

Chapter 1

Introduction

Thomas Hertel

The goal of this chapter is to introduce readers to the Global Trade Analysis Project (GTAP) and provide a bit of history on the data base and software which accompany this documentation. Those readers who are already familiar with GTAP may wish to briefly scan section 1.2 to get a feel for new developments in version 5, before moving on to chapter 2. However, individuals who have not had extensive contact with this Project will likely find that this material provides a useful lens through which to view the version 5 data base and associated software. By reading about the historical evolution of this unique data base, one acquires not only an appreciation of how far we have come, but also a sense of what might be possible in the future.

1.1 Overview of GTAP

1.1.1 What is GTAP?

GTAP was established in 1992, with the objective of lowering the cost of entry for those seeking to conduct quantitative analyses of international economic issues in an economy-wide framework. The Project consists of several components:

- a fully documented, publicly available, global data base,
- a standard general equilibrium modeling framework,
- software for manipulating the data and implementing the standard model,
- a global network of more than 1200 researchers (about 600 formal members and 600 informal members) with a common interest of multi-region trade analysis and related issues,

- a consortium of national and international agencies providing leadership and a base level of support for the Project, and
- a World Wide Web site for dissemination of data, software and Project-related information (www.gtap.agecon.purdue.edu).

1.1.2 Motivation for GTAP

As the world economy becomes more integrated, there is an increasing demand for quantitative analyses of policy issues on a global basis. Due to its economy-wide coverage, GTAP is particularly useful for analyzing issues that cut across many diverse sectors. This data base is particularly popular with researchers analyzing the potential impact of: (a) global trade liberalization under a future WTO round, (b) regional trade agreements, (c) economic consequences of attempts to reduce carbon dioxide (CO₂) emissions via carbon taxes, and (d) domestic impacts of economic shocks in other regions (e.g., the Asian crisis, or rapid growth in China). Sector-by-sector analyses of these questions can provide a valuable input into studies of these issues. However, by their very nature, these shocks affect all sectors and many regions of the world, so there is no way to avoid employing a data base which is exhaustive in its coverage of commodities and countries. The Global Trade Analysis Project is designed to facilitate such multi-country, economy-wide analyses.

1.1.3 GTAP Data Base

The central ingredient in GTAP's success has been the global data base. It combines detailed bilateral trade, transport and protection data characterizing economic linkages *among regions*, together with individual country input-output data bases which account for intersectoral linkages *within regions*. (See the glossary at the front of the document for a complete list of GTAP regions and commodities for the version 5 data base.) Construction and maintenance of this data base adheres to the following principles:

Public Availability. The data base is made available to anyone requesting it, at a modest fee. This prevents needless duplication of effort in creating this public good. By charging for the data base we are also able to cover some of the costs incurred in constructing it. However, the bulk of the costs are still covered by the GTAP consortium members through their annual contributions to the Project.

Regular Updates. The current release is the fifth (Version 5) since 1993. (The average life span of a release is two years.)

Broad Participation. The network of GTAP users represents an excellent resource for improvement of the data base. Another benefit from broad participation is the extensive scrutiny to which the data base is subjected. Those who identify areas for improvement or extension of the data

base are free to make this available to GTAP staff in order to have it considered for incorporation into the version 6 data base. The operational concept is "if you don't like it, help fix it!"

Comparative Advantage. By making the data base publicly available, and offering to incorporate improvements provided by members of the network, each individual is able to work to his/her own comparative advantage, while capitalizing on the contributions of others.

Documentation and Replicability. One requirement for new contributions to the GTAP data base is that the sources and procedures used to create them be provided along with the data. This publication represents the basic documentation for the version 5 data base. Additional detail -- GTAP region-level tables in particular -- may be found on the GTAP web site, under version 5 documentation. However, in spite of this extensive documentation, it is not always possible to make this *exhaustive*. Therefore, we refer interested users to the chapter authors themselves in those cases where this document is insufficiently detailed.

1.1.4 Model and Software

In order to operationalize this large data base, a standard modeling framework has been developed. The components of this multiregion, applied general equilibrium model are relatively standard. However, distinguishing features include: the treatment of private household behavior, international trade and transport activity, and global savings/investment relationships. Also, quite a number of auxiliary variables have been introduced in order to facilitate a variety of alternative closures. For a complete description of the GTAP modeling framework and selected applications, see: Hertel, T.W. (editor) *Global Trade Analysis Using the GTAP Model*, New York: Cambridge University Press, 1997.

A copy of the current version of the model and software for implementing it are freely available on the GTAP web site. In addition, a set of GTAP technical papers is available on the web site. Many of these offer extensions to the standard model incorporating features such as imperfect competition, technology spillovers, detailed treatment of energy demands and CO₂ emissions, as well as dynamics.

The model is implemented using the GEMPACK software suite, developed at the Centre of Policy Studies, Monash University, under the direction of Ken Pearson. This software permits the user to conduct simulations of the standard model in which changes in policy, technology, population and factor endowments are examined. The user specifies the split between exogenous and endogenous variables (i.e., model closure). Behavioral parameters may also be altered. Outputs include a complete matrix of bilateral trade, activity flows (and percentage changes) by sector and region, private and government consumption, regional welfare, and a variety of summary variables. Users with access to GEMPACK may also modify the theory of the model. The RunGTAP software environment greatly facilitates use of the standard model, as well as replication of the GTAP

technical papers. There are also GAMS models designed to run on top of the GTAP data base. For more information on these, see the GTAPinGAMS link on the GAMS web site.

There are currently more than 470 documented applications built upon some part of the GTAP framework (see the web site), and many more are currently underway. These are aimed at addressing a great variety of issues including: trade policy reform, regional integration, energy policy, global climate change, technological progress, and links between economic growth and trade.

1.1.5 Short Course in Global Trade Analysis

Each year, during the June - August period, a short course is offered, with the goal of introducing newcomers to the GTAP model, software and data base. This course begins with a seven-week, web-based introduction to the material, followed by an intensive, on-site component lasting one week. Emphasis on hands-on training, has proven to be an excellent way for interested researchers to become operational with GTAP. This course is occasionally offered overseas as well. More information about these offerings is available on the GTAP web site.

1.2 Historical Development of the Data Base

The version 5 data base documented in this monograph builds heavily on earlier work at Purdue, as well as research and data base development efforts at a number of national and international agencies. Indeed, early versions of the GTAP data base built very heavily on the SALTER Project which was undertaken at the Australian Industry Commission during the 1980s and early 90s. Version 1 of the GTAP data base used the same thirteen source input-output (I-O) tables as SALTER (and much of the software for processing them), while supplying new bilateral trade and protection data. Versions 2 and 3 added new regions to the data base, while gradually updating the original SALTER I-O tables. Version 4 replaced almost all of the original SALTER data bases and it added many more regions (reaching a total of 45).

Version 5 has 66 regions — four times the number in the original, version 1 data base. Virtually all of these additions to the data base have been provided by members of the GTAP network, often resident in the countries for which they are supplying data. With the expansion of the number of countries in the data base, the relative importance of the remaining, “composite regions” has become continually smaller, thereby greatly improving the accuracy with which the global economy is represented.

Two contributions stand out in particular with version 5. The first of these was a project undertaken by the Dutch Agricultural Economics Research Institute (LEI), with partial funding from the European Commission, aimed at fully disaggregating the 15 European Union member states. This

was successful and greatly enhanced the value of the GTAP data base for analysis of European policy issues. The second major project was funded by the UK Department for International Development (DFID) and undertaken by Mark Horridge of the Centre of Policy Studies, in conjunction with GTAP staff. This effort aimed to incorporate a set of Social Accounting Matrices constructed by the International Food Policy Research Institute (IFPRI) with support from the Danish government, under the so-called “MERISSA” project. The resulting product was seven new data bases for Southern Africa, which have given rise to a flurry of research on regional trade agreements and other important economic policy issues in Southern Africa.

In addition to increasing the regional coverage, versions 4 and 5 of the data base have added sectoral detail -- thereby departing from the original SALTER sectoral definitions. This is much more difficult than adding regions, since it requires revisiting the source data for all regions -- not just new additions. In version 4, we disaggregated the 37 sectors previously used to the level of 50 sectors. Much of the additional detail was provided in food and agriculture. While there was already a great deal of detail in this area, it was not particularly useful detail, from the viewpoint of the some GTAP Consortium members. In addition, version 4 broke out autos and parts and electronic equipment, in light of their dominance in world trade. Finally, electricity, gas and water are disaggregated in order to better serve the interests of those working on energy-environment issues. Version 5 aims to support improved analysis of services -- particularly in the context of future trade agreements, by disaggregating an additional seven service sectors, including communications, financial services, and insurance, as well as three modes of transportation: air, sea and land. With the advent of the transportation splits, the version 5 data base now disaggregates international transport margins by mode.

One of the most important contributions of GTAP has been in the area of improved bilateral trade data for economic modeling. This work was begun by Marinos Tsigas at Purdue University in the late 1980s, and it has emphasized the use of automated procedures for the reconciliation of reported imports and exports, in addition to the estimation of bilateral trade margins (Tsigas, Binkley and Hertel, 1992). Mark Gehlhar continued this tradition at Purdue and took this interest with him to the Economic Research Service of the USDA, where he has continued to refine his methods (Gehlhar, 1996). Gehlhar's approach to reconciling merchandise trade data is clever, yet simple. He compares a country's reported trade with that reported by its partners and computes a reliability index based on the percentage of the time that these two reports match one another. For any given trade flow, the reporter with the highest reliability index for that commodity is the one who is deemed to be accurate. Gehlhar's work shows that earlier approaches based on broad generalizations (e.g., imports are more reliable than exports, developed countries are more reliable than their developing counterparts) are unjustified. Gehlhar discusses this approach, and his findings for the version 5 data base, in chapters 15.B and 15.C. Gehlhar has also contributed to the estimation of bilateral trade and transportation margins by mode. This new work, undertaken jointly with Robert McDougall, is also documented in chapter 15.D.

Chapter 15.E covers the topic of trade in services. This work has been spearheaded by Robert McDougall. Here, the source data is much weaker – with bilateral flows incomplete and inconsistent in many cases. Therefore, version 5 concentrates on getting the total exports and imports right, with the bilateral flows determined in a mechanical fashion. We plan to incorporate the existing information on bilateral services trade into the version 6 data base using the information-theoretic approach proposed by McDougall (2000).

The GTAP protection data base has also evolved considerably since the project's inception. Most of the work for the version 1 data base was conducted by Bradley McDonald, while he was employed at ERS/USDA. Tariff data was drawn from the GATT Trade Policy Reviews, while support and protection data for agriculture was taken from a combination of OECD and ERS/USDA country studies of Producer Subsidy Equivalents (PSEs). The culmination of the Uruguay Round negotiations provided a rare opportunity to improve GTAP's protection data base. With individual countries submitting tariff schedules to the WTO, a rich data base emerged. In version 2, we were able to build on disaggregated tariff data provided by the US Trade Representative's office. These data, documented in chapter 2 of the GTAP book cited above, was aggregated up from the tariff line level using import weights. In this way, the GTAP data base was able to capture bilateral variation in tariffs for the same composite products. This variation, due to the composition of trade interacting with varying tariff rates, has been found to be quite significant in some cases. The version 2 data base also witnessed introduction of a variety of non-tariff barriers, including anti-dumping duties, countervailing duties, price undertakings, and export restraints on textiles and wearing apparel. Unfortunately these proved to be a one-time only contribution from the WTO and have since been dropped as they became severely outdated.

The version 3 protection data base capitalized on work done for the World Bank's 1995 conference on the Uruguay Round and the Developing Countries (Martin and Winters, eds., 1996). Pre- and post- Uruguay Round protection data compiled by the World Bank, based on the WTO's integrated data base, as well as other sources, made this a unique data base. Unfortunately, this was a one-time effort which was not updated. As a result, the version 4 tariffs data base was sourced from the UNCTAD TRAINS data base, via the World Bank, courtesy of Will Martin, Jerzy Rozanski and Emiko Fukase. This was supplemented on the agriculture side by estimates of market price support and subsidy rates based on the OECD's PSE data base. These were contributed by Marinos Tsigas, then at ERS/USDA. For the non-OECD regions we were forced to draw on earlier protection information, adjusted for changes in world prices and hence average protection levels.

Version 5 again draws on the same source for tariff data and chapter 16.B, authored by Aki Kuwahara, Jerzy Rozanski and Azita Amjadi, describes the WITS system used to produce the aggregated GTAP non-agricultural tariff rates. Unfortunately, at the time version 5 was produced, it was not yet possible to obtain the aggregated tariff preferences from WITS, and so the tariffs in version 5 are still just applied MFN tariffs. The only Free Trade Agreements reflected in version 5 are NAFTA, the Australia-New Zealand Closer Economic Relations Trade Agreement

(ANZCERTA), the European Union customs area, EU-EFTA, and the Southern Africa Customs Union (SACU).

An important new development in the Version 5 GTAP data base has been the use of information based on the Agriculture Market Analysis Database (AMAD) for agricultural tariffs. This project represents a joint effort between a number of national and international agencies aimed at providing consistent protection estimates for agriculture. Since these data were not all available in *ad valorem* form, additional work was required, and this was undertaken at ERS/USDA under the leadership of Paul Gibson. Details are provided in chapter 16.C, authored by Gibson, Wainio and Whitley.

Another big change in version 5 has to do with the treatment of agricultural subsidies. A decade of decoupling in many of the OECD countries had left the earlier approach of treating these as a simple output subsidy quite out-of-date. In version 5, we take advantage of the PSE classification of subsidies into more refined categories in order to better reflect the economic impact of farm subsidies, treating some as intermediate input subsidies, and others as subsidies on land and/or capital. This work was undertaken at the Danish Institute of Agriculture and Fisheries Economics (SJFI) and at the Economic Research Service of USDA. The allocation of these subsidies is not without controversy, and the two agencies ended up with somewhat different subsidy profiles. In the end, a standard formula was agreed upon and incorporated into the version 5 data base, and this work is documented in chapter 16.E, authored by Hans Jensen.

One point that needs to be strongly emphasized for users of this data base is that the protection data supplied in GTAP is intended to represent a *starting point* for analysis. Any researcher using GTAP to conduct analysis of a specific policy liberalization scenario must scrutinize these data carefully for the focus countries in her/his analysis. In many cases, some adjustment will be required to reflect improved information that is often available from country-specific sources. Having ascertained cases where the protection information in GTAP must be altered to better reflect reality, the user can take advantage of the ALTERTAX program to make these changes. This program is documented in GTAP Technical Paper No. 12 by Gerard Malcolm (1998), and it has been made into an easy-to-use feature of the RunGTAP software authored by Mark Horridge (2001).

A final area in which there have been some significant changes in the version 4 data base is that of primary factor splits; i.e. how much of the value-added in a given industry is attributable to labor, how much to capital, etc. Version 4 offered for the first time a split of labor payments into unskilled and skilled components for each of the sectors in each of the regions in the data base. This feature has been retained in version 5. The sector/region splits are done based on 15 national labor surveys which have been processed into a common format and definition by Tri Thanh Vo and Rod Tyers, with additional work by Jing Liu and Nico van Leeuwen. By fitting a regression model to these data, the authors are able to predict industry splits for the remaining GTAP regions, based on per capita GDP and the stock of tertiary educated labor in the economy as a whole (see chapter 18.D and GTAP Technical Paper No. 11 by Liu *et al.*, 1998, for more details).

In addition to the labor splits, versions 4 and 5 have broken out a natural resource input into the forestry, fisheries, coal, oil, gas, and other minerals sectors. The cost share of this natural resource input has been chosen (“calibrated”) in order to replicate a target elasticity of supply for each of these resource-constrained sectors (see chapter 18.C for more details). By constraining these natural resource-based sectors in this way, we hope to capture the appropriate price/quantity responses to economy-wide shocks in the absence of detailed modeling of resource stocks and their depletion.

Another important area of recent development for the GTAP data base has been the incorporation of energy volume data. These were first introduced in an interim version following release of version 4 – nicknamed GTAP 4E. This was widely used by those conducting research on climate change policy. The process of reconciling energy volume and price data obtained from the International Energy Agency (IEA) as well as other sources, with the value flows in the GTAP data base, has proven to be a challenging task. There remained substantial room for improvement after the release of GTAP 4E. For this reason we invested considerable effort in the energy data base and associated programs used in version 5. This work, led by Jean-Marc Burniaux, was quite successful at reducing the discrepancies between the two sources of information on energy usage. An overview of this work is provided in chapter 17, with more details available in the web-only documentation accompanying this volume.

1.3 Future Development of the Data Base

Future development of the data base is heavily influenced by the GTAP advisory board. This group is made up of representatives from each of the agencies in the GTAP Consortium. Based on feedback at the 2000 and 2001 board meetings, we are devoting considerable energy to improving the overall quality of the data base. There are several dimensions to these efforts. Firstly, we plan to work more intensively with national data base contributors. One way of doing this is to encourage earlier contributions of national data bases for version 6. Accordingly, we have instituted a process of interim releases of version 5, *i.e.* 5.1, 5.2, etc. These will occur on roughly a six month cycle and will involve a new version of the GTAP data base for which everything else (base year, international trade data, protection data, balancing programs, etc.) remains unchanged. The only difference is the introduction of the new I-O data. The latter could involve an update of an existing region, or an entirely new country to be broken out of one of the existing, composite regions in version 5. These interim releases will spread out more evenly the contribution of new data bases which have tended to become bunched up in the period prior to the pre-release of a new version.

Another aspect of quality improvement in version 6 will be increased communication with contributors regarding the adjustments required to make their data compatible with the rest of the GTAP data base – and seeking to understand where and why discrepancies arise in this process. This

may, in turn, suggest areas for improvement in the contributed data bases. Or alternatively, these discrepancies may point to needed improvements in the international data bases used in GTAP – or perhaps alternative means of reconciling the two. A good example of the latter that has arisen in the context of version 5, has to do with the overall size of the agricultural sector in many of the EU member states. Since the source data bases are out-of-date in many cases, and since the relative size of agriculture in many of these countries has been shrinking, a “neutral” update to match base year macroeconomic totals (see chapter 19 for a discussion of the update methodology) results in an overly large agriculture sector in many cases. Explicit targeting of agricultural output figures for the base year would overcome this problem.

This leads to the second area in which we hope to improve the quality of the GTAP data base, namely the multilateral data sets. Top on the list of priorities in this area is the goal of incorporating tariff preferences. With the rapid proliferation of regional trade agreements, the number of tariff preferences has skyrocketed. Picking these up and appropriately aggregating them to the level of GTAP commodities has been a purely technical problem which appears to have been finally resolved. Bringing this improved information into the version 6 GTAP data base will be a high priority. Other areas where the multilateral data base must be improved include: estimation of bilateral services trade flows, estimation of transport margins, by mode, and the treatment of some of the most obvious non-tariff barriers (e.g., anti-dumping duties).

Another promising direction for development of future GTAP data bases involves the incorporation of additional data on physical and environmental variables. Examples include emissions of non-CO₂ greenhouse gases from agriculture and industry, as well as data on land use by sector and region. These types of information would greatly enhance the value of GTAP for analysis of climate change policy and it appears that a project in this area will be forthcoming in the future. Recent GTAP board meetings have also suggested the desirability of improving the data on government revenues and expenditures, as well as international investment data – but to date we have not found external support for such an undertaking.

The GTAP data base is a dynamic entity which is evolving in response to the needs and support of individual users as well as public agencies with an interest in international trade, natural resources and the environment. We encourage you to become involved in this network, subscribing to our discussion list, possibly attending the short course, and using this data base. We look forward to your feedback!

1.4 Outline of this Document

The remainder of this documentation is made up of two parts. Part I, comprising chapters 2-10, presents summary data derived from the entire GTAP data base as well as an overall summary of some of the associated utilities. The data base overview chapters include macroeconomic data by region (chapter 2), disposition of sales and costs by commodity and by region (chapter 3), protection data (chapter 4), and I-O multipliers (chapter 5). Since the data base has now grown to the point that it is not possible to include summary data for all sectors and regions in the hard-copy document, some of these summary tables are placed on the web, in the form of supplementary tables. Chapter 6 computes effective protection rates while chapter 7 discusses the time series trade data available with each GTAP release. Chapters 8-10 describe the GTAP data base and the software used in aggregations and in viewing data.

Part II begins with a detailed documentation of the new input-output tables for the version 5. Documentation of the earlier contributions are available in the monographs documenting prior releases (see the web site for electronic versions). Chapters 12-14 discuss other important aspects of the domestic data base construction procedures, including: the supplementary food and agricultural data base, methods used for disaggregating source tables, and procedures for building the composite regions which cover the “rest of the world”.

The trade data base is documented in chapter 15, including an overview (15.A), treatment of re-exports (15.C), estimation of bilateral trade margins by mode (15.D) and estimation of services trade flows (15.E). Protection data is another large topic, covered in the multi-part chapter 16. This includes discussion of the merchandise tariff data, as well as protection data for agriculture in particular (tariffs, export subsidies and domestic support), and export tax equivalents associated with the Agreement on Textiles and Clothing.

Chapter 17 covers the sources and procedures used to build the energy data base that accompanies the version 5 data base. While only volume data are supplied with GTAP 5, price data were also required in order to reconcile these volume data with GTAP’s value-based energy expenditures. Chapter 18 describes the macroeconomic, capital stock, and energy data used in the data base. The chapter also documents the methodology used to derive labor splits for all sectors, and the primary factor shares for the natural resource-based industries. Chapter 19 summarizes the procedure (FIT) for updating the regional input-output tables to the base year 1997. Chapter 20 describes how the GTAP behavioral parameters were obtained. Finally, chapter 21 describes the process of assembling the full data base.

References

- Gehlhar, Mark. 1996. "Reconciling Trade Data for Use in GTAP," GTAP Technical Paper No. 10, Center for Global Trade Analysis, Purdue University.
- Hertel, Thomas W. (Editor). 1997. *Global Trade Analysis: Modeling and Applications*. New York; Cambridge University Press.
- Liu, Jing, Nico van Leeuwen, Tri Thanh Vo, Rod Tyers, and Thomas W. Hertel. 1998. "Disaggregating Labor Payments by Skill Level in GTAP," GTAP Technical Paper No. 11, Center for Global Trade Analysis, Purdue University.
- Malcolm, Gerard. 1998. "Adjusting Tax Rates in the GTAP Data Base," GTAP Technical Paper No. 12, Center for Global Trade Analysis, Purdue University.
- Martin, W. and L.A. Winters (Editors). 1996. *The Uruguay Round and the Developing Countries*. New York, Cambridge University Press.
- McDougall, Robert. 2000. "International Services Trade Data for CGE Modelers," paper presented at the Third Annual Conference on Global Economic Analysis, Melbourne Australia.
- Tsigas, Marinos E., Thomas W. Hertel, and James K. Binkley. 1992. "Estimates of Systematic Reporting Biases in Trade Statistics," *Economic Research*, Volume 4 4:297-310.