Chapter 5

GTAPAgg: Data Aggregation Program

Mark Horridge

GTAPAgg is a Windows program used to prepare data bases for the GTAP economic model. The full GTAP 7 Data Base covers 57 commodities and 113 regions. Nearly all applications of the GTAP Model use an aggregation of this data base. GTAPAgg helps you prepare an aggregation scheme and then uses this scheme to prepare an aggregated data base for the GTAP Model. The program is very simple to use and is accompanied by an online help file.

The main GTAPAgg window shows a column of eight buttons. Normally, you would use the upper buttons first and then work downwards. The buttons are:

— **Instructions and Help:** Use this button for online help.

— **Choose alternate source data folder:** By default, GTAPAgg uses the standard supplied GTAP Data Base; however, you occasionally might use this button if you have another GTAP Data Base that you want to aggregate.

— **Read aggregation scheme from file:** Use this button to start from an aggregation scheme that you previously prepared; 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:** Use this button to specify how the original regions are mapped into aggregated regions.

— **View/change sectoral aggregation:** Use this button to specify how the original sectors (tradeable commodities) are mapped into aggregated sectors.

— **View/change factor aggregation:** Use this button 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:** This button 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 and 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 10 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.hrx*, accompanies the GTAPAgg program. It is encrypted so that only GTAPAgg can extract data from it. The program also includes the global GTAP sets file (*sets.har*), global parameters file (*default.prm*), and perhaps energy use and time-series trade data files (interesting but not required by the standard GTAP Model).

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,

— sometimes, files of aggregated time-series data, *tstrade.har*, and energy 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 a Header Array or HAR file. A utility program, ViewHAR, is supplied to view, edit, or export data from HAR files.

5.4   GTAPAgg Versions

The full version, which lets you create a data base of any size, is available to those who purchase the full GTAP Data Base. They will receive a personalized license file, called GTAPAGG.LIC, which must be placed in the same folder as the file GTAPAgg.exe.

The constrained version, which is the default version available in the GTAPAgg package, allows no more than 10 aggregated regions or sectors. The free or demo version allows no more than 3 aggregated regions or sectors.

5.5   If You Want the Original Un-encrypted GTAP Data

The main input data file, *basedata.hrx*, containing flow values, is encrypted so that only GTAPAgg can extract data from it. However, you can examine its contents with the supplied GView program.

If you have the full version of GTAPAgg you can easily produce a full-sized, not encrypted, flows data file. Simply select the "one-to-one" option for the region and sector

mappings to make a decrypted flows data file called *basedata.har*. GTAPAgg can also aggregate this file: use the *Choose alternate source data folder* button.

5.6   *Aggregate Before or After Model Simulations or Data Modification*

The normal sequence of operations, shown in the left column of Figure 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 2001, 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.

5.7   *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.

5.8   *GTAPAgg and FlexAgg*

FlexAgg, documented in Chapter 4, is a command-line procedure which also aggregates the GTAP Data Base. Aggregation schemes are stored in files which users create with a text editor. These files are similar or identical to those used by GTAPAgg. Indeed, you could use GTAPAgg to prepare Agg files for use with FlexAgg, or you can prepare AGG files in a text editor for use with GTAPAgg. FlexAgg requires access to the full-size un-encrypted GTAP data, and so is only available to purchasers of the full GTAP Data Base. It uses a special format of input file, so you have to aggregate first, and then load into RunGTAP.

Figure 5.1   Aggregation at Different Stages