

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 the GTAP 6 Data Base, 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 GTAP 6 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 3,700 researchers in more than 100 countries with a common interest in global economic analysis of trade, resources and the environment,
- 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.org).

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 financial 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 inter-sectoral linkages within regions. (See the glossary at the front of the document for a complete list of regions and commodities for the GTAP 6 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 a portion 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 sixth (GTAP 6) since 1993. (The average life span of a release is about two years.)

Broad Participation. The network of GTAP users represents an excellent resource for extension 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 next release of the 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 core documentation for the GTAP Data Base. Additional detail may be found on the GTAP web site, under GTAP 6 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. Often it is only through bilateral correspondence of this sort that data base limitations can be properly identified and remedied. This takes us to the final ingredient of the GTAP Data Base, namely quality control.

Quality Assurance. As the GTAP Data Base has become more widely used and the policy analyses based on this platform have become more influential, the demands for improved quality control have also increased. In recent years, this has emerged as the top priority of consortium members and the Center has therefore devoted increasing resources and attention to this issue. There are a variety of ways in which the Center seeks to provide quality assurance. This begins with documentation and replication. If users don't know what is in the data base, and how it has been constructed, they cannot intelligently critique it.

Our approach to quality control differs somewhat between the international data bases (trade data, tariffs, energy volumes, etc.) and the domestic data bases (most notably the national input-output tables). Quality control for the international data base inputs is in some ways easier. Here we often have alternative data bases against which to compare the source data supplied to GTAP. For example, before the decision was made on where to source the tariff data for the GTAP 6 Data Base, we compared four different international data bases, each of which was maintained by a different consortium member. We also have a policy of placing a premium on continuity of suppliers. Recognizing that data bases only improve through use, revision, and further use, we do not change sources lightly. For example, the international trade data base in GTAP 6 is supplied by the same individual (Mark Gehlhar) who has supplied trade data for versions 1 – 5 of the data base. Over time, he has obtained considerable feedback from users and the methodology used for trade data reconciliation and construction has evolved considerably.

Quality assurance for the domestic data bases is more difficult. This begins with Center staff working closely with potential suppliers of national data and checking their submissions for basic accounting consistency. This process of iteration between the Center and individual contributors often takes several months and may involve a certain amount of informal training of the contributor in methods for sector disaggregation, adjustment and re-balancing. Once the data base satisfies the basic GTAP requirements, we move to the next stage of quality assurance. Here, we look for anomalies. The easiest way to identify these is to compare the new data base with the earlier version for that country and look for dramatic changes in basic economic relationships. This is quite effective, but it is not possible when a new region is being supplied. Furthermore, if the new I-O table is supplied by the same individual, this is just a check on consistency of procedures, not an

independent check on content. Therefore, we also compare the data base to a “representative” I-O table that represents a composite of those national data bases in which we have confidence. This often turns up errors in data processing and disaggregation. However, there is only so much quality assurance that can be provided at the Center. In the end, we rely heavily on national users to evaluate the quality of the data base. Indeed, before each release, the data base goes through a series of “pre-releases” in which consortium members and data base contributors evaluate the data prior to public release.

In summary, the Center for Global Trade Analysis places a high priority on quality assurance. However, ultimately, we rely on members of the network, with specific areas of expertise, to scrutinize the data base. In this sense, quality assurance in the global economic data base is an ongoing process.

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 multi-region, applied general equilibrium model are relatively standard. For a complete description of the GTAP modeling standard framework, 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.

The standard model is designed to be easy to modify and extend, and there are a variety of model extensions available on the GTAP web site, under the technical paper series. These extensions include features such as imperfect competition, technology spillovers, detailed treatment of energy demands and CO₂ emissions, agricultural commodity markets, 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 700 documented applications on the GTAP web site. This represents just a small fraction of the total number of applications undertaken. 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, among other topics.

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 GTAP 6 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, the earliest 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. Versions 4 and 5 replaced all the remaining original SALTER data bases and added many more regions, reaching a total of 66 in GTAP 5.

GTAP 6 has 87 regions — more than six times the number in the original, GTAP 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.

While it has become relatively routine to add new regions to the data base, the disaggregation of sectors remains a costly venture, as it involves revisiting all of the existing national and international data bases. Versions 4 and 5 of the data base added sectoral detail. In GTAP 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 supported improved analysis of

services 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 GTAP 5 and 6 data bases disaggregated international transport margins by mode.

The GTAP protection data base has evolved considerably since the project's inception and this continued to be a focal point in the GTAP 6 Data Base. Compared to version 1, the process of constructing a global protection data base has become infinitely more sophisticated. Most of the work for the first version of the 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 GTAP 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 GTAP 2 Data base also witnessed the introduction of a variety of non-tariff barriers (NTB), including anti-dumping duties, countervailing duties and price undertakings. Unfortunately these proved to be a one-time only contribution from the WTO and have since been dropped as they became severely outdated. (The one NTB continued through GTAP 6 refers to the export tax equivalent of textile and apparel quotas – abolished at the end of 2004 – but still relevant in 2001. See Chapter 16.F for detailed estimates.)

The GTAP 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 tariff data bases for GTAP 4 and 5 were sourced from the UNCTAD TRAINS data base, via the World Bank, courtesy of Will Martin, Jerzy Rozanski and Emiko Fukase. In version 4, this was supplemented on the agriculture side by estimates of market price support contributed by Marinos Tsigas, then at ERS/USDA. However, up to this point the treatment of specific tariffs remained a big problem. Since their aggregation requires conversion to ad valorem equivalent form, and since this conversion required additional price data, this was problematic. Fortunately, in GTAP 5 we were able to draw on the Agriculture Market Analysis Database (AMAD) for agricultural tariffs. And this data base was supplemented by estimates of the ad valorem equivalent of specific tariffs compiled by Paul Gibson and collaborators at ERS/USDA. This work revealed the great importance of specific tariffs to overall protection in agriculture in the OECD countries.

However, despite all of these improvements in the tariff data base for GTAP versions 2 - 5, we still lacked a proper treatment of preferences – particularly non-reciprocal preferences granted by industrial countries to developing country trade partners. This omission was particularly problematic in light of the prominent role played by the erosion of such preferences in the context of the Doha Development Agenda (DDA) negotiations at the WTO. Fortunately, this omission has been remedied in the GTAP 6 Data Base. Thanks to the outstanding work of one of the GTAP consortium members – CEPII (Paris) – along with the International Trade Center in Geneva, we have been able to build on the MACMap data base for tariffs and import protection. This includes a comprehensive treatment of trade preferences as well as the conversion of specific tariffs for both agriculture and non-agriculture commodities. Thus, for the first time, we have a comprehensive market access data base that treats agriculture and non-agriculture symmetrically. This is a great advance, and, coupled with detailed data on bound tariffs collected by CEPII, has permitted construction of a wide range of highly relevant policy scenarios for the Doha round of trade talks.

Another important topic of debate in the DDA negotiations has been the reform of domestic subsidies for agriculture. The fact that these had been included in WTO disciplines was a great advance under the Uruguay Round Agreement of the WTO. As increasing attention has been focused on the trade distortions caused by these subsidies, they have evolved from primarily output-based payments to increasingly decoupled payments based, for example, on historical land area. This required a more sophisticated treatment of the subsidies in the GTAP model. Therefore, in GTAP 5 we took 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), the Economic Research Service of USDA and most recently at the OECD. The allocation of these subsidies is not without controversy. In the end, a standard formula was agreed upon and the same one has been used for the GTAP 5 and 6 Data Bases. This work is documented in chapters 16.B and 16.C, authored by Hsin Huang and Hans Jensen, respectively.

Based on the July 2004 Framework Agreement of the DDA, it appears that export subsidies will be completely eliminated. Thus this policy measure is also very important. Chapter 16.E, authored by Aziz Elbehri reviews the 2001 estimates of export subsidies in food and agriculture, as notified to the WTO by member countries. These mainly affect the European Union and the United States.

One point that needs to be strongly emphasized for users of the GTAP data base for trade policy analysis 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 ALTERNAX 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). Users may also want to update the 2001 policy environment as reflected in the GTAP 6 Data Base to a more recent base year (e.g., 2005). This is most naturally done via a simulation in which the intervening policy reforms are implemented, using the model to predict what the resulting world economy would look like. This can be a useful starting point for forward-looking policy analysis (e.g., analysis of the DDA).

Another important area of recent development for the GTAP Data Base has been the incorporation of increasingly detailed data on the physical energy flows underlying the GTAP Data Base. These data have been obtained from the International Energy Agency (IEA). The energy volume data were first introduced in an interim version following release of GTAP 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 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 the Center invested considerable effort in the energy data base and associated programs used in GTAP 5. This work, involving major contributions by Jean-Marc Burniaux, Huey-Lin Lee, and Robert McDougall, 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.

1.3 Future Development of the Data Base

Judging from past experience, future development of the data base will be heavily influenced by the GTAP Advisory Board. This group is made up of representatives from each of the agencies in the GTAP Consortium. Their continued guidance ensures that the data base will evolve to meet the changing needs of policy makers concerned about global economic issues. As noted above, quality assurance is a top priority. Nonetheless, there are a few areas where we anticipate significant breakthroughs in the not-too-distant future.

The first area of anticipated improvement involves an update of the energy price series used in the GTAP Data Base. Energy prices have historically been much more problematic than volumes. Presently the energy price data are extrapolations of historical price data based on world price indexes. Incorporation of 2001 price estimates to go with the new energy volume data is highly desirable and a top priority.

As the trade, protection, and input-output data have improved in quality, attention has come to be focused more on the primary factor intensities by sector. For example, when trade reform favors the light manufacturing sector in a developing country, can this be expected to raise unskilled wages in the economy? This depends on the relative factor intensity of this sector. If it is relatively intensive in unskilled labor, then this will indeed be the case. Unfortunately, the current characterization of

primary factor intensities in non-agriculture sectors in GTAP is rather crude. It is based on the assumption that the portion of value-added not explicitly reported as wages and salaries in the input-output table represents payments to capital. This is problematic in sectors where there is a great deal of informal activity and where self-employed labor represents a large share of the activity. Since the I-O tables often do not impute a wage to self-employed labor, such sectors often show an erroneously high capital share in value-added, which in turn can lead to misleading conclusions about the likely wage impacts of trade reform. Efforts are currently underway to improve this aspect of the data base, drawing on household survey data and other sources of information on primary factor payments.

There is considerable scope for using GTAP to explore fiscal issues, but in order to do so, a more complete representation of taxes will be required. Robert McDougall and Jan Hagemeyer document a first step in this direction in Chapter 18.E in which they discuss how direct taxes have been incorporated into GTAP 6. The next step will be to reconcile indirect tax receipts reported in the data base and those reported by the International Monetary Fund. Once this improved treatment of taxes is in place, there will be a great many new issues that can be addressed with GTAP-based models.

One of the most exciting developments in the GTAP Data Base involves the incorporation of additional data on physical and environmental variables in support of global environmental analysis. With the support of the US-Environmental Protection Agency, Huey-Lin Lee has already constructed a supplementary data base on greenhouse gas emissions that can be used in conjunction with a modified GTAP model to examine the impact of alternative policies on the emission of carbon dioxide (CO₂) and non-CO₂ gases from agriculture and industry (http://www.gtap.agecon.purdue.edu/resources/res_display.asp?RecordID=1143). She is presently finalizing a data base on land-use and associated emissions, by agro-ecological zone that will permit authors to extend their analysis of climate change policy to include the impacts of land use change on global emissions.

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 three parts. Part I, comprising chapters 1 – 3, provides an overview of the GTAP 6 Data Base and highlights new features. Part II, chapters 4 - 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 4), disposition of sales and costs by commodity and by region (chapter 5), protection

data (chapter 6), and I-O multipliers (chapter 7). 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. Chapters 8-10 describe the GTAP Data Base and the software used in aggregations and in viewing data.

Part III begins with a detailed documentation of the new input-output tables for the GTAP 6 Data Base. 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 tariff data, as well as protection data for agriculture in particular (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 GTAP 6 Data Base. While only volume data are supplied with GTAP 6, price data were also required in order to reconcile these volume data with GTAP’s value-based energy expenditures. Chapter 18 describes the macroeconomic and capital stock 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 agriculture and for the natural resource-based industries. Chapter 19 summarizes the procedure for updating the regional input-output tables (FIT) to the 2001 reference year. Chapter 20 describes how the GTAP behavioral parameters were obtained. Finally, chapter 21 describes the process of assembling the full GTAP 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. <http://www.gtap.org>.
- 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.