CENTRE of POLICY STUDIES and the IMPACT PROJECT

Eleventh Floor, Menzies Building
Monash University, Wellington Road
CLAYTON Vic 3800 AUSTRALIA

Telephone: (03) 9905 2398, (03) 9905 5112
from overseas: 61 3 9905 2398 or 61 3 9905 5112

Fax: (03) 9905 2426
from overseas: 61 3 9905 2426

e-mail: impact@buseco.monash.edu.au

Internet home page: http//www.monash.edu.au/policy/

The TERM model and its data base

by

Mark HORRIDGE

Centre of Policy Studies, Monash University

General Paper No. G-219 July 2011

ISSN 1 031 9034 ISBN 978 1 921654 26 8

The Centre of Policy Studies (COPS) is a research centre at Monash University devoted to economy-wide modelling of economic policy issues.

The TERM model and its data base

by Mark Horridge,
Centre of Policy Studies, Monash University, Australia

Draft chapter for forthcoming book published by Springer:
"Economic Modelling of Water: The Australian CGE Experience", edited by Glyn Wittwer.

Abstract

TERM (The Enormous Regional Model) provides a strategy for creating a "bottom-up" multi-regional CGE model which treats each region of a single country as a separate economy. This makes it a useful tool for examining the regional impacts of shocks that may be region-specific. TERM is designed to allow quick simulations with many regions, so allowing for models of large countries with 30 to 50 provinces, such as USA or China. TERM also offers a standard procedure for preparing a database which requires, in addition to a national input-output or use-supply table, a minimal amount of regional data. More regional data can be used if available.

JEL classification: C68, D58, R12, R13
Keywords: Regional CGE

Contents

Tables

Figures

The TERM model and its data base[1]

by Mark Horridge
Centre of Policy Studies, Monash University, Australia

1. Introduction

TERM is a framework for CGE (computable general equilibrium) modelling of multiple regions within a single country. It was developed to address two common problems of multi-regional CGE models:

- as the number of regions increases, simulations became very slow, or require large amounts of memory.
- it is difficult to develop a database for such models; published data is usually quite sparse.

TERM offers a solution to both problems:

- the database and equation system are structured to allow fast solutions with small memory needs. An inbuilt automatic system to aggregate regions and/or sectors allows model size to be reduced to speed simulations, while preserving detail that is needed for a particular application.
- there is a standard procedure for preparing a database which requires, in addition to a national input-output or use-supply table, a minimal amount of regional data. More regional data can be used if available.

From the outset the TERM framework has been intended as a template which might be quickly applied to a variety of countries. Thus the standard version of TERM is fairly simple, avoiding mechanisms which might be specific to a particular country or application. Rather the emphasis is on allowing a basic multi-regional model to produce simulation results as soon as possible. Very often, analysis of results reveals shortcomings of the model or data , or suggests priorities for improvement. To arrive quickly at this stage is key to the quality of the final model.

TERM builds on the ORANI model (Dixon et al., 1982), which distinguished over 100 sectors, and introduced large-scale computable general equilibrium modelling in 1977. In particular, the minimal data requirements for constructing a TERM database scarcely exceed those for a "topdown" multi-regional version of ORANI, described below. In fact, the standard procedure for preparing a TERM assumes that a working "topdown" database has already been prepared and used for simulations. This allows most potential problems with regional data to be noticed and fixed at an early stage.

2. Progress in Australian regional economic modelling

Since ORANI, related models have developed in several new directions. ORANI's solution algorithm combined the efficiency of linearised algebra with the accuracy of multi-step solutions, allowing the development of ever more disaggregated and elaborate models. The GEMPACK software developed by Ken Pearson (1988) and colleagues since the mid-1980s simplified the specification of new models, while cheaper, more powerful computers allowed the development of computer-intensive multi-regional and dynamic models.

On the demand side, these advances have been driven by the appetite of policy-makers for sectoral, regional, temporal, and social detail in analyses of the effects of policy or external shocks. Since parliamentary representatives are elected by regions, demand for regional detail is particularly strong.

To meet this need, even early versions of ORANI (see Dixon *et al*. 1978) included a “topdown” regional module to work out the regional consequences of national economic changes: national results for quantity (but not price) variables were broken down by region using techniques

[1] Portions of this chapter draw on Horridge, Madden and Wittwer (2005) and on Wittwer and Horridge (2010)

borrowed from input-output analysis. The name "top-down" reflects the feature that national results drive regional results and are unaffected by the regional subsystem. Key assumptions are:

- for each sector the technology of production (ie, cost shares) is uniform across regions.
- for commodities that are heavily traded between regions (the "national" commodities), each region's share of national output is fixed or exogenous. So for these sectors, the percent change in output is uniform across regions.
- for the remaining, "local", commodities (that are little traded between regions) output in each region adjusts to meet demand in that region.

Using the top-down technique, from 8 to 100 regions can easily be distinguished. Region-specific demand shocks may be simulated, but, since price variables have no regional dimension, there is little scope for region-specific supply shocks[2]. On the other hand, the "top-down" approach requires little extra data or computer power.

A second generation of regional CGE models adapted ORANI by adding two regional subscripts (source and destination) to many variables and equations. In this "bottom-up" type of multiregional CGE model, national results are driven by (ie, are additions of) regional results. Liew (1984), Madden (1989) and Peter et al.(1996) describe several Australian examples. Dynamic versions of such models have followed (Giesecke 1997). The best-known example of this type of regional model is the Monash Multiregional Forecasting model, MMRF (Adams *et al.* 2002).

Bottom-up models allow simulations of policies that have region-specific price effects, such as a payroll tax increase in one region only. They also allow us to model imperfect factor mobility (between regions as well as sectors). Thus, increased labour demand in one region may be both choked off by a local wage rise and accommodated by migration from other regions. Unfortunately models like MMRF pose formidable data and computational problems—limiting the amount of sectoral and regional detail. Only 2 to 8 regions and up to 40 sectors could be distinguished[3]. Luckily, Australia has only 8 states, but size limitations have hindered the application of similar models to larger countries with 30 to 50 provinces, and have hitherto prevented us from distinguishing smaller, sub-state regions.

Finer regional divisions are desirable for several reasons. Policy-makers who are concerned about areas of high unemployment or about disparities between urban and rural areas desire more detailed regional results. Environmental issues, such as water management, often call for smaller regions that can map watershed or other natural boundaries more closely. Finally, more and smaller regions give CGE models a greater sense of geographical realism, closing the gap between CGE and LUTE (Land Use Transportation Energy) modelling.

The TERM[4] model adds to the ORANI/MMRF tradition by allowing greater disaggregation of regional economies than was previously available. For example, it allows us to analyse effects for each of 57 statistical divisions within Australia—which would be computationally infeasible using the MMRF framework.

3. The structure of TERM

A key feature of TERM, in comparison to predecessors such as MMRF, is its ability to handle a greater number of regions or sectors. The greater efficiency arises from a more compact data structure, made possible by a number of simplifying assumptions.

3.1. Defeating the curse of dimensionality

The database for a CGE model consists of matrices of flow values dimensioned by commodity, industry and region. The model will contain quantity and price variables for each of these flows, so the number of variables and equations tends to track database size. The computer resources (time

[2] Such limitations could be partially circumvented: see Higgs et al., 1988.

[3] More precisely, these 2nd-generation models (like MMRF) become rather large and slow to solve as the product: (number of regions) x (number of sectors) exceeds 300. TERM raises this limit to about 2500.

[4] TERM is an acronym for "The Enormous Regional Model".

and memory) needed to solve the model increase super-proportionately[5] as the size of the database increases. Indeed a doubling of database size may multiply solution time by 3. Sectoral or regional detail may have to be sacrificed to reduce computing problems.

To illustrate, the value of intermediate demands in a single-region CGE model (like ORANI) might be represented by a matrix V, with dimensions COM*IND, where, for example:

V("Coal","Steel") = value of Coal used by the Steel industry.

With 50 commodities and industries, V would contain 2500 elements.

In the MMRF framework, V would be dimensioned COM*IND*REG*REG, where the first regional subscript denotes the region of origin of some input, and the second regional subscript denotes the region where the input is used. Since MMRF distinguishes 8 Australian states, the V matrix would be 64 times bigger than in ORANI—leading to much larger (but just acceptable) solution times. A USA version of MMRF, distinguishing 50 states, would imply a database which was 70 [=$(50/8)^2$] times larger than the 8-state MMRF, leading to model solution times and memory requirements perhaps 500 times those of the Australian MMRF—which is quite impractical.

TERM's solution to this problem is to restructure the model database so that no matrix contains more than 3 of the "large" COM, IND or REG dimensions. For example, instead of the large 4-dimensional intermediate input matrix used by MMRF: V(COM, IND, REG, REG), we could instead use two 3-dimensional matrices:

V(COM, IND, REG) = value of commodities used by industries in region of use; and

T(COM, REG, REG) = value of commodities used, by regions of production and use.

which together are 25 times smaller (with 50 sectors and regions), leading to model solution times and memory requirements perhaps 125 times less. The cost is a small loss in generality: the sourcing or trade matrix T encapsulates the assumption that all users in a particular region of, say, vegetables, source their vegetables from other regions according to common proportions.

3.2. The TERM data structure

Figure 1 is a schematic representation of the model's input-output database. It reveals the basic structure of the model, which is key to its efficiency. The rectangles indicate matrices of flows. Core matrices (those stored on the database) are shown in bold type; the other matrices may be calculated from the core matrices. The dimensions of the matrices are indicated by indices (c, s, i, m, etc) which correspond to the following sets:

Table 1: Main sets of the TERM model

Index	Set name	Description	Typical size
s	SRC	(dom,imp) Domestic or imported (ROW) sources	2
c	COM	Commodities	40
m	MAR	Margin commodities (Trade, Road, Rail, Boat)	4
i	IND	Industries	40
o	OCC	Skills	8
d	DST	Regions of use (destination)	30
r	ORG	Regions of origin	30
p	PRD	Regions of margin production	30
f	FINDEM	Final demanders(HOU, INV,GOV, EXP);	4
u	USER	Users = IND union FINDEM	44

The sets DST, ORG and PRD are in fact the same set, named according to the context of use.

The matrices in Figure 1 show the value of flows valued according to 3 methods:

---

[5] A key stage in the model solution process is the solution of an N*N linear equation system, where N is the number of endogenous variables. Using conventional techniques, we can expect that the time for this step will follow the *cube* of N. GEMPACK's sparse matrix and automatic substitution techniques reduce this penalty substantially; we assume below that solution time and space requirements follow $N^{1.5}$.

1) Basic values = Output prices (for domestically-produced goods), or CIF prices (for imports)
2) Delivered values = Basic + Margins
3) Purchasers' values = Basic + Margins + Tax = Delivered + Tax

The matrices on the left-hand side of the diagram resemble (for each region) a conventional single-region input-output database. For example, the matrix USE at top left shows the delivered value of demand for each good (c in COM) whether domestic or imported (s in SRC) in each destination region (DST) for each user (USER, comprising the industries, IND, and 4 final demanders: households, investment, government, and exports). Some typical elements of USE might show:

USE("Wool","dom","Textiles","North") : domestically-produced wool used by the textile industry in North

USE("Food","imp","HOU","West") : imported food used by households in West

USE("Meat","dom","EXP","North") : domestically-produced meat exported from a port in North. Some of this meat may have been produced in another region.

USE("Meat","imp","EXP","North") : imported meat re-exported from a port in North

As the last example shows, the data structure allows for re-exports (at least in principle). All these USE values are "delivered": they include the value of any trade or transport margins used to bring goods to the user. Notice also that the USE matrix contains no information about regional sourcing of goods.

The TAX matrix of commodity tax revenues contains an element corresponding to each element of USE. Together with matrices of primary factor costs and production taxes, these add to the costs of production (or value of output) of each regional industry.

In principle, each industry is capable of producing any good. The MAKE matrix at the bottom of Figure 1 shows the value of output of each commodity by each industry in each region. A subtotal of MAKE, MAKE_I, shows the total production of each good (c in COM) in each region d.

TERM recognizes inventory changes in a limited way. First, changes in stocks of imports are ignored. For domestic output, stock changes are regarded as one destination for industry output (ie, they are dimension IND rather than COM). The rest of production goes to the MAKE matrix.

The right hand side of Figure 1 shows the regional sourcing mechanism. The key matrix is TRADE, which shows the value of inter-regional trade by sources (r in ORG) and destinations (d in DST) for each good (c in COM) whether domestic or imported (s in SRC). The diagonal of this matrix (r=d) shows the value of local usage which is sourced locally. For foreign goods (s="imp") the regional source subscript r (in ORG) denotes the port of entry. The matrix IMPORT, showing total entry of imports at each port, is simply an addup (over d in DST) of the imported part of TRADE.

The TRADMAR matrix shows, for each cell of the TRADE matrix the value of margin good m (m in MAR) which is required to facilitate that flow. Adding together the TRADE and TRADMAR matrix gives DELIVRD, the delivered (basic + margins) value of all flows of goods within and between regions. Note that TRADMAR makes no assumption about where a margin flow is produced (the r subscript refers to the source of the underlying basic flow).

Matrix SUPPMAR shows where margins are produced (p in PRD). It lacks the good-specific subscripts c (COM) and s (SRC), indicating that, for all usage of margin good m used to transport any goods from region r to region d, the same proportion of m is produced in region p. Summation of SUPPMAR over the p (in PRD) subscript yields the matrix SUPPMAR_P which should be identical to the subtotal of TRADMAR (over c in COM and S in SRC), TRADMAR_CS. In the model, TRADMAR_CS is a CES aggregation of SUPPMAR: margins (for a given good and route) are sourced according to the price of that margin in the various regions (p in PRD).

TERM assumes that all users of a given good (c,s) in a given region (d) have the same sourcing (r) mix. In effect, for each good (c,s) and region of use (d) there is a broker who decides for all users in d whence supplies will be obtained. Armington sourcing is assumed: the matrix DELIVRD_R is a CES composite (over r in ORG) of the DELIVRD matrix.

Index	**Set**	**Description**
c	COM	Commodities
s	SRC	Domestic or imported (ROW) sources
m	MAR	Margin commodities
r	ORG	Regions of origin
d	DST	Regions of use (destination)
p	PRD	Regions of margin production
f	FINDEM	Final demanders(HOU, INV,GOV, EXP)
i	IND	Industries
u	USER	Users = IND union FINDEM
o	OCC	Skills
h	HOU	Household types

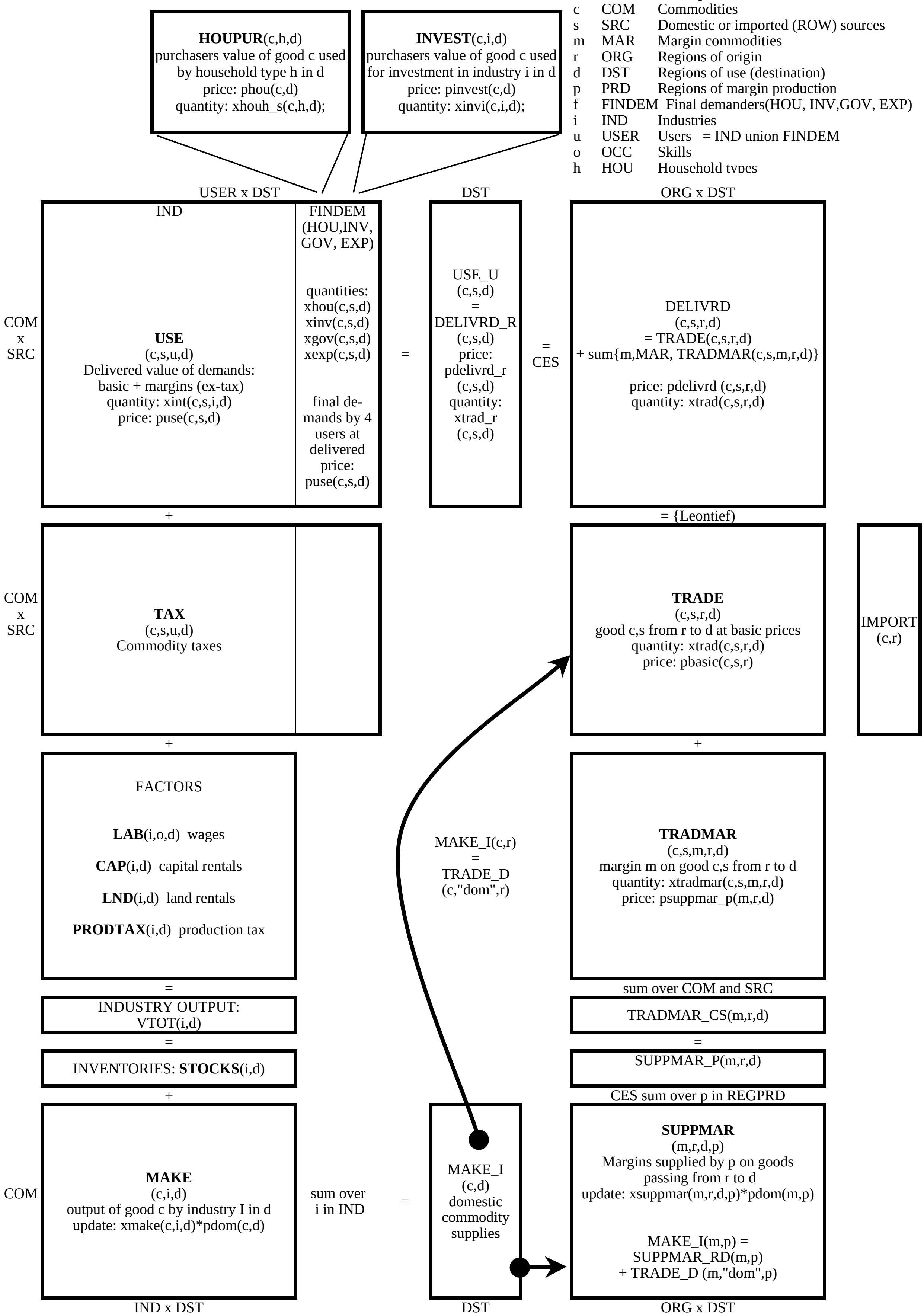

Figure 1: The TERM flows database

A balancing requirement of the TERM database is that the sum over user of USE, USE_U, shall be equal to the sum over regional sources of the DELIVRD matrix, DELIVRD_R.

It remains to reconcile demand and supply for domestically-produced goods. In Figure 1 the connection is made by arrows linking the MAKE_I matrix with the TRADE and SUPPMAR matrices. For non-margin goods, the domestic part of the TRADE matrix must sum (over d in DST) to the corresponding element in the MAKE_I matrix of commodity supplies. For margin goods, we must take into account both the margins requirement SUPPMAR_RD and direct demands TRADE_D.

At the moment, TERM distinguishes only 4 final demanders in each region:

(a) HOU: the representative household
(b) INV: capital formation
(c) GOV: government demand
(d) EXP: export demand.

For many purposes it is useful to break down investment according to destination industry. The satellite matrix INVEST (subscripted c in COM, i in IND, and d in DST) serves this purpose. It allows us to distinguish the commodity composition of investment according to industry: for example, we would expect investment in agriculture to use more machinery (and less construction) than investment in dwellings.

Similarly, another satellite matrix, HOUPUR, allows us to distinguish several household types with different budget shares. Both satellite matrices enforce the assumption that import/domestic shares and commodity tax rates are uniform across household (or investor) types: For example we assume that the tax rate on cigarettes is the same for rich and poor, as is the share of imports in cigarette consumption.

Missing from Figure 1 is an account of how factor incomes and tax revenue accrue to regional households and governments. Such data would be needed to convert the TERM data schema into a complete SAM. Australian versions of TERM typically assume that wage income generated in region A accrues to households in region A, while capital income goes into a national "pot" which is shared between regional households. Similarly, tax revenue accrues to a national authority which distributes it between regions. Such assumptions might be inappropriate if TERM were applied to other countries. For example, in the USA some taxes accrue directly to state governments, and wage income generated in Washington DC may well be spent by households in Maryland or Virginia. Hence the generic version of TERM enforces no default system: users must devise their own mapping of income to agents, appropriate to a particular country. Their decisions may be be influenced by the chosen level of regional detail. For example, some Australian versions of TERM distinguish 57 'statistical division' regions which do not entirely correspond with administrative regions -- so regional government incomes are not modelled in these versions.

4. The TERM equation system

The equations of the TERM model are broadly similar to those of other CGE models. Producers choose a cost-minimizing combination of intermediate and primary factor inputs, subject to production functions which are structured by a series of CES "nesting" assumptions, illustrated in Figure 2. Two high-level aggregates, of primary factors and of intermediate inputs, are each demanded in proportion to industry output (Leontief assumption). The primary factor aggregate is a CES composite of capital, land and a labour aggregate—the latter being itself a CES composite of labour by skill group. The aggregate intermediate input is again a CES composite of different composite commodities, which are in turn CES composites of commodities from different sources—as described in detail in the next section. Industry outputs are transformed into commodity outputs via a CET mechanism that is calibrated from the MAKE matrix of Figure 1.

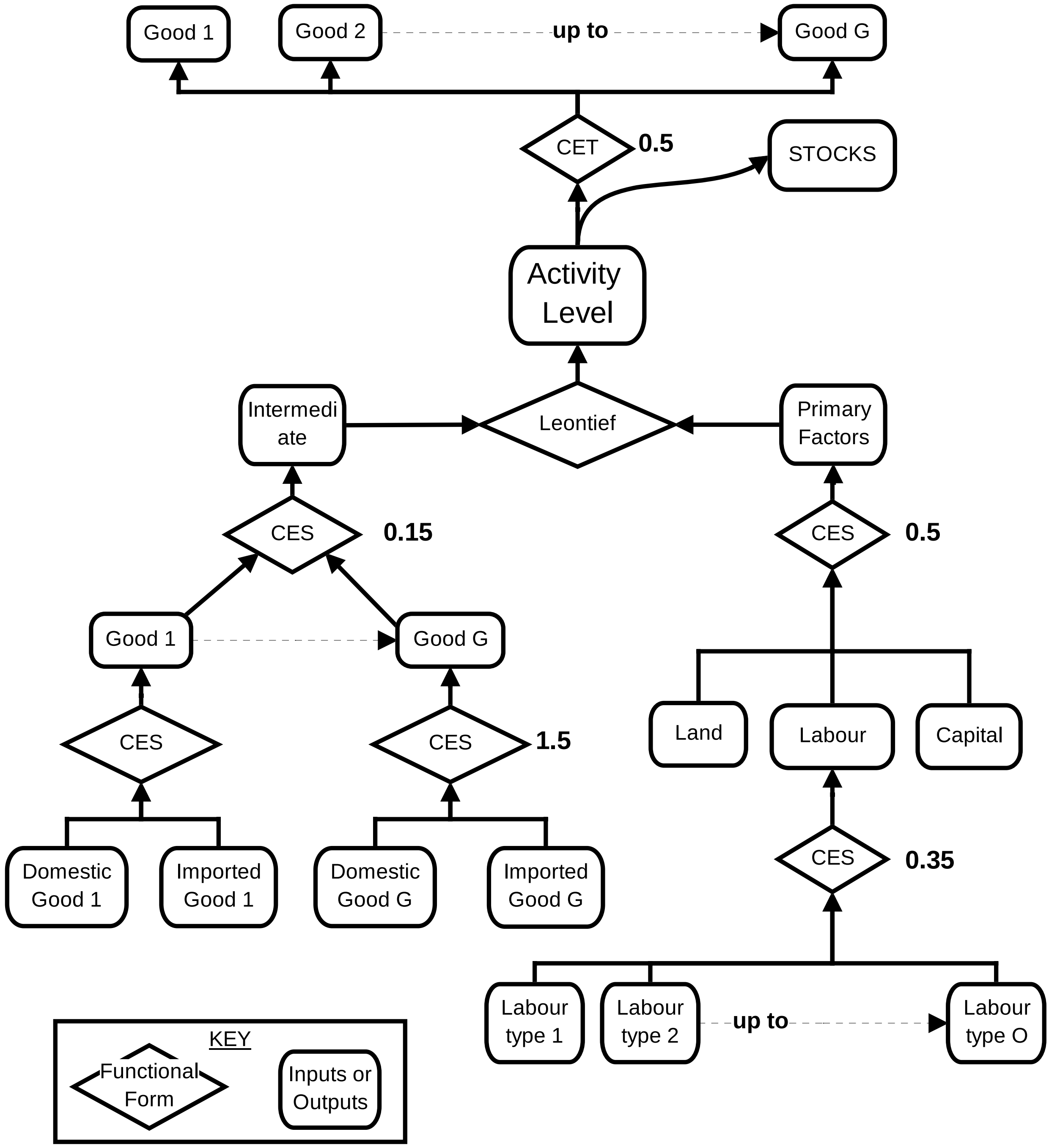

Figure 2: TERM production structure

4.1. TERM sourcing mechanisms

Figure 3 illustrates the details of the TERM system of demand sourcing. Although the figure covers only the demand for a single commodity (Vegetables) by a single user (Households) in a single region (North), the same diagram would apply to other commodities, users and regions. The diagram depicts a series of 'nests' indicating the various substitution possibilities allowed by the model. Down the left side of the figure, boxes with dotted borders show in upper case the value flows associated with each level of the nesting system. These value flows may also be located in Figure 1. The same boxes show in lower case the price (p....) and quantity (x....) variables associated with each flow. The dimensions of these variables are critical both to the usefulness of the model and to its computational tractability; they are indicated by subscripts c, s, m, r, d and p, as explained in Table 1. Most of what is innovative in TERM could be reconstructed from Figures 1 and 2.

At the top level, households choose between imported (from another country) and domestic vegetables. A CES or Armington specification describes their choice—as pioneered by ORANI and adopted by most later CGE models. Demands are guided by user-specific purchasers' prices (the purchasers' values matrix PUR is found by summing the TAX and USE matrices of Figure 1). 2 is a typical value for the elasticity of substitution.

Demands for domestic vegetables in a region are summed (over users) to give total value USE_U (the "_U" suffix indicates summation over the user index u). The USE_U matrix is measured in "delivered" values—which include basic values and margins (trade and transport), but not the user-specific commodity taxes.

Moving down, the next level treats the sourcing of USE_U between the various domestic regions. The matrix DELIVRD shows how USE_U is split between origin regions r. Again a CES specification controls the allocation; substitution elasticities range from 5 (merchandise) to 0.2 (services). The CES implies that regions which lower production costs more than other regions will tend to increase their market share. The sourcing decision is made on the basis of delivered prices—which include transport and other margin costs. Hence, even with growers' prices fixed, changes in transport costs will affect regional market shares. Notice that variables at this level lack a user (u) subscript—the decision is made on an all-user basis (as if wholesalers, not final users, decided where to source vegetables). The implication is that, in North, the proportion of vegetables which come from South is the same for households, intermediate, and all other users.

The next level down shows how a "delivered" vegetable from, say, South, is a Leontief composite of basic vegetable and the various margin goods. The share of each margin in the delivered price is specific to a particular combination of origin, destination, commodity and source. For example, we should expect transport costs to form a larger share for region pairs which are far apart, or for heavy or bulky goods. The number of margin goods will depend on how aggregated is the model database. Under the Leontief specification we preclude substitution between Road and Retail margins, as well as between Road and Rail. For some purposes it might be worthwhile to construct a more elaborate nesting which accommodated Road/Rail switching.

The bottom part of the nesting structure shows that margins on vegetables passing from South to North could be produced in different regions. The figure shows the sourcing mechanism for the road margin. We might expect this to be drawn more or less equally from the origin (South), the destination (North) and regions between (Middle). There would be some scope (σ=0.5) for substitution, since trucking firms can relocate depots to cheaper regions. For retail margins, on the other hand, a larger share would be drawn from the destination region, and scope for substitution would be less (σ=0.1). Once again, this substitution decision takes place at an aggregated level. The assumption is that the share of, say, Middle, in providing Road margins on trips from South to North, is the same whatever good is being transported.

Although not shown in Figure 3, a parallel system of sourcing is also modelled for imported vegetables, tracing them back to port of entry instead of region of production.

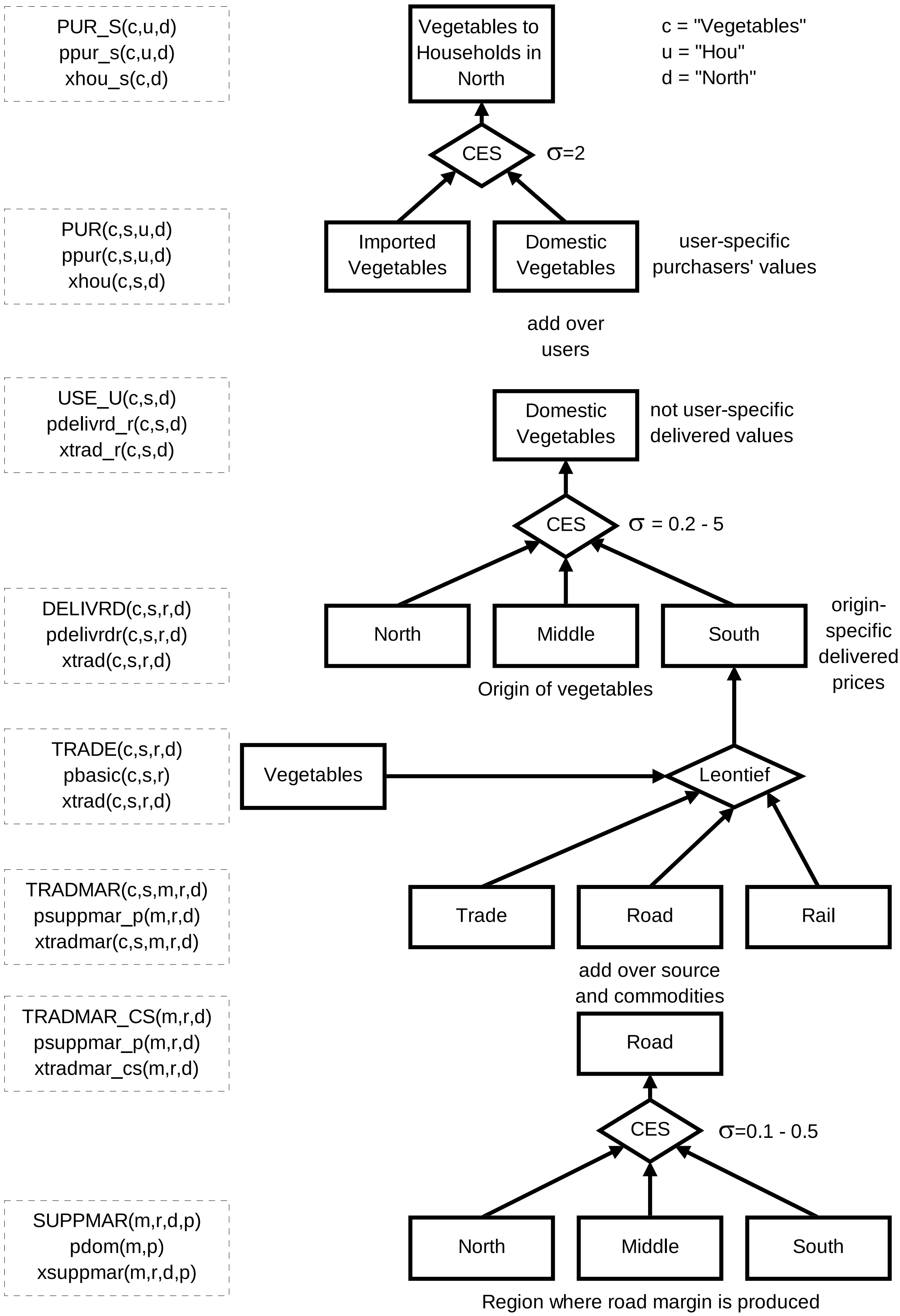

Figure 3: TERM sourcing mechanisms

4.2. Other features of TERM

The remaining features of TERM are common to most CGE models, and in particular to ORANI, from which TERM descends. Industry production functions are of the nested CES type: Leontief except for substitution between primary factors and between sources of goods. Exports from each region's port to the ROW face a constant elasticity of demand. The composition of household demand follows the linear expenditure system, while the composition of investment and government demands is exogenous. A variety of closures are possible. For the shortrun simulation we might hold fixed industry capital stocks and land endowments, whilst allowing labour to be fully mobile between sectors within a region and partially mobile between regions. At the regional level we may link household consumption to regional factor incomes.

4.2.1. National and regional macro closures

Closure flexibility in TERM applies separately at the national and regional levels. For example, we may wish to impose a balance of trade constraint at the national level, without however enforcing balanced trade for each region. We might stipulate that regional consumption C_r follows wage income W_r via a rule like:

$$C_r = F_r W_r \lambda$$

where F_r is a regional propensity to consume and λ is a slack variable which adjusts to satisfy the national balance of trade constraint.

Similarly we might relate government spending in each region to that region's GDP, while holding fixed national government spending.

4.3. Comparison with the GTAP model

GTAP, a well-known model of the world economy, has a fairly similar structure to TERM. The "regions" of GTAP, however, are countries or groups of countries, whilst in TERM they are regions within a single country. In GTAP, regional trade deficits must sum to zero [the planet is a closed system] whilst in TERM a national trade deficit is possible. There are also differences in data structures: GTAP has a far more detailed representation of bilateral trade taxes than does TERM, reflecting the freer trade that is usually possible within a nation. TERM can accommodate commodity tax rates that vary between regions (North might tax whisky more than South) but it does not allow for regional tax discrimination (such as a tax, in North, that applied only to whisky from West). Inter-regional labour movements, a rarity in GTAP, are usual in TERM. Finally, TERM has a more detailed treatment of transport margins. While GTAP identifies how much each country contributes to *world* shipping supply, the TERM data structure shows how much each region contributes to supply of transport between all separate pairs of source and destination regions (the matrix SUPPMAR in Figure 1).

5. Gathering data for 144 sectors and 57 regions

As formidable as the computational demands of regional CGE models, are the data requirements—which usually far exceed what is available.

5.1. The false allure of regional input-output tables

Newcomers to regional CGE often assume that the natural starting point is a published set of regional input-output tables. However, even if such tables are available, they may suffer from serious deficiencies:

- They typically distinguish far too few sectors to support serious CGE modelling.
- The regional coverage may be too coarse, or incomplete (China without Hongkong) or inconsistent (regional tables for different dates, or with different formats).
- They are typically not designed for use by CGE modellers.

Enlarging on the last point, regional IO modellers produce tables for their own purposes, which are different to those of CGE modellers. A central purpose of much regional IO analysis is to estimate the ‘multiplier’: the amount by which employment in some region will increase if autonomous demand (e.g. a construction project) increases by, say, a million dollars. For this purpose it is sufficient if the regional table shows, in a single row, the value of imports (from other regions or the rest of the world) used by each sector and final demander. Indeed, regional IO tables are often presented in this form. Such tables tell us, for example, how much locally-produced gasoline is used by the transport sector, but not how much gasoline (from any source) is used. They cannot therefore support the varied range of applications for which CGE models are designed. For example, we could not easily estimate by how much an increased gasoline tax would increase transport costs.

A regional IO specialist needs some criterion to choose between alternative techniques of constructing regional IO tables. The criterion will reflect the planned use of the tables. For example, Flegg et al. (1995) and Flegg & Tohmo (2009) refer to the effect (of different methods) on estimated multipliers as a way to compare techniques. For a CGE model, other criteria are important and so different techniques of generating regional tables may be preferred.

5.2. The TERM data strategy

By contrast TERM offers a strategy, depicted in Figure 6, to estimate its database from very limited regional data. We describe below some key features of this strategy as applied to the Australian TERM model with base year of 1997[6].

(a) The process starts with a national input-output table and certain regional data. The *minimum* requirements for regional data are very modest: the distribution between regions of industry outputs and of final demand aggregates. This distribution can be conceived as a set of regional shares, which may in turn be based on value data, or on physical units (eg, tonnes of wheat) or on numbers employed. This flexibility (regarding units) greatly increases the amount of data which may be used. Additional regional detail, such as region-specific technologies or consumption preferences may be added selectively, when available.

(b) The process is automated, so that additional detail can easily be added at a later stage.

(c) The database is constructed at the highest possible level of detail: 144 sectors and 57 regions. Aggregation (for computational tractability) takes place at the end of the process, not at the beginning. Perhaps surprisingly, the high level of disaggregation is often helpful in estimating missing data. When aggregated, the model database displays a richness of structure that belies the simple mechanical rules that were used to construct its disaggregated parent. For example, even though we normally assume that a given disaggregated sector has the same input-output coefficients wherever it is located, aggregated sectors display regional differences in technology. Thus, sectoral detail partly compensates for missing regional data[7].

Our technique of combining a national IO table with limited regional data to produce a detailed inter-regional table bears many similarities to methods developed over several decades by regional IO modellers. Indeed, published regional input-output tables may well be in part constructed rather than observed. Unfortunately the method of construction may be poorly documented or unrepeatable. The TERM data programs are downloadable and may easily be be customized to suit particular needs. They will appeal to the modeller who would prefer to construct a multi-regional database using known assumptions, rather than rely on data constructed somehow by others.

Of course, published regional input-output tables may well form part of the inputs to the TERM data process. But they should certainly not constrain the degree of regional or sectoral detail that we aim for.

---

[6] A more detailed description is given in Wittwer and Horridge (2010). However, that paper describes a later edition of the Australian TERM database, which distinguishes more sectors (172 rather than 144) and many more regions (206 rather than 57). The greater regional detail relies on census data that shows employment by sector and small region. The later edition also benefits from more detailed (four-digit) merchandise trade data for 60 ports.

[7] This point is enlarged below in section *5.4.1. Region-specific technology and output mix.*

5.3. The national input-out database

As shown in Figure 6, the TERM data process starts from the 1997 Australian input-output Tables, distinguishing 107 sectors. Our first step was to convert these tables to the file format of ORANI-G, a standard single-country CGE model. Next, working at the national level, we expanded the 107 sectors to 144. In choosing to split sectors, we hoped to avoid infelicities of classification that have caused problems in the past (such as the lumping together of exports of sugar, cotton and prawns) and also to split up sectors which showed regional differences in input mix or sales pattern. For example, we split up electricity generation according to the fuel used (which differs among Australian regions) and added considerable agricultural detail. The interests of one collabarator led to a remarkably detailed treatment of the wine and grape sectors, which were divided according to quality (some regions produce high-quality wine for export, others a cheaper brew for local drinking).

The main source for the sectoral split was unpublished ABS commodity cards data. Such data provide a split of sales for approximately 1,000 commodities to 107 industries, plus final users. However, the cards data do not always provide a desirable split from the 107 industries to the eventual 144 sectors of the disaggregated database. For example, there are significant sales of sugarcane to the other food products sector (107-sector aggregation). We allocated all sugarcane sales to refined sugar and zero sales to the seafood and other food products in our 144-sector disaggregation. When the intermediate sales split was less obvious, we used activity weights of the purchasing sectors for the split.

The 144-sector national database has an independent value for our modelling work (for example, it forms the bulk of the MONASH database). For TERM purposes it was converted to a simpler format prior to the addition of regional detail.

5.4. Estimates of the regional distribution of output and final demands

The next step was to obtain, for each industry and final demander, an estimate of each statistical division's share of national activity (these shares are the R001, R002, etc, of Figure 6). To develop a full input-output table for each region, we required estimates of industry shares (i.e., each region's share of national activity for a given industry), industry investment shares, household expenditure shares, international export and import shares, and government consumption shares.

The main data sources for the industry split were:

- AgStats data from ABS, which details agricultural quantities and values at the SD level;
- employment data by industry at the SD level prepared by our colleague Tony Meagher from ABS census data and surveys;
- published ABS manufacturing census data (state level); and
- state yearbooks (for mining, ABS 1301, and for grapes and wine, ABS 1329.0).

Our sectoral split included a split of electricity into generation by fuel type plus a distribution sector. We relied on the internet sites of various electricity and energy agencies for capacity levels, on which shares of national activity were based.

Manufacturing, mining and services data disaggregated at the statistical division level were in quantities rather than values. These were adjusted these to fit state account sector aggregates (ABS 5220.0), as wages and industry composition vary between states. Industry investment shares are similar to industry activity shares for most sectors. Exceptions include residential construction input shares, set equal to ownership of dwellings investment shares in each statistical division.

Published ABS data (Tables 4 and 5, ABS 6530.0) provide sufficient commodity disaggregation for the task of splitting regional consumption aggregates into commodity shares. Such data also provide a split between capital city regions and other regions within each state.

In compiling international trade data by region, we first gathered trade data by port of exit or entry. For this task, we used both unpublished ABS trade data available for each state and territory plus the annual reports of various ports authorities. Queensland Transport's annual downloadable publication *Trade Statistics for Queensland Ports* gives enough data to estimate exports by port of exit with reasonable accuracy for that state. For other states, port activity is less complex, with most manufacturing trade passing through capital city ports and regional ports specialising in mineral and grain shipments.

State accounts data provide aggregated Commonwealth and state government spending in each region (ABS 5220.0). Employment numbers by statistical division for government administration and defence provide a useful split for these large public expenditure items. For other commodities, population shares by statistical division were used to calculate the distribution of Commonwealth and state government spending across regions.

By applying these shares to the national CGE database, we were able to compute the USE, FACTOR, and MAKE matrices on the left-hand side of Figure 1. None of these matrices distinguish the source region of inputs.

5.4.1. Region-specific technology and output mix

By default, applying regional output shares to a national dataset leads to industry technologies that do not vary by region. That assumption would be very crude, *were it not for the fact that very many sectors are distinguished during data construction*. For example, it seems reasonable to assume that Bananas are grown in the same way in those (few) regions that they grow. It would be less reasonable to assume that "Agriculture" had the same technology.

Within Australia, some regions generate electricity with black coal (which is internationally traded), some with brown coal (which is too bulky to ship). The difference is important as brown coal emits far more CO2. To accommodate the regional technology difference, we distinguished separate brown-coal and black-coal electricity sectors—which each had uniform technology over regions. However, any one region used only one of the technologies. Prior to simulation we could aggregate together brown-coal and black-coal electricity sectors to produce a single electricity sector which burns brown coal in some regions, black coal in others.

A related strategy is used to capture regional variation in crop mix. During the data-building process we usually distinguish a large number of crops, each of which is a single product industry. Thus we avoid complications by assuming a diagonal MAKE matrix. Prior to simulation we may aggregate together various agricultural industries, whilst leaving the associated commodities separate. The effect is that the input technology of "Agriculture" varies by region, as does the crop mix. Moreover, inputs such as land and labour can be switched between between crops, facilitating the analysis of land use change.

5.5. The TRADE matrix

The next stage was to construct the TRADE matrix on the right-hand side of Figure 1. For each commodity either domestic or imported, TRADE contains a 57x57 submatrix, where rows correspond to region of origin and columns correspond to region of use. Diagonal elements show production which is locally consumed. As shown in Figure 6 we already know both the row totals (supply by commodity and region) and the column totals (demand by commodity and region) of these submatrices. For Australia, hardly any detailed data on inter-regional state trade is available. We used the gravity formula (trade volumes follow an inverse power of distance) to construct trade matrices consistent with pre-determined row and column totals. In defence of this procedure, two points should be noted:

- Wherever production (or, more rarely, consumption) of a particular commodity is concentrated in one or a few regions, the gravity hypothesis is called upon to do very little work. Because our sectoral classification was so detailed, this situation occurred frequently.
- Outside of the state capitals, most Australian regions are rural, importing services and manufactured goods from the capital cities, and exporting primary products through a nearby port. For a given rural region, one big city is nearly always much closer than any others, and the port of exit for primary products is also well defined. These facts of Australian geography again reduce the weight borne by the gravity hypothesis.

For a particular commodity the traditional gravity formula may be written:

$$V(r,d) = \lambda(r).\mu(d).V(r,*).V(*,d) / D(r,d)^2 \qquad r \neq d$$

where

V(r,d) = value of flow from r to d (corresponding to matrix TRADE in Fig. 1)

V(r,*) = production in r (known)

V(*,d) = demand in d (known)

D(r,d) = distance from r to d

The $\lambda(r)$ and $\mu(d)$ are constants chosen to satisfy:

$\Sigma_r V(r,d) = V(*,d)$ and $\Sigma_d V(r,d) = V(r,*)$.

For TERM, the formula above gave rather implausible results, especially for service commodities. Instead we set:

$$V(r,d)/V(*,d) \propto \sqrt{V(r,*)}/D(r,d)^k \qquad r \neq d$$

where K is a commodity-specific parameter valued between 0.5 and 2, with higher values for commodities not readily tradable. Diagonal cells of the trade matrices were set according to:

V(d,d)/V(d,*) = locally-supplied demand in d as share of local production

= MIN{ V(d,*)/V(*,d),1} × F

where F is a commodity-specific parameter valued between 0.5 and 1, with a value close to 1 if the commodity is not readily tradable.

The initial estimates of V(r,d) were then scaled (using a RAS procedure) so that:

$\Sigma_r V(r,d) = V(*,d)$ and $\Sigma_d V(r,d) = V(r,*)$.

Transport costs as a share of trade flows were set to increase with distance:

$$T(r,d)/V(r,d) \propto \sqrt{D(r,d)}$$

where T(r,d) corresponds to the matrix TRADMAR in Fig. 1. Again, the constant of proportionality is chosen to satisfy constraints derived from the initial national IO table.

All these estimates are made with the fully-disaggregated database. In many cases, zero trade flows can be known *a priori*. For example, ABS data indicate that rice is grown in only four of the 57 statistical divisions. At a maximum sectoral disaggregation, the load born by gravity assumptions is minimized.

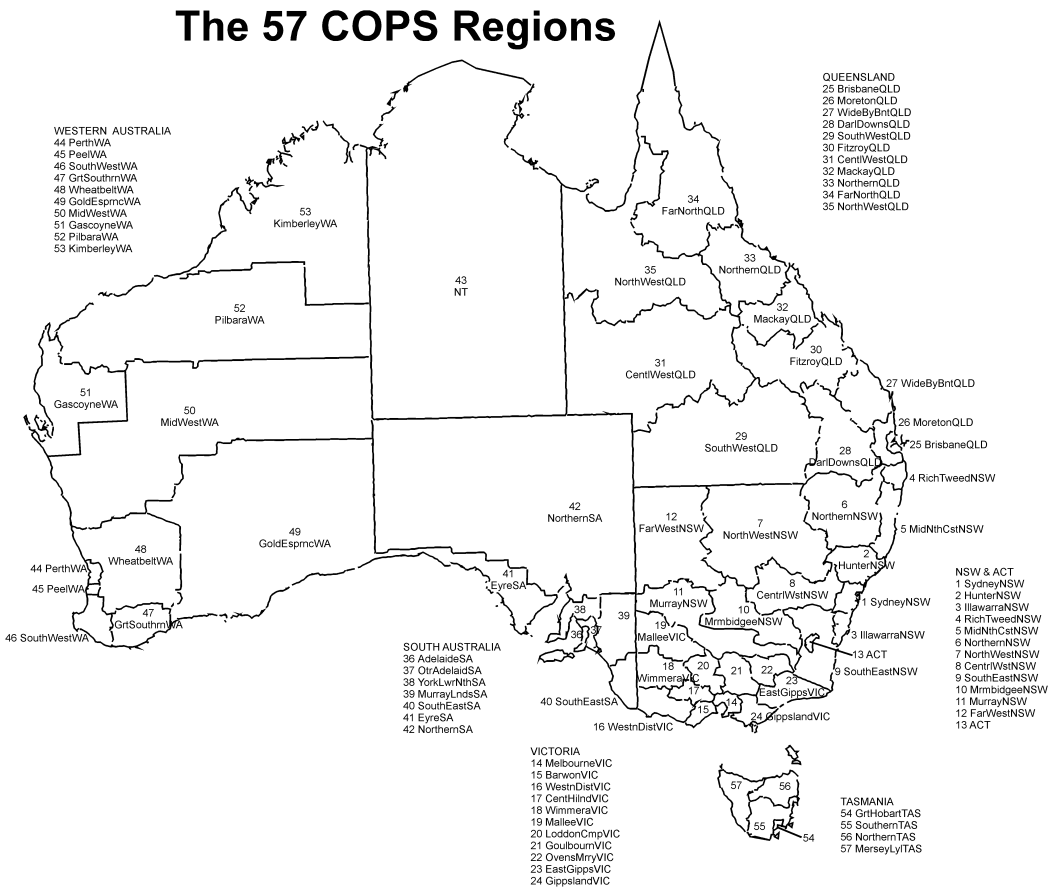

Figure 4: Statistical divisions in Australia

5.6. Aggregation

Even though TERM is computationally efficient, it would be slow to solve if a full 144-sector, 57-region database were used. The next stage in the data procedure is to aggregate the data to a more manageable size. This stage is automated and effortless. The aggregation choice is application-specific. For example, to analyse the effects of drought we might choose a sectoral aggregation that retained detail in the agricultural and agriculture-related sectors, while grouping manufacturing and service industries broadly.

Similarly, the regional aggregation will be tailored to the simulation. For example, to analyse water shortage we might aggregate the original 57 regions of Figure 4 to a smaller number of regions (see Figure 5) that approximately outline the watershed of the Murray-Darling river which feeds most of Australian agriculture. To analyse, say, the spread of dengue fever, a different regional aggregation would be appropriate.

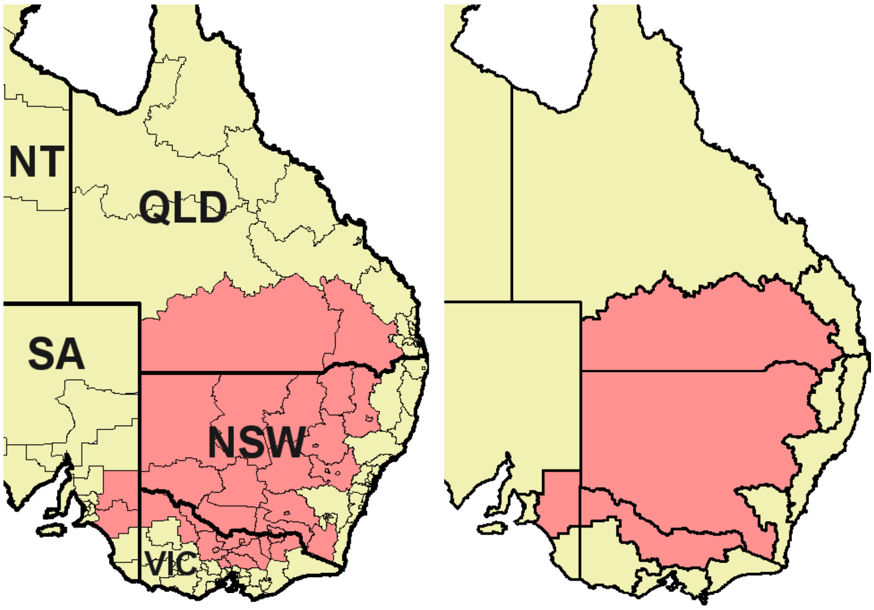

Figure 5: Aggregating from master database to policy simulation (watershed) regions

As Figure 6 shows, the TERM data process supports some other models used at CoPS, including the single region ORANI-G and MONASH models. By aggregating TERM's 57-region database down to the 8 Australian states, we obtain the kernel of the MMRF database. MMRF is still frequently used, since it incorporates features that TERM lacks, such as state government accounts and emissions modelling. TERM is needed when sub-state detail is required, especially if supply-side shocks must be imposed which differ amongst regions within a state.

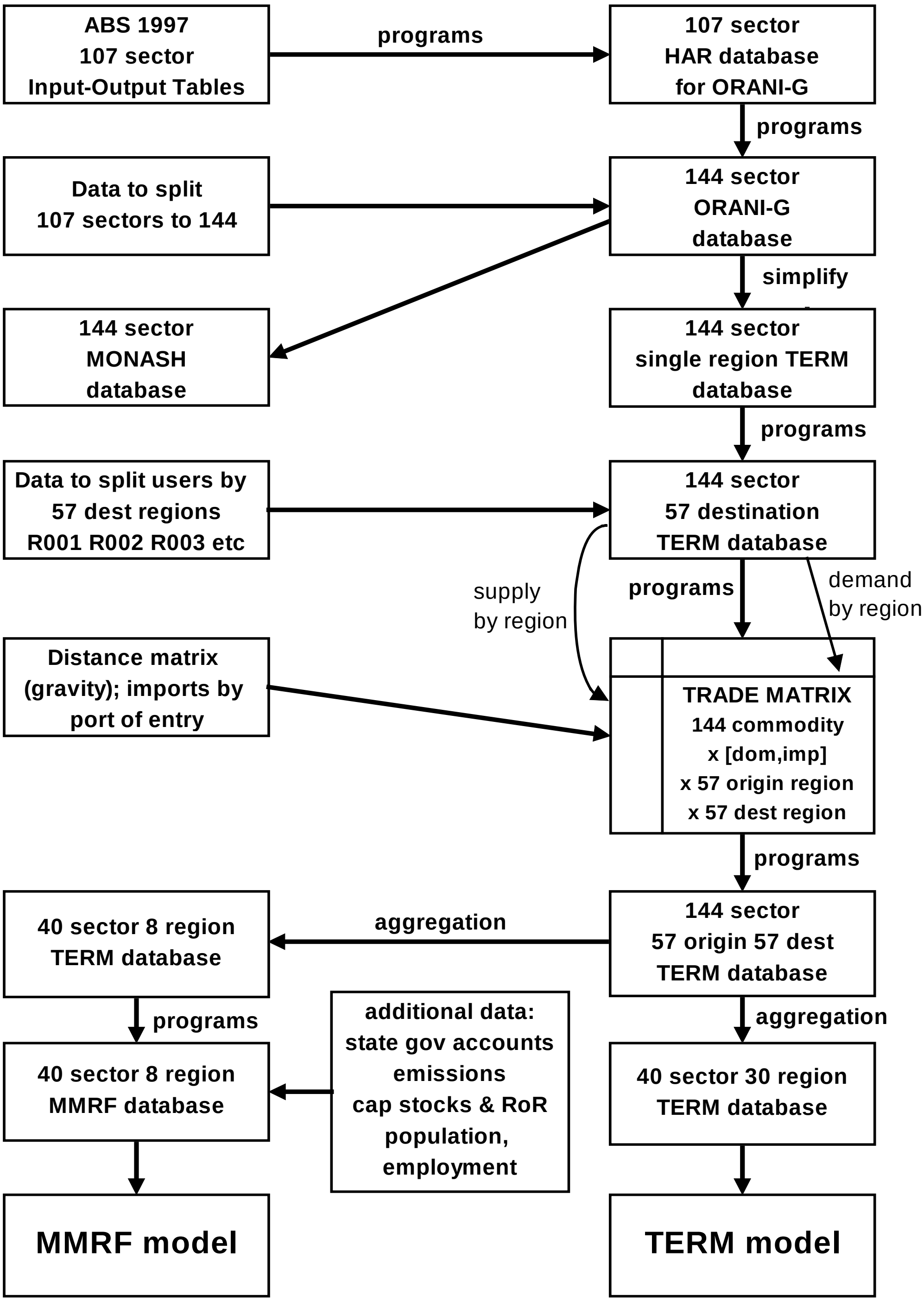

Figure 6: Producing regional databases for MMRF and TERM

6. Conclusion: applications and developments of TERM

The TERM framework, developed originally for comparative static analyses of Australian issues, has been extended in several directions.

- It has been applied to several other countries, including Brazil, China, Finland, Indonesia, South Africa, Poland, USA and Japan. An Italian version is planned.
- Dynamic or multiperiod versions of TERM have been constructed for Australia, Brazil and Finland.
- The South African and Brazilian versions distinguish several household types, to focus on income distribution. One Brazilian version drives a large microsimulation database, distinguishing 100,000 households.
- Another Brazilian version focuses on land use, dividing land into four main types in each region. The aim is to analyse whether increased export and biofuel demand for crops is compatible with preserving Amazon rainforests.
- TERM's capacity to model region-specific supply-side shocks in small regions has proved useful in modelling the effects of natural disasters such as earthquakes, droughts or crop diseases.

The web page http://www.monash.edu.au/policy/term.htm links to a number of these versions.

The wide applicability of the TERM framework rests on two key features: the ability of the model to solve quickly without compromising regional or sectoral detail; and a method to construct a database rapidly, so that the first simulations can be run soon, maximizing the scope for experience to suggest ways to improve the model or data.

References

ABS 2002 (and previous issues), *Australian National Accounts: State Accounts*, Catalogue 5220.0, Canberra.

Adams, P., Horridge, M. and Wittwer, G. 2002, *MMRF-Green: A dynamic multi-regional applied general equilibrium model of the Australian economy, based on the MMR and MONASH models*, Prepared for the Regional GE Modelling Course, 25-29 November 2002.

Dixon, P., B. Parmenter and D. Vincent 1978, "Regional Developments in the ORANI Model", in R. Sharpe (ed.), Papers of the Meeting of the Australian and New Zealand Section Regional Science Association, Third Meeting, Monash University, pp. 179-188.

Dixon, P., Parmenter, B., Sutton, J. and Vincent, D. 1982, *ORANI: A Multisectoral Model of the Australian Economy*, North-Holland, Amsterdam.

Flegg, A. & Tohmo, T. (2008), Regional input-output models and the FLQ formula: a case study of Finland, Discussion Paper 0808, Department of Economics, University of the West of England.

Flegg, A., Webber, C. & Elliott, M. (1995), On the appropriate use of location quotients in generating regional input-output tables, *Regional Studies*, 29, 547-561.

Giesecke, J. 1997, *The FEDERAL-F model*, CREA Paper No. TS-07, Centre for Regional Economic Analysis, University of Tasmania, November.

Higgs, P., B. Parmenter and R. Rimmer 1988, "A Hybrid Top-Down, Bottom-Up Regional Computable General Equilibrium Model", *International Science Review*, Vol. 11, No. 3, 1988, pp. 317-328.

Horridge, J.M., Madden, J.R. and G. Wittwer (2005), "The impact of the 2002-03 drought on Australia", Journal of Policy Modeling, Vol 27/3, pp. 285-308.

Liew, L. 1984, "Tops-Down" Versus "Bottoms-Up" Approaches to Regional Modeling, *Journal of Policy Modeling*, Vol. 6, No. 3, pp. 351-368.

Madden, J. 1990, FEDERAL: a two-region multi-sectoral fiscal model of the Australian economy, Ph.D. thesis, University of Tasmania, Hobart.

Maddock, R. and McLean, I. 1987, “The Australian economy in the very long run”, chapter 1 in Maddock, R. and McLean, I. (editors), *The Australian Economy in the Long Run*, Cambridge University Press.

Naqvi, F. and Peter, M. 1996, “A Multiregional, Multisectoral model of the Australian Economy with an Illustrative Application”, *Australian Economic Papers*, 35:94-113.

Pearson, K. 1988, “Automating the Computation of Solutions of large Economic Models”, *Economic Modelling*, 7:385-395.

Peter, M., Horridge, M. Meagher, G.A., Naqvi, F. and Parmenter, B.R.,1996, *The Theoretical Structure of MONASH-MRF*, CoPS/Impact Working Paper OP-85, June 1996 from: http://www.monash.edu.au/policy/elecpapr/op-85.htm

Wittwer, G. and Horridge, M. (2010) 'Bringing Regional Detail to a CGE Model using Census Data', Spatial Economic Analysis, 5: 2, 229 — 255