b, balancing alone cannot guarantee that the table will reflect the underlying production technologies reasonably well. If the cells of one of the two tables reflect the interindustry transaction correctly then the other one is wrong. They represent entirely different technologies. In what follows we will assume that table 1a is the correct one.

[1] Another check is to see whether the major aggregate components of the national income, such as aggregate expenditure on private consumption, government consumption, investment, exports and imports sum up to the GDP. In assembling the final table of GTAP v5 from regional databases, these conditions along with other constraints were imposed in the data adjustment procedure. For details on the conditions and the procedure see McDougall (2002).

The question now is how to differentiate between alternative tables that are also balanced such as Tables 1a and 1b. To deal with this issue in a global database, in which the commodity, industry and regional aggregation levels are already determined, is not an easy task. It may contain millions of cells to be checked. It is therefore practically necessary to be selective. We suggest the following procedures:

1. Identify the following two lists of transactions by examining the national and/or regional IO tables:

    - A negative list of interindustry transactions - which encompasses all insignificant transactions;

    - A positive list of interindustry transactions - which encompasses all dominant transactions. This list includes all transactions that involve either major inputs of industries (high cost shares) or the industry is a major buyer even if the input is not so important to the industry itself (higher sales share).

2. For all transactions covered in both the negative and the positive lists determine the threshold values of all cost and sales share based on the source IO tables and/or other published sources.

3. Identify problem cells by comparing the actual cell values of the current (assembled or obtained from some other source) global IO table against their threshold values. If the values of the cells that are in the negative list are greater than their corresponding threshold values then they are considered a problem. Similarly, if the values of cells that are in the positive list are less than either of the threshold value (cost or the sales share), then they are also a problem.

In the above example, the cells C_{12}, C_{21}, and C_{33} would be in the negative list with the threshold value of zero. The cells C_{13}, C_{31} and C_{32} would be in the positive list, while C_{11}, C_{22}, and C_{23}, can be in neither. The cell C_{13} has a sales share threshold of 12/17; C_{31} has a cost-share threshold of 10/15 and C_{32} has a cost share threshold of 15/22. It is clear that the positive list of cells and their threshold values identify the main destination as intermediate input for each commodity and major commodity inputs of each industry. The negative list excludes the possibility of an irrelevant transaction becoming an important one such as C_{33} in table 1b. By checking the cell values of table 1b against these criteria, one can immediately come to the conclusion that table 1b is very different from table 1a and therefore needs adjustments.

3 A proposed solution procedure

Now the problem can be restated in the following terms. We do not know the true IO table, such as the table 1a. But we have an IO table like 1b and the two negative and positive lists of the cells and their corresponding approximate threshold values. What can we do to recover a table like 1a from this information? As we are focusing on a subset of cells, and as we have to obtain the threshold values from secondary sources (or possibly from the original IO tables of the countries, which would need updating as well) the likelihood that we would be able to recover the correct IO table is negligible. By imposing the threshold conditions on the critical cell values of the table, however, we may be able to improve the quality of the IO table significantly. To impose these conditions we suggest the following steps:

1. Replace the existing cell values by their corresponding threshold values for cells on both the positive and the negative lists (the shares and the control total yield the corresponding cell values).

2. Calculate the discrepancies between new and control row and column totals.

3. For rows and columns that have large discrepancies, deduce the next possible cells from the list of excluded cells that can absorb the discrepancies and allocate a portion of it manually if such cells could be identified. Also check whether adjustments in the final demand categories may be necessary. The objective here is to leave undistributed errors to a minimum. By doing so we limit the possibility of discrepancies being allocated to cells that may alter the production technologies significantly.

4. Balance the table using any procedure such as manual process, entropy maximization or RAS.

5. Finally, check the new table again, using the steps outlined in the previous section to make sure that the critical cell values are within reasonable bounds of their threshold values.

4 An application: a partial revision of GTAP v5 database

In this section we apply the database checking and adjusting procedure described in the previous sections to GTAP database version 5. For illustrative purpose, we have chosen the interindustry transactions of some agricultural sector in some countries. We believe that this approach helps to focus on the issue more clearly.

One of the reasons for choosing agricultural sectors for this exercise is that it is most likely to have the problems in the cell representing its interindustry transactions. In input-output tables of some countries, the agricultural industries are represented as aggregate sectors (for example, the 1987 IO table of Malaysia; and Botswana's Social Accounting matrix, Horridge, 2002). Sometimes they do not distinguish between specific primary and processing sectors. For example, there is livestock sector in IO table of Bangladesh but there are no other meat products and cattle meat (Noman et al. 2002, Table 11-H-A1). With such aggregation it is likely that some important characteristics of the processing and/or the primary sectors disappear. For example, the processed and primary agricultural sectors have very distinct patterns of demand for inputs - paddy is intensive in land while rice milling is not. When the sectors in the database are highly aggregated, it is difficult to isolate the sectoral impacts of policy changes because the structures of industry supply and demand are a mixture of the characteristics of both primary and processing industries.

In GTAP version 5 database, however, primary and processing industries are explicitly represented for major agricultural industries (Table 2). This representation facilitates the description of the economic transactions that occur between the primary and processing industries and consequently, provides a better perspective of the agricultural sectors. It also permits the assessment of sector-specific impacts and responses that are not possible under highly aggregated agricultural industries. However, the disaggregated representation also results in higher demand for data, which may need some disaggregation in the national IO tables in the first place where some information may already have been lost.

For a global database, the provision of this data in this detail is an enormous task for two reasons. First, the primary and processing components of the agricultural sector are seldom disaggregated in most of the input-output tables available in different countries. Hence, the disaggregating task would require additional data. Secondly, in some countries, particularly in developing countries, the publication of input-output tables is usually outdated. Consequently, the task of updating the data would require additional information that may have an effect on the input-output coefficients.

Table 2. Representation of primary and processed agricultural industries in the GTAP database

Primary sector	*Processed sector*
Paddy rice	Processed rice
Cane and beet sugar	Sugar
Oil seeds	Vegetable oil
Cattle and sheep	Cattle meat
Other animals	Other meat products
Raw milk	Dairy products

When primary agricultural products and agricultural processing industries are represented separately, as in the GTAP v5 database, it is expected that the principal user of the primary sector output would be the processing industry. In addition, sales of primary agricultural products would occur mainly in the intermediate input transactions because raw products needs processing before it can be consumed or sold for other purposes. For example, the most logical user of paddy rice would be the rice processing industry (eg, rice millers). Naturally, most of the paddy rice sales would go to the rice processing industry[1]. This pattern also applies to other primary agricultural products such as cattle, pigs and poultry, and sugarcane/beet sugar. This intermediate demand patterns, such as those reported in column 3 of Table 4, can be extracted from published input-output tables. These cells, representing the strong forward and backward linkages, go into the positive list.

The dominance of the primary-processed industry transactions suggests that sale of a primary agricultural product outside of its corresponding processing industry would be negligible. In the same way, primary and agricultural processing industries are expected not to use *substantial amounts* of non-agricultural intermediate inputs. For example, the vegetable oil industry is not expected to use fisheries in producing vegetable oil. These

[1] Sale of paddy rice to the processed rice industry ranges from 91 to 98 per cent in the published input-output tables of Indonesia, Taiwan, Korea and the Philippines.

interindustry transactions of negligible sizes can also be inferred from the published IO tables (source) and these cells form the negative list of interindustry transactions. For example, we can take the entries with zero values in column 4 of Table 3 and Table 4.

Table 3. Sale of primary agricultural products for intermediate input use in selected countries, (%)

Country (1)	*Primary commodity* (2)	*User Industry* (3)	*Published IO table** (4)	*GTAP version 5 database* (5)	*Revised database* (6)
Malaysia	Paddy rice	Chemical, rubber and plastic	0	33	0
Malaysia	Sugarcane	Chemical, rubber and plastic	0	24	0
Malaysia	Paddy rice	Processed rice	NA	43	78
USA	Paddy rice	Other food industry	NA	75	0
USA	Paddy rice	Processed rice	NA	7	86**
India	Sugarcane	Sugar	NA	11	43
Taiwan	Cattle	Cattle Meat	74	5	86
Korea	Pigs and Poultry	Pigs and poultry meat	88	20	90
Japan	Pigs and poultry	Pigs and poultry meat	72	0.001	81

NA – Not available. Paddy rice is empty in the mapping between GTAP and US IO sectors (Hanson et al, 2002, Table 11.K.A1). This means that paddy rice is not separately identified in the US IO table used to generate the input data for the GTAP database.

*Source: International Input-Output Table, Taiwan-Japan, 1990; International Input-Output Table, Korea-Japan, 1990; 1987 Input-output table of Malaysia.

**GTAP version 5 average of paddy rice sales to processed rice industry in Australia, Japan, Thailand, Vietnam, Indonesia, Korea and China.

An assessment of the GTAP v5 database against these figures indicates that the regional input-output tables contain some problems in the intermediate input transactions in the sense that some primary agricultural products are sold in large quantities to non-agricultural industries as intermediate inputs. For example, sale of paddy rice and sugarcane to the chemical, rubber and plastic industries in Malaysia are 33 and 24 percents of their respective total intermediate sales (see column 5, Table 3). Furthermore, cattle, dairy, and vegetable oil industries in Indonesia, Malaysia and Philippines have high intermediate demands for fisheries products in GTAP database (column 5, Table 4), whereas these transactions are expected to fall in the negative list. On the other hand the 5 per cent of total intermediate sales of cattle output to cattle meat sector in Taiwan (see column 5 of Table 3) while the Taiwanese IO table showing a 74 per cent sale would be a violation of the threshold value for a cell that will be in the positive list. These are some examples of sales and purchases of inputs that are clearly not consistent with those in the published input-output tables.

Table 4. Shares in total intermediate input cost in selected countries, (%)

Country (1)	*Industry* (2)	*Intermediate Input* (3)	*Published IO Table** (4)	*GTAP version 5 database* (5)	*Revised database* (6)
India	Wheat	Electricity	7.5**	35	10
Indonesia	Cattle meat / other meat	Fisheries	0	47	0
Indonesia	Dairy	Fisheries	0	47	0
Philippines	Vegetable oil	Fisheries	0	31	0
Malaysia	Cattle	Fisheries	0	46	0

*Source: International Input-Output Table, Philippines-Japan, 1990; International Input-Output Table, Indonesia-Japan, 1990; and 1987 Input-Output Table of Malaysia

**Personal communication with D. Pratap, contributor of Indian Input-Output table to GTAP.

With the problems in the inter-industry transactions of the selected agricultural industries identified by comparing the cell values of the database with their corresponding threshold values, the GTAPv5 database was modified to conform with the respective sales and cost shares as shown in column 6 of table 3 and 4. The cell values were changed manually and the balance of the IO tables were maintained by proportionately allocating the discrepancies to all primary factor inputs of the respective industries and where necessary, to final demand. For the purpose of this paper, the target threshold values were deliberately chosen to reduce the burden of adjustment in the primary factor inputs.[1] If the exercise covers the evaluation and adjustment of the full database one may use some algorithmic routine based on RAS or entropy maximization with appropriate constraints imposed on the values of the critical cells to balance the table after the initial manual adjustments.

5 Four illustrative simulation of GTAP model with the original and the revised database

To evaluate the effect of database revision on simulation results the GTAP model version six (see Itakura and Hertel, 2003) was simulated with the two sets of databases, the original GTAPv5 database and the revised one (with corrected transactions) under standard (long-run) closure.

Industries in GTAP model produce their outputs by combining value-added (a CES combination of primary factors) and intermediate inputs (CES combination of domestically and foreign produced inputs) using a CES production function. On the basis of the knowledge of the production function and the nature of the database we can form some preliminary expectations.

As the original GTAP database contains cases in which a significant proportion of primary commodity is used by an irrelevant industry, not by its processing industry, such as paddy in Malaysia being used by the chemical, rubber and plastic industry, and by other food products in the USA, one can easily expect that the expansion in output of paddy rice might directly affect the output of the chemical, rubber and plastic industry in Malaysia while expansion in the other food products in the USA would imply a growth in the paddy production sector but not in the rice processing sector of the USA. This is not what is normally expected.

But if primary and processing industries are properly aligned in the database, as is aimed in the revision, input of the primary commodity should normally dominate the intermediate input cost of the processing industry and similarly, majority of the intermediate sale of the primary commodity should go to the processing industry. Hence, given the nature of the production function, in equilibrium, a growth in the processing industry necessarily implies a growth in the primary industry and a growth in the primary industry requires a concurrent growth in the processing industry. In other words, growths in processing and primary industries are expected to move together.

[1] Lahr (1998) also adopted the targeting approach in producing hybrid regional input-output tables.

In order to examine whether these effects can be observed we have designed the following four counter-factual experiments:

1. A 10 per cent output-augmenting technological change in rice production in the US and Malaysia.

2. Elimination of tariffs by all countries on imports of other food products from the USA.

3. Elimination of tariffs by all countries on rice imports from the USA.

4. An application of 50 per cent tariff on imports of pigs and poultry meat into Japan and on imports of cattle meat into Taiwan.

The first and the second experiments highlight the effects of violating the threshold values in the negative list while the third and the fourth experiments are to highlight the effect of violating the threshold values in the positive list.

5.1 A 10 per cent output-augmenting technological change in paddy production in the US and Malaysia.

The results of this simulation are presented in table 5. As expected, with the original GTAP database, productivity improvement in paddy rice production causes an increase in the paddy output in both countries. But it does not lead to in an increase in the output

Table 5. Output response to technological change in paddy rice production (%)

	United States		Malaysia	
	Revised database	*Version 5 GTAP database*	*Revised database*	*Version 5 GTAP database*
Paddy rice	4.8	9.9	4.7	4.5
Processed rice	5.7	0.6	5.4	**7.3**
Chemical, rubber and plastic	0	0	-0.4	4.4

Source: GTAP simulations

of the processed rice in the US and leads to a rise in the output of CRP industry in Malaysia. With the revised database, however, the output of processed rice in both countries goes up almost by the same rate as the output of paddy rice.

5.2 Elimination of tariffs by all regions on other food products imports from the USA.

The simulation results of this exercise are reported in table 6. When all other countries eliminate tariffs on other food products imported from the US, naturally exports of the other food products from the US rises. As a result, with the original database the output of paddy goes up in the US while the output of processed rice falls by 0.23 per cent. With the revised database, however, the output of both paddy rice and processed rice fall, as they have to compete for primary factors and other inputs with the expanding other food sector.

Table 6. Impact on US output of elimination of tariffs on processed rice and other food products imported from the US (%)

	Elimination of tariff on processed rice		Elimination of taiff on other food products	
	Revised database	*Version 5 GTAP database*	*Revised database*	*Version 5 GTAP database*
Paddy rice	108.2	9.6	-0.4	3.1
Processed rice	124.2	212.4	-0.4	-0.23
Wheat	3.2	6.8	1.4	1.3
Other food products	-0.1	-0.1	7.4	7.5
Price of land	6.5	4.0	1.6	1.7

Source: GTAP simulations

5.3 Elimination of tariffs by all regions on rice imports from the USA

The US is a major exporter of rough, brown and milled rice. In the crop year 2000-01, its rice exports were 3,082 thousand metric tons with regular milled rice accounting for a third (Rice situation and outlook yearbook 2001).

Being a major exporter, the US industry is an active proponent of the removal of trade barriers such as in the European Union where tariffs of approximately $370 per metric tons are imposed on milled rice imports, the tariff-rate quotas on US exports to Japan and the antidumping policies in Mexico against US exports. The import tariff on US rice ranges from 5 to 400 per cent in the GTAP database.

As shown in Table 6, there are substantial differences in the responses of the paddy and processed rice sectors to elimination of tariffs on the imports of the US rice when the model is simulated with the two different databases. With the original database, the output of the processed rice rises by 212 per cent while the output of the paddy rice rises by 9.6 per cent only. With the revised database, however, the output of processed rice rises by 124 per cent and that of the paddy rice goes up by 108 per cent. Furthermore, US wheat output increased by approximately half of that obtained using the original database. It is so because now the wheat sector has to compete with the paddy sectors for primary factors whose prices have gone up. For example, the price of land rose by 7 per cent with the new database compared to 4 per cent under the original database.

5.4 A 50 per cent tariff on imports of pigs and poultry meat into Japan and on cattle meat imports into Taiwan

Table 7 summarizes results of simulating the GTAP model for a rise in the tariffs in meat imports into Japan and Taiwan using the original and revised GTAPv5 database. As the tariff goes up, the meat processing sectors (cattle meat and pigs and poultry meat) are expected to expand. Together with these sectors, one would also expect their primary product suppliers to expand as well.

As can be seen from table 7, the output of pig and poultry meat expands in Japan in both databases. The outputs of both the primary and the processed sector moves in almost similar magnitude under the revised data base while the primary sector expands proportionately less than the processed sector in the original database. The same case applies to the application of tariff on cattle meat in Taiwan only that pigs and poultry meat increases in the original database because live pigs and poultry are being sold to cattle meat in large proportion. With the revised database, production of pigs and poultry declines as it competes for resources, notably land, with cattle production.

Table 7. Impact on output of a 50 per cent tariff on pigs and poultry meat imports in Japan and cattle meat imports in Taiwan, (%)

	Japan		Taiwan	
	Revised database	*Version 5 GTAP database*	*Revised database*	*Version 5 GTAP database*
Cattle	1.0	1.6	15.8	2.5
Cattle meat	1.0	1.7	18.4	48.1
Pigs and poultry	6.8	3.3	-0.3	0.3
Pigs and poultry meat	9.9	16.6	-0.2	-0.1

Source: GTAP simulations

Unlike with the original database, in all of the four counter factual experiments with the revised database we have seen that the responses of the primary sector and the processing sector display a consistent and intuitive pattern throughout. It is not only that the outputs of the primary and the processing sectors move together in the same direction at a comparable rate but also that the output of the processing sector grows at a faster rate than the output of the primary sector (see tables 5 –7). It is so because the processing sector is not the only buyer of the corresponding primary sector output. The sectors that buy the rest may not be growing as fast as the primary sector or may even be declining and therefore the processing sector output needs to grow at a bit faster rate to clear the market. Similarly the unusual growth rate of processed rice in Malaysia with the original GTAPv5 database is an indication of the anomaly that the share of the processed rice sector in the total sales of paddy rice in Malaysia was quite small see table 5).

6 Summary and conclusion

In this paper we have argued that the prevalent approach used in the development of databases for global general equilibrium model has some deficiencies. In particular the global database derived from various national and international data by imposing the balancing requirement and targeting the totals of some important aggregates only may end up with an IO structure that may not reflect the true structure of the production sectors of the regions. Model simulations undertaken with such databases may produce misleading results and thus may lead to wrong policy implications. It is therefore important that all users of CGE models check the database first before they run policy simulations.

We have also described a practical approach in identifying the database problem and suggested a systematic method to rectify them. This involved identifying a positive and a negative list of interindustry transactions and their threshold values from the published IO tables or other relevant sources. The negative list contains cells that would be much closer to zero while the positive list contains cells that are dominant in either as a proportion to total (intermediate) sale of the commodity or as a proportion of total intermediate cost of the industry. Comparing the cells of the existing tables against these threshold values were used to identify the problems and replacing the cells with their threshold values and rebalancing the table was the solution suggested to correct the problem. By doing so, we do not claim that the correct IO table will be recovered, but we believe that the quality of the table will be enhanced. The simulation results will accord well with intuition.

This approach was applied to identify and remove the problems from the interindustry transaction matrices of the agricultural sectors in the GTAP v5 database. Four stylised diagnostic simulations were run with the original and revised database. It was found that the quality of the simulation results was much better with the revised database than that with the original database. That is, the results with the revised database are those one would expect *a priori* for the types of the shocks applied. This was not the case with the original database.

References

Dimaranan, B. and McDougall, R. (eds) 2002, *Global Trade, Assistance, and Production*: *The GTAP 5 Data Base*, Center for Global Trade Analysis, Purdue University, Indiana.

Economic Research Service, 2001, Rice: Situation and Outlook Yearbook, United States Department of Agriculture, Washington, DC.

Hanson, K. and Somwaru, A. 2002, 'Sources and Procedures for the Version 5 Datatabase: United States', in Dimaranan, B. and McDougall, R. (eds) 2002, *Global Trade, Assistance, and Production*: *The GTAP 5 Data Base*, Center for Global Trade Analysis, Purdue University, Indiana.

Hazledine, T. and Macdonald, T.D. 1992, *A Critique of Computable General Equilibrium Models for Trade policy Analysis,* Working Paper No. 92-4. International Agricultural Trade Research Consortium, United States Department of Agriculture, Washington D.C.

Horridge, M. 2002, 'Sources and Procedures for the Version 5 Datatabase: Southern Africa', in Dimaranan, B. and McDougall, R. (eds) 2002, *Global Trade, Assistance, and Production*: *The GTAP 5 Data Base*, Center for Global Trade Analysis, Purdue University, Indiana.

Institute of Developing Economies 1996a, *International Input-Output Table, Indonesia Japan1990*, I.D.E Statistical Data Series No. 70,, Tokyo, Japan.

_____ 1996b, *International Input-Output Table, Korea-Japan 1990*, I.D.E Statistical Data Series No. 71, Tokyo, Japan.

_____ 1996c, *International Input-Output Table, Philippines-Japan 1990*, I.D.E Statistical Data Series No. 72, Tokyo, Japan.

_____ 1997 *International Input-Output table, Taiwan-Japan 1990*, I.D.E Statistical Data Series No. 78, Tokyo, Japan.

Itakura, K. and Hertel, T. 2003. *A note on changes since GTAP book model (version2.2a/GTAP94)*, see www.gtap.agecon.purdue.edu/resources.

Lahr, M.L. 1998, A strategy for producing hybrid regional input-output tables, Paper presented at Twelfth International Conference on Input-Output Techniques in New York City, May.

McDougall, R. 2002. 'Updating and Adjusting the Regional Input-Output Tables', in Dimaranan, B. and McDougall, R. (eds) 2002, *Global Trade, Assistance, and Production*: *The GTAP 5 Data Base*, Center for Global Trade Analysis, Purdue University, Indiana.

Mercenier, Jean and Yeldan Erinc, 1999. A plea for greater attention on the data in policy analysis. *Journal of Policy Modeling*, 21(7): 851-873.

Noman, A. and Jong, H. 2002, 'Sources and Procedures for the Version 5 Datatabase: Bangladesh', in Dimaranan, B. and McDougall, R. (eds) 2002, *Global Trade, Assistance, and Production*: *The GTAP 5 Data Base*, Center for Global Trade Analysis, Purdue University, Indiana.

The 1987 Malaysian Input- Output Table, Department of Statistics, Malaysia.

United States International Trade Commission, 2002, *The Economic Effects of Significant US Import Restraints*, Third Update 2002, Investigation No. 332-325, Washington DC.