

## *Chapter 5*

# *GTAPAgg2 Data Aggregation Program*

*Mark Horridge*

The new GTAPAgg is a Windows program used to prepare data bases for the GTAP economic model. Like its predecessor, it helps aggregate the GTAP Data Base, which version 10 covers 65 commodities and 141 regions. Nearly all applications of the GTAP Model use an aggregation of this data base. GTAPAgg2 helps you prepare an aggregation scheme, then uses your scheme to prepare an aggregated data base for the GTAP Model. What makes GTAPAgg2 different is that we can install one package for all years or one for each of the years 2004, 2007, 2011 and/or 2014. Another distinction of the new GTAPAgg is that it is also available for the GTAP satellite data sets. For each major GTAP release, say GTAP 9 or 10, there are several data extensions, for the multiple reference years, conveniently managed by the new GTAPAgg. This new version of GTAPAgg is similar to its previous versions in that it is very simple to use and is accompanied by an online help file.

Subscribers of the GTAP Data Base will receive a personalized licence file, called GTAPAGG.LIC, like they did before which must be placed in the same folder as the file GTAPAgg2.exe. Subscribers will also receive a link to the GTAP website from which they will download a data package that needs to be installed in GTAPAgg2. As in previous version, the main GTAPAgg window shows a column of eight buttons. Normally, you would use the upper buttons first, then work downwards. The buttons are:

- **Instructions and Help:** for online help.
- **Choose alternate source data folder:** By default, GTAPAgg uses the standard supplied GTAP Data Base. You would use this button to change to another version of the GTAP Data Base or a satellite account. You can also select the reference year that you are interested.
- **Read aggregation scheme from file:** Use this button to start from an aggregation scheme that you prepared previously. Otherwise, GTAPAgg starts up with a default 10 x 10 aggregation. By right-clicking, you can choose to load only the sector aggregation (or only the region aggregation) from an existing aggregation scheme.
- **View/change regional aggregation:** to specify how the original regions are mapped into aggregated regions.
- **View/change sectoral aggregation:** to specify how the original sectors (tradeable commodities) are mapped into aggregated sectors.
- **View/change factor aggregation:** to specify how the original primary factors (endowment commodities) are mapped into aggregated factors.
- **Save aggregation scheme to file:** When you are satisfied with your regional and sectoral mappings, you must save them in a text file, before proceeding to the next step.

- **Create aggregated database:** uses your aggregation scheme to prepare corresponding GTAP data files, suitable for use by the GTAP Model.

## 5.1 *Creating an Aggregation Scheme*

An aggregation scheme is a system of mappings between the many sectors and regions of the original GTAP Data Base and smaller numbers of sectors and regions in the aggregated data base that you want to create. GTAPAgg stores aggregation schemes in a text file with the extension or suffix AGG. An AGG file contains three mappings — for regions, sectors and primary factors. For regions, you must fill in the names of your new (aggregated) regions, and must specify to which new region each of the original GTAP regions corresponds. The same interface is used to choose mappings for sectors and factors.

Similar screens are used to view and edit both regional and sectoral (tradeable commodity) aggregations. For regions, you are able to: (1) use the bottom panel to change the number of new, aggregated, regions, their names and descriptions (try right clicking), and (2) use the top-right panel to show to which new regions each original region is mapped. Each new region must have a unique code (12 letters maximum); the longer "new region descriptions" are optional. Use the same techniques to edit the sectoral (tradeable commodity) aggregation.

For factors (endowment commodities), you must in addition specify the SLUG/ETRAE settings. Each factor is either "sluggish" or "mobile". Each sluggish factor has an ETRAE number less than or equal to zero. Convey your intentions by entering a negative ETRAE number (which implies sluggish) or "mobile" in the final column of the bottom panel. The GTAP Model insists that at least one of the aggregated factors is called "capital".

You can easily make a table documenting your aggregation. If you are editing either the regional or sectoral aggregations, use the *Copy* button to send a report of the aggregation to the clipboard. Then open your spreadsheet program and paste the report into a blank worksheet. Delete columns or rows you do not need, then paste the table into your word processing program for final formatting.

## 5.2 *Creating the Aggregation*

When your aggregation scheme is ready, click the **Save Aggregation Scheme** and then the **Create Aggregated Database** buttons. Wait a few seconds while the 5 million numbers of the GTAP Data Base are grouped and summed. A ninth button will appear, allowing you to view the aggregated data.

## 5.3 *Data Files*

A global GTAP data file, *basedata.har*, accompanies the GTAPAgg program. The program also includes the global GTAP sets file (*sets.har*), global parameters file (*default.prm*), energy volume data and CO<sub>2</sub> emissions and perhaps a time-series trade data files (while interesting, the latter 3 are not required by the standard GTAP Model). The metadata file that has information on IO table contributions, namely, *metadata.har*, is newly included in the package.

When GTAPAgg produces a new aggregated data base, five files are produced:

- a flows data file, *basedata.har*, in HAR format,
- a parameters file, *default.prm*, in HAR format,
- a sets file, *sets.har*, in HAR format,
- a data summary file, *baseview.har*, derived from *basedata.har*,

- a file of tax rates, *baserate.har*, derived from *basedata.har*,
- if available, files of aggregated time-series data, *tstrade.har*, CO<sub>2</sub> emissions data, *co2.har*, and energy volume data, *gsdvole.har*.

It is vital that these files are used together as a group. Therefore, GTAPAgg bundles all the above files together into a *ZIP* archive. The idea is to prevent files from different aggregations from getting mixed up. The aggregation scheme file is also included in the *ZIP* archive. Use a program such as *Unzip* (or *WinZip* or *PKUnzip*) to unpack the files into the subdirectory where they will be needed. Alternatively, if you have the *RunGTAP* program for running the GTAP Model, you can use its *New Version Wizard* command to load the *ZIP* archive directly.

Most of the files produced by GTAPAgg are stored in a special type of file, called *Header Array* or *HAR* file. A free utility program, *ViewHAR*, is supplied to view, edit, or export data from *HAR* files.

## 5.4 *Aggregate Before or After Model Simulations or Data Modification*

The normal sequence of operations, shown in the left column of Figure 5.1, is to first use GTAPAgg to make a data base, which is then used by the GTAP Model. Each model simulation produces an "update" file in the same *BASEDATA* format, showing post-simulation flow values. In fact, both the GTAP Model and GTAPAgg share the same format for both input and output files. This allows us to combine GTAPAgg and the GTAP Model in various ways.

For example, the second column starts by aggregating to 50 regions. The result is used by the GTAP Model in "AlterTax" mode -- a simulation designed to adjust tax rates implied by the flows data. The AlterTax simulation produces a modified data base which is again aggregated by GTAPAgg, to make data for a conventional policy simulation. A similar, more ambitious project would be for the first GTAP Model run to simulate all economic changes that have occurred since 2014 or any of the other reference years, to produce a data base representing current conditions.

The third column starts with a one-to-one GTAPAgg run which simply decrypts the supplied data. A user-written data program is then used to modify the full-size original data. The result is aggregated by GTAPAgg, prior to a conventional policy simulation. If the user-written program were coded in *GEMPACK* it could directly read and produce files in the *BASEDATA.HAR* format. A *GAMS* program could be used instead, in conjunction with the format-translation tools mentioned below.

## 5.5 *For GAMS Users*

GTAPAgg warns you if you choose a region or sector name that is a *GAMS* reserved word. Free command-line utilities, *Har2Gdx* and *Gdx2Har*, can be downloaded to rapidly convert the output *HAR* files to and from the *GAMS* *GDX* format.

## 5.6 *GTAPAgg and FlexAgg*

*FlexAgg*, documented in chapter 4, is a command-line procedure which also aggregates the GTAP Data Base. *FlexAgg* uses a special format of input file to store flows, so you have to aggregate first, then simulate.

Figure 5.1 Aggregation at Different Stages

