Russian Energy Taxes for GTAP 5.3

Robert McDougall

GTAP Research Memorandum No. 1

March 2003

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

Russian Energy Taxes for GTAP 5.3

Robert A. McDougall

March 2003

Contents

List of Tables

List of Figures

1 Introduction

This memorandum records work done to revise energy data for Russia for release 5.3 of the GTAP data base, more specifically, data for commodity taxes on energy. The main object of the work is to let us fit the Russian input-output (*IO*) table to the energy data. A secondary object is to improve the representation of energy taxes in the fitted table.

Release 5.3 is a minor release of the GTAP data base, to incorporate updated source IO data for several regions and, for the first time, a separately identified Russia. The IO data source for Russia is a table developed by Sergei Kiselev and Roman Romachkine, with support from the United States International Trade Commission, handled by Rob Koopman. After review and reformatting by Terrie Walmsley, this table was received by the Center for Global Trade Analysis on 2002–10–21.

Unfortunately, a first attempt to use this table failed, when the GTAP data base construction procedure aborted while processing it. This project arises from the need to remedy this failure.

Of the many modules into which the GTAP data construction procedure is organized, we are concerned with just two, one in which the energy data are assembled (the *energy module*), and one in which the IO tables are adjusted to fit the energy data and various other multiregional data sets (the *Fit module*).

As we describe in section 2, the old run aborts in the Fit module, in attempting to fit the Russian table. It does so because the targets assigned to the electricity sector for that table are mutually inconsistent. Specifically, the user cost of energy inputs exceeds producer revenue from electricity output. That implies that non-energy input costs must be negative; but that is forbidden by sign constraints on input costs.

This failure occurs despite precautions taken in the energy module to avoid just such infeasibilities. The precautions fail because of differences in tax data: the tax rates applied in the Fit module differ from those used in the feasibility calculations in the energy module. More specifically, the Fit module (in this case) uses the rather high tax rates incorporated in the contributed IO table, while the energy module (in this case) assumes zero taxes. The producer revenue of the electricity sector suffices to cover the market value of energy inputs but not the taxes.

This inconsistency in tax rate data between the modules arises from inconsistency in the treatment of missing values in the tax rate arrays within the energy module. In preparing final estimates for transmission to the energy module, the missing values are preserved, but in the infeasibility calculations they are replaced by zeros.

To remove the infeasibility therefore we aim to remove the internal inconsistencies within the energy module. Since we consider the tax rates in the contributed table relatively reliable, we aim to do this in a way that brings the energy module rates closer to those in the contributed IO table. Furthermore we aim to do this by changing only the source data not the energy module programs. This is because in the GTAP 5 series we observe a constraint, that the data base construction programs are not to be changed.

In section 3 we examine more deeply the energy module's handling of tax data, to learn why the missing value flags appear for the critical flows in the old run, and how to change them in a new run. We find that we cannot make the Fit module tax rates entirely consistent with the energy module rates, nor the energy module tax rates entirely consistent with the IO table rates. But we can substantially reduce the inconsistencies.

In the course of this project, on 2003–01–27, we received a revised version of the Russian table, from the same sources as the 2002-10 version. Section 4 reviews the energy tax data in this table. Finally, section 5 describes how we extract tax rate data from the new table and use it to modify the energy module source data, assesses our success in reducing the tax rate inconsistencies, records the successful completion of a new construction run, and notes suggestions for future work.

2 Why won't Russia fit?

We now validate and amplify the claims (section 1) that Russia fails to fit in the old run because the electricity sector targets are infeasible, that they are infeasible because of inconsistencies between the tax rate targets applied in the Fit module and those applied in the feasibility calculations in the energy module, and that those inconsistencies arise from inconsistencies in the treatment of missing value flags in the energy module. We provide first an overview, then a detailed discussion.

2.1 Overview

A key feature of GTAP release 5 is the sourcing of the energy sector to International Energy Agency (*IEA*) statistics. This entails overwriting energy-related components of data from other sources, including IO data. In the IO table fitting module, we impose targets for several energy-related data elements, including

- market value of domestic usage of energy commodities, by input and use,
- border values of exports of energy commodities,
- rates of export subsidy, by commodity,
- powers of taxes on energy commodity usage, and
- powers of production taxes on energy industries.

The energy module supplies values to the Fit module for all these data elements.

From these explicit targets, we can infer several implied targets. For each energy commodity, the market value of exports is equal to the border value, multiplied by the power of the export subsidy. The market value of total sales is equal to market values of export and domestic usage, summed over uses. Producer revenue is equal to total sales discounted for production taxes. For each energy industry, the user cost of each energy input is equal to the market value of purchases multiplied by the power of the commodity tax; and the total user cost of energy inputs is equal to user cost summed over energy inputs.

The IO tables observe an accounting identity, that the total cost of inputs into each industry is equal to total producer revenue from the corresponding commodity (this is

Table 1: Implied targets for Russia, in the failed Fit run (US$ billion)

	Producer revenue	Energy costs	Non-energy costs
`col`	6	0	6
`oil`	34	0	34
`gas`	35	1	33
`p_c`	25	22	2
`ely`	43	56	-13

Table 2: Estimated implied targets for Russia, from the energy module (US$ billion)

	Producer revenue	Energy costs	Non-energy costs
`col`	6	0	6
`oil`	34	0	34
`gas`	35	1	33
`p_c`	25	22	2
`ely`	43	39	3

the well-known *sectoral balance condition*, not in its most common formulation, but in an equivalent formulation more convenient for present purposes). They also observe a sign constraint, that for each input and use, user cost is non-negative (besides other constraints not relevant here). From these constraints it follows that for each energy sector, the user cost of energy inputs must not exceed producer revenue (since the user cost of non-energy inputs must not be negative). If the implied target for the user cost of energy inputs exceeds the implied target for producer revenue, we have a *non-energy cost infeasibility.*

Examination of the "`xac`" file from the failed Russian '`fit`' run suggests the presence of a non-energy cost infeasibility for the electricity industry (the symptoms are very small changes in most variables in the later steps of the run, but large percentage decreases in non-energy inputs into electricity). Table 1 confirms this. As it shows, the implied producer revenue target exceeds the implied energy cost target for each energy industry except the electricity industry.

The problem arises despite preventative measures in the energy module. There the energy use targets are adjusted to eliminate infeasibilities, using separate estimates of the implied targets. After the adjustments, those estimates of course show no infeasibility (table 2).

Comparing the two sets of estimates, we see that although they generally agree, they differ greatly on the energy costs of the electricity industry. The energy module measures them at $39 billion, whereas our estimate of the Fit implied target is $56 billion.

Table 3: Implied targets for energy inputs into the Russian electricity industry, in the failed Fit run (US$ billion)

	Market value	Tax	User cost
`col`	4	0	4
`oil`	0	0	0
`gas`	19	14	33
`p_c`	6	1	7
`ely`	11	0	11
Total	39	17	56

The reason for the difference in the energy cost estimates is different assumptions about energy taxes. The energy module assumes zero taxes for energy inputs into the Russian electricity sector, but the Fit module uses the rates far different from zero. As table 3 shows, the Fit targets agree with the energy module on the market value of energy inputs, $39 billion; the discrepancy in energy costs is accounted for entirely by taxes.

How is it that the energy module and the Fit module can make different assumptions about tax rates, when the energy module supplies tax rate data to the Fit module? The explanation is that those data are incomplete. The energy module supplies not only arrays of tax rates but also arrays of data status flags; each component of a status flag array indicates whether to treat the corresponding component of the corresponding data array as a known or a missing value.

In undertaking the feasibility adjustments, where tax rates are missing, the energy module assumes zero taxes. But the Fit module fills in the missing values from the contributed IO tables. For the Russian electricity industry, the energy module reports missing values, but the contributed table supplies high rates of tax to the Fit module.

In the remainder of this section, we examine the handling of the commodity tax data between the feasibility adjustments in the energy module and the Fit step. More specifically, section 2.2 describes the tax data passed from the energy module to the Fit module; sections 2.3 to 2.7, the handling of the tax data in the later stages of the energy module; and section 2.8, their handling in the early stages of the Fit module. In section 2.9 we discuss the derivation of the feasibility data presented in tables 1 to 3; finally, section 2.10 presents conclusions.

2.2 Data flow between the energy and Fit modules

The energy module writes several files that are copied to and used by the Fit module, but we are concerned in this paper with just one of them, the energy cost data file '`ec.har`'. Of the arrays in this file, we are concerned with the arrays relating to money values and tax rates (table 4). We note:

- None of the arrays distinguish between domestic product and imports.

Table 4: Array structure of the energy cost data file '`ec.har`'

Header	Dimensions	Description
`VFM`	$r \times e \times i$	market value of intermediate usage of commodity %2 by industry %3 in region %1
`VPM`	$r \times e$	market value of household consumption of commodity %2 in region %1
`OUTS`	$e \times r$	rate of subsidy on production by industry %1 in region %2, as fraction of basic value
`TF?`	$e \times s \times r$	tax data status flag for intermediate usage of commodity %1 by industry %2 in region %3
`TF`	$e \times s \times r$	power of tax on intermediate usage of commodity %1 by industry %2 in region %3
`TP?`	$e \times r$	tax data status flag for household consumption of commodity %1 in region %2
`TP`	$e \times r$	power of tax on household consumption of commodity %1 in region %2

e energy sectors in energy target sectoral classification
i energy target sectors
r regions
s GTAP sectors

- All of the arrays use the *energy target sectoral classification* (discussed below) for their commodity dimension.
- Of the arrays pertaining to intermediate usage, the money value array uses the energy target sectoral classification for their industry dimension, but the tax power and tax data status arrays use the GTAP sectoral classification.

A point of notation: in the descriptions of the array contents, here and elsewhere in the paper, we use '%1' to refer to the first index in the array, '%2' to refer to the second index, and so on.

The energy target sectoral classification is an aggregation of the GTAP sectoral classification; for GTAP 5, the GTAP sectoral classification distinguishes 57 sectors, but the energy target sectoral classification only 22. The part that most concerns us in this paper is the classification of energy sectors. This is almost as detailed as the GTAP classification; the GTAP classification distinguishes six energy sectors, and the energy target classification, five (table 5). Specifically, it combines natural gas extraction, "`gas`", and gas distribution, "`gdt`", into a single sector "gas", "`gas`".

Tables 6 to 12 show the actual energy sector data transferred from the energy module to the Fit module. In the tax data status flag arrays "`TF?`" and "`TP?`", a value of 0 indicates that the corresponding datum is available, a value of 1 that it is not available. We note that, as shown in table 9, intermediate usage tax data are available for all

Table 5: Energy target classification of energy commodities

Code	Description	Corresponding GTAP commodities
col	coal	col
oil	crude oil	oil
gas	gas	gas gdt
p_c	petroleum and coal products	p_c
ely	electricity	ely

Table 6: Energy industries part of "rus" slice of array "VFM" in file 'ec.har' (market value of energy inputs into energy industries, US$ million)

Commodity	col	oil	gas	p_c	ely
col	0.636	0.055	2.510	0.007	4087.861
oil	0.001	0.980	0.049	21547.334	73.317
gas	0.002	0.052	1418.924	0.010	18769.086
p_c	0.052	0.321	0.109	646.664	5628.938
ely	1.572	7.989	2.170	3.951	10659.244

energy industries except electricity, for all energy inputs except electricity. For non-energy industries (not shown here) intermediate usage tax data are not available. We note also that where intermediate usage tax data are available, the power of the tax is for each component equal to one; that is, the tax rate is zero.

2.3 Commodity taxes in the energy module: Overview of later stage processing

We discuss here the later stages of the handling of commodity tax rates and tax data status flags in the energy module. Our object is to determine the relation between the data used in addressing non-energy cost infeasibilities and the data passed to the Fit module; in general, and in particular for Russia in the aborted GTAP 5.3 run.

In summary: the latter part of the energy module constructs energy cost and tax data to be passed to Fit (section 2.2) using energy quantity data from 'evol.har',

Table 7: "rus" slice of array "VPM" in file 'ec.har' (market value of household consumption of energy, US $ million)

col	692.062
oil	14.459
gas	3400.293
p_c	2249.861
ely	6637.641

Table 8: "rus" slice of array "OUTS" in file 'ec.har' (rate of production subsidy, fraction of basic value)

col	-0.032765
oil	-0.029952
gas	-0.033266
p_c	0.010989
ely	0.109605

Table 9: Energy industries part of "rus" slice of array "TF?" in file 'ec.har' (data status for power of tax on intermediate usage)

Commodity	col	oil	gas	p_c	ely	gdt
col	0	0	0	0	1	0
oil	0	0	0	0	1	0
gas	0	0	0	0	1	0
p_c	0	0	0	0	1	0
ely	1	1	1	1	1	1

0 data available
1 data not available

Table 10: Energy industries part of "rus" slice of array "TF" in file 'ec.har' (power of tax on intermediate usage)

Commodity	col	oil	gas	p_c	ely	gdt
col	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
oil	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
gas	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
p_c	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000
ely	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000

Table 11: "rus" slice of array "TP?" in file 'ec.har' (data status for power of tax on household consumption)

col	1
oil	1
gas	1
p_c	1
ely	1

0 data available
1 data not available

Table 12: “rus” slice of array “TP” in file ‘ec.har’ (power of tax on household consumption)

col	1.000000
oil	1.000000
gas	1.000000
p_c	1.000000
ely	1.000000

energy price and commodity tax data from ‘epreg.har’, and production tax data from ‘prot.har’. From these the program ‘ecimk’ constructs initial unbalanced estimates ‘ecini.har’ of tax-free and tax-paid money values of energy flows; program ‘epadj’ balances them to generate final estimates ‘ecmain.har’ (section 2.4); ‘mketax’ extracts tax rates from ‘ecmain.har’ and computes tax status flags from ‘evol.har’ and ‘epreg.har’, writing all to ‘etax.har’ (section 2.5); and a ‘modhar’ job assembles arrays from ‘ecmain.har’ and ‘etax.har’ to generate the module output file ‘ec.har’ (section 2.6).

The energy module constructs money values of energy usage by combining price and quantity data. In so doing, it must ensure that the implied targets for the non-energy costs of energy industries are positive. To this end, it adjusts prices; if, for example, the energy costs of the electricity industry exceed its total costs, it adjusts the electricity price upward so that total sales (and therefore total costs) exceed energy costs. It adjusts prices rather than quantities because the price data are considered more uncertain than the quantity data.

In adjusting prices, the energy module cannot proceed on a sector-by-sector basis, because one energy sector’s output prices are another’s input prices. For example, raising the price of petroleum products to avoid a cost infeasibility in the petroleum products industry may create an infeasibility in the electricity industry. It is necesary therefore to treat the price adjustment as a simultaneous problem.

The energy module deals with this by turning the infeasibility elimination problem into a balancing problem. It estimates output values for energy industries that are not necessarily consistent with the data for usage of the industries’ product, but are consistent with strictly positive non-energy costs. It then rebalances the data so as to eliminate inconsistencies between output and usage while preserving signs of non-energy costs.

The energy tax data are incorporated implicitly into the energy cost data set by including therein both tax-paid and tax-free values. This simplifies the sign conditions to be preserved in rebalancing. After rebalancing, tax rates are calculated from the tax-free and tax-paid cost data set.

2.4 Energy module: Energy cost data construction and balancing

The program ‘ecimk’ combines quantity data from ‘evol.har’, price and commodity tax data from ‘epreg.har’, production tax data from ‘prot.har’, and various other

Table 13: Array structure of '`ecini.har`' and '`ecmain.har`'

Header	Dimensions	Description
`MB`	$r \times e$	basic value of imports of commodity %2 into region %1
`OB`	$r \times e$	basic value of output of industry %2 in region %1, estimated from the costs side
`OB1`	$r \times e$	basic value of output of industry %2 in region %1, estimated from the sales side
`OP`	$r \times e$	producer revenue from output of industry %2 in region %1, estimated from the costs side
`VFM`	$r \times e \times i$	basic value of intermediate usage of commodity %2 by industry %3 in region %1
`VFA`	$r \times e \times i$	tax-paid value of intermediate usage of commodity %2 by industry %3 in region %1
`VPM`	$r \times e$	basic value of household consumption of commodity %2 in region %1
`VPA`	$r \times e$	tax-paid value of household consumption of commodity %2 in region %1
`XB`	$r \times e$	basic value of exports of commodity %2 from region %1

e energy sectors in energy target classification
i energy target sectors
r regions

data from other files, to produce initial, unbalanced energy cost estimates, which it writes to '`ecini.har`'.

Table 13 shows the array structure of '`ecini.har`'. The array "`OP`" contains estimates of producer revenue, or equivalently, of total industry input costs, that is, total costs excluding production taxes. By construction, these estimated total costs exceed the total energy costs that may be calculated from the tax-paid intermediate usage values "`VFA`". The cost-based estimates for basic values of output, "`OB`", are consistent with producer revenue values "`OP`" and the output tax rate estimates derived from '`prot.har`'. If the data are balanced, the cost-based estimates "`OB`" agree with the sales-based estimates "`OB1`"; by construction, these latter are consistent with the basic sales value data in "`VFM`", "`VPM`", "`XB`", and "`MB`".

The program '`epadj`', following commands '`ecmain.cmf`', eliminates the imbalances in the energy cost estimates in '`ecini.har`' to produce the final energy cost data file '`ecmain.har`'. It eliminates the imbalances by adjusting energy prices, more specifically, the prices applying to energy commodities that are both produced and absorbed domestically. It does not adjust import prices, and it does not adjust export prices—much. It does adjust export prices slightly, to avoid potential infeasibilities in balancing data for export-oriented sectors. The resultant changes in export money values are not brought to account in the trade data, but simply discarded. So there are some slight

Table 14: The "rus" slice of "OB" and "OB1" in 'ecini.har' and 'ecmain.har' (estimates for basic values of output, US$ billion)

	OB in 'ecini.har'	OB1 in 'ecini.har'	OB and OB1 in 'ecmain.har'
col	5.885	5.885	5.886
oil	34.932	34.932	34.938
gas	35.699	35.699	35.700
p_c	24.193	19.278	24.376
ely	30.420	12.133	38.387

differences between the export data supplied to the Fit module and those used in the feasibility calculations; this does not appear to have been harmful.

The balancing procedure follows an entropy-theory approach. It works on a country-by-country basis. Within each country, the price of each energy commodity is adjusted in response to infeasibilities both in the producing industry and in the using industries. Cost infeasibilities lead to upward adjustment in output prices and downward adjustment in input prices. Overall, energy prices tend to rise, because the input price adjustments are weighted by industry shares in total sales, and sales to non-energy industries generate no downward adjustment.

After balancing, the cost-based and sales-based output estimates "OB" and "OB1" are consistent. In general, their final value tends to be closer to the initial value of "OB" than to the initial value of "OB1", since most of the price adjustments are made on the output side not the input side. Indeed, because of positive feedback from output to input prices, the final output values may exceed the initial values of both "OB1" and "OB". Table 14 shows the outcomes for Russia: no initial imbalance and very little change for coal, oil or gas; but large imbalances and changes for petroleum products and electricity. The final estimate for petroleum products is close to the initial costs-side estimate; the final estimate for electricity, $38 billion, is considerably higher even than the initial costs-side estimate, $30 billion, because of increases in prices of petroleum products and electricity inputs into the electricity industry.

In preparing table 2, we used data from 'ecmain.har': "OP" for producer revenue, and "VFA" for energy costs.

We discuss 'epadj' further in section 3.9, especially its handling of prices.

2.5 Energy module: Computing tax data and tax data status flags

The program 'mketax' produces the final data for taxes, and writes them to a file 'etax.har'. It computes the tax data from tax-exclusive and tax-inclusive money values of flows, read from 'ecmain.har', and the tax data status flags from various data from 'evol.har' and 'epreg.har'.

In calculating the intermediate usage tax data status flags "TF?", the program relies mostly on an array of specific tax rates, "RUTX", read from the file 'epreg.har'. In this array, as in many other specific tax rate arrays in this module, a value of minus one is

Table 15: IEA use classes

Code	Description	Corresponding GTAP use classes
`I`	industrial	intermediate usage by industries other than electricity
`H`	household	private consumption
`U`	utilities	intermediate usage by the electricity industry
`X`	export	exports

taken not as an actual tax rate but as a missing value indicator. The array ranges over region, commodity, and use. The commodity index ranges over energy target sectors, the use index over the *IEA use classes* listed in table 15.

In general the program sets each component of "`TF?`" equal to one if the corresponding component of "`RUTX`" is equal to -1.0, but equal to zero otherwise. It uses the "`U`" slice of "`RUTX`" for the electricity industry, and the "`I`" slice for all other industries. But it applies two exceptions to this rule. One relates to the availability of price data, the other is a special rule for energy industries.

To apply the price data exception, the program reads domestic market shares "`DSHR`" from the file '`evol.har`' and basic prices "`RGBP`" from '`epreg.har`'. The prices "`RGBP`" are initially constructed as source-generic prices, but later in the energy module—notably in program '`ecimk`'—they are taken as pertaining to domestic product only. In this as in many other price arrays in this module, a value of zero is taken not as a price datum but as a missing value indicator.

Then the price data exception is this: where the domestic product market share "`DSHR`" is strictly positive, but the domestic product price "`RGBP`" is not, "`RUTX`" is set equal to one. To explain this, we note that in program '`ecimk`', the computation of *ad valorem* tax rates uses cross-source average prices; but when the price data exception applies, those prices and consequently those tax rates are unreliable.

The other exception is that regardless of the tax and price data in '`epreg`', the tax data status flags for non-electricity energy inputs into non-electricity energy industries are set equal to zero. This reflects the fact that regardless of those data, '`ecimk`' sets those rates equal to zero; so in '`ecini`' and downstream they are in effect known zeros.

2.6 Energy module: Final assembly of tax data

The file '`ec.har`' is assembled by '`modhar`' following stored input '`ec.sti`'. This copies the money value data arrays "`VFM`" and "`VPM`" from '`ecmain.har`', and the tax data and tax data status arrays "`OUTS`", "`TF?`", "`TF`", "`TP?`", and "`TP`" from '`etax.har`'.

2.7 Energy module: Conclusions

Tax data are marked unavailable in "`TF?`" under either of these circumstances (section 2.5):

1. the specific tax rate is unavailable from the "`RUTX`", and the component does not pertain to a non-electricity energy input into a non-electricity energy industry; or

2. the domestic market share "`DSHR`" is strictly positive, but the price "`RGBP`" is not available.

For Russia, the first circumstance applies.

How does '`ecimk`' compute *ad valorem* tax rates in such cases? If (as for Russia) the specific tax rate is unavailable from "`RUTX`", it applies a rate of zero. If the specific tax rate is available but the domestic product price "`RGBP`" is not, it uses a cross-country average price computed over those countries where prices are available (section 2.4). In either case, the tax rates used in addressing non-energy cost infeasibilities in '`epadj`' are unrelated to those applied later in '`fit`'.

We note that for export levies, '`ecimk`' does not use the energy data tax rate array "`RUTX`", but instead uses the protection data set array "`XSUB`" (from a file '`prot.har`' derived mainly from a file '`protectn.har`' imported from the protection module).

2.8 Data flow in the Fit module

In the Fit module, the multi-region file '`ec.har`' is split into single-region files '*reg*`_ec.har`', one for each region "*reg*". The array structure of these files is similar to that of the original '`ec.har`' (table 4), except of course that the arrays lose their region dimension.

For each region, the program '`ftshegy`' calculates energy-sector shocks from the region-specific energy targets file '*reg*`_ec.har`' and various other region-specific files; these it writes to a number of shock files. These and other shocks are applied by the program '`fit`' to generate a first-cut rebalanced IO table '*reg*`_tbb.har`'. A program '`postcorr`' then applies some minor post-balance corrections, to produce the final rebalanced table '*reg*`_tb.har`'.

To calculate shocks to powers of commodity tax on intermediate usage of energy commodities, the program '`ftshegy`' reads the data status flag and the target value of the power of the tax from arrays "`TF?`" and "`TF`" in file '*reg*`_ec.har`'. Since the commodity index in these arrays ranges over energy target sectors, but over GTAP sectors in the '`fit`' shocks, '`ftshegy`' maps the data status flags and tax rates from energy target commodities to GTAP commodities. Note also that these arrays apply to domestic product and imports alike, whereas '`ftshegy`' calculates shocks for these separately.

For each of domestic products and imports, for each GTAP energy commodity and GTAP industry, if the data status flag for the corresponding energy target commodity is equal to zero, '`ftshegy`' computes the shock as the percentage difference between the initial value of the power of the tax and the target value for the corresponding energy target commodity. But if the data status flag is equal to one, it sets the shock equal to zero; in effect then the initial value of the power of the tax is also the target value.

The initial IO table read by programs '`fit`' and '`ftshegy`' is a file '*reg*`_ii0.har`'. But the initial IO table supplied to the Fit module is a different file, '*reg*`_rbt.har`'. Between these two tables occur two intermediate steps. First, a program '`smearvfd`' smears intermediate usage of domestic product, and makes offsetting adjustments to household consumption and labor employment. Second, a variant of the fit program, '`iielim`', eliminates changes in stocks. The smearing program adjusts tax rates (and other data) only minutely, and the inventory investment elimination program preserves tax rates exactly (subject to solution error); so these are practically the same in '*reg*_ii0.har' as in '*reg*_rbt.har'.

In conclusion, the target for the power of the tax on intermediate usage in the Fit run is either that which can be read from array "TF" of file 'ec.har', if "TF?" in 'ec.har' is equal to zero, or that which can be calculated from the initial IO table '*reg*_rbt.har', if "TF?" is equal to one.

For output subsidies, 'ftshegy' reads target rates from array "OUTS" in file '*reg*_ec.har'. As noted above (table 4), this array uses the energy target sectoral classification for the industry dimension. But the shocks to the power of the output tax in 'fit' are indexed by GTAP sector. So 'ftshegy' maps target rates from energy target industries to GTAP industries.

2.9 Problems in identifying implied targets

In discussing infeasibilities and implied targets in the Fit module (section 2.1), we have glossed over a problem, that not all the relevant implied targets are well defined. More specifically, the producer revenue targets reported in table 1 are well defined, but the energy cost targets are not. More specifically again, the energy cost targets may or may not be well defined, depending on circumstances.

The problem arises because the Fit module uses two sectoral classifications for explicit targets, the GTAP and the energy target classifications. It uses the GTAP classification for export and import values and for powers of commodity and production taxes. But it uses the energy target classification for money values of intermediate usage and household consumption of energy products. Furthermore, the tax targets discriminate domestic products and imports, but the usage targets do not.

Because 'fit' uses the energy target classification for energy usage values, we cannot define implied targets for total sales, total costs, or producer revenue of energy sectors at the GTAP sectoral level. Accordingly, in table 1, we use the energy target sectoral classification.

Using the energy target sectoral classification, we have well defined targets for total sales and therefore for total costs. To derive thence a target for total revenue, we need also targets for output tax rates by energy target industry. At first this might seem a problem, since the 'fit' targets are by GTAP industry. But by construction (section 2.8), these targets are uniform across energy target industries (that is, are the same for GTAP industries "gas" and "gdt"), so tax rates at the energy target industry level are well defined.

On the cost side however there are real problems. For each energy usage target, between two and eight distinct commodity tax rate targets apply. There are separate tax rate targets for domestic products and imports; for the energy target commodity "gas", there are separate tax rate targets for GTAP commodities "gas" and "gdt"; and similarly, for the industry "gas" there are separate tax rate targets for the industries "gas" and "gdt".

From these two to eight tax rate targets we can derive a single overall target under either of two conditions: if we know the weights to apply in calculating the overall target as an average of the individual targets, or if the individual targets are equal. But the required weights are usage value targets at the same level as the tax rate targets, and these as we have seen do not exist. So the overall target rates are well defined only if the individual target rates are equal.

Now if the individual target rates derive from the energy module, then indeed they are equal (section 2.8). They are uniform across domestic products and imports, and across GTAP commodities corresponding to the same energy target commodity, because of the shock construction procedure in '`ftshegy`' in Fit (section 2.8). And they are uniform across GTAP industries corresponding to the same energy target industries because of the target construction procedure in '`mketax`' in the energy module (section 2.3). But if they derive not from the energy module but instead from the IO table, then nothing in the construction procedure guarantees equality. And for the electricity industry in Russia, this is the case.

As it happens the problem for Russia is not serious (table 16). For most energy inputs into energy industries, the rates of tax on domestic product and imports are similar. There are indeed some exceptions, notably for the GTAP commodity "`gdt`". But these relate to very small import flows; we conjecture that they arise from IO table smearing in the robustification module, not from the contributed table. Rates for the GTAP commodities "`gas`" and "`gdt`" are also very similar. There are large differences between the corresponding two GTAP industries; but for these industries, '`ftshegy`' sources the target rates from the energy module not the IO table (table 9). Furthermore, as tables 1 and 6 show, target energy inputs into the corresponding energy target industry, "`gas`", are small enough to prevent any danger of infeasibility.

In the end then, the conclusions concerning the sources of infeasibility for the Russian table arrived at above (section 2.1) remain secure.

2.10 Conclusions

From the foregoing, we find that there is indeed a non-energy cost infeasibility (section 2.1) preventing us from fitting the Russian table, and that the source of this infeasibility lies in the commodity tax data, and more specifically in the combination of high tax rates in the IO table, missing data flags for energy inputs into the electricity industry in the tax rate target data, and the unsuccessful attempt in the energy module to eliminate non-energy cost infeasibilities before tax rates are fully determined. To remedy this, we need to change the source data input into the energy module, so as to remove those tax data flags. In so doing it is of course desirable to align the new target tax rates with the tax data in the contributed IO table; and to do that, we need a better knowledge of the early stages of the energy module.

3 Prices and commodity taxes in the energy module

From section 2 it is clear that the source of the fitting problems with Russia lies in the commodity tax data, and that the remedy lies in adjusting those data. That section also describes the flow of the tax rate data through the energy and Fit modules, from the construction of the initial energy cost estimates '`ecini.har`' in the energy module

Table 16: Taxes on energy inputs into energy industries in Russia, from contributed table (fraction of tax-free value)

		Industry					
Commodity	Source	col	oil	gas	gdt	p_c	ely
col	d	0.0196	-0.0082	0.0898	0.0379	-0.0082	0.0540
	m	0.0196	-0.0082	0.0870	0.0395	-0.0080	0.0540
oil	d	0.0159	0.1030	0.1515	0.1918	0.1194	0.1084
	m	0.0111	0.1031	0.1516	0.1912	0.1194	0.1084
gas	d	0.7296	0.6414	0.4107	0.2668	0.6415	0.7712
	m	0.7306	0.6415	0.4106	0.1041	0.6414	0.7472
gdt	d	0.7295	0.6413	0.4107	0.2567	0.6414	0.7704
	m	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
p_c	d	0.2224	0.1604	0.2412	0.3054	0.1217	0.2618
	m	0.2224	0.1604	0.2412	0.3054	0.1217	0.2618
ely	d	0.0243	0.0001	0.0268	0.1945	0.0000	0.0353
	m	0.0243	0.0001	0.0268	0.1950	0.0000	0.0353

d domestic production
m importation

to the calculation of the energy sector shocks into the 'fit' run in the Fit module. It identifies the file 'epreg.har' as the main immediate source both for the tax rates and for the tax data flags. In this section we examine the derivation of the tax data in 'epreg.har' from the initial data inputs into the energy module. We also examine some further aspects of the later energy module programs that would come into play with new source data.

Although the target tax rates eventually recorded in 'ec.har' are expressed as *ad valorem* rates, the tax data inputs into the energy module are expressed as specific rates. In deriving *ad valorem* rates from specific, the energy module requires price data. These price data are processed by the energy module in parallel with the tax data; they originate in the same data input files, and are processed by the same programs, and pass through the same intermediate data files. So we examine the derivation of the price data and the tax rate data together.

Previous documentation on energy data construction is available in two GTAP working documents, *Documentation for Energy Price/Tax Database* by Gerard Malcolm and Mustafa Babiker (1998), and *The Process of Incorporating Energy Data into GTAP* by Gerard Malcolm and Truong Phuoc Truong (1999). These are both accurate and extensive; but nonetheless insufficient for our purposes, since they do not contain all the detail we need, they do not of course document the Russian data in the aborted GTAP 5.3 run, and they do not reflect subsequent revision of the energy module (early-stage by Truong Phuoc Truong and later-stage by Jean-Marc Burniaux and Robert McDougall). They nevertheless remain valuable as an alternative exposition, as recording the views of the modules' original authors, and for their coverage of aspects of less concern to us here, including notably a much more detailed account of the source data.

3.1 Data flow overview

The input data for prices and taxes are read from several files, but principally from a file 'p_iea.har' of user prices and specific tax rates derived from IEA statistics. This file contains not a single array but several arrays for user prices, and likewise for tax rates. Each array covers a single (commodity, use) pair, where uses follow the IEA use classification (section 2.5), and commodities an *IEA product classification* (table 17). This is a set of 9 commodities that map to the 5 energy target commodities.

Each of these arrays in 'p_iea.har' is indexed by year and country. The country ranges are specific to each array, but are all subsets of a *price data country set*, a subset of 91 countries from the standard GTAP country set, hard-coded in a program 'ep', and recorded in an array "CTY" in 'eporig.har'. Russia is included. The year ranges are hard-coded, and are 1990–95 for the tax rate arrays but 1990–97 for the price arrays. The tax rates are given in local currency unit per original physical unit, where the local currency units are country-specific, and the original physical units product-specific; but the prices are all given in US$ per TOE.

Initial assembly, filling in and adjustment is done by program 'ep' (sections 3.2 to 3.6), writing to a file 'eporig.har'. This contains an array of basic prices, "CBP", and an array of specific tax rates, "CTAX", both in US$ per TOE. These arrays are indexed by price data country and by energy target energy commodity (section 2.2). The tax rates

Table 17: IEA products

Code	Description	Corresponding energy target commodity
S	steaming coal	col
C	coking coal	col
O	crude oil	oil
N	natural gas	gas
H	heavy fuel oil	p_c
L	light fuel oil	p_c
G	gasoline	p_c
D	diesel oil	p_c
E	electricity	ely

"CTAX" are also indexed by IEA use class.

Program 'epreg' (section 3.7) transforms the price and tax data in 'eporig.har', mapping from price data countries to GTAP regions. It writes the region-level price and tax data to 'epreg.har'.

Program 'ecimk' (section 3.8) combines price and tax data from 'epreg.har' and quantity data from a file 'evol.har' to construct initial unbalanced energy cost estimates 'ecini.har'. These contain implied *ad valorem* tax rates for intermediate usage and household consumption. In constructing these estimates, 'ecimk' expands the use dimension of the 'epreg.har' tax data from IEA use classes to energy target industries.

Program 'epadj' (section 3.9) adjusts prices to balance the cost estimates, maintaining specific tax rates and allowing *ad valorem* rates to vary, and writes the balanced cost estimates to 'ecmain.har'. Program 'mketax' (section 2.5) calculates commodity tax rates from tax-free and tax-paid usage values in 'ecmain.har', expands the industry dimension for taxes on intermediate usage from energy target sectors to GTAP sectors, and writes them to 'etax.har'. These then are assembled by 'modhar' with other data to form the main output file from the energy module, 'ec.har' (section 2.6).

Table 18 summarizes this main line of data flow for prices and taxes in the energy module.

A second line of processing derives from the need for consistency between the energy data and the trade data. The energy module achieves this by reading the initial GTAP trade data set 'trade.har', revising it to match part-processed energy data from 'epreg.har', and thenceforward maintaining consistency with the revised trade data 'tradee.har'.

Accordingly, although the basic price data in 'eporig.har' and 'epreg.har' pertain to domestic product and imports alike, program 'ecimk' applies them to domestic product only. For imports, it seeks prices compatible with the revised trade data 'tradee.har'; it then computes source-generic prices as an average of domestic-product and import prices. As described in section 3.10, this leads to cross-country price interdependencies: each country's import prices in 'ecimk' depend not on its own price data in 'p_iea.har', but on its trading partners' export price data.

Table 18: Data flow overview for commodity taxes in energy module

File	Written by	Data
`p_iea`	...	(IEA product, IEA use)-specific arrays indexed by array-specific country set and year: • user prices in US$ per TOE • specific tax rates in local currency unit per original physical unit
`eporig`	`ep`	data in US$ per TOE, indexed by energy target commodity and price data country: • basic prices • specific tax rates, indexed additionally by IEA use class
`epreg`	`epreg`	data in US$ per TOE, indexed by energy target commodity and GTAP region: • basic prices • specific tax rates, indexed additionally by IEA use class
`ecini`	`ecimk`	data in US$ per year, indexed by energy target commodity and GTAP region: • intermediate usage at tax-free prices, indexed additionally by energy target industry • intermediate usage at tax-paid prices, indexed additionally by energy target industry • household consumption at tax-free prices • household consumption at tax-paid prices
`etax`	`mketax`	powers of taxes, indexed by energy target commodity and GTAP region, on: • intermediate usage, indexed additionally by GTAP industry • household consumption

Table 19: Sources for “`UUPRC`” and “`UUTAX`”

	USR			
PRD	I	H	U	X
S	S*N	S*H	S*U	...
C	C*N	...	...	...
O	...	...	...	...
N	N*N	N*H	N*U	...
H	H*N	...	H*U	...
L	L*N	L*H	...	...
G	...	G*H	...	...
D	D*N	D*H	...	...
E	E*N	E*H	...	...

arrays in ‘p_iea.har’
‘...’ indicates ‘no array’
For UUPRC, replace ‘*’ in array name with ‘P’;
for UUTAX, replace with ‘T’.

Some special features of the data apply from the initial data input files for the energy module through to ‘`epreg`’. Zeros in the price arrays typically indicate missing values; in the tax arrays, zeros mean zero, and missing values are indicated by the value -1.0. Aside from these missing value indicators, the tax rate arrays are non-negative.

3.2 Initial handling: Overview

The program ‘`ep`’ assembles price and tax data from the energy module's data input files, performs various transformations and adjustments, and writes the results to ‘`eporig.har`’.

The input data are read from many files, but principally a file ‘`p_iea.har`’ of IEA user prices and specific tax rates. This file contains not a single array but several arrays for prices, and likewise for tax rates. Each array covers a single IEA product and IEA use class; but many (product, use) pairs are not covered. Whether by design or accident, the (product, use) pair coverage is the same for taxes as for prices (table 19). Each array is indexed by year and country. The country ranges are specific to each array, and are read from a file ‘`p_ieaset.har`’, where they are stored in arrays of the same name as the corresponding data arrays in ‘`p_iea.har`’. All the country ranges are subsets of the price data country set. The tax rates are given in local currency unit per original physical unit, where the local currency units are country-specific, and the original physical units product-specific; but the prices are all given in US$ per TOE.

The data handling procedures for prices and tax rates are broadly similar:

- The product- and use-specific source arrays are assembled into arrays indexed over product, country, use, and year: “`UUPRC`” for prices and “`UUTAX`” for taxes.
- The tax data only are converted to from original to standard units—from local

currency unit per original physical unit to US$ per TOE—and stored in an array "AUTAX".

- The multi-year arrays "UUPRC" and "AUTAX" are reduced to arrays without a year index, "UPRC" and "UTAX", by selecting data from the most recent available year; for the price data only, time series price indices are applied to compensate for differences in data reference year.
- User prices "UPRC" are converted to use-specific basic price estimates "BPE", taking account of margin rates "MARG" and tax rates "UTAX".
- The commodity dimension of the price and tax arrays "BPE" and "UTAX" is aggregated from IEA product to energy target commodity. At the same time the use dimension of the price array is eliminated. The results are written to arrays "CBP" and "CTAX" in 'eporig'.

There are also some filling-in or adjustment steps applying only to prices or only to taxes, not covered in this list.

Sections 3.3 and 3.4 describe the initial assembly and construction of user price and basic price data in program 'ep'. Section 3.5 does the same for the tax data. As they show, although the tax rates and prices are handled mostly in parallel, there is some limited direct interaction between them. There is also some weak indirect interaction, through the margin rates "MARG"; as section 3.6 shows, this creates cross-country dependencies in the price data. This is in addition to cross-country dependencies arising from filling-in. Figure 1 summarizes the data flow within 'ep'.

3.3 Initial assembly: User prices

The program 'ep' defines a user price array "UUPRC", measured in US dollars per TOE, and indexed by energy product, country, IEA use class, and year. It first initializes all components of "UUPRC" to zero. It then reads country ranges for (product, use) pairs from 'p_ieaset.har', and associated data arrays from 'p_iea.har'.

The program also defines a price index array "INDEX", based at 1990 = 100, in domestic currency units, indexed over energy product, country, and year. First, all components of the 1990 slice are set equal to 100. Then, available data are read from various arrays in a file 'index.har'. Several different product-specific arrays are read, but the product classification is an aggregation of the IEA energy product classification. The country coverage is incomplete; a separate covered country set is provided for each source data array. Finally, missing values are supplied by using the US price as a proxy, and applying exchange rates "USXR".

The program then reduces the multi-year price array "UUPRC" to an array with no year dimension, "UPRC". It takes the component of "UUPRC" from the most recent available year, and then updates it to 1997 by applying local currency price indices "INDEX" and US exchange rates "USXR". Table 20 shows the Russian price data at this point.

Then follow adjustments for various countries for which the IEA provides export spot prices. One of these concerns Russia: export prices for crude oil are obtained

Figure 1: Data flow for prices and tax rates in ‘`ep`’

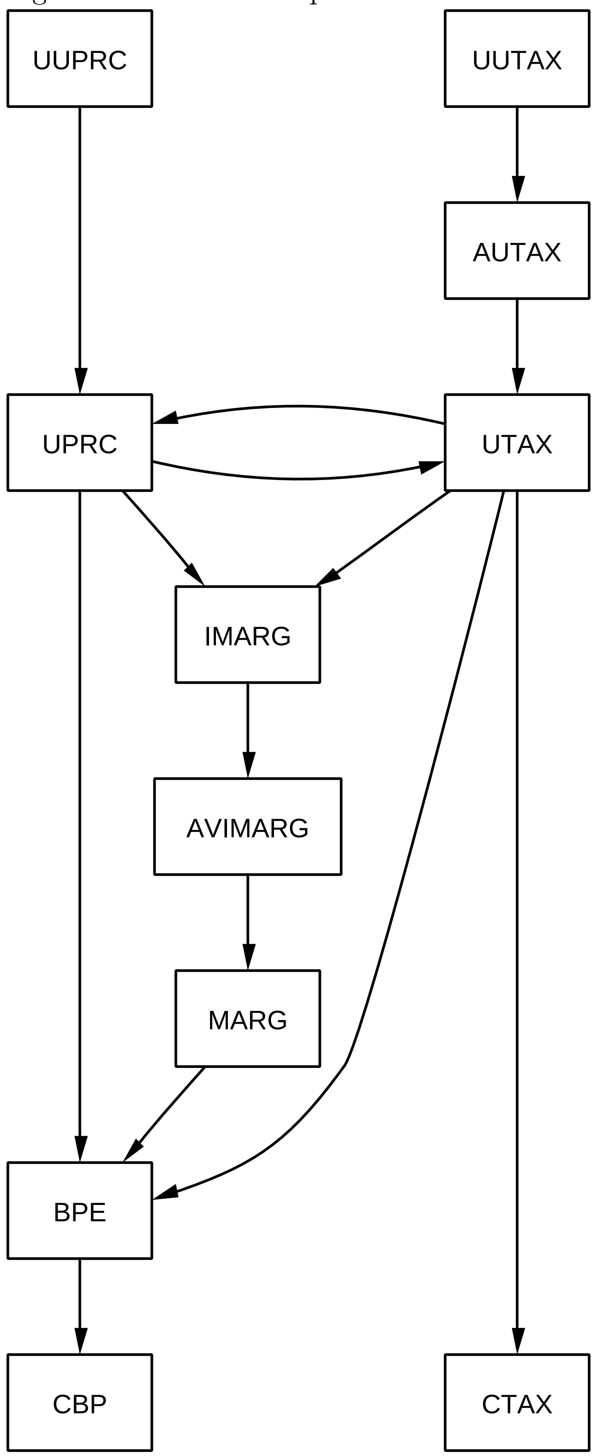

Table 20: "`rus`" slice of "`UPRC`", after initial reads (user prices, US$ per TOE)

	USR			
PRD	I	H	U	X
S	0.0000	0.0000	0.0000	0.0000
C	0.0000	0.0000	0.0000	0.0000
O	0.0000	0.0000	0.0000	0.0000
N	0.0000	0.0000	0.0000	0.0000
H	34.9098	0.0000	34.9098	0.0000
L	82.5350	98.9495	0.0000	0.0000
G	0.0000	123.4557	0.0000	0.0000
D	95.7129	0.0000	0.0000	0.0000
E	223.1053	35.0905	0.0000	0.0000

from an array "`OPX`" in '`p_iea.har`'. For Russia, this yields `UPRC("O","rus","X") = 124.1225`.

Then follow adjustments that do not concern Russia, for various south and central American countries covered in an OLADE document.

The program '`ep`' then applies price data from a DOE publication. The data cover 86 countries including Russia, but only five (product, use) pairs: `("H","I")`, `("L","I")`, `("D","I")`, `("L","H")`, and `("G","H")`; and only one year, 1995. For the covered countries and (product,use) pairs, the program fills in missing values of "`UPRC`", then adjusts them with physical conversion factors and price indices. Since none of these values are missing from the Russian slice, none are filled in here.

Then follow several adjustments that do not affect Russia: for various countries covered in an Asian Development Bank document, for China, for various countries covered in a World Bank survey of Asian energy prices, and for India.

The program "`ep`" then categorizes countries as high-price energy exporters, high-price energy importers, low-price energy exporters, and low-price energy importers; Russia is categorized as a low-price exporter. Within each group, it calculates cross-country average prices; for low-price exporters, these are stored in a coefficient "`LPEXAVPRC`". The average prices are calculated as simple averages of the available non-zero components "`UPRC`". The program then uses them to fill in missing values in "`UPRC`". For Russia, this fills in just one price, `UPRC("O","rus","I") = 272.7944`.

It is at this point in its construction that "`UPRC`" is compared with commodity tax rates "`UTAX`", and "`UTAX`" adjusted to ensure that it does not exceed "`UPRC`" (section 3.5).

The program '`ep`' then fills in any still-missing values in the export slice of "`UPRC`". First, it replaces zeros in the export slice of "`UPRC`" with the corresponding components of the cross-country average price "`AVPRICE`". Then, if any components are still zero (in practice, this occurs for diesel oil "`D`" and electricity "`E`") it sets the export price equal to the industrial price less tax; using the industrial price as a proxy for the export price, but assuming that the tax rate for exports is zero. This is done after adjusting commodity tax rates "`UTAX`" to ensure that they do not exceed "`UPRC`". If the "`UTAX`"

Table 21: "`rus`" slice of "`UPRC`", after filling in export prices (user prices, US$ per TOE)

	USR			
PRD	I	H	U	X
S	92.9032	58.4257	89.8256	54.4194
C	33.3582	0.0000	0.0000	66.5068
O	272.7944	0.0000	0.0000	124.1225
N	93.6423	217.9793	135.9953	112.0428
H	34.9098	0.0000	34.9098	113.2057
L	82.5350	98.9495	0.0000	152.7430
G	0.0000	123.4557	0.0000	202.0023
D	95.7129	391.5232	0.0000	95.7129
E	223.1053	35.0905	0.0000	223.1053

Table 22: "`rus`" slice of "`UPRC`", after filling in with cross-country average prices (user prices, US$ per TOE)

	USR			
PRD	I	H	U	X
S	92.9032	58.4257	89.8256	54.4194
C	33.3582	71.1429	71.1429	66.5068
O	272.7944	121.7220	121.7220	124.1225
N	93.6423	217.9793	135.9953	112.0428
H	34.9098	146.3001	34.9098	113.2057
L	82.5350	98.9495	347.8018	152.7430
G	202.0021	123.4557	202.0021	202.0023
D	95.7129	391.5232	446.4728	95.7129
E	223.1053	35.0905	804.4903	223.1053

entry is a missing value, the calculation assumes a zero rate. After these adjustments, the Russian slice of "`UPRC`" is as shown in table 21.

At this point, '`ep`' uses the user prices "`UPRC`" to fill in missing values in import margins "`IMARG`" and to calculate the use-specific basic price estimates "`BPE`" (section 3.4).

The program '`ep`' then fills in zeros in "`UPRC`" with cross-country average prices "`AVUPR`". The "`AVUPR`" are constructed as simple cross-country averages of the non-zero "`UPR`". Apparently because for some (product, use) pairs no non-zero "`UPR`" components are available, some components of "`AVUPR`" are then set equal to others; for example, the average price of coal in household consumption is set equal to the average price in industrial intermediate usage (`AVUPR("C","H") = AVUPR("C","I")`). After filling in, the Russian slice of "`UPRC`" is as shown in table 22.

At this point, with all components of "`UPRC`" filled in, '`ep`' averages "`UPRC`" over energy products within energy target commodities to construct the array "`CUP`", user price of energy by energy target commodity, country, and IEA use class. It is from

Table 23: Data sourcing for “rus” slice of “UPRC”

	I	H	U	X
S	0	0	0	3
C	0	4	4	3
O	2	4	4	1
N	0	0	0	3
H	0	4	0	3
L	0	0	4	3
G	4	0	4	3
D	0	0	4	3
E	0	0	4	3

0 read from main IEA data set
1 read from IEA spot price data set
2 filled in from low-cost exporting country average “LPEXAVPRC”
3 filled in from cross-country average “AVPRICE”
4 filled in from cross-country average “AVUPR”

these prices that the program ‘epreg’ later calculates FOB export prices “FOBP” (section 3.10).

Table 23 shows the various methods used to construct individual elements of the “rus” slice of “UPRC”.

3.4 Initial construction: Basic prices

Initially, values are assigned to “BPE” from an import-price-based estimate “BPI”. The “rus” slice of “BPI” contains only zeros. So at this point, the “rus” slice of coefficient “BPE” contains only zeros.

If no non-zero value has yet been assigned, “BPE” is calculated as user price “UPRC” less specific tax rate “UTAX” deflated by a margin rate “MARG”. But if “UTAX” holds a missing value flag, then “BPE” is calculated as user price “UPRC” deflated according to the margin rate. This is done after the positive-basic-price adjustment to “UTAX” (section 3.5). At this point, the “rus” slice of “UTAX” contains only missing value flags, so for Russia, “BPE” is just user price “UPRC” deflated by the margin rate. This yields the estimates shown in table 24.

If no non-zero value has yet been assigned, “BPE” is set equal to a cross-country average price “AVBPE” (table 25). The coefficient “AVBPE” is calculated as a simple cross-country average of the available non-zero “BPE” components. After filling in, the “rus” slice of “BPE” is as shown in table 26.

With all components of “BPE” now filled, ‘ep’ averages them across uses, and across IEA products within energy target commodities, to obtain use-generic energy-target-commodity-indexed basic price estimates “CBP”. The averages are calculated using volume weights “TPRODWT” (table 27 shows the “rus” slice). The results are written to

Table 24: Coefficient "`BPE`" in program '`ep`', before filling in from cross-country averages

	I	H	U	X
S	68.96	13.14	70.32	50.58
C	29.25	0.00	0.00	61.33
O	99.95	0.00	0.00	121.09
N	51.08	72.66	104.48	102.14
H	24.72	0.00	25.57	107.56
L	42.18	67.75	0.00	119.18
G	0.00	67.06	0.00	156.97
D	52.15	212.66	0.00	94.62
E	223.11	35.09	0.00	223.11

use-specific basic price estimates, US$ per TOE

Table 25: Coefficient "`AVBPE`" in program '`ep`'

	I	H	U	X
S	62.78	40.90	57.52	52.74
C	62.28	79.10	79.10	62.94
O	141.22	144.25	144.25	125.50
N	109.56	115.74	117.06	102.25
H	101.39	140.21	86.21	111.14
L	152.93	151.82	156.94	123.72
G	212.96	248.17	212.96	165.09
D	180.91	266.16	273.18	325.70
E	773.37	922.65	773.37	773.37

use-specific basic price estimates, US$ per TOE

Table 26: Coefficient "`BPE`" in program '`ep`', after filling in from cross-country averages

	I	H	U	X
S	68.96	13.14	70.32	50.58
C	29.25	79.10	79.10	61.33
O	99.95	144.25	144.25	121.09
N	51.08	72.66	104.48	102.14
H	24.72	140.21	25.57	107.56
L	42.18	67.75	156.94	119.18
G	212.96	67.06	212.96	156.97
D	52.15	212.66	273.18	94.62
E	223.11	35.09	773.37	223.11

use-specific basic price estimates, US$ per TOE

Table 27: "`rus`" slice of coefficient "`TPRODWT`" in program '`ep`'

	USR			
PRD	I	H	U	X
S	0.038264	0.167326	0.587209	0.106815
C	0.000000	0.000000	0.052234	0.048152
O	0.009398	0.000459	0.000000	0.990143
N	0.140888	0.116619	0.206008	0.536485
H	0.027872	0.027217	0.000000	0.263369
L	0.036678	0.015452	0.084660	0.245188
G	0.062683	0.136718	0.000000	0.025823
D	0.074339	0.000000	0.000000	0.000000
E	0.624078	0.295100	0.000000	0.080822

(product, use) shares in volume of energy usage by GTAP commodity

Table 28: "`rus`" slice of array "`CBP`" in file '`eporig`'

	value (US$/TOE)	share derived from "`AVBPE`"
col	58.62	0.05
oil	120.90	0.00
gas	91.99	0.00
p_c	108.38	0.15
ely	167.62	0.00

use-generic basic price estimates

array "`CBP`" in '`eporig.har`'.

One implication of tables 24 and 27 is that most of the weight in calculating "`CBP`" comes from (product, use) pairs for which data are available before filling in with "`AVBPE`" (table 28).

3.5 Initial assembly: Tax rates

The initial tax rate array, "`UUTAX`", is indexed by commodity, country, use, and year. All elements are first set to the missing value indicator, -1.0; then commodity- and use-specific slices are read in from various arrays in '`p_iea.har`', as shown in table 19. The country ranges for the various source arrays are taken from from '`p_ieaset.har`'. The set of years is hard-coded as 1990–95. As noted above (section 3.2), the rates are given in local currency unit per original physical unit, where the local currency units are country-specific, and the original physical units product-specific. Coverage of Russia is as shown in table 29.

After the main data arrays are read in, some missing elements of "`UUTAX`" are supplied from other sources. Specifically, data for certain American countries, drawn

Table 29: Data availability for “rus” in source arrays for “UUTAX”

	USR			
PRD	I	H	U	X
S	1	0	1	2
C	1	2	2	2
O	2	2	2	2
N	1	0	1	2
H	1	2	1	2
L	1	0	2	2
G	2	0	2	2
D	0	1	2	2
E	1	0	2	2

0 data for “rus” in array
1 no data for “rus” in array
2 no array

from an OLADE publication, are read from ‘p_ola.har’; and two data points for India are hard-coded.

The assembly of the initial tax rate array “UUTAX” is now complete. For “rus”, the most recent non-missing values from “UUTAX” are 1994 values for several components of household consumption (table 30), and 1993 values for industry usage of diesel oil and household consumption of steaming coal (table 31).

The tax rates in “UUTAX” are next converted from local currency units to US dollars, and from original physical units to TOE, and written to an array “AUTAX”. A separate formula is applied for each IEA product. Each formula involves “unadjusted user tax” “UUTAX”, a United States exchange rate “USXR”, and zero or more physical conversion factors.

Most of the formulas apply an physical unit conversion factor “DCONFAC”. This factor notionally applies to domestic product only; it is however copied from another coefficient “CONFAC” in which no distinction is made between domestic products and imports. The formulas fall into several groups.

- No formula is provided for product “O”, crude oil. (As shown in table 19, no source data are read from ‘p_iea.har’ for crude oil; neither are any read from ‘p_ola.har’; but one of the two data points hard-coded for India pertains to crude).

- For natural gas “N”, no energy unit conversion factor is applied, but only the currency conversion factor “USXR”.

- For electricity there are applied both the currency exchange factor “USXR” and, to convert from gigawatt hours, an energy unit conversion factor “EGYCF”.

Table 30: `("rus", "y94")` slice of "`UUTAX`"

	USR			
PRD	I	H	U	X
S	-1	-1	-1	-1
C	-1	-1	-1	-1
O	-1	-1	-1	-1
N	-1	873	-1	-1
H	-1	-1	-1	-1
L	-1	37463	-1	-1
G	-1	73	-1	-1
D	-1	-1	-1	-1
E	-1	2	-1	-1

Table 31: `("rus", "y93")` slice of "`UUTAX`"

	USR			
PRD	I	H	U	X
S	-1	165	-1	-1
C	-1	-1	-1	-1
O	-1	-1	-1	-1
N	-1	232	-1	-1
H	-1	-1	-1	-1
L	-1	10940	-1	-1
G	-1	25	-1	-1
D	14	-1	-1	-1
E	-1	0	-1	-1

Table 32: `("rus", "y94")` slice of "`AUTAX`"

	USR			
PRD	I	H	U	X
S	-1	-1	-1	-1
C	-1	-1	-1	-1
O	-1	-1	-1	-1
N	-1	246	-1	-1
H	-1	-1	-1	-1
L	-1	12291	-1	-1
G	-1	24409	-1	-1
D	-1	-1	-1	-1
E	-1	7243	-1	-1

Table 33: `("rus", "y93")` slice of "`AUTAX`"

	USR			
PRD	I	H	U	X
S	-1	210	-1	-1
C	-1	-1	-1	-1
O	-1	-1	-1	-1
N	-1	186	-1	-1
H	-1	-1	-1	-1
L	-1	10204	-1	-1
G	-1	24021	-1	-1
D	12316	-1	-1	-1
E	-1	4193	-1	-1

- For steaming coal "`S`", coking coal "`C`", and heavy fuel oil "`H`", both the exchange rate "`USXR`" and the quantity unit conversion factor "`DCONFAC`" are applied. For these products, "`CONFAC`" is TOE per tonne: 0.6300 for steaming coal, 0.7090 for coking coal, and 0.8591 for heavy fuel oil.
- For light fuel oil "`L`", gasoline "`G`", and diesel oil "`D`", there are applied not only the currency and quantity conversion factors "`USXR`" and "`DCONFAC`", but also a volume conversion factor "`VOLCF`".

After these conversions, the most recent non-missing values of the rates array for Russia are as shown in table 32 for 1994 and table 33 for 1993.

The multi-year array "`AUTAX`" is now reduced to an array "`UTAX`" with no year index. The most recent non-missing value from "`AUTAX`" is assigned to "`UTAX`". Table 34 shows the Russian slice.

The array "`UTAX`" is now adjusted to ensure that tax rates do not exceed user

Table 34: "`rus`" slice of "`UTAX`": initial values

	I	H	U	X
S	-1	210	-1	-1
C	-1	-1	-1	-1
O	-1	-1	-1	-1
N	-1	246	-1	-1
H	-1	-1	-1	-1
L	-1	12291	-1	-1
G	-1	24409	-1	-1
D	12316	-1	-1	-1
E	-1	7243	-1	-1

Table 35: "`rus`" slice of user price "`UPRC`" when positive basic price condition is checked

	I	H	U	X
S	93	58	90	0
C	33	0	0	0
O	273	0	0	124
N	94	218	136	0
H	35	0	35	0
L	83	99	0	0
G	0	123	0	0
D	96	392	0	0
E	223	35	0	0

prices. The adjustment is applied to flows for which "`UPRC`" is strictly positive (i.e., not a missing value) and "`UTAX`" exceeds "`UPRC`": in such cases, the tax rate is changed to the tax rate missing value indicator, -1.0. This is done when some but not all missing values in "`UPRC`" have been filled in (section 3.3); the "`rus`" slice of "`UPRC`" is then as shown in table 35. Remarkably enough, this adjustment converts all non-missing elements of the Russian slice to missing value indicators, but no elements pertaining to other countries. The enormous disproportion between the tax rates and prices for some flows (`("D","I")`, `("L","H")`, `("G","H")`) suggests some gross error in the initial input data.

From this point on, the Russian slice of the specific tax rate array consists entirely of missing value indicators.

Next '`ep`' aggregates from IEA products to energy target commodities. The aggregate rates are calculated as weighted averages over IEA products within energy target commodities, using quantity weights "`UPRODWT`" (table 36), and are written to an array "`CTAX`" in '`eporig.har`'. Missing values in the data input require some care. The treatment adopted is such that, for each energy target commodity:

- If some corresponding IEA product has both an actual data value and a non-zero

Table 36: "rus" slice of "UPRODWT"

	I	H	U	X
S	1.00	1.00	0.92	0.69
C	0.00	0.00	0.08	0.31
O	1.00	1.00	1.00	1.00
N	1.00	1.00	1.00	1.00
H	0.14	0.15	0.00	0.49
L	0.18	0.09	1.00	0.46
G	0.31	0.76	0.00	0.05
D	0.37	0.00	0.00	0.00
E	1.00	1.00	1.00	1.00

weight "UPRODWT", then the average is calculated over all the corresponding IEA products, with missing value indicators replaced by rates of zero.

- Otherwise, the energy target commodity has a missing value indicator.

3.6 Initial construction: Margin rates

As described in section 3.4, margin rates "MARG" are used in converting user prices "UPRC" to basic prices "BPE". These margin rates depend in turn on "UPRC" and on specific tax rates "UTAX". The dependency is not just within countries but across countries; each country's margin rates and basic prices depend (or are liable to depend) on each other country's user prices and taxes. The dependency however is typically weak.

To estimate import margins "IMARG", 'ep' compares user prices "UPRC" less tax rates "TAX" with import prices "BPI". The tax rates "TAX" are the familiar "UTAX" with missing value indicators -1.0 replaced by zero; thus in effect missing values are set to zero for the import margin calculation. But the import margin is calculated only if certain conditions are met: "UPRC" and "BPI" are non-zero (that is, are not missing values); and imports account for more than 20 per cent of total quantity supplied. If these conditions are not met, then no value is supplied (a zero is supposedly provided by default).

As a comment in the 'ep' code notes, the initial calculation yields some "pretty wild margins"; a floor and ceiling are therefore imposed, so that the rates lie between zero and 5.0. The resultant array "IMARG" is indexed by IEA product, country, and use. For Russia, all components of "BPI" are zero (section 3.4), so the "rus" slice of "IMARG" is not filled.

From these 'ep' derives average import margin rates "AVIMARG", indexed by country and use only (table 37). These are calculated as simple cross-country averages of "IMARG", taken over those countries with non-zero "IMARG" (thus zeros on "IMARG" do not reduce averages in "AVIMARG"). Margin rates on industry and household usage of diesel oil ("AVIMARG" components ("D","I") and ("D","H")) are then set equal to the

Table 37: Coefficient “ AVIMARG ” in program ‘ep’

	USR			
PRD	I	H	U	X
S	0.382	3.950	0.222	0.000
C	0.122	0.000	0.000	0.011
O	2.793	0.000	0.000	0.004
N	1.300	2.000	0.402	0.094
H	0.328	0.000	0.333	0.051
L	1.235	0.284	0.000	0.338
G	0.000	0.835	0.000	0.344
D	0.835	0.835	0.000	0.000
E	0.000	0.000	0.000	0.000

import margin rate as fraction of basic value

margin on household usage of gasoline (AVIMARG("G","H")) (since no non-zero components of “ IMARG ” are available). Finally, margin rates on industry and household usage of natural gas are hard-coded, with values 1.3 and 2.0 respectively. This feature was introduced by Jean-Marc Burniaux, to “put basic prices more in line with import prices” (that is, these high margin rates are needed to obtain sufficiently low basic prices “ BPE ”).

Final margin rates “ MARG ”, again indexed by IEA product, country, and use, are calculated as weighted averages of import margin rates “ AVIMARG ”, US margin rates “ USAMARG ”, and Japanese margin rates “ JPNMARG ”. Like “ AVIMARG ”, these are indexed by IEA product and use. Their values are hardcoded in “ ep ” (tables 38, 39).

The weights, “ AVWGT ”, “ USAWT ”, and “ JPNWT ”, follow these rules:

- For each (product, use) pair, if “ USAMARG ” is equal to zero (expected for missing values), then “ USAWT ” is set at zero; likewise if “ JPNMARG ” is equal to zero, then “ JPNWT ” is set at zero;
- Otherwise, for the United States, “ USAWT ” is equal to one, and “ JPNWT ” to zero; likewise for Japan, “ JPNWT ” is one and “ USAWT ” zero.
- Otherwise, “ USAWT ” and “ JPNWT ” are set at 0.2.
- “ AVWGT ” is calculated as a residual.

Tables 40, 41, and 42 show the Russian slices. Table 43 shows the Russian slice of the resultant “ MARG ” array.

Because of this procedure, each country’s basic prices are in general liable to depend on each other country’s user prices and tax rates. The dependency is weak, because of the cross-country averaging in “ AVIMARG ”, and also because “ AVIMARG ” contributes only part of the weight for “ MARG ”.

Table 38: Coefficient "`USAMARG`" in program '`ep`'

	USR			
PRD	I	H	U	X
S	0.375	1.430	0.375	0.380
C	0.000	1.430	0.000	0.380
O	0.215	0.000	0.215	0.110
N	0.215	0.000	0.215	0.110
H	0.677	0.000	0.677	0.000
L	0.677	0.950	0.677	0.000
G	0.000	1.200	0.000	0.000
D	0.000	1.200	0.000	0.000
E	0.000	0.000	0.000	0.000

US margin rate as fraction of basic value

Table 39: Coefficient "`JPNMARG`" in program '`ep`'

	USR			
PRD	I	H	U	X
S	0.214	0.000	0.345	0.000
C	0.214	0.000	0.345	0.000
O	0.051	0.000	0.086	0.000
N	0.051	0.000	0.086	0.000
H	0.402	0.000	0.151	0.058
L	0.402	0.499	0.151	0.058
G	0.000	0.499	0.000	0.058
D	0.000	0.499	0.000	0.058
E	0.000	0.000	0.000	0.000

Japanese margin rate as fraction of basic value

Table 40: “rus” slice of coefficient “AVGWT” in program ‘ep’

	USR			
PRD	I	H	U	X
S	0.6	0.8	0.6	0.8
C	0.8	0.8	0.8	0.8
O	0.6	1.0	0.6	0.8
N	0.6	1.0	0.6	0.8
H	0.6	1.0	0.6	0.8
L	0.6	0.6	0.6	0.8
G	1.0	0.6	1.0	0.8
D	1.0	0.6	1.0	0.8
E	1.0	1.0	1.0	1.0

share of average import margin data in formula for “MARG”

Table 41: “rus” slice of coefficient “USAWT” in program ‘ep’

	USR			
PRD	I	H	U	X
S	0.2	0.2	0.2	0.2
C	0.0	0.2	0.0	0.2
O	0.2	0.0	0.2	0.2
N	0.2	0.0	0.2	0.2
H	0.2	0.0	0.2	0.0
L	0.2	0.2	0.2	0.0
G	0.0	0.2	0.0	0.0
D	0.0	0.2	0.0	0.0
E	0.0	0.0	0.0	0.0

US data share in formula for “MARG”

Table 42: "rus" slice of coefficient "JPNWT" in program 'ep'

	USR			
PRD	I	H	U	X
S	0.2	0.0	0.2	0.0
C	0.2	0.0	0.2	0.0
O	0.2	0.0	0.2	0.0
N	0.2	0.0	0.2	0.0
H	0.2	0.0	0.2	0.2
L	0.2	0.2	0.2	0.2
G	0.0	0.2	0.0	0.2
D	0.0	0.2	0.0	0.2
E	0.0	0.0	0.0	0.0

Japanese data share in formula for "MARG"

Table 43: "rus" slice of coefficient "MARG" in program 'ep'

	USR			
PRD	I	H	U	X
S	0.347129	3.445835	0.277388	0.076000
C	0.140487	0.286000	0.069000	0.084475
O	1.729271	0.000000	0.060200	0.025053
N	0.833200	2.000000	0.301666	0.096914
H	0.412340	0.000000	0.365254	0.052495
L	0.956832	0.460468	0.165600	0.281619
G	0.000000	0.841048	0.000000	0.286850
D	0.835414	0.841048	0.000000	0.011600
E	0.000000	0.000000	0.000000	0.000000

margin rate as fraction of basic value

3.7 From countries to regions

The program '`epreg`' maps price and tax rate data from countries to regions. It reads from '`eporig.har`' data indexed by price data country, maps these from price-data to standard GTAP countries, aggregates from standard GTAP countries to standard GTAP regions, and writes to '`epreg.har`'.

'`epreg`' reads the price data country set from array "`CTY`" and basic prices of domestic product (US$ per TOE) from array "`CBP`" in file '`eporig.har`'. It then maps from the price data countries to standard GTAP countries:

- 91 standard GTAP countries are themselves price data countries.
- 93 standard GTAP countries are mapped to proxies (Ghana and Brazil are frequent choices).
- 27 standard GTAP countries are ignored.

The set of standard countries mapped to price data countries is read from "`SSTD`" in '`epset.har`'; the mapping itself is read from "`SSTP`" in the same file. Russia of course maps to itself since it is one of the price data countries.

The program then aggregates from standard countries to regions. It calculates regional prices as weighted averages of country prices. The weights are quantities of energy absorbed domestically. It then writes the country-level price data to array "`RGBP`" in file '`epreg.har`'.

Likewise, it reads in data for tax rates (US$ per TOE) for price data countries from array "`CTAX`" in '`eporig.har`', transforms them first to standard countries and then to GTAP regions (again using averages weighted by domestic absorption quantities), and writes out the region-level data to "`RUTX`" in '`epreg.har`'. And for export prices (section 3.10), it reads in "`CUP`" from '`eporig.har`' and writes out "`FOBP`" to '`epreg.har`'. In this last case, it not only transforms the country dimension from price data countries to GTAP regions, but also removes a use dimension, retaining just the slice pertaining to IEA use class "`X`", exports.

3.8 Initial cost estimates

The program '`ecimk`' produces initial estimates '`ecini.har`' of money values ("costs") of energy flows. It relies mostly on quantity data from '`evol.har`' and price and tax data from '`epreg.har`'. Section 2.4 provides an overview. As discussed there, '`ecini.har`' contains commodity tax data not explicitly, but impliedly through tax-free and tax-paid money value data, "`VFM`" and "`VFA`" for intermediate usage and "`VPM`" and "`VPA`" for household consumption. It reads in and modifies specific tax rates from '`epreg.har`', and constructs tax-free price estimates "`AVPRICE1`" from various sources; it uses "`AVPRICE1`" to construct the tax-free value estimates "`VFM`" and "`VPM`"; and both "`AVPRICE1`" and specific tax rate estimates to construct the tax-paid value estimates "`VFA`" and "`VPA`".

'`ecimk`' reads initial estimates for specific tax rates from array "`RUTX`" in file '`epreg`'. The use class index in "`RUTX`" ranges over IEA use classes, whereas the industry index in "VFM" and "VFA" ranges over energy target industries. So 'ecimk' must expand the IEA use class "I" to all energy target industries except electricity.

- For intermediate usage by industries other than electricity:
  - It takes rates from the "I" slice of "RUTX".
  - It sets rates of tax on use of non-electricity energy commodities by non-electricity energy industries equal to zero.
  - For "col", "gas", and "p_c", it makes an upward adjustment to the rate of tax on use by non-energy industries such as to maintain tax revenue summed over industries. For "oil" it makes no such adjustment (by construction the "oil" slice of "RUTX" contains only missing value flags).
- For intermediate usage by the electricity industry, it takes rates from the "U" slice of "RUTX".
- For household consumption:
  - It takes rates from the "H" slice of "RUTX".
  - It sets the rate of tax on crude oil equal to zero.
- It sets the tax rate equal to zero wherever the initial value read from "RUTX" is -1.0 (the missing value indicator).

The results are stored in arrays "REGUTAX" and "TFSPEC"; "REGUTAX" covers both intermediate usage and household consumption, but the rates for intermediate usage are not final. "TFSPEC" contains the final rates for household consumption. Tax rates for exports are derived and stored separately.

For Russia, this yields tax rates of zero for all domestic use: for non-electricity energy inputs into non-electricity energy industries, by imposition within 'ecimk'; and for all other intermediate usage and household consumption flows, by replacement of missing value flags in "RUTX".

To convert specific tax rates to *ad valorem*, "ecimk" uses a tax-free price (US$ per TOE) "AVPRICE1", pertaining to domestic absorption of domestic product and imports alike. To obtain "AVPRICE", "ecimk" reads basic prices from array "RGBP" in file 'epreg'. Non-positive values in that array indicate missing values; "ecimk" fills these in with cross-country averages.

Although the prices "RGBP" derive from source-generic user prices in program 'ep', the program 'ecimk' applies them to domestic product only. For imports, it seeks prices compatible with the GTAP trade data set. Accordingly, it reads in duty-free source-specific import values from "VIWS" in 'tradee.har'. It also reads import quantities (in TOE) from array "MVOL" in file 'evol.har', and duty rates from "PDIM" in 'prot.har'. From these it derives cross-partner-average duty-paid import prices "IMPRICE", as described in section 3.10. It then computes the domestic absorption prices "AVPRICE1" as weighted averages of these import prices and domestic product prices, reading the domestic product share from array "DSHR" in 'evol.har'. Table 44 shows some of these data for Russia. Finally it computes tax-free and tax-paid money value estimates from "AVPRICE1", "REGUTAX" and "TFSPEC", and quantity data.

Table 44: Price data for Russia in ‘ecimk’

	Domestic price	Import price	Domestic share	Average price
col	58.6171	77.4745	0.9287	59.9614
oil	120.9008	133.0791	0.9596	121.3932
gas	91.9920	117.4068	0.9939	92.1459
p_c	108.3818	164.6577	0.9721	109.9507
ely	167.6221	680.7742	0.9913	172.0662

Source:

- domestic price: “RGBP” in ‘epreg.har’
- import price: “MP” in ‘epini.har’
- domestic share: “DSHR” in ‘evol.har’
- average price: “AVP1” in ‘ecirep.har’

All prices are in US$ per TOE

3.9 Price adjustment

The program ‘epadj’ reads initial unbalanced estimates of money values of energy production, trade, and usage from ‘ecini.har’. It rebalances these, and writes the rebalanced estimates to ‘ecmain.har’. It accounts for the changes in money values through a price model; all quantities are held fixed, so money value changes are strictly proportional to price changes. In ‘ecini.har’ and ‘ecmain.har’, commodity tax rates are not recorded explicitly, but are given impliedly by tax-free and tax-paid usage values: for intermediate usage, “VFM” and “VFA”, and for household consumption, “VPM” and “VPA”.

We give a broader overview of ‘epadj’ in section 2.4; in this place we focus on its handling of prices.

In updating the energy cost arrays, ‘epadj’ preserves specific not *ad valorem* tax rates. So the price changes in general imply changes in *ad valorem* tax rates. Table 45 shows prices for Russia before and after adjustment. As it shows, there is a relatively moderate increase in the price of petroleum products, and a large increase in the price of electricity. Since the initial specific tax rates are zero, both the initial and final *ad valorem* tax rates are likewise zero.

Note that table 45 shows changes in basic prices of domestic product. The price model in ‘epadj’ distinguishes prices between domestic product and imports, even though its usage data do not. The usage arrays “VFM”, “VFA”, “VPM”, and “VPA” are valued at domestic absorption prices, these being an average of domestic product and import prices. More precisely, there are two prices for domestic products, one for domestic absorption and one for exports; these are averaged to obtain an average price for domestic production. The prices reported in table 45 are the prices for domestically absorbed domestic product. The price adjustments are made mainly to these prices, to a much smaller extent to export prices, and not at all to import prices.

Table 45: Estimated tax-free prices of domestically produced and absorbed energy commodities in Russia, before and after price adjustment (US$ per TOE)

	before	after	relative change
col	58.62	58.63	0.00
oil	120.90	120.93	0.00
gas	91.99	92.00	0.00
p_c	108.38	150.87	0.39
ely	167.62	544.95	2.25

Source: Array "DP" in files 'epini.har' and 'ep.har' in energy module from GTAP 5.3 trial run

3.10 Import prices

As noted in section 3.8, the commodity prices applied by 'ecini' are an average of domestic product and import prices. The import prices are calculated within 'ecini' from quantity, money value, and import tariff data. We now discuss the antecedents of the money value data.

The energy module constructs the trade quantity data as follows:

- A program 'eca' reads in bilateral FOB trade value data, indexed by GTAP commodity, from a file 'trade.har', and FOB prices from 'epreg.har', and writes bilateral trade quantity data, with separate arrays for each energy target commodity, to files 'col-vol.har', 'oil-vol.har', and so on.

- A program 'smeartrd' reads these estimates, and writes adjusted estimates to files 'col-vols.har', 'oil-vols.har', and so on. The adjustment removes zeros that would be troublesome in the next balancing step.

- Programs 'rase' and 'btvmrecon' read these estimates, and write rebalanced estimates to files 'col-vol.har', 'oil-vol.har', and so on (residing in a different directory than the earlier files of the same name). These rebalanced estimates are consistent with non-bilateral export and import quantity data read from a file 'evol.har'.

- A program 'cif' reads these rebalanced estimates (still indexed by energy target commodity), FOB prices from 'epreg.har', and margin data from 'trade.har', and writes bilateral CIF trade value data, indexed by GTAP commodity, to a file 'tradee.har'.

- As described in section 3.8, the program 'ecimk' reads those CIF trade data, and import tariff data from 'prot.har', to compute energy-target-commodity-indexed duty-paid import values, summed over sources.

- Combining those duty-paid import value data with the import quantity data from 'evol.har', 'ecimk' thence derives estimates of cross-partner average duty-paid import prices.

Some comments and further details on this procedure:

- It is important that the same import quantities, array "`MVOL`" in file '`evol.har`', are used by '`rase`' to balance the bilateral trade quantity data and by '`ecimk`' to pass from money values to prices. This ensures that those prices are in effect averages of the prices used by '`cif`' in constructing the CIF trade value data in '`tradee.har`'.

- Those prices are cross-country averages of the FOB prices read from array "`FOBP`" in '`epreg.har`', uplifted by margins.

- The margin rates are calculated from margins "`VTWR`" and FOB values "`VXWD`" in the original trade data file '`trade.har`'.

- Each country's import prices, as computed in '`ecimk`', depend not on its own import prices "`BPI`" assembled by '`ep`', but on its partner countries' export prices. These derive from a the export slice of another array, user prices "`CUP`", constructed by the same program.

This yields an alternative view of the procedure:

- The program '`ep`' assembles user price estimates "`UPRC`", as described in section 3.3, and writes them to "`CUP`" in '`eporig.har`'. These are indexed by energy target commodity, price data country, and IEA use class. The slice for use class "`X`", exports, contains FOB export prices.

- The program '`epreg`' reads those price-data-country-indexed prices and writes GTAP-region-indexed export prices to "`FOBP`" in '`epreg.har`'.

- The program '`cif`' reads those energy-target-commodity-indexed prices and, in effect, derives GTAP-commodity-indexed prices.

- The same program in effect combines those FOB prices with margin rate data from '`trade.har`' to derive CIF prices.

- The program '`ecimk`' in effect combines those CIF prices with import tariff data from '`prot.har`' to obtain duty-paid prices.

- The programs '`cif`' and '`ecimk`' in effect combine those partner-specific GTAP-commodity-indexed prices with energy volume data from '`col-vol.har`' *etc.* to obtain cross-partner-average energy-target-commodity-indexed import prices.

In the foregoing, "in effect" means that the transformations described are performed not in fact on the price data, but on corresponding money value data. Table 46 presents yet another view of the process, following the same thread as the last list, but referencing actual data arrays.

Table 46: Data flow overview for trade prices in energy module

Program or file	Array	Data
`p_iea.har`	`SPN`, *etc.*	user price in US$ per TOE, in (IEA product, IEA use)-specific arrays, indexed by country and year
`eporig.har`	`CUP`	user price in US$ per TOE, indexed by energy target commodity, price data country, and IEA use class
`epreg.har`	`FOBP`	FOB price of exports in US$ per TOE, indexed by energy target commodity and GTAP region
`cif`	`ECOMM_VXWD`	FOB value of international trade in US$ per year, indexed by energy target commodity, source (GTAP region), and destination (GTAP region)
	`NEW_VXWD`	FOB value of international trade in US$ per year, indexed by GTAP commodity, source (GTAP region), and destination (GTAP region)
`tradee.har`	`VIWS`	CIF value of international trade in US$ per year, indexed by GTAP commodity, source (GTAP region), and destination (GTAP region)
`ecimk`	`IMBAS`	duty-paid value of imports, in US$ per TOE, indexed by energy target commodity and GTAP region
	`IMPRICE`	duty-paid price of imports, in US$ per TOE, indexed by energy target commodity and GTAP region
	`AVPRICE1`	basic price, in US$ per TOE, indexed by energy target commodity and GTAP region

3.11 Conclusions

We gather here some conclusions from the foregoing, relating to commodity taxes for Russia in the abortive GTAP 5.3 run, and commodity tax and price processing generally in the energy module.

As noted in section 2.2, the tax rate data supplied for Russia by the energy module show:

- missing values for electricity usage by energy industries, and for energy usage by the electricity industry;
- elsewhere, zero rates for energy usage by energy industries; and
- elsewhere, missing values.

We now see how this comes about:

- As shown in table 29, for 20 (product, use) pairs the program '`ep`' reads no tax data.
- As shown also in table 29, for 10 more (product, use) pairs the source data arrays contain no entries for Russia.
- As noted in section 3.5, for the remaining 6 (product, use) pairs, the source data tax rate exceeds the user price (often grossly), and is overwritten with a missing value indicator.
- As noted in section 2.4, in program '`ecimk`', zero rates are supplied in all cases in lieu of missing values.
- As noted in section 2.5, for non-electricity energy inputs into non-electricity energy industries, regardless of the input data, the data availability flags are set to indicate that data are available.

To change the tax rates and tax rate data flags supplied by the energy module, we must change the tax rate source data. But for several reasons, control over the source data does not give us full control over the outputs:

- As noted in section 3.1, the programming assumes that actual tax rates (not missing value flags) are positive. This prevents us from introducing commodity subsidies (more precisely, we are not confident that the existing code would handle them correctly).
- The source data are organized by IEA use class (table 15). This aggregates intermediate usage to two classes, usage in electricity generation and everything else. So we cannot differentiate tax rates across non-electricity industries. The program '`ecimk`' however imposes its own differentiation: zero for non-electricity industries, and a common rate for non-electricity industries, such as to preserve the cross-non-electricity-industry average rate (section 3.8). So tax rates are differentiated across non-electricity industries, but not in a way that we can control.

- As a corollary to the above, we have no control over taxes on non-electricity energy inputs into non-electricity energy industries; these come out as zero whatever we put in.

- For 20 (product, use) pairs, as noted above, '`ep`' reads no source data. When later it comes to calculate "`CTAX`", this leads to either unduly low tax rates or missing value indicators for many (commodity, use) pairs. In particular, it leads to missing value indicators for several components of the "`U`" slice of "`CTAX`"; this is liable to lead to inter-module tax rate inconsistencies just as in the old run.

- The source data are specific tax rates, but the output data and the data available from the IO table are *ad valorem* tax rates. So we must convert the desired *ad valorem* tax rates to specific rates, and to do that we need the applicable prices. Now the energy module deals with both basic prices and user prices. But the basic prices depend on the tax rates, so we cannot take them as given; and the user prices are available only in the earlier stage of processing, so if we use them we do not take account of later-stage adjustments. So we must either solve a system incorporating both those adjustments and the price-tax relations, or else neglect the adjustments.

Between the final state of the user prices "`UPRC`" in '`ep`' and the final fixing of the *ad valorem* tax rates in '`mketax`', prices are adjusted as follows:

- converted from user to basic level in '`ep`',

- averaged over uses, and over IEA product classes within energy target commodities, also in '`ep`' (section 3.4),

- mapped from countries to regions in '`epreg`' (section 3.7),

- averaged with import prices in '`ecimk`' (section 3.8), and

- adjusted to avoid non-energy cost infeasibilities in '`epadj`' (section 2.4).

All these steps, except the country-to-region mapping, affect Russian prices.

4 A new contributed table

Between the detection of the problems discussed above and their remedy, in January 2003, a new version of the Russian table was received from the same contributors as the previous (Sergei Kiselev and Roman Romachkine). As table 47 shows, this new version provides much the same overall energy commodity tax profile as the previous, but with substantially lower tax rates on gas and on exports. These tax rate changes reflect changes in gas and export valuation.

Table 48 shows some of the variation in tax rates within the averages reported in table 47. The observations included in the table are those that contribute most to variances within groups. Points worthy of note include:

Table 47: Average rates of tax in contributed tables for Russia, by energy target commodity and IEA use class (per cent)

	I		H		U		X	
	old	new	old	new	old	new	old	new
col	4	4	24	24	5	5	7	3
oil	12	12	0	0	11	11	15	11
gas	51	25	51	36	77	40	138	10
p_c	23	23	64	64	26	26	22	11
ely	-2	-2	-46	-46	4	4	8	8

- the 22 per cent subsidy on government services use of electricity, contrasting with taxes on other major uses;
- the unusually high rates of tax on government services use of coal, "`p_c`" and "`gas`", compared to other uses;
- the generally lower rates of tax on "`gdt`" compared to "`gas`".

As noted in section 3.11, these rate variations cannot be retained in the energy module output.

5 Revised tax rate source data

Following the diagnosis and research described in sections 2 and 3 for the infeasibility problem, we plan and execute a remedy. In section 5.1 we develop a plan for revising the tax rate source data using information for the latest contributed IO table. In section 5.2 we report on its execution, and in section 5.3, on its application in a new GTAP data base construction run. In section 5.4 we assess our success in achieving our goals of letting Russia fit and exploiting the tax rate data in the contributed table. Finally in section 5.5 we note lessons for future work.

5.1 Revising the source data: Plan

In broad terms, our plan is to:

- extract *ad valorem* tax rates from the contributed table,
- convert them to specific rates, and
- write those rates into the energy module source data.

As noted in section 3.11, however, our control of the source data tax rates does not give us full control of the output data rates.

In developing a detailed plan, several design issues arise:

Table 48: Selected tax rates in later contributed table for Russia (per cent)

Average			Actual			
Commodity	Use	Rate	Commodity	Source	Use	Rate
col	I	4	col	dom	osg	23
gas	I	25	gas	dom	gas	88
					crp	39
					nmm	43
					i_s	40
					otp	57
					wtp	85
					atp	85
					obs	61
					ros	32
					osg	61
					dwe	72
			gdt	dom	gas	64
					crp	5
					nmm	5
					i_s	4
					nfm	4
					mvh	5
					trd	4
					ros	3
					osg	9
	U	40	gas	dom	ely	44
			gdt	dom	ely	23
	X	10	gdt	dom	export	5
p_c	I	23	p_c	dom	p_c	12
					otp	30
					osg	53
					dwe	48
				imp	osg	53
ely	I	-2	ely	dom	ofi	14
					osg	-22
					dwe	13

Note: The “actual rates” selected for reporting are those that contribute most to the within-group variances.

- Do we revise tax rate data only, or price or quantity data also? We choose to revise tax rate data only.

- Do we allow negative tax rates (subsidies), or do we filter them out? We choose to filter them out.

- Do we revise tax rates for the electricity industry only, or for all uses? We choose to revise tax rates for all uses.

- Do we revise tax rates for all years, or for the latest year only? We choose to revise rates for the latest year only.

- Do we take account of adjustments to prices subsequent to the finalisation of user price estimates "`UPRC`" in '`ep`' (section 3.11)? We choose not to take account of those adjustments.

- Which state of the contributed table do we use? The initial table input into the data base construction procedure (the "`ini`" file), the "robustified" table input into the Fit module (the "`rbt`" file), the table with inventory investment eliminated, input into the '`fit`' run (the "`ii0`" file), or some other state? We choose the robustified table.

- In aggregating tax rates from GTAP to IEA use classes, do we use contributed table or target weights? We choose to use contributed table weights.

- How do we access the price data? Do we read the energy module source data, and replicate that module's procedures to construct a user price array, or do we use a user price array constructed by the energy module in a previous run? We choose to use a previously constructed array. Similarly, we use a previously constructed array for margin rates.

To convert tax rates from specific to *ad valorem*, given user prices, we need also to know the margin rate. We may then use the formula

$$\begin{aligned} s &= ab \\ &= \frac{au}{u/b} \\ &= \frac{au}{1+m+a} \end{aligned}$$

where s denotes the specific tax rate, a the *ad valorem* tax rate, as a fraction of the basic value, b the basic value price, u the user price, and m the margin rate, as a fraction of the basic value. But for exports, the *ad valorem* tax rate is provided as a fraction of a margin-inclusive but tax-free price f; so there we use

$$\begin{aligned} s &= af \\ &= \frac{f}{u}\, au \\ &= \frac{(1+m)\, b}{(1+m+a)\, b}\, au \\ &= \frac{(1+m)\, au}{1+m+a} \end{aligned}$$

Of our design choices, some are based on clear cost/benefit considerations, and some have little effect on results. There are some however that offer scope for improved results in future work:

- We might try allowing negative tax rates (subsidies).
- In aggregating tax rates from GTAP to IEA use classes, we might use energy target weights.
- We might take account of some downstream price adjustments—in the first instance, of the averaging of prices across uses, and across IEA products with energy target commodities.

In detail, our plan for the revising the tax rate source data is:

1. Prepare a revised version of the price data country set information file '`p_ieaset.har`' including Russia in the country ranges for all tax data arrays.
2. Run the earlier stages of the GTAP data base construction procedure, using the latest IO table contributed for Russia, to produce a robustified Russian IO table ('`rus_rbt.har`').
3. For basic values and commodity taxes:
   (a) Extract data from the robustified table,
   (b) Sum over sources (domestic production, importation), and aggregate from GTAP to IEA use classes and from GTAP to energy target commodities.
4. Calculate *ad valorem* tax rates.
5. Filter out negative tax rates (subsidies).
6. Map the tax rates from energy target commodities to IEA products.
7. Convert from *ad valorem* to specific rates, using user prices and margin rates read from an '`eporig.har`' file from an old energy module run.
8. Convert the specific rates from US dollars to local currency, and from TOE to the physical units used in the source data arrays.

Table 49: Target tax rate, by IEA product and IEA use class (per cent)

	I	H	U	X
S	3.5508	24.1627	5.4002	2.9061
C	3.5508	24.1627	5.4002	2.9061
O	11.8052	0.0000	10.8438	11.0737
N	25.1594	35.8204	40.1073	9.8099
H	23.4516	63.8434	26.1824	10.5335
L	23.4516	63.8434	26.1824	10.5335
G	23.4516	63.8434	26.1824	10.5335
D	23.4516	63.8434	26.1824	10.5335
E	0.0000	0.0000	3.5320	7.9602

9. Expand those rates from a single-year to a multi-year array, filling years other than the latest with missing value indicators.

10. Revise the source data arrays "`STN`" *etc.* for tax rates to incorporate the new data.

11. Assemble a new version of the source data file '`p_iea.har`', incorporating the revised tax rate arrays.

5.2 Revising the source data: Execution

We now report on results obtained following the plan described in section 5.1 for revising the Russian tax rate source data.

Following plan steps 3 and 4, we compute average rates of tax from the new contributed table, by energy target commodity and IEA use class. These are just the "new" rates reported in table 47 above. Following step 5, we replace negative rates with zeros; this affects electricity use by electricity and other industries. And following step 6, we map from energy target commodities to IEA products. The rates are then as shown in table 49.

Following step 7, we then obtain the specific tax rates, in US$ per TOE, shown in table 50. These may be compared with the "`UTAX`" values from the aborted GTAP 5.3 run reported in table 34.

- By construction, the new data set has no missing values.
- Where the old data set has non-zero rates, the new rates are much lower (it may be recalled that the old rates were later replaced by zeros because they exceeded estimated user prices).
- In once case, electricity usage by households, there is a zero new rate but a strictly positive old rate (the zero new rate resulting from the overwriting of an initial negative rate).

Following step 8, we obtain rates in local currency units and product-specific physical units as shown in table 51 (these may be compared with the old run rates in tables

Table 50: Target tax rate, by IEA product and IEA use class (US$ per TOE)

	I	H	U	X
S	2.3859	3.0117	3.6434	1.5399
C	1.0072	11.2528	3.4211	1.8823
O	11.3103	0.0000	11.2946	12.4049
N	11.3008	23.2508	32.0331	10.0890
H	4.9712	57.0075	5.6176	10.8397
L	8.8328	30.0980	63.7951	14.8672
G	38.3736	31.7882	41.9147	19.6679
D	10.8440	100.8120	92.6414	9.1311
E	0.0000	0.0000	27.4453	16.4501

Table 51: Target tax rate, by IEA product and IEA use class (local currency unit per product-specific physical unit)

	I	H	U	X
S	6.9608	8.7867	10.6297	4.4927
C	3.3071	36.9472	11.2327	6.1804
O	7.0134	0.0000	7.0037	7.6922
N	52.3339	107.6745	148.3453	46.7221
H	22.1009	253.4415	24.9743	48.1905
L	35.1414	119.7446	253.8084	59.1491
G	0.1483	0.1229	0.1620	0.0760
D	0.0447	0.4155	0.3818	0.0376
E	0.0000	0.0000	0.0109	0.0066

30 and 31). Finally, following plan steps 9 and 10, we incorporate the new rates into the source data arrays in a revised "`p_iea.har`"; or rather, we incorporate as many of the new rates as the limitations of the data structure permit (section 3.2).

5.3 Applying the revised data

Having revised the tax rate source data file '`p_iea.har`' as described in section 5.2, we use the revised file in a new GTAP data base construction run. This results in new tax data status arrays and tax rate arrays in the file '`ec.har`' generated by the energy module, and in the Russian table successfully passing '`fit`'.

In '`ec.har`', most of the tax rate status flags for Russia change from 1 (indicating data absent) to 0 (data present). The flags that remain non-zero are:

- "`oil`" use by non-electricity energy industries and by private households, and
- "`oil`", "`p_c`", and "`ely`" use by the electricity industry.

Table 52: Target rates of tax for energy use in Russia, supplied by the energy module to the Fit module, in the new GTAP data base construction run (per cent)

	I	E	U	H
col	4	0	6	5
oil	..	0	..	..
gas	17	0	38	27
p_c	4	0	..	18
ely	0	0	..	0

I : non-energy industries
E : non-electricity energy industries
U : electricity industry
H : household consumption
.. : missing value

Source: arrays "TF" and "TP" in file 'ec.har' in the new run

Crude oil use by non-energy industries and private households is of small consequence. But "p_c" and "ely" use by the electricity industry is considerable, and moreover has strategic significance for the removal of the non-energy cost infeasibility (section 2.1).

Table 52 summarizes the Russian tax rates in arrays "TF" and "TP" in 'ec.har'. These rates are passed to the Fit module; using them, 'fit' avoids the non-energy cost infeasibility, and the run completes successfully.

We now trace the new tax rates through the energy module. We also trace the price data, since that is affected and affects the tax rates.

Of the 36 Russian tax rates calculated from the IO table, we supply the allowed maximum of 16 to the revised source data file, up from the 6 rates incorporated in the older version. All 16 new observations satisfy the requirement that the tax rate not exceed the user price, so all 16 are retained in the final version of "UTAX", up from zero in the previous run (section 3.5). At this point, the non-missing values in "UTAX" agree with the rates computed in the data revision process and reported in table 50.

As expected, the user price data in 'ep' are much as before, but price-tax interaction does lead to change in one component of "UPRC", that for exports of diesel oil. This falls from US$96 per TOE in the old run (table 22) to $85 per TOE in the new run. The price falls in the new run because it depends on the rate of tax for industry use (section 3.3), and this changes from a missing value indicator (effectively a zero) in the old run to $11 per TOE in the new run (table 50).

The basic prices "BPE" are generally lower in the new run (table 53) than in the old (table 26), since the user prices are mostly unchanged but the tax rates non-zero (in calculating basic prices in the old run, 'ep' treats missing values in "UTAX" as zeros). The lower use-specific IEA-product-indexed prices "BPE" lead in turn to lower cross-use average energy-target-commodity indexed prices "CBP" (table 54).

After finalizing "UTAX", 'ep' averages tax rates over IEA products with energy target commodities to produce a more aggregate array "CTAX". Of the 20 elements of the "rus"

Table 53: Coefficient "BPE" in program 'ep', after filling in from cross-country averages, in the new run

	I	H	U	X
S	67.19	12.46	67.47	50.58
C	28.37	79.10	79.10	61.33
O	99.95	144.25	144.25	121.09
N	44.92	64.91	79.87	102.14
H	21.20	140.21	21.46	107.56
L	37.66	47.14	156.94	119.18
G	212.96	49.79	212.96	156.97
D	46.24	157.91	273.18	83.90
E	223.11	35.09	773.37	223.11

use-specific basic price estimates, US$ per TOE

Table 54: "rus" slice of array "CBP" in file 'eporig', in the old and new runs

	old	new
col	58.62	56.76
oil	120.90	120.90
gas	91.99	85.15
p_c	108.38	105.00
ely	167.62	167.62

use-generic basic price estimates, US$ per TOE

Table 55: Missing value share in aggregating Russian tax rates from IEA products to energy target commodities in the new energy module run

	I	H	U	X
col	0.00	0.00	0.08	1.00
oil	1.00	1.00	1.00	1.00
gas	0.00	0.00	0.00	1.00
p_c	0.31	0.15	1.00	1.00
ely	0.00	0.00	1.00	1.00

Table 56: Tax rates, by energy target commodity and IEA use class, from array "CTAX" in file 'eporig.har' in the new energy module run (US$ per TOE)

	I	H	U	X
col	2.3859	3.0117	3.3458	-1.0000
oil	-1.0000	-1.0000	-1.0000	-1.0000
gas	11.3008	23.2508	32.0331	-1.0000
p_c	6.2938	26.8196	-1.0000	-1.0000
ely	0.0000	0.0000	-1.0000	-1.0000

slice of "CTAX", 10 are tax rates and 10, missing value indicators. There are missing values for exports of all commodities, for all uses of "oil", and electricity industry usage of "p_c" and "ely". Most of these match blocks of missing values in "UTAX"; the one exception is electricity industry use of "p_c", where "UTAX" has an observation for heavy fuel oil. But "CTAX" still has a missing value indicator for ("p_c","rus","U") because heavy fuel oil has a weight "UPRODWT" of zero (all the weight for ("p_c","rus","U") is in light fuel oil, but UTAX("L","rus","U") is a missing value indicator).

Other components of "CTAX" are affected by the substitution of zeros for missing values in averaging over IEA products. As table 55 shows, the affected rates (in decreasing order of severity) are those for ("p_c","I"), ("p_c","H"), and ("col","U"). The "rus" slice of the new "CTAX" array is shown in table 56 (compare the rates from the target revision program reported in table 50).

As before, the transformation from energy data countries to GTAP regions does not affect prices or tax rates for Russia.

The program 'ecimk' computes import prices and averages them with domestic product prices to generate the initial price estimates "AVPRICE1". The import prices hardly change from the old run. So the new "AVPRICE1", like the new "CBP", are close to but somewhat lower than their counterparts from the old run (table 57).

Program 'ecimk' also adjusts tax rates on industry use, zeroing out some rates for energy industries, and increasing some rates for non-energy industries (table 58). Combining price and tax data, it then generates *ad valorem* tax rates as shown in table 59 and incorporates them implicitly into 'ecini.har'.

The program 'epadj' adjusts prices as shown in table 60. The price adjustments

Table 57: “`rus`” slice of array “`AVPRICE1`” in program ‘`ecimk`’, in the new and old runs

	old	new
`col`	59.96	58.24
`oil`	121.39	121.39
`gas`	92.15	85.35
`p_c`	109.95	106.66
`ely`	172.07	172.06

basic price estimates averaged over domestic product and imports, US$ per TOE

Table 58: Rates of tax on industry use of energy in ‘`ecimk`’ (US$ per TOE)

		Adjusted	
	Initial	`I`	`E`
`col`	2.3859	2.3934	0.0000
`oil`	`..`	0.0000	0.0000
`gas`	11.3008	14.5236	0.0000
`p_c`	6.2938	6.7054	0.0000
`ely`	0.0000	0.0000	0.0000

`I` : non-energy industries
`E` : non-electricity energy industries
`..` : missing value

Table 59: Target rates of tax for energy use in Russia, from the initial money value estimates ‘`ecmain`’ in the new energy module run (per cent)

	`I`	`E`	`U`	`H`
`col`	4	0	6	5
`oil`	0	0	0	0
`gas`	17	0	38	27
`p_c`	6	0	0	25
`ely`	0	0	0	0

`I` non-energy industries
`E` non-electricity energy industries
`U` electricity industry
`H` household consumption

Table 60: Estimated tax-free prices of domestically produced and absorbed energy commodities in Russia, before and after price adjustment in the new run (US$ per TOE)

	before	after	relative change
`col`	56.76	56.78	0.00
`oil`	120.90	120.95	0.00
`gas`	85.15	85.16	0.00
`p_c`	105.00	150.78	0.44
`ely`	167.62	645.23	2.85

Source: Array "`DP`" in files '`epini.har`' and '`ep.har`' in energy module from GTAP 5.3 trial run

are larger than in the old run (table 45). In particular the electricity price increases by 285 per cent, compared to 225 per cent previously, as the electricity price must now cover taxes on gas inputs. Since specific tax rates are conserved, the *ad valorem* tax rates adjust, assuming the final values reported in table 52.

5.4 Assessment

In assessing the new run, we have to appraise its success in removing the non-energy cost infeasibility that aborted the old run, and in better aligning tax rates in the fitted IO table with those in the contributed table.

To assess the alignment of target and contributed table tax rates, we examine the target rates passed in '`ec.har`' from the energy to the Fit module. But the contributed table rates in table 47, are not fully comparable with the target rates in table 52, since the former are averaged over all industries, while the latter distinguish between energy and non-energy industries. In table 61 therefore we present contributed table and target rates both aggregated to IEA use classes. We use contributed table weights to aggregate. We show the contributed table rates after the censoring of negative rates, so any differences between the two sets of rates represent unplanned changes imposed by the energy module. We observe:

- With just one exception, `("col","U")`, the target rates are lower than or equal to the contributed table rates.
- The target rates show only missing values for crude oil.
- Between the contributed IO table and the energy module output, tax rates decrease greatly for "`p_c`", less for "`gas`", and less again for "`col`".

In table 62 we trace the evolution of the tax rates through the energy module. For comparability we express all rates as *ad valorem*; so the reported changes in rates reflect changes both in specific tax rates and in prices. Also for comparability, we average tax rates as necessary at each stage to report them for energy target commodities

Table 61: Target rates of tax on energy use in Russia, as obtained from the contributed IO table and as supplied by the energy module, by energy target commodity and IEA use class (per cent)

	I		H		U	
	IO table	energy module	IO table	energy module	IO table	energy module
`col`	4	3	24	5	5	6
`oil`	12	..	0	..	11	..
`gas`	25	14	36	27	40	38
`p_c`	23	4	64	18	26	..
`ely`	0	0	0	0	4	..

.. : no rate supplied

and IEA domestic classes. We report only for domestic use classes, since the energy module provides no rates for export taxes. We handle missing values according to their treatment in the energy module; some we convert to zeros, others we retain.

Column 6 of table 62 derives from the same data as the "energy module" columns of table 61. The numbers reported differ slightly, because of different weights used in aggregation. Table 61 uses only weights from the contributed table, whereas table 62 uses a combination of contributed table and energy target weights (it uses contributed table weights to aggregate from GTAP to energy target industries, then energy target weights to aggregate to IEA use classes).

Much of what we gather in table 62 under the heading of changes in rates involves correcting convenient but false assumptions made in calculating the new rates. For example, in calculating the new rates we apply use-specific prices "`UPRC`" to convert *ad valorem* to specific rates. As columns 2 and 3 in table 62 show, applying use-generic prices makes a great difference. But these use-generic prices become available only some way into the process; so it is convenient to represent this effect as a change in tax rates, from those calculated with the use-generic prices to those calculated with the use-specific.

From table 62 we see that the most important changes occur at stage 3. Here the switch from use-specific to use-generic prices leads to reductions in rates for many flows, notably `("col","H")`, `("gas","I")`, `("p_c","I")`, and `("p_c","H")`. At stage 2, the omission of many (product, use) pairs from the energy module source data leads to the conversion of several high tax rates to missing values; this occurs for `("oil","I")`, `("oil","U")`, and `("p_c","U")`. These missing values are restored to the original IO table rates at the fitting stage (stage 6). But in other cases missing values at the IEA commodity level lead to reductions in non-missing values at the energy target commodity level; so we see moderate unreversed reductions in rates for `("p_c","I")` and `("p_c","H")`. Stage 1, elimination of negative tax rates, makes a large difference to one flow, `("ely","H")`. Stage 5, price adjustment in '`epadj`', causes a further moderate reduction in the rate for `("p_c","H")`. The effects of stage 4, domestic-import price

Table 62: Tax rates in the new GTAP data base construction run, by energy target commodity, IEA domestic use class, and stage of processing (per cent)

		Stage of processing						
Comm.	Use	0	1	2	3	4	5	6
col	I	4	4	4	4	4	4	4
col	H	24	24	24	5	5	5	5
col	U	5	5	5	6	6	6	6
oil	I	12	12	..	..	..	..	0
oil	H	0	0	..	..	..	..	0
oil	U	11	11	..	..	..	..	11
gas	I	25	25	25	13	13	13	13
gas	H	36	36	36	27	27	27	27
gas	U	40	40	40	38	38	38	38
p_c	I	23	23	16	6	6	4	4
p_c	H	64	64	54	26	25	18	18
p_c	U	26	26	..	..	..	..	26
ely	I	-2	0	0	0	0	0	0
ely	H	-46	0	0	0	0	0	0
ely	U	4	4	..	..	..	..	4

.. : missing value

stages of processing:
0 : as calculated from the contributed IO table
1 : after removing negative tax rates (subsidies)
2 : as input into the energy module (calculated from "UTAX" and "UPRC")
3 : after applying use-generic prices (from "UTAX" and "CBP")
4 : after averaging domestic and import prices (from "TFSPEC" and "AVPRICE1")
5 : after adjusting prices to avoid non-energy cost infeasibilities (from "VFA" and "VFM" in 'ecmain.har')
6 : in the fitted table

averaging, are barely discernible.

In conclusion, concerning the matching of the contributed table tax rates, we find:

- Because of energy module restrictions, we lose the energy use subsidies in the contributed table.
- For the same reason, we lose the cross-industry tax profile of the contributed table. As table 48 shows, the profile imposed by the energy module bears little resemblance to that found in the contributed table.
- Likewise we impose the same rate of tax on the GTAP commodities "`gas`" and "`gdt`", although the contributed table very plausibly records higher rates for "`gas`" than for "`gdt`".
- Even allowing for those constraints, the average rates delivered by the energy module are far from a faithful representation of those found in the contributed table (table 61). Indeed, only in one case, `("gas","U")`, is a high rate in the contributed table well preserved in the energy module output.
- The main cause of divergence between the contributed table and target rates is divergence between the IEA-product-indexed use-specific prices "`BPE`" and the energy-target-commodity-indexed use-generic prices "`CBP`".
- Despite all these imperfections, we do succeed in retaining a significant part of the contributed table energy taxes in the fitted table.

We now consider the removal of the non-energy cost infeasibility, In this we do indeed succeed, but as much through good fortune as good management. Because of energy module program restrictions, the feasibility calculations neglect taxes on crude oil, petroleum product, and electricity inputs into electricity production. Fortunately that neglect does not lead to failure in this case; the neglected taxes reduce non-energy input costs by more than half, from US$3.9 billion in the feasibility calculation to $2.0 billion in the fitted table, but they do not eliminate them. But there is no guarantee that this method would succeed in future cases. More generally, we find:

- For this purpose, what matters is that '`epreg.har`' show a non-missing value; it does not matter what the value is.
- Also for this purpose, the only rates that matter are those for use in the electricity industry.
- Of those, the rate for "`oil`" matters little for Russia, since usage is very low.
- Of those that do matter, we cannot obtain a non-missing value for "`oil`" or "`ely`", but we can for "`col`" and "`gas`".
- We can obtain a non-missing value for "`p_c`" for some countries, those with a non-zero `("H","U")` component of "`UPRODWT`". Russia is not one of those countries.
- Infeasibilities are likely to arise in cases where the contributed table shows high rates of tax, and the energy data show high quantities, of "`oil`", "`ely`", or (for some countries) "`p_c`" use by the electricity industry.

5.5 Future

We gather together here implications for future work. These are of two kinds: further data revisions, to avoid infeasibilities in "`fit`" or to align target tax rates better with contributed table rates; and revision of the energy module programs, to gain more control over the target rates.

Concerning data revisions, we find:

- The goal of ensuring non-energy cost feasibility can be met only partially. It is liable to be unachievable in cases where the contributed table shows significant taxes on "`oil`", "`p_c`", or "`ely`" usage by "`ely`", and the energy module source data shows considerable quantities of such usage.
- The goal of ensuring consistency between fitted and contributed table tax rates can best be achieved by supplying less rather than more tax rate data to the energy module.
- Overall, the best strategy is to supply tax rate data for electricity industry usage only, but in calculating those data to take account of the cross-use averaging of energy prices in '`ep`'.
- Finally, it is to be hoped that these findings are not used, since data-only solutions are flawed and limited, and revision of the programs is desirable on many grounds.

We do not discuss here all the reasons for revising the energy module programming, much less seek to develop a plan for revision. We do however note some lessons emerging from or reinforced by the present exercise:

- Programs should be written to handle complete data sets, even if the data currently available are incomplete. In particular, the energy module should be able to accept source data on tax rates for all (product, use) pairs, even if not data are not currently available for all such pairs.
- The early stage programs in the energy module should embody less special information about the source data. At present they embody knowledge about the various sources of data (the IEA, OLADE, TERI, *etc.*), the ranges (for example, which (product, use) pairs may have entries in OLADE, which in TERI), and their units (which sources use US gallons, which use barrels, and so on). In a few cases, individual data points are hard-coded. The energy module should be shielded from this detail; the source data should be regularised in pre-processing.
- In some programs ('`ep`', '`eporig`') there is an unsafe assumption that '`tablo`' implicitly initializes coefficients to zero. This should be corrected.
- Where data arrays are incomplete, this should be recorded not through the use of magic numbers as missing value indicators in the data arrays themselves but in separate arrays of data availability flags.

- Where possible, incomplete data sets they should be filled in before the data are used downstream. This simplifies the downstream programming, and allows for fully effective checking for internal consistency. In particular, the tax rate arrays should be fully determined in the energy module; no missing value flags should pass to the Fit module.

- Methods of filling in should be reviewed, to see whether we are using the best available proxies. In particular, the use of cross-country averages in "`BPE`" in '`ep`' appears questionable. As tables 24–26 show, it leads in the present case to some very uneven cross-use profiles.

- Currently some filling in is done implicitly in '`epreg`'; the map from GTAP standard countries to energy data countries involves some use of proxies (for example, it maps many GTAP African countries to Ghana). But the filling in is incomplete and seemingly unsystematic. This should be reviewed and regularized.

- For coal and for petroleum products, the energy module source data distinguish several different product classes. Yet from an early stage, in '`ep`', it averages prices and tax rates across those classes. It would be desirable to retain the distinction to a later point, so that for example steaming-coal-intensive and coking-coal-intensive users face different coal prices.

- As remarked before, in GTAP Board meetings and elsewhere, we should make more use of the contributed IO tables as energy data sources.

- Exploitation of the IO tables as a source of energy data should take place explicitly and in the energy module itself, not implicitly and downstream.

- If we use IO table tax rate data in the energy module, we should consider allowing tax rate differentiation across non-energy industries.

- We should drop the assumption of zero taxes on non-electricity energy industry use of non-electricity energy commodities. The Russian table shows taxes on "`p_c`" use of "`oil`"; this is perfectly plausible and should not be disallowed.

- The varying dimensions of the industry index in the energy targets are confusing (the targets for money value of use are indexed by energy target sector, the tax rate targets by GTAP sector). This might best be resolved by standardizing on energy target sectors.

- One of the greatest difficulties in interfacing the energy module with the rest of the GTAP data base construction process relates to gas. The energy source data recognize just one sector, "`gas`", whereas the GTAP sectoral classification also recognizes a gas distribution sector, "`gdt`". The energy module results for "`gas`" are used for both GTAP "`gas`" and "`gdt`", even though they are clearly inappropriate for "`gdt`". For example, "`gdt`" should clearly be priced higher than "`gas`". One way to handle this is to introduce a new IO data structure in which gas distribution is treated as a margin activity. We would use this structure in fitting the IO tables to the energy data; we would then convert the fitted tables back to the old structure before incorporating them into the GTAP state data file.