GTAPAdjust

A program to balance or adjust a GTAP database

Mark Horridge

Centre of Policy Studies,
Monash University, Melbourne, Australia

March 2011

Contents

1 INTRODUCTION 3
1.1 Prerequisites 3
2 MOTIVATION 3
3 INSTALLING GTAPADJUST 3
3.1 Examining the results of the supplied example 4
4 OVERVIEW 4
5 STARTING A NEW PROJECT 5
6 THE "NORMALIZED" DATA FORMAT 6
6.1 Contents of normalized data file 7
6.2 The balance conditions 8
7 INSTRUMENTS AND TARGETS 9
8 PROGRAMS AND FILES OF ADJUST.BAT 9
9 GTAPADJUST.TAB 10
10 USING THE DIAGNOSTIC FILES 11
11 YOUR SHOCKS 12
12 TROUBLESHOOTING 12
13 AFTER GTAPADJUST IS FINISHED 12
14 IMPLEMENTATION NOTES 12
14.1 GEMSIM or Fortran? 12
14.2 Limits 13
14.3 Solution method 13
14.4 Group Policy 13

Google: *GTAPAdjust GTAP* to find the latest version.

1 Introduction

Sometimes the need arises to alter a GTAP database, either because some more recent data is available, or because you have a different view about some magnitudes. But, as explained in Section 2, merely editing the standard GTAP data files is not straightforward, and may lead to errors.

GTAPAdjust is a TABLO program which scales GTAP data, either to

- eliminate existing imbalances in the data; or
- fit the data to targets such as output, exports, or imports by sector, or national GDP.

You might use the supplied version or GTAPAdjust, or adapt/extend it.

1.1 Prerequisites

- You need a GEMPACK licence, either source-code or executable image.
- You should already be familiar with the GTAP model and its data structures, and the use of ViewHAR to examine data.
- You'll need intermediate GEMPACK skills, and a little knowledge of how BAT script files work[1].

2 Motivation

There are a number of reasons why it may be hard to achieve satisfactory results by directly editing the main GTAP flows data file (often called basedata.har):

- If you wanted, say, to increase the size of the French food sector, you would need to increase all of the costs incurred by that sector, and all of its sales. You would need to alter many entries in many arrays.
- Taxes on inputs do not appear directly, but must be inferred as the difference between two arrays.
- It is necessary to ensure that certain accounting identities remain satisfied. For example, the sum of all imports of food used in France must equal the sum, from all sources, of food exports to France (if you value all these flows in the right way).
- Production taxes do not appear in the data; they are treated as a residual, determined by the identity that each sector's cost must equal its sales. This mean that if you increased wages in the French food sector by 1 million dollars (and did nothing else), production tax would (implicitly) fall by 1 million dollars and industry costs and sales would remain unchanged.

These features complicate the task of manual editing, and also make it harder to write programs to alter the database. The approach offered here is to first convert the GTAP data to a more tractable format; then use the supplied GtapAdjust program to alter the flows; then convert the result back to the standard GTAP format.

The supplied GtapAdjust program is capable of handling many tasks. However, you could easily edit GtapAdjust.TAB to add more features.

3 Installing GTAPAdjust

The GTAPAdjust package is supplied as a ZIP archive—unzip this into an empty new folder with a short and simple name (such as C:\GTAPAdjust). Then open a command prompt (DOS box) in that folder and type:

---

[1] For example, see http://www.monash.edu.au/policy/gp-bat.htm

```
adjust.bat
```

To run the supplied example.job. You should see a series of messages like:

```
running Gtp2Norm with Gtp2Norm.cmf
running normchek with normchk0.cmf
running GtapAdjust with input\GtapAdjust.cmf
running DiffHar to compare adjusted with unadjusted data
running NormChek with NormChk1.cmf
running Norm2Gtp with Norm2Gtp.cmf
running GtpVew with GtpVew.cmf

BATCH JOB SUCCESSFUL -- see following files in output folder:
basedata.har
default.prm
gtapview.har
sets.har
taxrates.har
```

If instead you see error messages, it probably means that you do not have GEMPACK installed properly on your system (it needs to be on the system path), or that your GEMPACK is too old (you need Release 10.002 or later[2]).

3.1 Examining the results of the supplied example

(a) In the Work folder (probably C:\GtapAdjust\work) use ViewHAR to open the file InitialRpt.har (summarizing original data) and examine the header CST0, showing original industry outputs. Australian output of omn is 39151. Header ACK0 says that the sum of all errors in acounting identities is around 10 (which is very small compared to world GDP of 56 million).

(b) In the Input folder (probably C:\GtapAdjust\input) use TABmate to open the file GtapAdjust.cmf and find near line 58 the statements:

```
!    old exog                new exog   !
swap qdem("omn","Australia")=vCOSTS("omn","Australia");
final_level  vCOSTS("omn","Australia")= 50000;
```

These lines say that output is to be increased to 50000, by endogenizing the sales scaling variable qdem.

(c) In the Work folder use ViewHAR to open the file FinalRpt.har (summarizing adjusted data) and examine the header CST1, showing new industry outputs. Australian output of omn is 50000, as desired. Header PCST shows that this corresponds to 27.7% output increase.

4 Overview

Figure 1 gives a broad overview of the GTAPAdjust sequence. The program works with 3 sub-folders of the GTAPAdjust folder: Input, Work, and Output:

- Input folder must contain the original GTAP database, consisting of files: sets.har, basedata.har and default.prm. GTAPAdjust insists that the input files have these names and are located in this folder..
- Work folder stores various temporary and diagnostic files. For example, NormChek0.har contains diagnostics showing whether the input data is balanced (ie, whether accounting

[2] The supplied example will run with the limited-size Executable-image version of GEMPACK. For larger databases you may well need one of the Unlimited Executable image version, or the Source-code version of GEMPACK.

identities are satisfied. Headers CK1A, CK2A, CK3A and CK4A should contain zeros or tiny numbers[3].

- Output folder will contain the new adjusted GTAP database.

The basic plan is that:

- Program Gtp2Norm converts your original data to the simpler "normalized" format, orig.har.
- Using your instructions in input\GtapAdjust.cmf, the GtapAdjust program produces scaled data, adjusted.har.
- Program Norm2Gtp converts the scaled adjusted data back into the standard GTAP format.

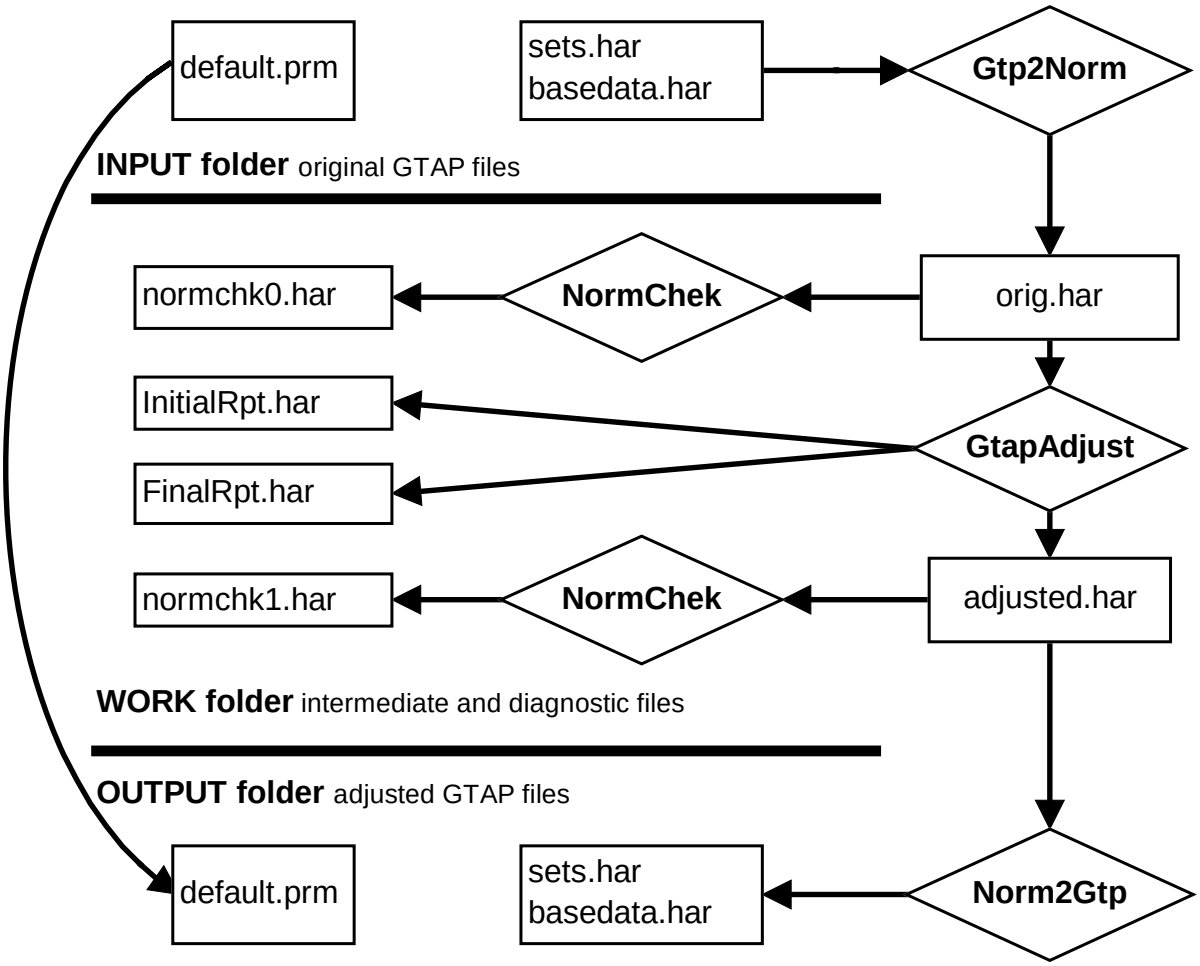

Figure 1 Overview of GTAPAdjust files and programs

5 Starting a new project

When you are starting to adjust a new dataset you should:

1. Read all of this document, unless you read it all recently.
2. Make and preserve a zip of the input folder if it contains previously used material that you may need again later.
3. Delete all files in the input folder.
4. Locate the GTAP database which is to be your starting point. Place this data (with names sets.har, basedata.har and default.prm) in the input folder.

[3] By "zeros or tiny numbers" is meant: small relative to the original flows. Use headers CHK1, CKH2, CHK3 and CHK4 to see differences beside (or as a percentage of) the original flows.

5. Edit GTAPAdjust.CMF to comment out all user-specified shocks below line 55.
6. From command prompt, type: **`ADJUST.BAT`**
   Check that it ran OK. This job merely adjusts the original data to make it balanced—probably it is already balanced
7. Edit GTAPAdjust.CMF to impose new targets.
8. From command prompt, type: **`runjob GtapAdjust input\GtapAdjust.cmf`**
9. Examine outputs. If happy, proceed to next step. Otherwise, return to step 7.
10. From command prompt, type: **`ADJUST.BAT`**

The majority of time will be spent in repeating steps 7 to 9.

6 The "normalized" data format

At an early stage, Adjust.bat reformats the GTAP flows data to a simplified, so-called "normalized", format. The main features of the normalized format are:

- The parameters file (default.prm) is the same as in the standard GTAP format.
- A single file replaces the GTAP flows file (basedata.har) and the sets file (sets.har).
- Taxes are represented explicitly, not as the difference between other flows.
- All the original matrices with two regional subscripts are consolidated in one giant TRADE array.
- Other original matrices with a sectoral subscript are consolidated in one giant NATIONAL array.

At the end, Adjust.bat translates the normalized data back to the standard GTAP format.

The normalized format helps to simplify the GtapAdjust program which adjusts the data, and the NORMCHEK programs which check and summarize data.

TABLO programs GTP2NORM and NORM2GTP convert data between the normalized and standard GTAP formats.

Table 1 Normalized data file

Header	Type	Dimension	Coeff	Name
Sets				
REG	1C	14 length 12		Set REG Regions
COM	1C	26 length 12		Set COM Commodities
IND	1C	26 length 12		Set IND Industries
MAR	1C	1 length 12		Set MAR International trade margin commodities
FAC	1C	5 length 12		Set FAC Primary factors
INV	1C	1 length 12		Set CGDS_COMM Investment good
TRD	1C	4 length 12		Set TRADES Basic, exptax, imptax, margins
COST	1C	32 length 12		Set COST All industry costs
USER	1C	29 length 12		Set USER All local users
Core				
TRAD	RE	TRADES*COM*REG*REG	TRADE	GTAP trade data
NATL	RE	COST*SRC*USER*REG*TYP	NATIONAL	GTAP national data
Other				
EVOA	RE	ENDW_COMM*REG	EVOA	Endowments - Output at Agents' Prices
SAVE	RE	REG	SAVE	Savings - Net Expenditure at Agents' Prices
VKB	RE	REG	VKB	Capital Stock - Value at Beginning-of-Period
VDEP	RE	REG	VDEP	Capital Stock - Value of Depreciation
DPSM	RE	REG	DPARSUM	Sum of Distribution Parameters in Household Demand System
VST	RE	MARG_COMM*REG	VST	Trade - Exports for International Transportation, Market Prices

6.1 Contents of normalized data file

Table 1 lists headers of a normalized GTAP data file[4]. First come various sets. Sets COM and IND are each the same as the GTAP TRAD_COMM set. Investment is regarded as a type of final demand. In this example, there is only one margin good (svces); in other data there might be more margins.

Next come two large arrays, TRADE and NATIONAL, which contain nearly all the flows data used by the GTAP model.

The TRADE array shows, for each COM, all bilateral trades (first REG index is source; the second is destination). The set TRADES (see Table 2) corresponds to different layers of this array, namely BASIC, EXPTAX, IMPTAX, followed by members of the MAR set (in this case, Transport). To value trade at FOB prices we would add together the BASIC and EXPTAX layers. Further addition of the margin layers would give us trade at CIF prices. Addition of the final, IMPTAX layer would give duty-paid imports.

The NATIONAL array shows, for each REG, all expenditures by non-export demanders. The USER set consists of the sectors (=COM=IND) followed by final demanders HOU, INV, and GOV. The COSTS set consists of the commodities (=COM=IND) followed by the primary factors and the

[4] Look at files orig.har or adjusted.har (in the work folder) to see examples of this file format.

PTAX output tax. The SRC index shows whether goods are produced locally or elsewhere. Yet another index, TYP, divides costs into basic or tax components.

Table 2 Members of special sets

USER	COSTS	TRADES	TYP	SRC
FoodFrsFsh	FoodFrsFsh	BASIC	BAS	dom
Coal_Oil_Gas	Coal_Oil_Gas	EXPTAX	TAX	imp
…..*other*	*sectors…..*	IMPTAX		
svces	svces	Transport		
cns	cns			
HOU	Land			
INV	UnSkLab			
GOV	SkLab			
	Capital			
	NatRes			
	PTAX			

6.2 The balance conditions

The GTAP database must satisfy 3 accounting identities which we can express in terms of the TRADE and NATIONAL arrays. For each identity, flows are measured at "basic" prices, which include production tax (for domestic goods) and tariff tax (for imported goods). For each constraint, GtapAdjust applies a scale factor which has the same dimensions as the constraint.

6.2.1 Industry costs = sales of locally made goods

Using TABLO notation, costs and sales are calculated as:

```
(all,i,IND)(all,r,REG) COSTS(i,r) =
     sum{k,COST, sum{s,SRC, sum{t,TYP, NATIONAL(k,s,i,r,t)}}};
(all,c,COM)(all,r,REG) DOMSALES(c,r)=sum{u,USER, NATIONAL(c,"dom",u,r,"BAS")};
(all,c,COM)(all,r,REG) EXPSALES(c,r)=sum{t,REG, TRADE("BASIC",c,r,t)};
(all,c,COM)(all,r,REG) SALES(c,r) = DOMSALES(c,r) + EXPSALES(c,r);
(all,m,MAR)(all,r,REG) SALES(m,r) = SALES(m,r) + VST(m,r);
```

To ensure COSTS=SALES, GtapAdjust scales COSTS, ie, cost columns of the NATIONAL array. The scale factor is called domslack.

6.2.2 Supply = demand for imports

Again in TABLO notation:

```
(all,c,COM)(all,r,REG)
      IMPSUP(c,r) = sum{t,REG, sum{x,TRADES, TRADE(x,c,t,r)}};
(all,c,COM)(all,r,REG)
      IMPDEM(c,r) = sum{u,USER, NATIONAL(c,"imp",u,r,"BAS")};
```

To ensure IMPSUP=IMPDEM, GtapAdjust scales IMPSUP, ie, corresponding columns of the TRADE array. Eg, other countries' exports to Australia are adjusted to sum to Australia's import needs. The scale factor is called impslack.

6.2.3 Supply = demand for international shipping margins

Again in TABLO notation:

```
Formula (all,m,MAR) MARDEM(m)
  = sum{c,COM,sum{t,REG,sum{f,REG,TRADE(m,c,f,t)}}};
Formula (all,m,MAR) MARSUP(m) = sum{r,REG, VST(m,r)};
```

To ensure MARSUP=MARDEM, GtapAdjust scales MARSUP, ie, the VST matrix. The scale factor is called marslack.

7 Instruments and targets

GtapAdjust also defines additional scaling variables, prefixed "q", which may be endogenized to hit particular targets. Instruments and targets are listed in Table 3 below.

Table 3 Instruments and targets for GtapAdjust.cmf

Instrument	Size	Description	Target
qnatsiz(r)	REG	scale to target national GDP	vGDPEXP
qdem(c,f)	COM*REG	scales ALL use of c made in f	vSALES or VCOSTS
qexp(c,f)	COM*REG	scale factor for exports	vEXPDEM
qimp(c,t)	COM*REG	scale factor for imports	vIMPCIF or vIMPDEM
qexptot(f)	REG	scale factor for aggregate exports	vGDPEXPS("Exp",*)
qimptot(t)	REG	scale factor for aggregate imports	vGDPEXPS("Imp",*)
qexpend(d,r)	FIN*REG	absorption factors: to target C, I, G	vGDPEXPS(FIN,*)
qTRDTAX(c,t,r)	COM*TRD*REG	scales trade taxes accruing to r	vTRDTAX
qTRDTAXTOT(t,r)	TRD*REG	scales trade taxes accruing to r	vTRDTAXTOT
qNATTAX(c,r)	COM*REG	scales commodity taxes	vNATTAX
qNATTAXTOT(r)	REG	scales commodity taxes	vNATTAXTOT
qfac(f,r)	FAC*REG	scales primary factor use	vGDPFACS

Note: For regional subscripts, f means "from" and t means "to". Set TRD has members (EXP,IMP); set FIN is (HOU,INV,GOV).

The GtapAdjust.cmf contains several (commented out) examples of instrument-target swaps and shocking of targets.

8 Programs and files of Adjust.bat

Adjust.bat consists of a number of separate operations or programs, which may be divided into 3 stages:

Stage 1 is concerned with converting the orginal GTAPdata to the normalized format:

- GTP2NORM converts the original data to the normalized format (orig.har).
- The NORMCHEK program is used to check and summarize orig.har.

Stage 2 consists of the actual data adjustment:

- GtapAdjust scales the data, so that accounting identities are satisfied, and (optionally) so that specified target totals are reached. The scaled output is stored as adjusted.har. Report file InitialRpt.har summarizes the input data. Report file FinalRpt.har summarizes the adjusted data, and reports various aggregate percent changes between old and new data.

Stage 3 is the "wrap-up" phase:

- DiffHAR computes ordinary and percent differences between all flows in orig.har and adjusted.har (showing changes made by GtapAdjust).
- The NORMCHEK program is used to check and summarize adjusted.har.
- NORM2GTP converts adjusted.har into the standard GTAP format (output\basedata.har).
- The standard GTPVEW (aka GTAPVIEW) program produces a summary of the new GTAP data.
- Finally (and optionally) GTAP.TAB is used to run a test simulation of the new data.

Table 4 lists files produced at these stages.

Table 4 Input, work and output files

INPUT folder	Description
basedata.har	original GTAP data files
sets.har	
default.prm	
GtapAdjust.cmf	specifies adjustments to be made
WORK folder	*in order of creation:*
orig.har	the original GTAP data in normalized format
gtp2norm.har	diagnostic file produced by GTP2NORM
normchk0.har	balance and other diagnostics for orig.har. See header SMRY.
InitialRpt.har	Summary of initial data from GtapAdjust
Adjusted.har	the adjusted data in normalized format
FinalRpt.har	Summary of adjusted data from GtapAdjust, together with old-to-new percent changes.
PctDiffs.har	Diffhar output, comparing Adjusted.har to Orig.har (percent changes)
OrdDiffs.har	Diffhar output, comparing Adjusted.har to Orig.har (oridinary changes)
normchk1.har	balance and other diagnostics for adjusted.har. See header SMRY.
OUTPUT folder	
default.prm	expanded GTAP data files
sets.har	
basedata.har	
taxrates.har	GTAPVIEW output
gtapview.har	

9 GtapAdjust.TAB

You will need to study GtapAdjust.TAB carefully to see how it works. It is divided into the following sections:

Sets

Necessary sets are read in or defined.

Main data arrays

The NATIONAL, TRADE, and VST arrays are declared and read in. Each array is updated by a percent change variable prefixed "u".

Helper variables

This long section defines a number of aggregates, such as GDP and sectoral output, imports and exports. These can potentially be used as targets in scaling. Each aggregate has a corresponding percent change variable, prefixed "v".

Costs=sales domestic accounting constraint
Imports accounting constraint
Margins accounting constraint

For each of these constraints, the supply, demand, and the difference between the two is defined, both in the levels and as a change variable. GtapAdjust.cmf includes statements to move the differences to zero (or, more usually, to keep them at or near zero).

Scaling instruments to meet accounting constraints

This defines the three scaling variables (domslack, impslack, and marslack) used to enforce the accounting constraints.

Scaling instruments to meet targets

This section list the instrument variables, prefixed "q", which may be used to enforce various targets.

Optional link from gdp to absorption

Please see the comments in the TAB file.

Scaling rules

This section defines the scaling rules, ie, it shows which "q" instrument variables scale which bits of the NATIONAL, TRADE, and VST arrays

Condensation

Some larger "u" variables are substituted out. Note that targets and instruments cannot be substituted, since their closure status (exogenous or endogenous) may change. Hence, to target very detailed flows (eg, import use by commodity, industry and region) you would need to add large variables for both instruments and targets. Such variables are not included (or are condensed out) in the standard supplied version (to ensure that the example works even with limited-size versions of GEMPACK).

Report files

Instructions to create the InitialRpt.har and FinalRpt.har files.

Updating minor data arrays

The remaining arrays of orig.har are updated using some simple rules.

10 Using the diagnostic files

You can examine diagnostic files for troubleshooting or quality control. The more important ones are listed below, with some hints on using them.

work\normchk0.har

NormChek0.har contains diagnostics showing whether the original GTAP data is balanced (ie, whether accounting identities are satisfied). Headers CK1A, CK2A, CK3A and CK4A contain demand-supply differences: they should contain zeros or tiny numbers[5]. Headers CHK1, CHK2, CHK3 and CHK4 allow you to view demand and supply as well as either ordinary or percentage differences between the two. Your input GTAP data should (and normally will) be accurately balanced. The header SMRY provides a useful overview.

work\InitialRpt.har

This is useful to see initial values of the available target variables.

work\FinalRpt.har

This is useful to see percent changes and final values of target variables.

The final headers [DSLK, ISLK, MSLK] show how much work GtapAdjust had to do to enforce accounting identities on adjusted.har. Huge values for these multipliers may indicate implausibility in the targets (or closure) specified in GtapAdjust.cmf.

work\normchk1.har

NormChek1.har contains diagnostics measuring the size of imbalances in the file adjusted.har produced by GtapAdjust. Imbalances should be small. The header SMRY provides a useful overview.

---

[5] By "zeros or tiny numbers" is meant: small relative to the original flows. Use ViewHAR's File...Options command to cause the minimum, maximum and total values and the number of negatives to appear on the Contents screen.

work\PctDiffs.har and work\OrdDiffs.har

These 2 files contain arrays showing the percent[6] and ordinary differences between each flow in files orig.har and adjusted.har, ie, they show the changes made by GtapAdjust. Once again, large changes may indicate implausibility in the targets imposed by GtapAdjust.cmf.

output\GTAPView.har and output\TaxRates.har

These are two diagnostic files produced by the standard GTAPView program from the output database. They contain useful summary information, and flag some potential problems.

11 Your shocks

GtapAdjust is likely to work very well if you enforce only one or a few changes. But if you attempt to enforce multiple targets for several regions, problems are possible. For example:

- You cannot make exports exceed output.
- You cannot set *both* exports by sector and aggregate exports.
- You cannot set *all* components of *both* sides of GDP.
- You cannot set *both* world cheese exports and world cheese imports.

Singular matrix errors, huge "slack" multipliers, or other solution problems may occur if you try the above. To avoid or diagnose these problems, apply targets in stages.

12 Troubleshooting

GTAPAdjust may stop with an error message, or display a list of log files. Examine the end of the last (most recent) log file. In many cases the message in the log file will be self-explanatory.

Note that the input folder must contain files: sets.har, basedata.har and default.prm in the correct GTAP format. "Correct format" means, the format used by the GTAP model (ie, you could use the files in RunGTAP).

13 After GTAPAdjust is finished

When you are satisfied with the GTAPAdjust output you should:

- Make and preserve a zip of the input folder so you can repeat the whole procedure later.
- Make a zip of the output folder: you can load this zip into RunGTAP via the New Version wizard.

14 Implementation notes

14.1 Dynamic GTAP data

The procedures are designed to work with the data and format used by the standard GTAP model. Dynamic GTAP uses some extra headers (YQHF,YQHT,YQTF, etc). Programs Gtp2Norm, GTAPAdjust, and Norm2Gtp are designed to transfer these headers if they are present in the input file basedata.har -- if not present in the input, they will be created with dummy values in the output file. You could copy the same technique to process other optional data.

[6] It makes no sense to add % changes, so make sure to "slice" not "sum" in ViewHAR when viewing % changes in the TRADE or NATIONAL arrays.

14.2 GEMSIM or Fortran?

To reduce download size, the programs used in Adjust.bat are supplied as TAB files which are run with GEMSIM. If you have the source-code version of GEMPACK (and are scaling a very large database) you might prefer to run EXE versions of the programs. To do so:

- Run the batch file COMPALLF.BAT to compile the programs
- Edit the file RunJob.bat (instructions within) to use Fortran-based exes.

14.3 Limits

You may need a source-code or unlimited Exe-image GEMPACK licence if the dataset is large. If very large, you may even need the 64-bit version of GEMPACK.

14.4 Solution method

Since the accounting constraints are linear, you should achieve perfect balance even with a a one-step, Johansen solution. In this case, flows are updated with formulae like:

F(new) = F(old)*[1 + a/100 + b/100 + ...]

where a and b are percent change variables. A problem arises if a=b=-70%: F would change sign! With 6 or more Euler steps, the updating formulae are more like:

F(new) = F(old)*[1 + a/100]*[1 + b/100]* .....

so that sign reversal is much less likely.

14.5 Group Policy

If your system administrator does not allow you to open a "DOS box" (CMD console or command prompt) on your PC, you will be unable to use GTAPAdjust (or accomplish much else of use).

<end of document>