

Chapter 4

FlexAgg: Data Aggregation Program

Robert A. McDougall

This chapter describes the command line data aggregation program, also known as FlexAgg, which is included in the GTAP 7 Data Package. It is based on user documentation shipped with the package.

4.1 System Requirements

The package has been tested on Windows XP. It will likely work with other WIN32 operating systems.

4.2 Preparing Aggregations

To prepare an aggregation, you need to provide the data base to be aggregated, prepare an aggregation mapping file and run the aggregation program. Note that the basic aggregation program does not aggregate the time-series trade data; this is so that it will work with GTAP Data Bases not accompanied by such data.

Providing the Data Base to be Aggregated

The program aggregates a data base consisting of four files:

- GSDSET.HAR : sets
- GSDPAR.HAR : parameters
- GSDDAT.HAR : data
- GSDVOLE.HAR : energy volumes

These files should be present in the directory where you install this package. Upon installation, the files comprise the standard GTAP 7 Data Base; you can however substitute any data base with the GTAP 7 structure (so, for example, you may use the aggregation program to aggregate an already aggregated data base).

Preparing an Aggregation Mapping File

To prepare an aggregation mapping file, make a copy of the file INPUT.TEM, and edit it to suit your requirements. For details on layout, see the instructions included in INPUT.TEM itself. Save the edited file with a name ending in the extension .TXT, for example, MY_AGG.TXT.

Running the Aggregation Program

The aggregation program is DATA-AGG.BAT. To run it:

Make the directory with the GTAP Data Package your current directory. For example, if you have installed the GTAP Data Package in the directory C:\GTAP, you might enter the command

```
cd C:\GTAP
```

Invoke the program, using either:

syntax 1:

```
DATA-AGG map-file
```

Or

syntax 2:

```
DATA-AGG [map-dir\]map-file [output-dir] [work-dir]
```

Note:

- For *map-file* substitute the name of your aggregation mapping file, omitting the extension .TXT. For example, if your mapping file is MY_AGG.TXT, substitute MY_AGG for *map-file*.

To use syntax 1, your mapping file must be in the same directory as the GTAP Data Package. With syntax 2 you can specify the location of your mapping file, the aggregated data files and intermediate files. You exercise this control through the optional arguments shown in square brackets. (Don't type the brackets in your command line; they're there only to indicate that the arguments are optional.) Substitute as follows:

map-dir	the directory containing your mapping file
output-dir	the directory to contain the aggregated data files
work-dir	the directory to contain all intermediate working files.

For reasons relating to (lack of) program maintenance, it is safest to restrict the directory and files names to the old DOS 8.3 style.

For example, to use a mapping file *test335.txt* in the local subdirectory *tstin*, writing the aggregated data files to a local subdirectory *tstout*, and the working files to *tstwrk*, enter

```
data-agg tstin\test335 tstout tstwrk
```

Besides the aggregated data files, the program will create report files *gview.har* and *gtax.har* containing summary information on the data base. To inspect these files we recommend the GEMPACK utility ViewHAR.

Further details:

With syntax 1, the program behaves as follows.

- The program will look in the GTAP Data Package directory for a mapping file with the extension .TXT.
- It will create a subdirectory with the same base name as your mapping file. For example, if your mapping file is *my_agg.txt*, it will create a subdirectory *my_agg*.
- It will write the aggregated data files *gset.har*, *gpar.dat*, *gdat.har*, *gvole.har* to the new subdirectory. It will also write there the derived informational files *gview.har* and *gtax.har*.
- It will leave intermediate working files in the current directory.
- After running an aggregation with syntax 1, you can remove intermediate files from your GTAP directory with the command:

```
aggclean
```

With syntax 2, the program behaves as follows:

- It looks for the mapping file in the directory indicated by the argument *map-dir*. If you do not supply a value for *map-dir*, it looks in the GTAP Data Package directory. If you supply a file name but the file does not exist, it stops with an error.
- It writes the final output files *gset.har*, etc. to the directory indicated by *output-dir*. If you do not supply a directory for *output-dir*, it writes the files to the GTAP Data Package directory. If you supply a directory name but the directory does not exist, it stops with an error.
- It writes intermediate working files to the directory indicated by *work-dir*. If you do not supply a value for *work-dir*, it writes the working files to the same directory as the aggregated data files. If you supply a directory name but the directory does not exist, it stops with an error.

Trouble Shooting*error message "error in map file"*

This message can be generated by an error in your mapping file. An error describing the error in the mapping file may be found in the intermediate working file gmap.txt. This is in the main GTAP directory, if you used syntax 1, or in your selected work directory, if you used syntax 2. See also below under the heading "Mapping file errors".

error message "user-specified directory or file does not exist"

This message is written if the program cannot find the mapping file (syntax 1 and 2), or if it cannot find the output or work directories (syntax 2 only).

One way to generate this error is to include the extension .TXT in the *map-file* parameter. For example, the command

data-agg MY_AGG.TXT WRONG!

would generate this error.

error message "can't extract disaggregate margin header information"

This message may be written if you supply, for the working files directory, a name that does not follow the DOS 8.3 style, for example, "test.work".

Mapping File Errors

(This is taken from earlier documentation by Randy Wigle.)

The following error checks are made on the input map (the .TXT file.)

- If an aggregated commodity is referred to in section 2 of the mapping, but it has not been defined in section 1, an error is noted on the first occurrence and the program exits
- If an aggregated region is referred to in section 4 of the mapping, but it has not been defined in section 3, an error is noted on the first occurrence and the program exits.
- If a line or section of the file is not set up properly, an error is noted and the current line is printed. NOTE that this will not always be the incorrect line! (In particular, if you leave out one of the "= = =" lines, the line where the error prints will usually be later in the file.
- If an aggregated commodity is duplicated in section 1, an error will be generated, but the program usually chokes in section 3 (it is finding two matching aggregated commodities for one disaggregated commodity.)
- If an aggregated commodity or region name violates GEMPACK rules (e.g. if it is longer than 12 characters, or if it includes characters other than letters, digits and underscores, or if

it doesn't start with a letter), then an error will be generated and the program terminates. See the discussion in the file INPUT.TEM.

4.3 *Aggregating Time-Series Trade Data*

To aggregate the time-series trade data, use the program TST-AGG.BAT. To use the program, you must have the disaggregate time-series data files *tstrade0.har* and *tstrade1.har* in your working directory, and (as with the main aggregation program) you must have a mapping file with filename extension txt. The syntax is the same as for the main aggregation program:

TST-AGG map-file

or

TST-AGG [map-dir\]map-file [output-dir [work-dir]]

The output file is *tstrade.har*, in a directory named after the mapping file (with the first syntax) or in the <output-dir> directory (with the second syntax).

4.4 *Legalese*

As always, to the extent permitted by applicable law, THE GTAP DATA PACKAGE COMES WITH ABSOLUTELY NO WARRANTY.

4.5 *Contents*

Unpacked, the distribution contains the following files:

User Information

README this file

GTAP Standard Data Base

- gsddat.har
- gsdpar.dat
- gsdset.har
- gsdvole.har
- tstrade0.har
- tstrade1.har

The files *tstrade0.har* and *tstrade1.har* contain the time-series trade data.

To limit the size of the distribution, we have not included the corresponding GTAPView files.

Bundled Freeware

A number of programs from outside sources are needed for unpacking and aggregating the data base:

awk.exe	from DJGPP, stub for gawk
gawk.exe	from DJGPP, text processing language
m4.exe	from DJGPP, macro processor
modhar.exe	from COPS/IMPACT, modifies HA files
rd1hmd.exe	from COPS/IMPACT, reads header from GEMPACK-formatted text file
seehar.exe	from COPS/IMPACT, converts HA files to text

In accordance with the license for awk, gawk, and m4, users seeking documentation and source code can obtain it through the DJGPP website: <http://www.delorie.com/djgpp/>.

These programs should run with all reasonably current versions of MS-DOS and MS-Windows (and some less-than-reasonably current); if nevertheless you do encounter problems, please advise the Center for Global Trade Analysis.

Aggregation Mapping Files

Some files are included to assist you in specifying mappings for data base aggregations:

.\INPUT.TEM	template for aggregation mapping files
-------------	--

Aggregation Program Files

The aggregation program lets you reduce the size of the data base by aggregating across sectors and regions. For user instructions, see section 4.2.

aggdat. (axs, axt, exe)
 AGGMAP.AWK
 aggpar. (axs, axt, exe)
 aggvole. (axs, axt, exe)
 DATA-AGG.BAT
 ENVSPCER.MSG
 G_USAGE.MSG
 GBARUN.BAT
 GTARUN.bat
 GSET.PRE
 gtapview. (axs, axt, exe)
 MAPERROR.MSG
 NODIR.MSG
 NOTBASE.AWK
 PREDIM.AWK
 PREEND.AWK
 PREETR.AWK
 PREMARG.AWK

PREREG.AWK
PRESEC.AWK
PRESLG.AWK
tst-agg.bat
tst-agg. (axs, axt, exe)

Miscellaneous Files

.\AGGCLEAN.BAT cleans up after an aggregation run

Additional Documentation

COPYING copyright notice for awk

Source Files for Aggregation Program

src\aggdat.tab
src\aggpar.tab
src\aggvole.tab
src\gtapview.tab