Initial Preparation of Energy Volume Data for GTAP 7

Robert McDougall

Angel Aguiar

GTAP Research Memorandum No. 10

November 2007

Center for Global Trade Analysis, Department of Agricultural Economics, Purdue University,
1145 Krannert Building, West Lafayette, IN 47907

Initial Preparation of Energy Volume Data for GTAP 7

Robert McDougall
Angel Aguiar

November 2007

This memorandum covers initial preparation of energy volume data for release 7 of the GTAP data base.

1 Outline of work

For release 7 of GTAP, as for release 6, we rely mainly on the "extended energy balances" for OECD and non-OECD countries, published by the International Energy Agency. The IEA has kindly allowed us free access to these through its on-line data service.

We download the data from the IEA web site as "Beyond 2020" files, one for OECD countries and one for non-OECD. For release 7, we update from the 2003 to the 2006 edition ([1], [2]).

The next step is to extract data from the "Beyond 2020" files into CSV files. When we did this, we knew only how to extract two-dimensional slices. Accordingly, we extracted (product, year) slices, of which there were 124 for each year, 62 each from the OECD and non-OECD files. As we did this manually, this step was somewhat laborious, and, worse, error-prone. We confined ourselves therefore to the previous and current GTAP data reference years, 2001 and 2004.

Since then, we have learned how to extract more than two dimensions at once (see "Nesting Dimensions" in the browser help). In future we should extract just one CSV file each from the OECD and non-OECD browser files.

Also manually, we record set and mapping information in various text files, and certain special data adjustments, to be discussed shortly below. The rest of the work is done programmatically. We convert the CSV files to GEMPACK format. At this point, we convert the notations "`c`" for "confidential and "`..`" for "not available" to zero values. In so doing, we lose information that might serve to improve our data quality; investigating its exploitation may be work for the distant future.

We now apply the special data adjustments mentioned above. These arise from one of the internal consistency conditions of the data set. The flow `NECHEM`, "Memo: Feedstock Use in Petchemical Industry" is by definition a subset of the flow `NEINTREN`, "Non-Energy Use Industry/Transformation/Energy", and its recorded volume should therefore be lower. In a few instances, listed below, this condition is not satisfied. Remarkably enough, each instance pertains to the product `ONONSPEC`, "Non-Specified Petroleum Products"; more remarkably, in each instance, the excess of feedstock over total reported usage is equal to the value reported for the flow CHEMICAL, "Chemical and Petrochemical". We conjecture therefore in these instances, some usage of ONONSPEC has been recorded both as energy use (under CHEMICAL) and as non-energy use (under NECHEM). Furthermore, the CHEMICAL reports appear abnormal, as they appear only in years in which the condition is unsatisfied, but not in earlier or later years. We conjecture therefore that this usage has been recorded under CHEMICAL in error, and belongs under NEINTREN; and emend the data accordingly. The instances, and amounts in ktoe, are:

2001	Slovakia	42.08
2004	Czech Republic	9.60
2004	France	0.96

Next, we reorder indices in the OECD and non-OECD data sets. This is for compatibility with preprocessing programs inherited from past releases. In future, we should avoid the need for this, by switching dimensions within the 2020 browser so as to extract the data in the preferred index ordering.

Next, in a program procexeb, one of the inherited programs mentioned just above, we combine the OECD and non-OECD data in a single data set, check for internal consistency, delete total and sub-total elements from the country, product, and flow sets, and disaggregate and fill in the country classification to the standard GTAP countries. In another inherited program ev_iea2gtap, we remap flows to something closer to the GTAP sectoral classification, and restructure the data to something closer to a Leontief input-output structure.

For this release, we make some changes to those programs. We regularize and simplify the incoming data structure, update the country, product, and flow classifications, and change from using single- to multi-year GDP data, to facilitate handling of multi-year energy volume data (the GDP data are used in expanding the country classification).

Although the old versions naturally read in data in the IEA country, product, and flow classifications, they used their own peculiar codes for set elements. They then mapped from IEA to GTAP countries, as described above, and from IEA flows to a classification intermediate between the IEA flow classification and the Energy Data Set flow classification described in the GTAP energy documentation ([3]). For products, they retained the initial classification, merely dropping classes representing totals or subtotals of other classes. In this release, we retain the IEA element codes for the IEA classifications; for products, though not for countries and flows, these persist through to the GTAP-ready data output.

Finally, with a new utility hasplit, we split the data set into single-year slices. Thus we generate at once two single-year data sets, one for 2001 and one for 2004.

2 Changes in IEA classifications

The IEA country, product, and flow classifications change slightly between the 2003 and 2006 editions. In particular:

- Botswana is split out from "Other Africa".

- Anthracite is split out from "Other Bituminous Coal".
- New products "Coal Tar" and "Elec/Heat Output from Non-Spec. Manuf. Gases" are added.
- New flows "Chemical Heat for Electricity Production" and "For Blended Natural Gas" are added.
- The old transformation flow "Liquefaction Plants" is split into "Coal Liquefaction Plants" and "Gas-to-Liquids Plants", as is the corresponding energy sector flow.
- Fishing is split out from "Agriculture/Forestry".
- The flow "From Other Sources—Primary Energy" is removed.

Besides these changes in classification, changes in concepts and practices pertaining to various individual classes, not detailed here, required internal revisions to the preprocessing programs.

3 Comparison of results for 2001 between the 2003 and 2006 editions

We have taken some pains to adapt the old preprocessing programs to changes in IEA classes and concepts, so that changes in data output for 2001 reflect only changes in the content of the source data between the 2003 and 2006 editions. We have revised the GTAP energy module to use the new product classification, so that we can compare energy module outputs across the two editions.

Results for the cross-edition comparison appear in table 1. Before considering the substance of the results, we discuss the method of comparison, which we shall use again.

We use an entropy divergence measure to identify the greatest divergences between the two data sets. The measure takes account of both the relative and the absolute difference between the two comparands. Since it is not well defined when one comparand is zero and the other non-zero, before comparing, we mutually contaminate the two data sets, so that, for instance, we replace the 2003 result with a weighted average of the 2003 and 2006 results, with a much greater weight on 2003 and a much smaller but non-zero weight on 2006. So in the first line of table 1, to an initial 2003 estimate of 5158 corresponds an adjusted estimate of 5157.

Having obtained comparable arrays, we then compute entropy divergence measures element by element. We do this both for the arrays themselves and for their marginal totals, as, for instance, the region-by-industry totals obtained by summing over commodities, the region-by-commodity totals obtained by summing over industries, and the region totals obtained by summing over both commodities and industries. We then pool all divergences together and select the greatest; in the present case, that is for the element representing region Japan (`jpn`), commodity electricity (`ely`), and industry transport equipment (`teq`) in the full array. This supplies the first line of the table.

In the initial comparison run, after the triple `(jpn,ely,teq)`, the second greatest divergence is observed for `(jpn,*,teq)`. Here the character `*` indicates summation of

Table 1: GTAP energy usage estimates for 2001: comparison between the 2003 and 2006 editions of the IEA's extended energy balances (USD M)

			Adjusted		Initial	
Region	Comm.	Ind.	2003	2006	2003	2006
jpn	ely	teq	5157	421	5158	420
xme	p_c	ser	5863	1093	5865	1092
xnf	*	ser	1866	49	1866	49
fra	gas	ser	1138	8	1138	8
xme	gas	ser	1708	137	1708	136
jpn	ely	ome	8229	14458	8227	14459
usa	ely	ely	15642	9164	15644	9162
ita	gas	ser	8	712	7	712
twn	p_c	p_c	6458	3003	6459	3002
ind	p_c	fpr	224	1509	224	1510
jpn	*	twl	2473	649	2473	649
xme	gas	gas	2485	700	2486	699
tha	oil	crp	0	243	0	243
ind	gas	gas	0	235	0	235
tha	p_c	p_c	107	937	107	938
xme	gas	oil	1929	542	1930	542

the corresponding index; so this comparison pertains to energy usage by the transport equipment industry in Japan, summed over all energy commodities. But, we may suspect, this largely repeats the information in the first line, which also pertains to the Japanese transport equipment industry, though only to its electricity usage. Likewise, the twenty-second greatest divergence observed is for `(*,ely,teq)`, and this too we may suspect is driven largely by the `(jpn,ely,teq)` divergence.

For our reported results, therefore, we take only the first line from the initial comparison run. We then rebalance the 2003 edition results, so as to eliminate the greatest single divergence from the 2006 edition in the initial comparison run, that is, the divergence for the element `(jpn,ely,teq)`. Since this is an element of the full array, this is accomplished simply by replacing the 2003 value (5157) by the 2006 value (421). In case of a subtotal, we would rescale the corresponding slice of the table.

That done, we undertake a second comparison run, again comparing both atomic elements and marginal totals. We find that the rebalancing undertaken between the first and second comparison runs has largely eliminated the divergence for the `(jpn,*,teq)`. The greatest remaining divergence is that for `(xme,p_c,ser)` (that is, for services industry usage of petroleum products in "Rest of Middle East"), and that supplies the second line of our table. And so we continue, alternately comparing and rebalancing.

We turn now to the substance of table 1. We find, first of all, and as noted above, a sharp decrease in electricity usage in the transport equipment industry in Japan, from US \$5.2 to \$0.4 billion. This reflects revisions supplied to the IEA by the Japanese administration ([1]). In fact, between the 2003 and 2006 editions, reported usage by "Transport Equipment" sinks from 2.3 Mtoe to zero. The $0.4 billion in the GTAP 2006 edition estimate derives from the 9.0 Mtoe reported for the IEA category "Non-Specified (Industry)", a portion of which we allocate over manufacturing industries including the transport equipment industry. In the revised Japanese report, presumably, usage in transport equipment manufacturing is reported under "Non-Specified (Industry)", which, accordingly, has a large transport equipment share; but our allocation procedure takes no account of that. It is likely therefore that our 2006 edition results understate the Japanese transport equipment industry's usage.

We note next a decrease in usage of petroleum products in "Rest of Middle East" (`xme`) service activities, from US $5.9 to $1.1 billion. This results from improved reporting by Saudi Arabia and Syria. Both countries remove entries of gas/diesel oil under "Non-Specified (Other)" and create entries under "Road" and "Non-Specified (Industry)". For Saudi Arabia, the amount removed from "Non-Specified (Other)" is 13 Mtoe, for Syria, 4.5 Mtoe.

Third, we find a sharp decrease in total energy usage in service activities in "Rest of North Africa" (`xnf`), from US $1.9 billion to $49 million. This reflects the replacement by more specific entries of entries under "Non-Specified (Other)" for electricity in Libya and gas/diesel oil in Algeria. This and the previous row bear on our treatment of "Non-Specified (Other)". As the IEA notes, "Non-specified includes all fuel use not elsewhere specified as well as consumption in the above-designated categories for which separate figures have not been provided." Our treatment is appropriate where the reported volumes are "fuel use not elsewhere specified", but, for the cases just examined, in the 2003 edition, the volumes are rather "consumption . . . for which separate figures have not been provided". In the 2006 edition, of course, those particular entries disappear, but it may be worthwhile in future to reconsider the treatment of the remaining "Non-Specified (Other)" entries.

Fourth, in France, usage of gas in service activities falls from US $1.1 billion to $8.0 million. Again, we find a revision by the reporting country: the French authorities have moved the entire volume of natural gas previously reported under "Commercial and Public Services", 9.4 Mtoe, to "Residential". Admitting our ignorance of French practice, it seems curious that the French commercial and public sector should, as the 2006 edition reports, use no natural gas, while other sectors of the economy do; certainly this diverges from usage patterns in other European countries. Italy, meanwhile, has made an opposite revision, moving 5.9 Mtoe from "Residential" to "Commercial and Public Services", as reflected in the eighth row of the table.

Without pursuing the comparison further, we conclude:

- The changes in GTAP estimates resulting from updating to the 2006 edition are highly specific to individual countries, industries, and commodities.
- The changes in GTAP estimates can be explained by changes in countries' reports to the IEA.
- Some of those changes in reports, such as those noted for Saudi Arabia, Syria, Libya, and Algeria, constitute improvements in reporting in less developed countries.

Table 2: IEA energy usage: comparison between 2001 and 2004 in the 2006 edition of the extended energy balances (Mtoe)

			Adjusted		Initial	
Region	Flow	Product	2001	2004	2001	2004
usa	*	PatFuel	0	149	0	149
chn	*	*	2924	4262	2924	4263
usa	P_C_Transfm	BitCoal	2	74	2	74
*	*	*	37060	39447	35501	39447
usa	MilitaryUse	*	0	15	0	15
ita	*	Heat	0	9	0	9
*	Elect_Gen	AntCoal	0	7	0	7
ukr	IronXSteel	NatGas	0	7	0	7
*	Dom_Prod	AntCoal	0	6	0	6
ukr	FidStok4CRP	*	0	5	0	5
pan	*	CrudeOil	5	0	5	0
ukr	NonSpecfInd	NatGas	14	1	13	1
usa	Elect_Gen	BitCoal	295	197	275	197
nor	Exp	NGL	4	0	4	0
rus	*	CrudeOil	740	904	688	904
usa	Elect_Gen	RefinGas	0	4	0	4

- Major changes in reports from industrialized economies, such as those noted for Japan and France, may or may not be improvements. Expert comment on these would be welcome.

4 Comparison between 2001 and 2004 in the 2006 edition

At the time of writing, we are not ready to generate 2004 results from the energy module, as we do not yet have GTAP-ready energy price or tax data. But we can compare the 2001 and 2004 data at the point of input into the energy module; this we do in table 2. Both year's data come from the extended energy balances' 2006 edition.

Table 2 follows the same method as table 1, but provides some new points of discussion. First, the data sets at the point of comparison contain negative values for certain flows, "International Marine Bunkers", "Stock Changes", and "Statistical Differences"; as our method does not accommodate negative flows, and as these are relatively minor flows, we drop them from the data sets.

Second, the fourth line of the table, indexed to `(*,*,*)`, represents change in total energy usage summed over countries, flows, and products. At this point the 2001 data have already been balanced to match the 2004 data for the three previous items, and this, especially the balancing for China, affects the array total. So we see that in the fourth line, the adjusted 2001 value differs considerably from the initial value, reflecting the adjustments already made to energy usage in China. In turn, the rebalancing of the array total performed after the fourth comparison affects the adjusted value in every subsequent line.

Third, from the fourth rebalancing on, there are interactions between our target totals. Rescaling the array to adjust the array total of course affects the elements and subtotals already targeted, `(usa,*,PatFuel)` and so on. We use a simultaneous balancing procedure, so as to preserve the earlier targets while achieving the current one. So the values already achieved for `(usa,*,PatFuel)`, `(chn,*,*)`, and so on, do not change in subsequent adjustments.

Turning now to the substance of the table, we may expect to find general growth in world energy volumes to hold a prominent place in the table, and this it does, though not the highest place, but the fourth. More notable, according to our measure, is growth specifically in China, from 2.9 to 4.3 Gtoe over three years (the country total, though perhaps a useful measure of general tendency, is not economically meaningful; total final consumption grows at a rapid but lower rate, from 798 to 1038 Mtoe; the more rapid growth in the country total perhaps indicates increasing indirection in energy use). More notable even than that is a phenomenon peculiar to the United States, the emergence of "Patent Fuel", as a major energy commodity.

The IEA notes describe patent fuel as "a composition fuel manufactured from hard coal fines with the addition of a binding agent". According to the 2006 edition of the balances ([1]), output from patent fuel plants in the United States rose from zero in 2001 to 52 Mtoe in 2002, and then to 74 Mtoe by 2004. Almost all (73 Mtoe) of this was consumed in electricity generation, representing about 12 per cent of total fuel inputs therein; small amounts were consumed in coke ovens, chemicals manufacture, and other minor uses. The only significant energy input into patent fuel production is bituminous coal, and accordingly, in the third line of the table, we find bituminous coal usage by the United States petroleum and coal transformation industry (to which `ev_iea2gtap` maps the flow "Patent Fuel") rising from 2 to 74 Mtoe. Again, in the thirteenth line, we find bituminous coal usage in United States electricity generation falling from 275 to 197 Mtoe.

In brief, then, some bituminous coal, instead of passing immediately into electricity generation, is first crushed and bound into patent fuel and then consumed. Whether this represents a change in practice or in reporting, we cannot say. It does highlight an inconsistency in the preprocessing procedure, noted in the energy documentation ([3]), that whereas we map the commodity "Patent Fuel" to coal, we map the corresponding "flow" to petroleum and coal processing. From the ISIC ([4]), it appears that coal is the right match; changing the flow mapping therefore becomes a high priority on our energy data agenda.

At the fifth line, the entry `(usa,MilitaryUse,*)` is misleading. `MilitaryUse` is our perhaps inappropriate code for the IEA flow `ONonSpec`, "Non-Specified (Other)", a class which does indeed include military fuel use, but also, as noted above, "fuel use not elsewhere specified as well as consumption in the above-designated categories for which separate figures have not been provided". For the United States in 2004, it comprises just one product, electricity. It seems doubtful that this represents military use, and certain that military use of other products has gone unreported here. In the energy module, we assign `MilitaryUse` to the general service sector (excluding transport, construction, and electricity supply), and on present information this seems harmless (the amount assigned from `MilitaryUse`, 15 Mtoe, is much less than that assigned from "Commercial and Public Services", 106 Mtoe).

For Ukraine, the growth in natural gas use in iron and steel making and as feedstock into the petrochemical industry, together with the decline natural gas use in "Non-Specified (Industry)", evidently reflects change in reporting more than in practice.

In general, we conclude:

- The changes reported between 2001 and 2004 appear generally plausible.
- We are unsure whether the most notable change, the interposition of patent fuel into the energy chain in the United States, represents change in practice or in reporting.
- Growth in energy volumes in China is, plausibly, rapid.
- While various flows appear (for instance, heat production in Italy) or disappear (for instance, exports of natural gas liquids from Norway), altogether the structure of energy production and use appears stable.

As always, comment is welcome.

5 Future work

In further work on energy data for release 7, we hope to address various agenda flagged in the release 6 documentation ([3]), some of which pertain to the volume data preprocessing. In particular, as noted above (section 4), we should remove the inconsistency between the flow and product mappings for patent fuel.

References

[1] International Energy Agency. Energy balances of OECD countries (2006 edition), extended energy balances. `http://wds.iea.org/wds/`, 2006.

[2] International Energy Agency. Energy balances of non-OECD countries (2006 edition), extended energy balances. `http://wds.iea.org/wds/`, 2006.

[3] Robert McDougall and Huey-Lin Lee. An energy data base for GTAP. In *Global Trade, Assistance, and Production: The* GTAP 6 *Data Base*, chapter 17. Center for Global Trade Analysis, Purdue University, 2006.

[4] United Nations. *International Standard Industrial Classification of All Economic Activities, Third Revision.* United Nations Publishing Division, New York, 1998. Series M No. 4, Rev. 3.