

Chapter 9

Ancillary software: aggregation programs

Robert McDougall, Betina Dimaranan

This chapter describes GTAP data package, and more especially the aggregation program included in the GTAP data package. It is based on user documentation shipped with the package.

9.1 System requirements

The package is designed to run on MS-DOS 5.0 or higher. At least 15 Megabytes of RAM must be available (more is recommended). You need 22 megabytes of disk space to install the package, and more to prepare aggregations. A memory manager such as EMM386 or QEMM386 is necessary.

One item in the package, the *VIEWHar* header array file viewer, runs on MS-Windows 3.1 or higher, and MS-Windows 95.

9.2 *Annotated packing list*

The package contains 6 3.5" DOS-formatted high-density floppy disks containing the following files:

DISK1

- COPYING.AWK licence conditions for awk program
- COPYING.UZP licence conditions for unzip program
- DATA.ZIP GTAP 3 data base
- README this document (text)
- UNZIP.EXE PK-compatible unarchiving program

DISK2

- AGG1.ZIP GTAP 3 aggregation program, part 1 of 4

DISK3

- AGG2.ZIP GTAP 3 aggregation program, part 2 of 4

DISK4

- AGG3.ZIP GTAP 3 aggregation program, part 3 of 4

DISK5

- AGG4.ZIP GTAP 3 aggregation program, part 4 of 4

DISK6

- TS.ZIP Trade data, with *TSView* time series viewer

DISK7

- VIEWHAR.ZIP ViewHAR header array file viewer
- SOURCES.ZIP selected GTAP programs

ViewHAR is from Mark Horridge, Center of Policy Studies, Monash University. TSView is from the Economic Research Service of the United States Department of Agriculture: the data files are by Mark Gehlhar, the program by Karl Gudmunds. AWK.EXE (wrapped with the aggregation program) is a DOS port of the Free Software Foundation's Gawk (Gnu awk). UNZIP.EXE is from Info-ZIP.

9.3 *Installing the package*

To install, copy all files from diskettes 1 to 7, then unzip all files with extension ZIP.

An unzip utility UNZIP.EXE is included in the UTIL directory in diskette 1. To unzip a file BAR.ZIP in directory F00, type

```
unzip F00\BAR
```

from some directory containing the UNZIP.EXE file. You may wish to install the TSView trade viewer in a separate directory. You may wish to install the ViewHAR header array file viewer in your GEMPACK directory.

To prepare to use ViewHAR under MS-Windows or MS-Windows 95, go into Program Manager, select "New" from the "File" menu, select "Program Item" from the "New Program Object" dialog box, and fill out the "Program Item Properties" box as indicated in the "Help".

Installing ViewHAR is unnecessary if you have already installed the GEMPACK graphical user interface WinGEM, since ViewHAR is bundled with it.

9.4 *Preparing aggregations*

To prepare an aggregation, you need to prepare an aggregation mapping file, and then run one of the aggregation programs.

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. For an example of a completed mapping file, see the file TEST3X3.TXT. Save the edited file with a name ending in the extension .TXT, for example, MY_AGG.TXT.

Running the aggregation program

There are two aggregation programs, `DATA-AGG.BAT` and `GDAT-AGG.BAT`. The first program, `DATA-AGG`, produces a complete aggregated data base. The second program, `GDAT-AGG`, produces an aggregated sets file and data file, but no parameters file.

If you want to conduct GTAP simulations with the aggregated data base, you must use `DATA-AGG`. If you do not want to run GTAP simulations, but want to use the aggregated data base only as a data source, you can use the faster-running `GDAT-AGG`.

The instructions for using the two programs are similar. We give here instructions for `DATA-AGG`. To use `GDAT-AGG`, simply substitute the program name `GDAT-AGG` for `DATA-AGG`. Make the directory with the GTAP data package your current directory. For example, if you have installed the GTAP data package in the the directory `C:\GTAP`, you might enter the command `cd C:\GTAP`.

To run the aggregation program, you can use one of two syntaxes. The syntaxes differ in functionality: with syntax 1 you do not have control over the location of your mapping file, intermediate files, or output files. With syntax 2 you do.

Syntax 1 is:

```
DATA-AGG mapfile
```

To use syntax 1, your mapping file must be in the same directory as the GTAP data package. For `mapfile` 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 `mapfile`. The program will create a sub-directory named after the mapping file, and put the output files there. For example, if your GTAP data directory is `C:\GTAP`, and your mapping file is `MY_AGG.TXT`, it will put the output files in a directory `C:\GTAP\MY_AGG`.

Syntax 2 is

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

With this syntax you can specify the location of your mapping file, the aggregated data files and intermediate files. You do this through the optional arguments shown in square brackets (do not type the brackets in your command line; they are merely 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 example, to use a mapping file `MY_AGG.TXT` in the directory `C:\MY_APP`, writing the aggregated data files to `C:\MY_APP\DAT`, and the working files to `C:\MY_APP\TEMP`, enter

```
data-agg C:\MY_APP\MY_AGG C:\MY_APP\DAT C:\MY_APP\TEMP
```

The aggregated data files are `GSET.HAR`, `GPAR.DAT` and `GDAT.HAR`. You may also find a file `GDAT51.HAR`. `GDAT.HAR` uses the new GEMPACK 5.2 format for real number arrays, while `GDAT51.HAR` uses the old format. If you use `GDAT-AGG` instead of `DATA-AGG`, the output files are `GSET.HAR`, `GDAT.HAR`, and (possibly) `GDAT51.HAR`.

The program will not make `G DAT51 . HAR` if GEMPACK 5.2 is installed.

Further details: syntax 1

With syntax 1, the program behaves as in GTAP release 2.

- 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` and `G DAT . HAR` to the new subdirectory.
- It will leave intermediate working files in the current directory. You can remove these if you wish using the command
`cleanup`

Further details: syntax 2

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.
- It writes the aggregated data files `GSET . HAR` 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.
- 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.

Header array format for the output data

`G DAT . HAR` uses the new GEMPACK 5.2 format for real number arrays, while `G DAT51 . HAR` uses the old real number format. The new format has the advantage of providing row and column labels. Unfortunately, programs from GEMPACK 5.1 or earlier cannot read it. We recommend using `G DAT . HAR` with GEMPACK 5.2 programs, and `G DAT51 . HAR` with GEMPACK 5.1 or earlier. The program will create `G DAT51 . HAR` automatically, unless it perceives an indication that you have installed GEMPACK 5.2: that you have a GEMPACK licence file in a directory `C:\GP`, or have set the environment variable `GPDIR`.

trouble shooting*error message “environment space insufficient”*

You may need to edit the batch file (DATA-AGG.BAT or GDAT-AGG.BAT). The last line of the file is a command with an option /e:1024. The command creates a temporary environment, and the options sets its size at 1024 bytes. If this is insufficient, specify a larger size, e.g. by substituting the option

/e:1536.

error message “error running a TABLO generated program”

In some cases, you may need only to rerun the job. The calibration programs ECAL and ALPHACAL sometimes fail on the first run but succeed on the second. For an indication of the source of the problem, view the latest .LOG file created by the program.

error message “error in MAP file”

This message can be generated by an error in your mapping file.

A message describing the error in the mapping file may be found at the bottom of the intermediate working file <mapfile>.MAP, i.e. the file with the same name as your mapping file but extension .MAP. See below under the heading “Mapping file errors”.

error message “user-specified directory or file does not exist”

This message may be generated 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 mapfile parameter. For example, the command

data-agg MY_AGG.TXT

WRONG!

would generate this error.

Mapping file errors

Here we describe the error checks the program makes on the input mapping file. The description is are taken from program documentation written by Randy Wigle.

- 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 b) 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 is not always the incorrect line. (In particular, if you leave out one of the lines beginning “= = =”, the line printed with the error message is usually later in the file.
- If an aggregated commodity is duplicated in section 1, an error is generated, usually in section 1 but in section 3 (when the program finds two matching aggregated commodities for one disaggregated commodity).
- If an aggregated commodity or region name violates GEMPACK rules (if it is longer than 12 characters, or includes characters other than letters, digits and underscores, or does not begin

with a letter), an error is generated and the program terminates. See the discussion in the file `INPUT.TEM`.

GTAPCHK info

One of the last things that the aggregation program does is run `GTAPCHK.TAB`. This is a GEMPACK program designed to report on the data base. It computes a host of summary measures including regional income, gross tax receipts, taxes net of subsidies, budget shares, and so on. These are displayed in the file `GTAPCHK.DIS`. We encourage you to examine this file before proceeding to work with the data.

9.5 *GEMPACK version issues*

The GEMPACK programs included in the package are prerelease versions from GEMPACK 5.2. We understand that the Impact Project will keep versions of these particular programs (`SEEHAR`, `MODHAR` and `RWHAR`) publicly available. Thus final release versions should become available shortly from the Impact Project (<http://www.monash.edu.au/policy/gempack.htm>).

If you need the `GDAT` file in the old GEMPACK 5.1 real array format, but for some reason the program does not make `GDAT51.HAR`, you can convert `GDAT.HAR` to the old format by using the included program `RWHAR`:

- make the GTAP data package directory the current directory,
- enter the command:

`rwhar`
- at the first menu, select `C51`, and
- follow the prompts to make an old-format file.

6. *Contact*

If you experience problems with this package that you cannot resolve by reference to the documentation, please contact us by email at gtap@ftp.purdue.edu.