GTAPinGAMS: The Dataset and Static Model

Thomas F. Rutherford
University of Colorado *

October 23, 1998

1 Introduction

The Global Trade Analysis Project (GTAP) is a research program initiated in 1992 to provide the economic research community with a global economic dataset for use in the quantitative analyses of international economic issues. The Project's objectives include the provision of a documented, publicly available, global, general equilibrium data base, and to conduct seminars on a regular basis to inform the research community about how to use the data in applied economic analysis. GTAP has lead to the establishment of a global network of researchers who share a common interest of multi-region trade analysis and related issues. The GTAP research program is coordinated by Thomas Hertel, Director of the Center for Global Trade Analysis at Purdue University. As Deputy Director of this Center, Robert McDougall oversees the data base work. Software development within the GTAP project has been assisted greatly by the efforts of Ken Pearson and other Australian researchers from Centre of Policy Studies, Monash University. (See Hertel [1997], McDougall [1998]. A list of applications based on the GTAP framework can be found at the GTAP home page (http://www.agecon.purdue.edu/gtap/apps/).)

The standard programming language for GTAP data and modeling work has been GEMPACK (Harrison and Pearson [1996]). In the GEMPACK framework the model is solved as a system of nonlinear equations. The present paper describes a version of the GTAP model which has been implemented as a nonlinear complementarity problem in the GAMS programming language. Along with the core model I have developed several ancillary programs for dataset management. I call the package "GTAPinGAMS". These programs should be useful to economists who program in GAMS and wish to use GTAP in applied work. These programs include tools for translation of the GTAP files into GAMS readable form, GAMS programs for dataset aggregation, filtering and the imposition of alternative tax rates on trade or domestic transactions.

The GTAP version 4 database represents global production and trade for 45 country/regions, 50 commodities and 5 primary factors. The data characterize intermediate demand and bilateral trade in 1995, including tax rates on imports and exports. The core static model described in this paper does not have precisely the same structure as the GTAP model implemented in GEMPACK. There are several immediate differences between the standard GTAP model and the GTAPinGAMS model. First, the units of account are different by a factor of 10000. GTAP mea-

*Development of these tools has been supported by the Electric Power Research Institute and the United States Department of Energy in joint work with David Montgomery, Paul Bernstein and Thomas Hertel. I have had excellent research assistance from Mustafa Babiker and Miles Light. Robert McDougall, Gerard Malcolm, and Ken Pearson were helpful answering my questions about the GTAP database, and Alex Meeraus has sorted out many subtle points in GAMS. Earlier work with GTAP was supported by the World Bank in collaborative work with David Tarr and Glenn Harrison. Renger van Nieuwkoop, Randy Wigle, Joseph Francois and Jesper Jensen provided some very helpful editorial suggestions on an earlier draft. The usual disclaimers apply.

sures all transactions in millions of 1995$. GTAPinGAMS measure transactions in tens of billions of 1995$.[1]

Second, there is a potentially important difference concerns the structure of final demand. In the GEMPACK model final demand is represented by a constant-difference-elasticity (CDE) demand system whereas in the GAMS model final demand is Cobb-Douglas. Given differences in functional forms, even if benchmark value shares and reference prices are identical the two models may produce somewhat different estimates of policy changes due to differences in income and substitution elasticities.

A third set of differences concerns the representation of investment demand and global capital markets. The standard GTAP model assumes that a "global bank" allocates international capital flows in response to changes in regional rates of return. The GTAPinGAMS model makes the simplest possible assumptions regarding investment demand, international capital flows and the time path of adjustment: all of these variables are exogenously fixed at base year levels. [2]

I chose to design the core model with simple Leontief, Cobb-Douglas and constant-elasticity-of-substitution (CES) functional forms so that it's structure could be as transparent as possible. These choices reflect my belief that any application of the GTAP data to a specific policy question should involve the development of a model tailored to the issues, and therefore the purpose of the core model is largely to illustrate how the benchmark data are organized.

This paper consists of three sections following this overview. Section 2 presents the core static model using Mathiesen's format for the Arrow-Debreu model. This section provides notation and equations describing technology, preferences and equilibrium conditions.

Section 3 describes how the GTAP model can be expressed in GAMS, either as an MPSGE model or as a system of algebraic equations. This material provides a short but complete overview of how the technology and preferences are calibrated along with GAMS code which performs this task.

Section 4 has a practical perspective with step-by-step instructions on how to install the GTAPinGAMS package. The intent of this material is to provide as short as possible a learning curve for economists who wish to perform a few calculations using the GTAP dataset. This section describes ancillary GAMS programs which have been developed for use with the GTAP 4 dataset. These include GAMS libinclude programs which read and write GTAP header-array files[3]; FILTER.GMS, a GAMS program which removes small trade flows and intermediate demands from a GTAP dataset to increase sparsity and provide improved computational performance in large scale models; IMPOSE.GMS, a GAMS program which permits arbitrary adjustment of benchmark tax rates with least-squares recalibration; and GTAPAGGR.GMS, a GAMS program which aggregates any GTAP dataset to a smaller number of goods, factors or regions.

Distribution files for GTAPinGAMS are located as follows:

- A self-extracting archive with executable code.
  (http://robles.colorado.edu/˜tomruth/gtapingams/utils.exe.)
- A zip archive with GAMS code in the GTAPinGAMS directory.
  (http://robles.colorado.edu/˜tomruth/gtapingams/gtapgams.zip.)
- A PDF version of this document.
  (http://robles.colorado.edu/˜tomruth/gtapingams/gtapgams.pdf.)

[1] Scaling units assures better numerical precision in equilibrium calculations.

[2] In extensions of the core static model, the GTAPinGAMS framework can be readily employed to study adjustment paths, but a description of these techniques lies beyond the scope of the present paper. See Rutherford, Lau and Pahlke [1998] for a pedagogic introduction to dynamic general equilibrium analysis within the GAMS framework.

[3] These tools have been implemented with the assistance of Ken Pearson using modified versions of his SEEHAR.EXE and MODHAR.EXE programs.

- A GZIPped-Postscript version of this document.
(http://robles.colorado.edu/˜tomruth/gtapingams/gtapgams.psz)

2 GTAP in Mathiesen's Equilibrium Format

An Arrow-Debreu model concerns the interaction of consumers and producers in markets. Lars Mathiesen [1985] proposed a representation of this class of models in which two types of equations define an equilibrium: zero profit and market clearance. The corresponding variables defining an equilibrium are activity levels (for constant-returns-to-scale firms) and commodity prices.[4]

Commodity markets merge primary endowments of households with producer outputs. In equilibrium the aggregate supply of each good must be at least as great as total intermediate and final demand. Initial endowments are exogenous. Producer supplies and demands are defined by producer activity levels and relative prices. Final demands are determined by market prices.

Economists who have worked with conventional textbook equilibrium models can find Mathiesen's framework to be somewhat opaque because many quantity variables are not explicitly specified in the model. Variables such as final demand by consumers, factor demands by producers and commodity supplies by producers, are defined implicitly in Mathiesen's model. For example, given equilibrium prices for primary factors, consumer incomes can be computed, and given income and goods prices, consumers' demands can then be determined. The consumer demand functions are written down in order to define an equilibrium, but quantities demanded need not appear in the model as separate variables. The same is true of inputs or outputs from the production process: relative prices determine conditional demand, and conditional demand times the activity level represents market demand. Omitting decisions variables and suppressing definitional equations corresponding to intermediate and final demand provides significant computational advantages at the cost of a somewhat more complex model statement.

For concreteness I now turn to specific features of the GTAP model. The core model described here is a static, multi-regional model which tracks the production and distribution of goods in the global economy. In GTAP the world is divided into regions, and each region's final demand structure is portrayed by a representative agent who allocates expenditure across goods so as to maximize welfare, with fixed levels of investment and public output. Production incorporates intermediate inputs, and primary factors include skilled and unskilled labor, land, resources and physical capital. The dataset includes a full set of bilateral trade flows with associated transport costs, export taxes and tariffs.

In the following section, before writing down the equilibrium conditions per se, I describe production technology and producer choices. I then outline the structure of private and public final demand. Finally, I write down the zero profit and market clearance equations.

2.1 Production

In the GTAP model there are two types of produced commodities, goods produced for domestic markets and goods produced for export. In the base GTAPinGAMS model these goods are assumed to be imperfect substitutes produced as joint products with a constant elasticity of transformation.[5] Specifically, if D_{ir} is domestic output and X_{ir} is exports, then

$$Y_{ir} = \left[\alpha_{ir}^{Y} D_{ir}^{1+1/\eta} + \beta_{ir}^{Y} X_{ir}^{1+1/\eta}\right]^{1/(1+1/\eta)}$$

[4] Under a maintained assumption of perfect competition, Mathiesen may characterize technology as CRTS without loss of generality. Decreasing returns are accommodated through introduction of a specific factor, while increasing returns are inconsistent with the assumption of perfect competition. In this environment zero excess profit is consistent with free entry for atomistic firms producing an identical product.

[5] Model files in the GTAPinGAMS distribution accomodate an infinite elasticity of transformation between domestic and export markets as they are treated in the GTAP implementation in GEMPACK. For simplicity, my algebraic exposition in this paper focuses on the case in which the elasticity of transformation is finite.

where Y_{ir} is the activity level for good i in region r. Producers are competitive, implying that given a value of Y_{ir}, supplies to the domestic and export markets are given by:[6]

$$D_{ir} = Y_{ir}\ a_{ir}^D(p_{ir}^D, p_{ir}^X)$$

and

$$X_{ir} = Y_{ir}\ a_{ir}^X(p_{ir}^D, p_{ir}^X).$$

Inputs to production include primary factors and intermediate inputs. Intermediate demands are proportional to the level of activity, so the total intermediate demand for good i in region r is:

$$ID_{ir} = \sum_j Y_{jr} a_{ijr}.$$

In the core model we assume that all intermediate input coefficients (a_{ijr}) are fixed, unresponsive to price.[7]

Following Armington [1969] intermediate demand is represented as a composite of imported and domestic goods as imperfect substitutes. Thus, we have:

$$ID_{ir} = \left[\alpha_{ir}^I DI_{ir}^\rho + \beta_{ir}^I MI_{ir}^\rho\right]^{1/\rho}$$

in which DI_{ir} is domestic intermediate and MI_{ir} is imported intermediate demand.

A Cobb-Douglas production function relates activity level and factor inputs. Producers minimize unit cost given factor prices and applicable taxes. The factor demands solve:

$$\min \sum_f p_{fr}^F (1 + t_{fir}^F) FD_{fir} \quad \text{s.t.} \quad \psi_{ir} \prod_f FD_{fir}^{\theta_{fir}} = Y_{ir}$$

taking Y_{ir} as given. Linear homogeneity of the production function implies that factor demands may be expressed as the product of an activity level and compensated demand function depending on factor prices and factor taxes:

$$FD_{fir} = Y_{ir}\ a_{fir}^F(p_r^F, t_{ir}^F)$$

2.2 Public and Private Demand

Public sector output is assumed to represent a Cobb-Douglas aggregation of market commodities:

$$G_r = \Gamma_r \prod_i GD_{ir}^{\theta_{ir}^G}$$

[6] For the sake of brevity, I present functional forms explicitly but represent unit demand and supply functions in reduced form, e.g. $a_{ir}^D(p_{ir}^D, p_{ir}^X)$. The next section of the paper presents detailed specific functions in the GAMS/MCP implementation.

[7] There is no reason that this functional form should be employed in every study. For example, when we use the GTAP dataset to study energy and environmental issues, it is important to account for the nature of substitution possibilities among energy carriers as well as between energy and non-energy inputs to production; so in those applications a nested CES function is employed in which energy trades off against value-added with a non-zero elasticity of substitution.

As is the case for intermediate demand, an Armington aggregation of domestic and imported inputs defines public sector demand:

$$GD_{ir} = \left[\alpha_{ir}^{G} DG_{ir}^{\rho} + \beta_{ir}^{G} MG_{ir}^{\rho}\right]^{1/\rho}$$

Public sector output is exogenous, however the composition of public sector inputs responds to relative prices, gross of applicable tax, hence:

$$GD_{ir} = \overline{G}_r \; a_{ir}^{G}(p_{ir}^{D}, p_{ir}^{M}, t_{ir}^{G})$$

A representative agent determines final demand in each region. These consumers are endowed with primary factors, tax revenue, and an exogenously-specified net transfer from other regions. This income is allocated to investment, public demand and private demand. Investment and public output are exogenous while private demand is determined by utility maximizing behaviour. The utility function is Cobb-Douglas:

$$U_r = \sum_i \theta_{ir}^{C} \log(CD_{ir})$$

As in the case of intermediate and public demand, an Armington aggregation of domestic and imported inputs defines each commodity, so

$$CD_{ir} = \left[\alpha_{ir}^{C} DC_{ir}^{\rho} + \beta_{ir}^{C} MC_{ir}^{\rho}\right]^{1/\rho}$$

Aggregate final demand is then defined by regional expenditure and the unit price of aggregate of domestic and imported goods, gross of applicable tax:

$$CD_{ir} = \frac{\theta_{ir}^{C} \; M_r}{p_{ir}^{C}(1 + t_{ir}^{C})}$$

Regional expenditure (M_r) includes factor income, net capital flows and tax revenue, net of the cost of investment and public expenditure.

2.3 Bilateral Trade

There are three types of imports in the model: imports to intermediate demand (MI_{ir}), imports to public sector demand (MG_{ir}) and imports to final consumer demand (MC_{ir}). The maintained assumption is that while the aggregate import share may differ between these three functions, each of these shares have the same regional composition within the import aggregate. A CES aggregation across imports from different regions s forms the total import composite:

$$MI_{ir} + MG_{ir} + MC_{ir} = \left[\sum_s \alpha_{isr}^{M} M_{isr}^{\rho}\right]^{1/\rho}$$

Two tax margins and a transportation cost apply on bilateral trade in the model. Real transport costs are proportional to trade:

$$T_{irs} = \tau_{irs} M_{irs}$$

and these inputs are defined by a Cobb-Douglas aggregate of international transport inputs supplied by different countries:

$$\sum_{irs} T_{irs} = \psi_T \prod_{i,r} TD_{ir}^{\theta^T_{ir}}$$

It is helpful to think of international transportation margins as transportation services which are provided by perfectly competitive producers from different regions with an Armginton aggregation across services from different countries and an elasticity of substitution equal to unity. The technology providing transportation services exhibits constant returns to scale, so we can specify a price p^T representing the unit cost of transportation on all commodity trade flows.[8]

Bilateral trade flows are determined by cost-minimizing choice, given the fob export price from region r, p^X_{ir}, the export tax rate, t^X_{ir}, and the import tariff rate, t^M_{ir}.[9] We then may write the demand for bilateral imports as:

$$M_{irs} = M_{is}\ a^M_{irs}(p^X_{ir'}, t^X_{ir's}, p^T, t^M_{ir's})$$

2.4 Income and Expenditure

Consumer expenditure for a representative agent are the sum of factor earnings and tax revenue, net the cost of investment, public sector output and net capital outflows:

$$\begin{array}{rll}
M_r = & \sum_f p^F_{fr}\ F_{fr} & \text{! factor income} \\
& + \sum_i t^Y_{ir}\ (p^D_{ir} D_{ir} + p^X_{ir} X_{ir}) & \text{! indirect taxes} \\
& + \sum_{ij} t^{ID}_{ijr}\ p^{ID}_{ir}\ Y_{jr}\ a_{ijr} & \text{! taxes on intermediate goods} \\
& + \sum_{fi} t^F_{fir}\ p^F_{fr}\ FD_{fir} & \text{! factor tax revenue} \\
& + \sum_i t^G_{ir}\ p^{GD}_{ir}\ GD_{ir} & \text{! public tax revenue} \\
& + \sum_i t^C_{ir}\ p^{CD}_{ir}\ CD_{ir} & \text{! consumption tax revenue} \\
& + \sum_{is} t^X_{irs}\ p^X_{ir}\ M_{irs} & \text{! export tax revenue} \\
& + \sum_{is} t^M_{isr}\ (p^X_{is}\ M_{isr}\ (1 + t^X_{isr}) + p^T T_{isr}) & \text{! tariff revenue} \\
& - \sum_i p^D_{ir} I_{ir} & \text{! investment demand} \\
& - \sum_i p^G_{ir}\ (1 + t^G_{ir})\ GD_{ir} & \text{! public sector demand} \\
& - p^C_n B_r & \text{! current account balance}
\end{array}$$

Capital flows in the base year are represented by B_r in this expression, and in a counterfactual equilibrium these are held fixed and denominated in terms of the numeraire price index, the consumer price level in region n (USA).

2.5 Market Clearance

Having defined technology, preferences and compensated demand functions, we now may turn to the market clearance conditions.

[8] There are some simplifications here. For example, the regional composition of transportation services is identical across all bilateral trade flows. Furthermore, while the dataset incorporates explicit trade and transport margins on international trade flows, wholesale and retail margins on domestic sales are ignored in the dataset, so there is some asymmetry in the database's price level.

[9] The model formulation assumes that the export tax applies on the fob price (net of transport margins), while the import tariff applies on the cif price, gross of export tax and transport margin.

- Domestic Output

  Domestic output equals demand for intermediate inputs to production, public sector use, final consumer demand plus domestic investment:[10]

$$
\begin{array}{lllll}
D_{ir} & = DI_{ir} & +DG_{ir} & +DC_{ir} & +I_{ir} \\
 & = ID_{ir}\ a_{ir}^{D,I} & +GD_{ir}\ a_{ir}^{D,G} & +CD_{ir}\ a_{ir}^{D,C} & +I_{ir}
\end{array}
$$

  in which $a_{ir}^{D,I}$, $a_{ir}^{D,G}$, and $a_{ir}^{D,C}$ represent the compensated demands for domestic inputs by submarket, each of which are functions of p_{ir}^{D} and p_{ir}^{M}.

- Imports

  Aggregate supply of imports, defined by the Armington aggregation across imports from different regions must equal aggregate import demand for intermediate, public and private consumption:

$$
\begin{array}{llll}
M_{ir} & = MI_{ir} & +MG_{ir} & +MC_{ir} \\
 & = ID_{ir}\ a_{ir}^{M,I} & +GD_{ir}\ a_{ir}^{M,G} & +CD_{ir}\ a_{ir}^{M,C}
\end{array}
$$

  in which $a_{ir}^{M,I}$, $a_{ir}^{M,G}$, and $a_{ir}^{M,C}$ represent compensated demands for imported inputs by submarket, each functions of p_{ir}^{D} and p_{ir}^{M}.

- Exports

  Export supplies equals import demand across all trading partners plus demands for international transport:[11]

$$
\begin{array}{lll}
X_{ir} & = \sum_s M_{irs} & +TD_{ir} \\
 & = \sum_s M_{is} a_{irs}^{M} & +T\ a_{ir}^{T}
\end{array}
$$

  In the second equation a_{irs}^{M} represents the unit demand for region r output per unit of region s aggregate imports.

- Armington Aggregate Supply

  The model includes supply-demand conditions for the Armington composite goods entering intermediate demand, public and private demand, as has already been specified above in the equations defining ID_{ir}, GD_{ir} and CD_{ir}.

- Primary Factors

  Primary factor (labor, capital, land, resource) endowment equals primary factor demand:

$$
F_{fr} = \sum_i Y_{ir}\ a_{fir}^{F}
$$

[10] Within the dataset investment inputs flow to the `cgd` sector, and demand for `cgd` sectoral output appears as the sole non-zero in the I_{ir} vector for each region r.

[11] When the elasticity of transformation between goods produced for the domestic and export markets is infinte, the market clearance conditions for D_{ir} and X_{ir} are merged, i.e.

$$
Y_{ir} = DI_{ir} + DG_{ir} + DC_{ir} + I_{ir} + \sum_s M_{irs} + TD_{ir}.
$$

and prices p_{ir}^{D} and p_{ir}^{X} are replaced throughout the model by a single price index, p_{ir}^{Y}.

2.6 Zero Profit

- Production

  Competitve producers operating constant-returns technology earn zero profit in equilibrium. For the GTAP producer, the value of output to the firm equals the value of sales in the domestic and export markets net of applicable indirect taxes. Costs of production include factor inputs (taxed at rate t^F) and intermediate inputs (taxed at rate t^{ID}):

$$(p_{ir}^D\ a_{ir}^D + p_{ir}^X\ a_{ir}^X)(1 - t_{ir}^Y) = \sum_f a_{fir}^F p_{fr}^F (1 + t_{fir}^F) + \sum_j a_{jir} p_{jr}^{ID} (1 + t_{jir}^{ID})$$

- Imports

  Zero profit conditions apply to trade activities as well as production. In equilibrium, the value of imports at the domestic cif price therefore equals the fob price gross of export tax, the transportation margin and the applicable tariff:

$$p_{ir}^M = \sum_s a_{irs}^M \left[p_{is}^X (1 + t_{isr}^X) + \tau_{irs} p^T \right] (1 + t_{isr}^M)$$

- Intermediate, Public and Private Demand

  Armington aggregation functions transform domestic and imported goods into composite goods for intermediate demand, public sector demand and private demand. Zero profit for these activites provide the following equilibrium identities:

$$p_{ir}^I = c(p_{ir}^D, p_{ir}^M, \alpha_{ir}^I, \beta_{ir}^I)$$

$$p_{ir}^G = c(p_{ir}^D, p_{ir}^M, \alpha_{ir}^G, \beta_{ir}^G)$$

$$p_{ir}^C = c(p_{ir}^D, p_{ir}^M, \alpha_{ir}^C, \beta_{ir}^C)$$

  in which

$$\begin{aligned} c(p^D, p^M, \alpha, \beta) &\equiv \min_{D,M} \quad p^D D + p^M M \quad \text{s.t.} \quad (\alpha D^\rho + \beta M^\rho)^{1/\rho} = 1 \\ &= \left(\alpha^\sigma p_D^{1-\sigma} + \beta^\sigma p_M^{1-\sigma} \right)^{1/1-\sigma} \end{aligned}$$

  is the unit cost function defined by the constant-elasticity-of-substitution aggregate of domestic and imported inputs.

3 The Dataset and Static Model

This section provides an overview of the GTAP model as it is represented in GAMS. I begin by describing the dataset including parameters which are stored and those which are assigned. Users accustomed to working with the GTAP dataset in its original GEMPACK implementation should be forewarned of significant differences between the GTAP dataset as it is stored in GEMPACK and how it is represented in GAMS. I have made a number of changes in the dataset structure to give greater prominence to benchmark tax rates, as well as to minimize the number of bytes required to stored the data on disk.

After describing the benchmark data, I go through the MPSGE model. This presentation includes a fair amount of explanatory text so that I hope it will be comprehensible to non-MPSGE programmers. Thereafter follows a description of the model as it may be expressed in GAMS algebra as an MCP model.[12]

3.1 The Dataset

An important feature of GTAPinGAMS package is that datasets may be freely aggregated into fewer regions, fewer sectors and even fewer primary factors. This feature permits a modeller to do preliminary model development using a small dataset to ensure rapid response and a short debug cycle. After having implemented a small model, it is then a simple matter to expand the number of sectors and/or regions in order to obtain a more precise empirical estimate.

All GTAP datasets are defined in terms of three primary sets: i, the set of sectors and produced commodities, r the set of countries and regions, and f the set of primary factors. Table 1 presents the identifiers for the 45 GTAP 4 sectors in their most disaggregate form. These are more-or-less identical to the GEMPACK model apart from the replacement of `for` in the original dataset by `frs` in this one. ("`for`" became a reserved keyword beginning with GAMS release 2.25.) These sectors may be aggregated freely to produce more compact datasets with one restriction: sector `cgd` must appear as a distinct sector in any aggregation.

Table 2 presents regional identifiers in the full dataset many of which correspond to standard UN three-character country codes[13]. Table 3 presents the three-character identifiers used for primary factors. Note that these differ from the primary factor names employed in the GEMPACK model.

GAMS code which declares all parameters in a GTAP dataset is shown in Table 4. The parameters beginning with `v` are base year (1995) value data, most of which are from the original GEMPACK implementation of GTAP. Not all value data from the original dataset are included here. The principal difference is that this dataset stores tax *rates* rather than gross and net of tax transaction values. The tax parameters, beginning with `t` are not in the original GEMPACK dataset.

Whenever a GTAP dataset is read additional intermediate parameter values are assigned. Declarations for the computed parameters are presented in Table 5. Table 6 lists the GAMS parameter assignment statements for the computed items. Briefly, this is done as follows: (i) aggregate exports at market prices (`vxm`) are defined from the matrix of bilateral trade flows; (ii) aggregate imports at market prices (`vim`) are defined by bilateral exports, export taxes, transportation margins and tariff rates; (iii) domestic output (`vdm`) is determined as a residual through the zero profit condition; (vi) domestic supply to the intermediate demand (`vdfm`) is defined as a residual given domestic production and other demands for domestic output; (vii) import supply to intermediate demand (`vifm`) is also defined as a residual given aggregate imports, private and public import demand. This sequence of assignments implies that any imbalance in the dataset shows up as either a discrepancy in the demand and supply for intermediate inputs or as an imbalance between

[12] The distribution files provide representations of the core model as a constrained nonlinear system (CNS) and a square system of nonlinear constraints within a conventional nonlinear program (NLP).

[13] Users can define their own aggregations of the GTAP data and use any labels to describe regions. For technical reasons, if a GTAP dataset is to be used with MPSGE, then regional identifiers can have at most 4 characters.

Table 1: Sectoral Identifiers in the Full GTAPinGAMS Dataset

```
SET   i   Sectors /

PDR     Paddy rice,
WHT     Wheat,
GRO     Grains (except rice-wheat),
V_F     Vegetable fruit nuts,
OSD     Oil seeds,
C_B     Sugar cane and beet,
PFB     Plant-based fibers,
OCR     Crops n.e.c.,
CTL     Bovine cattle,
OAP     Animal products n.e.c.,
RMK     Raw milk,
WOL     Wool,
FRS     Forestry,
FSH     Fishing,
COL     Coal,
OIL     Oil,
GAS     Natural Gas,
OMN     Other Minerals,
CMT     Bovine cattle meat products,
OMT     Meat products n.e.c.,
VOL     Vegetable oils,
MIL     Dairy products,
PCR     Processed rice,
SGR     Sugar,
OFD     Other food products,
B_T     Beverages and tobacco,
TEX     Textiles,
WAP     Wearing apparel,
LEA     Leather goods,
LUM     Lumber and wood,
PPP     Pulp and paper,
P_C     Petroleum and coal products,
CRP     Chemicals  rubber and plastics,
NMM     Non-metallic mineral products,
I_S     Primary ferrous metals,
NFM     Non-ferrous metals,
FMP     Fabricated metal products,
MVH     Motor vehicles,
OTN     Other transport equipment,
ELE     Electronic equipment,
OME     Machinery and equipment,
OMF     Other manufacturing products,
ELY     Electricity,
GDT     Gas manuf. and distribution,
WTR     Water,
CNS     Construction,
T_T     Trade and transport,
OSP     Other services (private),
OSG     Other services (public),
DWE     Dwellings,
CGD     Savings good/;
```

Table 2: Regional Identifiers in the Full GTAPinGAMS Dataset

```
SET   r   Regions /

AUS   Australia,                         ARG   Argentina,
NZL   New Zealand,                       BRA   Brazil,
JPN   Japan,                             CHL   Chile,
KOR   Republic of Korea,                 URY   Uruguay,
IDN   Indonesia,                         RSM   Rest of South America,
MYS   Malaysia,                          GBR   United Kingdom,
PHL   Philippines,                       DEU   Germany,
SGP   Singapore,                         DNK   Denmark,
THA   Thailand,                          SWE   Sweden,
VNM   Vietnam,                           FIN   Finland,
CHN   China,                             REU   Rest of EU,
HKG   Hong Kong,                         EFT   European Free Trade Area,
TWN   Taiwan,                            CEA   Central European Associates,
IND   India,                             FSU   Former Soviet Union,
LKA   Sri Lanka,                         TUR   Turkey,
RAS   Rest of South Asia,                RME   Rest of Middle East,
CAN   Canada,                            MAR   Morocco,
USA   United States of America,          RNF   Rest of North Africa,
MEX   Mexico,                            SAF   South Africa,
CAM   Central America and Caribbean,     RSA   Rest of South Africa,
VEN   Venezuela,                         RSS   Rest of Sub-Saharan Africa,
COL   Columbia,                          ROW   Rest of World  /;
RAP   Rest of Andean Pact,
```

Table 3: Primary Factor Identifiers in the Full GTAPinGAMS Dataset

```
SET    f      Primary factors /
              LND     Land,
              SKL     Skilled labor,
              LAB     Unskilled labor,
              CAP     Capital,
              RES     Natural resources /
```

Table 4: Parameters Explicitly Represented in a GTAPinGAMS Dataset

```
alias (i,j), (r,s);

PARAMETER
        ty(i,r)          Output tax
        ti(j,i,r)        Intermediate input tax
        tf(f,i,r)        Factor tax
        tx(i,s,r)        Export tax rate (defined on a net basis)
        tm(i,s,r)        Import tariff rate
        tg(i,r)          Tax rates on government demand
        tp(i,r)          Tax rate on private demand

        vafm(j,i,r)      Aggregate intermediate inputs
        vfm(f,i,r)       Value of factor inputs (net of tax)
        vxmd(i,r,s)      Value of commodity trade (fob - net export tax)
        vtwr(i,r,s)      Transport services
        vst(i,r)         Value of international transport sales
        vdgm(i,r)        Government demand (domestic)
        vigm(i,r)        Government demand (imported)
        vdpm(i,r)        Aggregate private demands (domestic)
        vipm(i,r)        Aggregate private demands (domestic);
```

Table 5: Computed Benchmark Parameters

```
parameter
        vim(i,r)         Total value of imports (gross tariff)
        vxm(i,r)         Value of export (gross excise tax)
        vdm(i,r)         Value of domestic output (net excise tax)
        vdfm(i,r)        Aggregate intermediate demand (domestic)
        vifm(i,r)        Aggregate intermediate demand (imported)
        vom(i,r)         Aggregate output value (gross of tax)
        vgm(i,r)         Public expenditures
        vpm(i,r)         Private expenditures
        vg(r)            Total value of public expenditure
        vp(r)            Total value of private expenditure
        vi(r)            Total value of investment
        vt               Value of international trade margins
        vb(*)            Net capital inflows
        market(*,*)      Consistency check for calibrated benchmark
        evoa(f,r)        Value of factor income
        va(d,i,r)        Armington supply
        vd(d,i,r)        Domestic supply
        vm(d,i,r)        Imported supply;
```

Table 6: Assignments for Computed Benchmark Parameters

```
vxm(i,r) = sum(s, vxmd(i,r,s)) + vst(i,r);

vim(i,r) = sum(s,(vxmd(i,s,r)*(1+tx(i,s,r))+vtwr(i,s,r))*(1+tm(i,s,r)));

vdm(i,r) = ( sum(j, vafm(j,i,r)*(1+ti(j,i,r)))
           + sum(f,  vfm(f,i,r)*(1+tf(f,i,r)))) / (1-ty(i,r)) - vxm(i,r);

vdfm(i,r) = vdm(i,r)  - vdgm(i,r) - vdpm(i,r) - vdm(i,r)$cgd(i);

vi(r) = sum(cgd, vdm(cgd,r));

vifm(i,r) = vim(i,r) - vipm(i,r) - vigm(i,r);

vom(i,r) = vdm(i,r) + vxm(i,r);

vgm(i,r) = vigm(i,r)+vdgm(i,r);

vpm(i,r) = vipm(i,r)+vdpm(i,r);

vg(r) = sum(i, vgm(i,r) * (1 + tg(i,r)));

vp(r) = sum(i, vpm(i,r) * (1 + tp(i,r)));

vt = sum((i,r), vst(i,r));

evoa(f,r) = sum(i, vfm(f,i,r));

vb(r) = vp(r) + vg(r) + vdm("cgd",r)
        - sum(f, evoa(f,r))
        - sum(i,     ty(i,r)   * vom(i,r))
        - sum((i,j), ti(j,i,r) * vafm(j,i,r))
        - sum((i,f), tf(f,i,r) * vfm(f,i,r))
        - sum((i,s), tx(i,r,s)  * vxmd(i,r,s))
        - sum((i,s), tm(i,s,r)  * (vxmd(i,s,r)*(1+tx(i,s,r)) + vtwr(i,s,r)) )
        - sum(i,     tg(i,r)*vgm(i,r))
        - sum(i,     tp(i,r)*vpm(i,r));

vm("c",i,r) = vipm(i,r);        vd("c",i,r) = vdpm(i,r);
vm("g",i,r) = vigm(i,r);        vd("g",i,r) = vdgm(i,r);
vm("i",i,r) = vifm(i,r);        vd("i",i,r) = vdfm(i,r);
va(d,i,r) = vm(d,i,r) + vd(d,i,r);
market(r,i) = vdfm(i,r) + vifm(i,r) - sum(j, vafm(i,j,r));
market("world","t") = vt - sum((i,r,s), vtwr(i,r,s));
```

Table 7: Benchmark Prices

```
parameter

        pc0(i,r)        Reference price index for private consumption
        pf0(f,i,r)      Reference price index for factor inputs
        pg0(i,r)        Reference price index for public
        pi0(j,i,r)      Reference price index for intermediate inputs
        pt0(i,s,r)      Reference price index for transport
        px0(i,s,r)      Reference price index for imports;

px0(i,s,r) = (1+tx(i,s,r))*(1+tm(i,s,r));
pt0(i,s,r) = 1+tm(i,s,r);
pc0(i,r)   = 1+tp(i,r);
pg0(i,r)   = 1+tg(i,r);
pi0(j,i,r) = 1+ti(j,i,r);
pf0(f,i,r) = 1+tf(f,i,r);
```

demand and supply of transportation services. The parameter `market` is created to generate a report of consistency of the benchmark data. (Primary factor markets always balance because endowments are computed residually given benchmark factor demands across sectors. Likewise, regional current account balances are computed from the income-expenditure identity.)

Table 7 lists declarations and assignments of reference prices for each of the benchmark transactions which are subject to tax. These parameters are used in the MPSGE and MCP models as part of the calibration of demand functions. Share parameters used solely in the MCP model are not included here.

It is a matter of personal taste in mathematics and computing, but I generally use one or two character identifiers in an algebraic exposition while employing GAMS parameters with as many as 10 characters. In order to avoid potential confusion due to differences in notation, Table 8 gives a cross-reference of symbols used in the algebraic formulation in this paper to the GAMS parameters which define the benchmark value of these variables in the GTAPinGAMS dataset.

Table 8: Algebraic Symbols and Related Benchmark Parameters

Symbol	**GAMS Parameter**
t^{Y}_{ir}	`ty(i,r)`
t^{ID}_{ir}	`ti(j,i,r)`
t^{F}_{ir}	`tf(f,i,r)`
t^{X}_{isr}	`tx(i,s,r)`
t^{M}_{isr}	`tm(i,s,r)`
t^{G}_{ir}	`tg(i,r)`
t^{C}_{ir}	`tp(i,r)`
$Y_{ir}a_{jir}$	`vafm(j,i,r)`
FD_{fir}	`vfm(f,i,r)`
M_{irs}	`vxmd(i,r,s)`
T_{irs}	`vtwr(i,r,s)`
TD_{ir}	`vst(i,r)`
DG_{ir}	`vdgm(i,r),`
MG_{ir}	`vigm(i,r)`
DC_{ir}	`vdpm(i,r)`
MC_{ir}	`vipm(i,r)`
X_{ir}	`vxm(i,r)`
D_{ir}	`vdm(i,r)`
DI_{ir}	`vdfm(i,r)`
MI_{ir}	`vifm(i,r)`
CD_{ir}	`vpm(i,r)`
GD_{ir}	`vgm(i,r)`
MI_{ir}	`vm("i",i,r)`
MG_{ir}	`vm("g",i,r)`
MC_{ir}	`vm("c",i,r)`
DI_{ir}	`vm("i",i,r)`
DG_{ir}	`vm("g",i,r)`
DC_{ir}	`vm("c",i,r)`
B_{r}	`B(r)`
τ_{irs}	`vtwr(i,r,s)/vxmd(i,r,s)`

Table 9: Variable Declarations for GTAP Implemented in MPSGE

```
$MODEL:GTAP

$SECTORS:
        Y(i,r)          ! Output
        A(d,i,r)        ! Armington aggregation of domestic and imports
        M(i,r)          ! Import aggregation
        C(r)            ! Private consumption
        G(r)            ! Public provision
        YT              ! Transport

$COMMODITIES:
        PG(r)           ! Public provision
        PC(r)           ! Private demand
        PD(i,r)         ! Output price
        PX(i,r)         ! Export price
        PM(i,r)         ! Import price
        PA(d,i,r)       ! Armington composite price
        PF(f,r)         ! Factor price
        PT              ! Transport services

$CONSUMERS:
        RA(r)           ! Representative agent
```

3.2 The MPSGE Formulation

Table 9 contains variable declarations for the GTAPinGAMS model as implemented in MPSGE.[14] The model includes sectors related to production by commodity and region (`Y(i,r)`); Armington aggregatation across imports from different trading partners (`M(i,r)`); Armington aggregation between domestic and imported varieties by market segment (`A(d,i,r)` in which `d` refers to intermediate, public and private demand); public demand by region (`G(r)`); private demand by region (`C(r)`); and the provision of international transport margins (`YT`).

The production activities for public and private demand are associated with outputs which represent the marginal cost of public and private consumption, `PG(r)` and `PC(r)`. For each commodity and region there are six different price indices. `PD(i,r)` represents the cost index for a unit of domestic output; `PX(i,r)` represents the cost index for exports; `pm(i,r)` represents the cost of a unit of imports (aggregated across all trading partners), and `PA(d,i,r)` represents the cost index of a unit of composite Armington supply by submarket. Primary factor prices are represented by `PF(f,r)`, and the market price of a unit of international transport services is represented by `PT`.

The final class of variables in the MPSGE model are consumers, and in this model there is one representative consumer for each region. In a solution `RA.L(r)` returns the equilibrium expenditure on household consumption by region `r`.

An MPSGE model is specified by a sequence of function "blocks", one for each production sector and consumer in the model. In this model there are six *classes* of production sectors. (See Table 9.) The first of these refers to production by commodity and region, `Y(i,r)`. This production

[14] I have omitted exception operators from the variable and function declarations to make the code easier to read. In most aggregations of the dataset, the model shown here is operational. In highly disaggregate models, however, not all goods are produced in all regions, and it is necessary to specify, for example, `Y(i,r)$(vdm(i,r)+vxm(i,r))`.

activity has a nested-CES cost function with a Leontief aggregation across intermediate inputs at the top level (see `s:0`) and unity within the value-added aggregate (`va:1`), and an elasticity of transformation across outputs equal to 2 (see `T:2`). There are inputs and outputs associated with the `Y(i,r)` production function. Outputs correspond to production for the domestic market, `O:PD(i,r)`, and production for the export market, `O:PX(i,r)`. The reference quantity entries for these coefficients are the benchmark values of domestic and export sales. A tax at an ad-valorem rate `ty(i,r)` is applied to both domestic and export sales.[15]

There are two types of inputs to the `Y(i,r)` production function, corresponding to goods and factors. Intermediate inputs are taken from the market for Armington aggregates into production. Substitution between factor inputs is created by assigning those inputs to the `va:` input aggregate.[16]

Taxes are levied on intermediate demand inputs at net rate `ti` and taxes apply to primary factor inputs at net rate `tf`. For example, the market value of primary factor services purchased by firms is `vfm(f,i,r)`, but the total cost to firms equals `vfm(f,i,r)*(1+tf(f,i,r))`, of which `vfm(f,i,r)` is paid as wages or dividends to factor owners while `vfm(f,i,r)*tf(f,i,r)` is paid as a tax to `RA(r)`.[17]

The Armington aggregation activity `A(d,i,r)` generates three functions for each commodity in each region. For simplicity I have specified a domestic-import elasticity of substitution equal to 4 for all goods, commodities and Armington submarkets.

The import aggregation activity, `M(i,r)`, is the most complex component of the model. First, it defines the aggregation of imports by trading partner. Second, it applies export taxes and import tariffs on all bilateral trades.[18] Third, it applies transportation margins which are proportational to quantities traded. The output market `PM(i,r)` serves as an input to the Armington aggregation sectors. There are two types of inputs to the `M(i,r)` activity. The `I:PX(i,s)` input represents fob payments to producers in region `s`.

The `I:PT#(s)` input represents multiple inputs of transportation services, one for each element of set `s`. There are multiple inputs of transportation services into each imported good simply because every bilateral trade flow demands its own transportation services. Using a Leontief aggregate on each bilateral trade flow assures that transport costs and imports remain strictly proportional to the base year level, $\tau(i, r, s) = \mathrm{vtwr(i,s,r)/vxmd(i,s,r)}$.

The function declaration indicates a top-level substitution elasticity equal to `esubmm` (`S:esubmm`), and it also indicates a *vector* of second level input nests, each with an elasticity of substitution equal to zero (`s.tl:0`).[19]

Final consumption by consumers and producers in region `r` is characterized by production activities `c(r)` and `g(r)`, respectively. The elasticity of substitution across goods in final demand is specified to be unity (`S:1`).

The `yt` production activity provides international transportation services as a Cobb-Douglas

---

[15]The output tax is defined on a *gross basis*. For example, the value of sales in the domestic market gross of tax equals `vdm(i,r)` of which `(1-ty(i,r))*vdm(i,r)` is returned to producers and `ty(i,r)*vdm(i,r)` is paid to the government.

[16]"`va:`" is a nesting idetifier. These names are arbitary and may have from one to four characters. Two reserved names are "`s:`" which represents the elasticity of substitution at the root of the inputs tree and "`T:`" which represents the elasticity of transformation at the root of the output tree.

[17]Readers unfamiliar with the MPSGE model representation may wish to refer back to the algebraic equilibrium conditions. The specification of the `$PROD:Y(I,R)` block automatically generates a zero profit condition for Y_{ir}. It also generates terms in the market clearance equations for all associated inputs and outputs. In this function the affected markets include the domestic output market, the market for export of good i from region r, markets for Armington composites entering intermediate demand, and primary factors markets. For this reason the tabular format is very compact – in essence, the user only needs to specify the dual (zero-profit) conditions and the modeling language automatically generates the primal (market clearance) equations.

[18]Note that export taxes on sales from region `s` in region `r` are accrue to the representative agent in region `s` (`A:RA(s)`) while import tariffs are paid to the representative agent in region `r` (`A:RA(r)`).

[19]A `.tl` suffix alerts MPSGE that a set of nests are being declared. When an input is to be associated with one of these nests, the set label flag must be specified on the input line.

Table 10: Function Declarations for GTAP Implemented in MPSGE

```
$PROD:Y(i,r)  S:0  T:eta  va:1
        O:PD(i,r)       Q:vdm(i,r)    A:RA(r) T:ty(i,r)
        O:PX(i,r)       Q:vxm(i,r)    A:RA(r) T:ty(i,r)
        I:PA("i",j,r)   Q:vafm(J,i,r) P:pi0(j,i,r) A:RA(r) T:ti(j,i,r)
        I:PF(f,r)       Q:vfm(f,i,r)  P:pf0(f,i,r) A:RA(r) T:tf(f,i,r) va:

$REPORT:
        V:FD(f,i,r)     I:PF(f,r)       PROD:Y(i,r)
        V:YD(i,r)       O:PD(i,r)       PROD:Y(i,r)
        V:YX(i,r)       O:PX(i,r)       PROD:Y(i,r)

$PROD:A(d,i,r)          S:esubdm
        O:PA(d,i,r)     Q:va(d,i,r)
        I:PD(i,r)       Q:vd(d,i,r)
        I:PM(i,r)       Q:vm(d,i,r)

$PROD:M(i,r)            S:esubmm   s.TL:0
        O:PM(i,r)       Q:vim(i,r)
        I:PX(i,s)       Q:vxmd(i,s,r)  P:px0(i,s,r)     s.TL:
+                       A:RA(S) T:TX(i,s,r) A:RA(r) T:(tm(i,s,r)*(1+tx(i,s,r)))
        I:PT#(s)        Q:vtwr(i,s,r)  P:pt0(i,s,r)     s.TL:
+                       A:RA(r) T:tm(i,s,r)

$PROD:G(r)  S:1
        O:PG(r)         Q:vg(r)
        I:PA("g",i,r)   Q:vgm(i,r)  P:pg0(i,r)  A:RA(r) T:tg(i,r)

$PROD:C(r)  S:1
        O:PC(r)         Q:vp(r)
        I:PA("c",i,r)   Q:vpm(i,r)  P:pc0(i,r)  A:RA(r) T:tp(i,r)

$PROD:YT  S:1
        O:PT            Q:vt
        I:PX(i,r)       Q:vst(i,r)

$DEMAND:RA(r)
        E:PF(f,r)       Q:evoa(f,r)
        E:PC(num)       Q:vb(r)
        E:PD(cgd,r)     Q:(-vi(r))
        E:PG(r)         Q:(-vg(r))
        D:PC(r)         Q:vp(r)
```

Table 11: Computing Demand Quantities from an MPSGE Equilibrium

```
parameter       cd(i,r) Private demand
                gd(i,r) Public demand
                td(i,r) Transportation demand;

cd(i,r) = vpm(i,r) * C.L(r) * PC.L(r) * pc0(i,r)
                        / ( PA.L("c",i,r) * (1 + tp(i,r)) );
gd(i,r) = vgm(i,r) * G.L(r) * PG.L(r) * pg0(i,r)
                        / ( PA.L("g",i,r) * (1 + tg(i,r)) );
td(i,r) = vst(i,r) * YT.L * PT.L / PX.L(i,r);
```

composite of goods provided in the domestic markets of each region.

The model statement concludes with a specification of endowments and preferences for each region's representative agent (`$DEMAND:RA(r)`). Each agent is endowed with primary factors and capital inflows. They are also "endowed" with a fixed negative quantities of the domestic `cgd` commodity and public sector outputs representing exogenously-specfied demands for investment and public sector output. All remaining income is allocated to private consumption.

The closure adopted here in which investment and public demand are both exogenous is adopted for simplicity, and also because the welfare estimates from this closure seem to most closely match the infinite-horizon model (see Rutherford and Tarr 1998). In the MPSGE model it is quite simple to adopt alternative assumptions regarding investment. For example, investment could be modeled by a constant marginal propensity to save as:

```
$DEMAND:RA(r)  s:1
        E:PF(f,r)       Q:evoa(f,r)
        E:PC(num)       Q:vb(r)
        E:PG(r)         Q:(-vg(r))
        D:PD(cgd,r)     Q:vi(r)
        D:PC(r)         Q:vp(r)
```

As stated above, the MPSGE formulation of an equilibrium model follows Mathiesen's modeling format in which intermediate demand and supply functions can be captured as functions of prices and activity levels. The computational advantage of this approach is that fewer variables are needed, and it is considerably less costly to solve the resulting smaller system of equations.[20]

Intermediate demands and supplies from MPSGE models can be computed by the modeller using equilibrium prices and activity levels. For example, in this model it is quite simple to compute private, public and transportation demands from the solution of an MPSGE model, because all of these activities are Cobb-Douglas. (See Table 11.) It is, however, possible to extract equilibrium demands directly from the MPSGE function evaluation through use of the `$REPORT:` statement, listed in Table 10 immediately after the Y_{ir} production block. In this model three demand and supply quantity variables are declared, representing primary factor demand by sector, supply to the domestic market and supply to the export market. These values are returned in `FD.L(f,i,r)`, `YD.L(i,r)`, and `YX.L(i,r)` respectively.

[20] In terms of compuational complexity, the cost of solving a system of equations increases somewhere between the square and the cube of the number of dimensions, although in large-scale implementations such as the GAMS/MCP solver PATH or MILES, computational complexity depends on both the number of equations and their density.

3.3 The Algebraic Formulation

Compactness of representation leaves fewer opportunities to make mistakes. For this reason alone I prefer to implement equilibrium models with MPSGE. I use algebraic models for teaching, or occasionally I write out a model in full algebraic form in order to introduce functional forms which are unavailable in MPSGE or in order to apply joint maximization as a solution method for large scale models with an imbedded linear program.

I recognize, however, that the tabular MPSGE syntax can be impenetrable for many competent modellers, therefore I conclude my discussion of the core GTAPinGAMS model by going through the implementation in GAMS algebra. I pose the model here as a mixed complementarity problem, but in this formulation all of the market clearance and zero profit conditions will hold with equality, so the model can be solved in GAMS as a nonlinear program (NLP) with a vacuous objective or a constrained nonlinear system (CNS).

In order to write out GTAP in algebraic form it is helpful to introduce some additional benchmark data structures which simplify demand function algebra. The extra parameters include benchmark value shares for all of the nonlinear demand and supply functions in the model. (See Table 12.)

The variables used in the GAMS/MCP model are listed in Table 13, separated by blank lines into four subsets. The first set of variables are unit demand and supply functions, corresponding to the symbols from the algebraic formulation above. (See Table 8.) Demand and supply functions are represented implicitly in the MPSGE model, but for simplicity these are introduced as separate symbols in the algebraic model. The remaining three subsets of variables in the model correspond precisely to the sectors, commodities and consumers in the MPSGE model.

Table 14 presents equations defining the unit demand and supply functions. Using benchmark quantities, prices and value shares, I use the calibrated share form to express demands as a function of input prices. (See Rutherford 1998, Chapter 3). In order to keep track of what is what, I am following Michael Saunders' suggestions for GAMS program style, listing variables and GAMS variables in upper case, parameters and sets in lower case. The three symbols in Table 14 have not yet been defined. `eta=2` represents the elasticity of transformation between production for the domestic and export markets, `esubdm=4` is the domestic-import Armington elasticity of substitution, and `esubmm=8` is the import-import Armington elasticity.

Having defined compensated demands, it is then straightforward to write down zero profit equations. For sector Y_{ir} this means that the cost of inputs to production (intermediate demand plus primary factors, gross of tax) must equal the value of outputs (domestic sales plus exports, net of tax).

The zero profit conditions for `A(d,i,r)`, and `YT` are based on CES and Cobb-Douglas cost functions because the associated unit demand functions are not defined explicitly in the model.[21] Market clearance equations are displayed in Table 16. The only tricky equation here is `MKT_PA(d,i,r)` in which I use three different subsets to equate submarket supply with intermediate, public and private demand. The subsets are declared:

```
set i_d(d)/i/, c_d(d) /c/, g_d(d) /g/;
```

This notation permits me to define a different right-hand-side expression for each element of set d in the model definition.

[21] Of course it is mathematically equivalent to use the cost function or an expression for cost based on the unit demand functions, i.e. if:

$$c(p) \equiv min_x \sum p_i x_i \quad \text{s.t.} f(x) = 1$$

then $c(p) = \sum_i p_i x_i^*(p)$ where $x_i^*(p)$ is the unit demand function.

Table 12: Benchmark Share Parameters used in the Algebraic Model

```
PARAMETER
        vad(i,r)        Sectoral value-added,
        tau(i,r,s)      Unit transport cost coefficient

        thetaf(f,i,r)   Value added,
        thetad(i,r)     Domestic output,
        thetag(i,r)     Government demand,
        thetap(i,r)     Private demand,
        thetat(i,r)     Transport,
        thetam(d,i,r)   Import value share,
        beta(i,s,r)     Value share of bilateral imports,
        gamma(i,s,r)    Goods share of unit import cost;

vad(i,r) = sum(f, vfm(f,i,r)*pf0(f,i,r));
tau(i,r,s)$vxmd(i,r,s) = vtwr(i,r,s) / vxmd(i,r,s);

thetam(d,i,r)$va(d,i,r) = vm(d,i,r) / va(d,i,r);
thetaf(f,i,r)$vad(i,r) = pf0(f,i,r) * vfm(f,i,r) / vad(i,r);
thetad(i,r)$vom(i,r) = vdm(i,r) / vom(i,r);
thetag(i,r) = pg0(i,r) * vgm(i,r) / vg(r);
thetap(i,r) = pc0(i,r) * vpm(i,r) / vp(r);
thetat(i,r)$sum((j,s), vst(j,s)) = vst(i,r) / vt;
beta(i,s,r)$vxmd(i,s,r) =
           (vxmd(i,s,r)*pmx0(i,s,r)+vtwr(i,s,r)*pmt0(i,s,r) )  / vim(i,r);
gamma(i,s,r)$vxmd(i,s,r) = vxmd(i,s,r)*pmx0(i,s,r) /
           (vxmd(i,s,r)*pmx0(i,s,r) + vtwr(i,s,r)*pmt0(i,s,r) );
```

Table 13: Variables in the MCP Model

```
VARIABLES

        A_G(i,r)        Public sector unit demand
        A_C(i,r)        Private unit demand
        A_F(f,i,r)      Factor unit demand
        A_X(i,r)        Export unit supply
        A_D(i,r)        Domestic unit supply
        A_M(i,r,s)      Import unit demand

        C(r)            Private consumption
        G(r)            Public provision
        Y(i,r)          Aggregate Output
        M(i,r)          Import aggregation
        A(d,i,r)        Armington aggregations
        YT              Transport

        PC(r)           Private demand
        PG(r)           Public provision
        PD(i,r)         Domestic Output price
        PX(i,r)         Export price
        PM(i,r)         Import price
        PA(d,i,r)       Armington composite price
        PF(f,r)         Factor price
        PT              Transport services

        RA(r)           Representative agent income;
```

Table 14: Compensated Unit Demand and Supply Functions

```
DEF_G(i,r)..

  A_G(i,r) =E= vgm(i,r) *
        PROD(j, (PA("g",j,r)*(1+tg(j,r))/pg0(j,r))**thetag(j,r)) /
        ( PA("g",i,r)*(1+tg(i,r))/pg0(i,r) );

DEF_C(i,r)..

  A_C(i,r) =E= vpm(i,r) *
        PROD(j, (PA("c",j,r)*(1+tp(j,r))/pc0(j,r))**thetap(j,r)) /
        ( PA("c",i,r)*(1+tp(i,r))/pc0(i,r) );

DEF_F(f,i,r)..

  A_F(f,i,r) =E= vfm(f,i,r) *
        PROD(ff, (PF(ff,r)*(1+tf(ff,i,r))/pf0(ff,i,r) )**thetaf(ff,i,r)) /
        ( PF(f,r)*(1+tf(f,i,r)) / pf0(f,i,r) );

DEF_X(i,r)..

  A_X(i,r) =E=  vxm(i,r) * (PX(i,r) /
        (thetad(i,r)  * PD(i,r)**(1+eta) +
        (1-thetad(i,r)) * PX(i,r)**(1+eta))**(1/(1+eta)) )**eta;

DEF_D(i,r)..

  A_D(i,r) =E=  vdm(i,r) * (PD(i,r) /
        (thetad(i,r)  * PD(i,r)**(1+eta) +
        (1-thetad(i,r)) * PX(i,r)**(1+eta))**(1/(1+eta)) )**eta;

DEF_M(i,r,s)..

  A_M(i,r,s) =E= vxmd(i,r,s) * ( PM(i,s) /
        ( gamma(i,r,s)  * PX(i,r)*(1+tx(i,r,s))*(1+tm(i,r,s))/pmx0(i,r,s)
        + (1-gamma(i,r,s)) * PT*(1+tm(i,r,s)) / pmt0(i,r,s) ) )**esubmm;
```

Table 15: Exhaustion of Production Equations in the MCP Formulation

```
*       Production:

PRF_Y(i,r)..

  SUM(j, vafm(j,i,r) * PA("i",j,r) * (1+ti(j,i,r)) ) +
        SUM(f, A_F(f,i,r) * PF(f,r) * (1 + tf(f,i,r)) )
        =E= (1 - ty(i,r)) * (PD(i,r) * A_D(i,r) + PX(i,r) * A_X(i,r));


*       Armington aggregation across imports from different countries:

PRF_M(i,r)..

  SUM(s, (1 + tm(i,s,r)) * A_M(i,s,r) *
        (PX(i,s) * (1 + tx(i,s,r)) + PT * tau(i,s,r))) =E= PM(i,r) * vim(i,r) ;

*       Public output:

PRF_G(r)..      SUM(i, PA("g",i,r) * (1+tg(i,r)) * A_G(i,r)) =E= PG(r) * vg(r);

*       Private consumption:

PRF_C(r)..      SUM(i, PA("c",i,r) * (1+tp(i,r)) * A_C(i,r)) =E= PC(r) * vp(r);


*       Import-domestic aggregation by submarket:

PRF_A(d,i,r)..

  ((1-thetam(d,i,r)) * PD(i,r)**(1-esubdm) +
        thetam(d,i,r)  * PM(i,r)**(1-esubdm) )**(1/(1-esubdm)) =E= PA(d,i,r);

*       Inter-national transport services (Cobb-Douglas):

PRF_YT..        PROD((i,r), PD(i,r)**thetat(i,r)) =E= PT;
```

Table 16: Market Clearing Equations in the MCP Formulation

```
*       Exports:

MKT_PX(i,r)..

  YX(i,r)*Y(I,R) =E= SUM(s, A_M(i,r,s)*M(i,s)) + VST(i,r)*YT*(PT/PX(i,r));

*       Domestic supply:

MKT_PD(i,r)..

  YD(i,r) * Y(I,R) =E=
        SUM(d, A(d,i,r) * vd(d,i,r) * ( PA(d,i,r)/PD(i,r) )**esubdm )
        + vi(r)$cgd(i);

*       Imports:

MKT_PM(i,r)..

  vim(i,r) * M(i,r) =E=
        SUM(d, A(d,i,r) * vm(d,i,r) * ( PA(d,i,r)/PM(i,r) )**esubdm );

*       International transport:

MKT_PT..        YT * vt =E=  sum((i,r,s), A_M(i,r,s) * M(i,s) * tau(i,r,s));

*       Armington supply:

MKT_PA(d,i,r)..

  va(d,i,r) * A(d,i,r) =E= sum(j, vafm(i,j,r) * Y(j,r))$i_d(d)
        + (A_C(i,r) * C(r))$c_d(d) + ( A_G(i,r) *  G(r))$g_d(d);

*       Government provision:

MKT_PG(r)..     G(r) =E= 1;

*       Factor market:

MKT_PF(f,r)..   evoa(f,r) =E= sum(i, A_F(f,i,r) * Y(i,r));


*       Private demand:

MKT_PC(r)..     C(r) * vp(r) =E= RA(r) / PC(r) ;
```

The final equation for this model is an expression defining regional income as a function of factor prices, transfers, and tax revenue. The complexity in this equation concerns accounting for revenue from each of seven different tax instruments.[22] Not to belabor the point, but the income expression in Table 17 illustrates the usefulness of MPSGE for tax policy analysis.[23]

[22]There is a subtle but important point with regard to the complex system of taxes in GTAP. Users should not assume that because the dataset has a tax *instrument* the associated tax *rates* have a strong empirical basis. The research work in putting together GTAP has tended to focus on trade taxes (import tariffs and export taxes), and all other tax rates come directly from the national input-output tables. If you undertake an analysis in which the structure of the domestic tax system plays an important role, it is highly recommended to collect and update the benchmark tax rates. For an example of how a domestic tax system may be introduced in a GTAP model, see Harrison, Rutherford and Tarr [1997].

[23]In the MPSGE model a single entry in the import activity introduces both the import and export taxes, and given a description of taxes applying to the producer, the modelling language automatically generates the appropriate income entries, greatly reducing the likelihood of an accounting error.

Table 17: Income Balance Equations in the MCP Formulation

```
INC_RA(r)..

  RA(r) =E= sum(f,   PF(f,r) * evoa(f,r))
        + sum(num, PC(num) * vb(r))
        - sum(cgd, PD(cgd,r) * vdm(cgd,r))
        - PG(r) * vg(r)

* Output tax:

        +    sum(i, ty(i,r)*(PX(i,r)*A_X(i,r)+PD(i,r)*A_D(i,r))*Y(I,R) )

* Tax on intermediate demand:

        +    sum((i,j), ti(j,i,r) * PA("i",j,r) * vafm(j,i,r) * Y(i,r) )

* Taxes on factor use:

        +    sum( (i,f),  tf(f,i,r) * PF(f,r) * A_F(f,i,r) * Y(i,r))

* Export tax:

        +     sum((i,s), tx(i,r,s) * PX(i,r) * A_M(i,r,s) * M(i,s) )

* Import tariff applies to mechandise gross of export tax
* and transport cost:

        +    sum((i,s), tm(i,s,r) * A_M(i,s,r) * M(i,r) *
                (  PX(i,s) * (1+tx(i,s,r)) + PT * tau(i,s,r) ) )

* Taxes on goverment consumption:

        +    sum(i, tg(i,r) * PA("g",i,r) * A_G(i,r) * G(r))

* Taxes on private consumption:

        +    sum(i, tp(i,r) * PA("c",i,r) * A_C(i,r) * C(r));
```

4 Practicalities

4.1 System Requirements

You will need to have the following GAMS system components:

- GAMS compiler version 2.50[24]
- PATH complementarity solver
- CONOPT nonlinear optimizer and nonlinear system solver (optional)
- MPSGE subsystem (optional)
- A Pentium computer running Windows 95 or NT with more than 100 MB of free disk space.

4.2 Getting Started

The GTAPinGAMS package is distributed in two archives, one containing executable utility programs, a second containing the directory structure and GAMS programs. The first is to be extracted either in the GAMS system directory or in another directory on the path. The second unzips into a new directory.

GAMS source code and two free aggregate datasets are provided with the distribution directory. There is no original GTAP data distributed with the GTAPinGAMS system. In order to generate large scale models with the GTAPinGAMS tools, it is necessary to obtain the GTAP 4 distribution data file GSDDAT.HAR.

The datasets included in the GTAPinGAMS distribution have been provided by the MobiDK Project at the Ministry of Business and Industry in Denmark and by the EPRI-funded analysis of the trade effects of the Kyoto agreement. The MobiDK dataset focuses on Denmark's international trade in goods and energy. It includes an investment composite, five energy goods (crude oil, refined oil, natural gas, coal and electricity), chemicals, other energy intensive goods, services and other non-energy intensive. The regions identified in this dataset include the European countries from GTAP, North America, Asia and Rest of World. This dataset should provide a good initial aggregation for analysts working on energy-environment issues in any of the GTAP EU countries (Denmark, Sweden, Germany or UK).[25]

The second dataset is intended to focus on competitiveness impacts of measures intended to reduce global carbon dioxide emissions. The regional aggregation for the DOE dataset is motivated by the nature of Kyoto agreement with separate representation of the US, Japan, the EU and China. Aggregate regions in the model include Other OECD, Former Soviet Union, Central European Associates, Other Asia, Mexico plus OPEC and Rest of World. Commodities in the DOE dataset include the investment aggregate, five energy goods, metals-related industry, other energy-intensive, other manufactures and services.[26]

Here are the steps involved in installing GTAPinGAMS under Windows 95/NT:

[24] These programs should work with GAMS 2.25.089 or later, but the matrix balancing relies on some significant improvements in robustness which Michael Ferris and his students have achieved with the latest release of PATH. If you are running GAMS with version 2.25 and encounter problems with the rebalancing routine, you could try unzipping a newer version of PATH (http://robles.colorado.edu/~tomruth/gtapingams/path.zip) into the GAMS system directory, first renaming your existing GAMS2PTH.EXE to GAMS2PTH.BAK.

[25] Note that sectoral and regional identifiers in these models are all three characters in length. If you are planning to use the GAMS/MPSGE version of the GTAPinGAMS model, then regional identifiers are limited to at most four characters and sectoral set labels may have at most 10 characters.

[26] This dataset is compatible with the GTAP-E satelite energy tables which represent all energy-related transactions in physical units. A subsequent paper will describe how GTAP-E data can be introduced into the GTAPinGAMS model.

1. Unpack the GTAPinGAMS Utility Programs

   These programs may be placed in any subdirectory on the path. A convenient choice would be the GAMS system directory. If GAMS were located in drive `d:`, subdirectory `gams`, you could enter:

   ```
   d:\>cd  \gams
   d:\gams>x:\utils
   ```

   where `x:` is the subdirectory location of your GTAPinGAMS distribution files.

2. Create a GTAPinGAMS Subdirectory

   This subdirectory can have any name. I name the directory GTAP:

   ```
   d:\>mkdir gtap
   d:\>cd gtap
   ```

3. Unpack the GTAPinGAMS Directory

   Connect to the directory you have just created and unzip the archive containing directory structure and GAMS source code:

   ```
   d:\gtap>unzip x:gtapgams
   ```

   NB: This statement requires that the utility program UNZIP.EXE installed in step 1 are accessible along the path.

4. Install the GTAP4 Data (optional)

   This step is required only if you have a copy of the the GTAP4 datafile GSDSMALL.ZIP.

   ```
   d:\gtap>cd gtapdata
   d:\gtap\gtapdata>unzip x:gsdsmall
   ```

   where x: is the subdirectory location of your GTAP distribution files.

5. Run the Build Script (optional)

   This step is required only if you have installed the GTAP4 datafile GSDSMALL.ZIP.[27]

   Connect to the BUILD subdirectory and execute the BUILD.BAT batch command file:

   ```
   d:\gtap>cd build
   d:\gtap\build>build
   ```

[27] N.B. The BUILD script only works properly with the distribution header array file for the full GTAP database, GSDDAT.HAR. This program is not designed to work with aggregated GTAP datasets which have been constructed in GEMPACK.

6. Test the Aggregation Routine

   Connect to the BUILD subdirectory and execute the GTAPAGGR.BAT batch command file to generate a test aggregation of the sample dataset:

```
d:\gtap>cd build
d:\gtap\build>gtapaggr doemacro
```

   The GTAPAGGR command generates a dataset named DOEMACRO.ZIP from dataset DOE.ZIP. If you examine the ZIP file for this model using UNZIP, you will find that GTAPAGGR has included in the model archive the set definition file (DOEMACRO.SET), the mapping file (DOEMACRO.SET) and the dataset itself (DOEMACRO.GMS); along with set and map files for the parent dataset DOE.[28]

```
        d:\gtap\data>unzip -l doemacro

Archive:  doemacro.zip

 Length    Date    Time    Name
 ------    ----    ----    ----
    576  09-03-98  10:45   doemacro.set
   2124  09-03-98  10:51   doemacro.map
    736  09-03-98  10:46   doe.set
   2780  09-03-98  15:49   doe.map
  87710  09-05-98  13:10   doemacro.gms
 ------                    -------
  93926                    5 files
```

   You can remove files from the model archive manually using UNZIP, or you can read these files into a GAMS program using the MRTDATA and UNZIP utilities provided in the INCLIB subdirectory.

7. Test the Recalibration Routine

   Connect to the BUILD subdirectory and execute the IMPOSE.BAT batch command file to impose a new set of benchmark tax rates on the DOE dataset, creating a new dataset named NOTAX:

```
d:\gtap>cd build
d:\gtap\build>impose notax doe
```

   The IMPOSE command generates a dataset named NOTAX.ZIP from DOE.ZIP using information from a benchmark revision file NOTAX.DEF in the DEFINES directory. This command also copies the set definition file DOE.SET to NOTAX.SET and the mapping file DOE.MAP to NOTAX.MAP, all in the DEFINES directory.[29]

[28]If you are using GAMS/MPSGE, you need to restrict regional identifiers to 4 or fewer characters. Commodity and factor names may have at most 10 characters.

[29]The mapping file is copied if one can be found. This is done to assure that it is always possible to trace the aggregation definitions for any dataset.

8. Check Benchmark Consistency and Generate an Echoprint

```
d:\gtap>cd build

d:\gtap\build>chkdata doe
d:\gtap\models>more <doe.ech
```

9. Run Some Tests with the Core Models

```
d:\gtap>cd models
d:\gtap\models>gams mrttest
```

   The MRTTEST.GMS program runs the model in four different formulations and compares the results. This GAMS program processes four other programs, MGETEST.GMS, MCPTEST.GMS, CNSTEST.GMS, and NLPTEST.GMS each of which executes a benchmark replication check, a benchmark clean-up and a single counter-factual scenario. Output from these four programs is collected in MRT.SOL and the results are compared by MRTTEST.GMS. If your installation is correct, you should see the message:

```
No difference detected in the solutions from alternative solvers.
```

   displayed on the screen.[30]

   These tests are conducted with the DOE dataset. If you edit file DATASET, you can chose to use a different dataset for the test.

10. Transfer a Dataset into HAR Format

    Any dataset constructed with the GTAPinGAMS tools can be written in a format which is readable by GTAP in GEMPACK. The commands required to translate a data set are as follows:

```
d:\gtap>cd build
d:\gtap\build>zip2har doe
```

    This command generates the file DOE.HAR in the DATA subdirectory which you can send to friends in Australia.

11. Set up the GTAPinGAMS Libinclude Routines (optional)

    This step is required only if you want to build a GTAPinGAMS model in a subdirectory which is not located immediately below your GTAPinGAMS root directory. All of the GAMS programs in BUILD and MODELS use batinclude programs in the subdirectory `../inclib`. These statements are invalid if you move the files outside the GTAPinGAMS directory structure.

    In order to produce a portable image of the model, copy files from INCLIB into the INCLIB subdirectory of you GAMS system directory. This might typically be directory C:/GAMS/INCLIB, in which case you would install the libinclude files by:

[30] The first calculation which is performed is a benchmark replication check in which a solver may report "INFEASIBLE". This simply means that there is some imprecision in the data, as is subsequently reported in the listing as "Benchmark tolerance". Any number on the order of 1.e-4 or smaller indicates a reasonably precise dataset.

```
d:\gtap>cd inclib
d:\gtap\inclib>copy *.* c:\gams\inclib
```

After having installed the GTAPinGAMS libinclude files in the GAMS INCLIB subdirectory, you may then replace, for example,

```
$INCLUDE ..\inclib\mrtdata DOE
```

by the statement

```
$LIBINCLUDE mrtdata DOE
```

Alternatively, you can read only the SET definitions (DOE.SET) for the data using the statement:

```
$LIBINCLUDE unzip DOE set
```

I have included a sample stand-alone model which you can test once the INCLIB files have been installed. If you have the MPSGE subsystem, execute the following:

```
d:\gtap>cd models
d:\gtap\models>cd standalone
d:\gtap\models\standalone>gams mrtmge
```

If you do not have MPSGE, you can instead run `gams mrtnlp`.

4.3 Directory Structure

When you install GTAPinGAMS, the root directory is empty, and all files reside in one of following six second-level subdirectories:

- BUILD

  Contains GAMS programs for dataset extraction, aggregation, rebalancing and translation. Contains batch files BUILD.BAT, GTAPAGGR.BAT, IMPOSE.BAT, CHKDATA.BAT and ZIP2HAR.BAT.

- DEFINES

  Contains set definition files, mapping files and tax rate assignment files for alternative aggregations of the data. There are three types of files in this subdirectory:

  – Files with extension .SET contain GAMS code defining the sets of commodities, regions and primary factors for a GTAPinGAMS model.

  - Files with extension .MAP define an aggregation in terms of the source dataset and mappings from sets in the source to sets in the target.
  - Files with extension .DEF contain definitions for a dataset which has been constructed by imposing a new set of tax rates or benchmark accounts on another dataset.

- MODELS

  Contains template GAMS programs illustrating how the GTAP data can be GAMS programs. These include four model files:

  - MRTMGE.GMS The static multiregional model specified as an mixed complementarity model using an MPSGE representation of demand and supply functions.
  - MRTMCP.GMS The static multiregional model specified as an mixed complementarity model with GAMS algebra.
  - MRTCNS.GMS The static multiregional model specified as an constrained nonlinear system with the GAMS CNS model type.
  - MRTNLP.GMS The static multiregional model specified as an square nonlinear system within a GAMS nonlinear program.

  A subdirectory under models STANDALONE contains a model which may be processed only after the INCLIB routines have been copied into the GAMS system INCLIB subdirectory.

- INCLIB

  Contains batinclude routines accessed by GAMS programs in subdirectories BUILD and MODELS. A few of these routines are distributed from my web page at the University of Colorado (http://robles.colorado.edu). I have provided these routines in a separate subdirectory in order to assure that the GTAPinGAMS package is completely self-contained and to avoid potential conflicts with user's personal collection of libinclude programs.

- DATA

  Contains constructed data files. Files with the *.ZIP extension are GAMSinGTAP datasets for alternative aggregations. Files with a *.HAR extension are GTAP datasets in the original format readable in GEMPACK. If you have the GTAP distribution file, all files in the DATA subdirectory may be regenerated.

- GTAPDATA

  Contains original files from the GTAP distribution archive GSDSMALL.ZIP. GAMSinGTAP uses only the GTAP data file GSDDAT.HAR from this directory.

4.4 Dataset Contents and File Formats

Data files may be stored either as compressed GAMS source or as a GEMPACK-readable Header Array (HAR) file. A HAR file may be examined with the GEMPACK utilities VIEWHAR.EXE or SEEHAR.EXE.

The two file formats provide a trade-off between speed and size. The GAMS-Zip format is used in all GTAPinGAMS computations because it is considerably faster to read. The HAR format, in which subarrays are stored without labels, provides a more compact data format, particularly when the HAR file is further compressed using PKZip or Infozip. A zipped HAR archive for the full GTAPV4 aggregation is 1.6 MB as compared with 4.4 MB for a GAMS-Zip file. The GAMS-Zip format is, however, considerably faster to access (40 seconds versus nearly 3 minutes for dataset GTAPV4).

4.5 Batch Files

Most economic modelers using the GTAP database will want to build their own model, making decisions about the structure of technology, preferences and policy parameters. The tools provided here are intended to simply facilitate the use of GTAP data. A modeler would typically run these programs once to produce a dataset as part of a modelling exercise.

All of the dataset aggregation and recalibration tools provided here are packaged in DOS batch files. The command files include:

- BUILD

  Unpack the GTAP V4 distribution data into GAMS-readable formats, and generate a filtered version of the full dataset suitable for large-scale computation. The "filtering" step rounds all values in the dataset to the nearest 100,000 $; and all tax rates are rounded to the nearest percent.

- GTAPAGGR

  Aggregate a larger GTAP dataset into a smaller GTAP dataset.

- IMPOSE

  Generate a new dataset by imposing an exogenous set of tax rates on an existing GTAP dataset. This permits adjustment of tariffs, export taxes, sales taxes and factor taxes.

- CHKDATA

  Read a dataset and check benchmark consistency, producing an echoprint of base year GDP and trade shares.

- ZIP2HAR

  A utility routine to move GTAPinGAMS datasets between GAMS ZIP and GEMPACK header-array formats.[31]

4.5.1 BUILD.BAT

This program is typically run once to generate a GAMS-readable dataset from the original GEMPACK distribution file GSDDAT.HAR. This begins by translating the full GTAP dataset into GAMS-readable format (GTAPV4.ZIP). This is done using the GEMPACK utility SEEHAR.EXE, a small Fortran program REWRITE.EXE and a GAMS program SCALE.GMS. The last of these programs scales trade and production data from billions of dollars to tens of billions of dollars.

The next step in the translation is to "filter" the GTAPV4 dataset removing all very small coefficients, extreme tax rates and various other inconsistencies. The default filter tolerance is 0.001 (one tenth of one percent), defined in FILTER.GMS. I use this tolerance to name the filtered dataset GTAP4001. When using GTAP version 4 data, I would normally aggregate using the GTAP4001 dataset as a source. The filtering process improves numerical robustness in large-scale models while introducing very small changes in the results. If you are working with a highly aggregate model, however, it should be possible to aggregate directly from the unfiltered dataset GTAPV4.

Specific steps in this program are as follows:

1. Round tax rates to the nearest percentage, and round all values to the nearest \$100,000.

2. Remove small intermediate inputs in production and trade. All of the following filtering steps involve the scalar parameter TOLERANCE which equals 0.001 in the central application.

[31] See CONVERT.GMS for details on conversion from har to zip format, and see GAMS2HAR.GMS and HAR2GAMS.GMS in the INCLIB directory for general-purpose tools for data transfers between GAMS and header array files.

3. Filter the Trade Matrix

   Eliminate any imports of a good into a region where the total value of imports is less than TOLERANCE times the combined value of import and domestic demand.

   Define the MAGNITUDE of a trade flow as maximum of the ratio of the trade flow net of tax to the associated aggregate export level or the trade flow gross of tax to the associated aggregate import level.

   Drop all trade flows which have MAGNITUDE less than TOLERANCE.

   Rescale remaining trade flows to maintain consistent values of aggregate imports and aggregate transport cost.

4. Filter the Production Matrices

   Define the MAGNITUDE of an intermediate input as the maximum of the ratio of the input value gross of tax to total cost, or the ratio of the input value net of tax to total domestic supply.

   Define the MAGNITUDE of a factor input as the ratio of the factor payment gross of tax to the value of output gross of tax.

   Drop all intermediate inputs and factor inputs which have MAGNITUDE less than TOLERANCE.

   Even with a very small TOLERANCE (0.1%), the filtering just described generates a substantial reduction in the number of nonzeros:

```
PARAMETER DENSITY       summary of changes in matrix density

                  BEFORE       AFTER
        TRADE     53.074      43.357
        PROD      81.942      46.824
```

5. Filter Final Demand

   Finally, we do the same thing with final demand (private and public), filtering both imports and domestic demand. We also filter inputs to the international transport activity. This removes all tiny coefficients from the dataset.

6. Recalibrate the Dataset

   The foregoing assignments represent a large number of small changes to the model data, and it is certain that we have introduced some inconsistencies which show up as violations of the profit and market clearance conditions defined in chkeq. For this reason, at this point we use a modified least-squares procedure to restore consistency, holding the international trade matrices fixed and recalibrating each of the regional economic flows.

   This is the step where it is very helpful to use a complementarity formulation and the PATH solver, as the solution is extremely difficult with MINOS or CONOPT due to the large number of accumlated superbasics. I have include model definitions here for an equivalent nonlinear programming approach, but I have not included this as a standard feature because I have found the NLP codes to be somewhat unreliable. If you own an NLP solver but don't have PATH, then it will be necessary to convert the `SOLVE` statements from MCP to NLP. If this proves difficult, contact GTAP and we can arrange for you to get a copy of GTAP4001.ZIP, the filtered dataset.

   Ferris and Rutherford [1998] present details of how we have set up constraints and objective function which are interesting but not essential to understanding the program. The key

Table 18: Files Referenced by BUILD.BAT

```
Inputs:
        ..\gtapdata\gsddat.har
        ..\defines\gtapv4.set
        ..\defines\gtap4001.set

Outputs:

        ..\data\gtapv4.zip
        ..\data\gtap4001.zip
```

point is that at this point we have changed some of the base year value flows to reinstate equilibrium, holding all tax rates fixed.

7. Reconstruct the input-output flows.

   For energy-related analysis, I find it helpful to maintain a process-oriented representation of the oil sector. For this purpose, I have included code which routes all crude oil flows in each region through the refined oil sector. This involves some careful programming to assure that tax payments and all base year transactions remains constant.

4.5.2 GTAPAGGR.BAT

Once you have built the initial GTAPinGAMS dataset GTAP4001 (or GTAPV4), you can begin to think about a particular application and which aggregations of the original GTAP data would be appropriate for studying those issues. I typically create two aggregations for any new model, one with a minimal number of regions and commodities and another with a larger number of dimensions. I use the small aggregation for model development and bring out the larger dataset whenever I am confident that the model is running reliably and producing sensible results.

The GTAPAGGR.BAT program is used to aggregate a GTAPinGAMS dataset. A command line argument defines the name of the target aggregation. You only need to provide the batch file with the target because the target's mapping file defines the source. Before running GTAPAGGR.BAT, you must create two files, one defining the sets of commodities, regions and primary factors in the target dataset, and another defining the name of the source dataset and a correspondence between elements of the source and target. The aggregation routine produces a brief report of GDP and trade shares in the new dataset. This is written to a file in the build directory.

The SET and MAP files for a new dataset are GAMS-readable files located in the **defines** subdirectory.

Table 20 a sample set file defining dataset DOEMACRO. The file defines the sets of goods, regions, and primary factors which are in the model. Commodity CGD, the investment-savings composite, must be included in every aggregation:

Table 21 presents the associated mapping file, DOEMACRO.MAP. The file provides a definition of the source dataset together with mapping definitions for commodities and factors. When no mapping is defined for the set of regions, the aggregation routine retains the same set as in the source data.

Here are a couple of exercises which could help a new user learn about the error messages returned by GTAPAGGR: (i) Comment out the line with `MFR.Y` mapping and run GTAPAGGR. (You will get an error message indicating the `MFR` has not been mapped). (ii) Change the `COL.COL`

Table 19: Files Referenced by GTAPAGGR.BAT

```
Inputs:

        Command line argument: target

        ..\defines\%target%.set
        ..\defines\%target%.map       (defines source)
        ..\data\%source%.zip

Output:

        ..\data\%target%.zip
        ..\build\%target%.ech
```

Table 20: Set Definitions for a GTAPinGAMS Aggregation DOEMACRO

```
$TITLE   An Aggregation of the DOE Dataset

SET   I   Sectors/
  Y     Aggregate output
  COL   Coal
  OIL   Petroleum and coal products (refined)
  CRU   Crude oil
  GAS   Natural gas
  ELE   Electricity
  CGD   Savings good /;

SET R Aggregated Regions /
  USA   United States
  JPN   Japan
  EUR   Europe
  OOE   Other OECD
  CHN   China
  FSU   Former Soviet Union
  CEA   Central European Associates
  ASI   Other Asia
  MPC   Mexico plus OPEC
  ROW   Other countries /;

SET   F   Factors of production /
  LAB   Labor,
  CAP   Capital /;
```

Table 21: Mapping Definitions for DOEMACRO

```
$SETGLOBAL source doe

*       -----------------------------------------------------------------
*       The target dataset has fewer sectors, so we need to specify how
*       each sector in the source dataset is mapped to a sector in the
*       target dataset:

SET   MAPI   Sectors/
        MTL.Y   Metals-related industry (IRONSTL & NONFERR)
        EIS.Y   Other energy intensive (CHEMICAL & PAPERPRO)
        MFR.Y   Other manufactures
        SER.Y   Other Services
        COL.COL Coal
        OIL.OIL Petroleum and coal products (refined)
        CRU.CRU Crude oil
        GAS.GAS Natural gas
        ELE.ELE Electricity
        CGD.CGD Savings good /;

*       The following statements illustrate how to aggregate
*       factors of production in the model.  Unlike the aggregation
*       of sectors or regions, you need to declare the set of
*       primary in the source as set FF, then you can specify the
*       mapping from the source to the target sets.

*       The reason for this special treatment is to permit the
*       aggregation program to operate with both GTAP version 4 and
*       GTAP version 3 data.  Sorry for the inconvenience!  TFR

set ff /LND,SKL,LAB,CAP,RES/;
SET MAPF mapping of primary factors /LND.CAP,SKL.LAB,LAB.LAB,CAP.CAP,RES.CAP/;

*       NB: There is no need to specify a MAPR array for generating
*       DOEMACRO from the DOE dataset.  This implies that the
*       source and target datasets have the identical sets of
*       regions, so there is no need to specify an set named MAPR.
*       The aggregation routine will automatically assign a
*       one-to-one mapping from the source to the target regions.
```

Table 22: Benchmark Tax Definitions File: NOTAX.DEF

```
*       Set up a benchmark equilibrium in which we eliminate all domestic taxes:

ty(i,r) = 0;
tp(i,r) = 0;
tf(f,i,r) = 0;
tg(i,r) = 0;
ti(i,j,r) = 0;
```

Table 23: Files Referenced by IMPOSE.BAT

```
Inputs:

        Command line arguments: source target

        defines\%source%.set
        defines\%source%.map
        defines\%target%.def
        data\%source%.zip

Outputs:
        defines\%target%.set
        defines\%target%.map
        data\%target%.zip
        build\%target%.ech
```

mapping to `COL.OIL` and run GTAPAGGR. You will get an error message indicating that sector `COL` in the target dataset has no sector mapped to it.

4.5.3 IMPOSE.BAT

This program is used principally to create a new dataset by imposing a new set of benchmark rates on an existing GTAP dataset. Two command line arguments define the target and source datasets. The source dataset must be in the DATA subdirectory, and a file defining benchmark tax rates for the target dataset is specified in the DEFINES (see Table 22). This program also generates a summary echo-print of trade and GDP shares for the new dataset and places this file in the BUILD subdirectory.

When you write the definitions file for adjusting tax rates, bear in mind that a gross basis tax (TY) is defined as a percentage of the gross-of-tax price, hence these tax rates have a maximum value of 100% and no minimum. A net basis tax, such as TF, TP, TG, TX or TM is defined as a percentage of the net-of-tax price, hence these tax rates have no maximum value and a minimum value of -100%.

Table 24: Files Referenced by ZIP2HAR.BAT

```
Inputs:
        Command line argument: dataset
        data\%dataset%.zip

Outputs:
        data\%dataset%.har
```

4.5.4 ZIP2HAR.BAT

This utility routine reads a GTAP dataset in GAMS-ZIP format and writes the data in a self-extracting compressed header array format.

References

[1] Armington, P., "A Theory of Demand for Products Distinguished by Place of Production," *IMF Staff Papers*, 16, 1969, 159-178.

[2] Ferris, M.C. and T.F. Rutherford, "Matrix Balancing: A Practical Approach", *University of Colorado Working Paper*, 1998.

[3] Harrison, W.J. and K.R. Pearson, "Computing solutions for large general equilibrium models using GEMPACK", *Computational Economics*, 9:83-127, 1996.

[4] Harrison, G.W., T.F. Rutherford, and D. Tarr, "Opciones de poltica comercial para Chile: una evaluacin cuantitiva" *Cuadernos de Economa*, August 1997. A related working paper in English is available on the web (see http://theweb.badm.sc.edu/glenn/addreg.pdf).)

[5] Hertel, T.W. (ed.) *Global Trade Analysis: Modeling and Applications*, Cambridge University Press, Cambridge and New York, 1997.

[6] Mathiesen, L. "Computation of Economic Equilibrium by a Sequence of Linear Complementarity Problems", *Mathematical Programming Study 23*, North-Holland, 1985, pp. 144-162.

[7] McDougall, R. "The GTAP Database", *Draft documentation*. See the GTAP 4 release (http://www.agecon.purdue.edu/gtap/database).

[8] Rutherford, T.F. "Economic Equilibrium Modeling with GAMS: An Introduction to GAMS/MCP and GAMS/MPSGE," *Draft monograph*, University of Colorado Department of Economics, 1998.

[9] Rutherford, T.F., M. Lau and A. Pahlke, "A Primer in Dynamic General Equilibrium Analysis", Working Paper, University of Colorado Department of Economics, 1998.

[10] Rutherford, T.F., and D. Tarr, "Regional Trading Arrangements for Chile: Do the Results Differ with a Dynamic Model?", University of Colorado Department of Economics, 1998.