Chapter 8.D

Disaggregation of Input-Output Tables

Robert A. McDougall

8.D.1 Overview

In this chapter we document the disaggregation of the input-output (I-O) tables. This is the major step undertaken in our in-house processing of the tables (for an overview of the processing, see Chapter 7). It takes place after some initial clean-ups and adjustments, and before the synthesis of the composite region tables (Chapter 8.F), further minor adjustments, and the fitting of the I-O tables to external data (Chapter 15.A).

While we use a 57-sector classification for the GTAP 7 Data Base, many countries' input-output source statistics do not support such detail. In particular, few I-O sources match the relatively ambitious requirements for agriculture and food processing, where GTAP distinguishes 20 sectors, while some sources distinguish as few as two (agriculture and food processing separately). Accordingly, we accept contributed I-O tables using any reasonable aggregation of this classification, (Huff, McDougall and Walmsley, 2000). Where the contributed table uses an aggregated classification, we must disaggregate it to the standard 57 sectors in-house.

Altogether, in GTAP 7, 43 regions require sectoral disaggregation. Of these four require disaggregation only in non-agricultural sectors, seven require disaggregation only in agricultural sectors, and 32 require disaggregation in both. For non-agricultural sectors we use a disaggregation procedure that is relatively undemanding in data, basing the I-O structure of the disaggregate industries on a *representative table* representing a kind of average of the I-O structures for the regions where we have fully disaggregated contributed I-O tables (Chapter 8.F). Obviously this is less than ideal, but to the extent that industry structures reflect patterns of consumption and technology common across countries, we may hope that it generates reasonable results. For agriculture however, geographic factors such as climate and soil influence the industry structure heavily; here accordingly we use country-specific data, namely the agricultural I-O data compiled by Everett Peterson (Chapter 8.A). This comprises a collection of country I-O tables using a sectoral classification heavily oriented toward agriculture. We perform the agricultural disaggregation before the non-agricultural.

8.D.2 Agricultural Disaggregation using Agricultural I-O Data

In the first stage we disaggregate the agricultural sectors in the contributed I-O tables, using outside information on the I-O structure of the agricultural sector. We apply this procedure to the

39 regions listed in table 8.D.1. At the end of this stage, the contributed tables are fully disaggregated in agriculture, but are in their original classification outside agriculture. Briefly, we follow this procedure:

- Make small adjustments to the agricultural I-O table, designed to replace zeros with small non-zero values in cases where the corresponding value in the contributed table is non-zero. Such zeros would otherwise render the next step infeasible.
- Rebalance the agricultural table to make it consistent with the contributed table, using a scaling factor approach.
- Mutually disaggregate the rebalanced agricultural table and the contributed table, by pro-rating.
- Clean up any minor problems in the disaggregated table.

At the beginning of this procedure we have two I-O tables: the contributed I-O table and an agricultural I-O table. The contributed table is somewhat aggregate in agriculture, and perhaps also outside agriculture, but typically contains many non-agricultural sectors. The agricultural table is drawn from the agricultural I-O dataset; it is fully disaggregated in agriculture, but highly aggregated outside agriculture.

To bring these two tables into contact with each other we use a sectoral classification that constitutes the *finest common aggregation* (*FCA*) of those for the two initial I-O tables. This is a *common* aggregation, because it is an aggregation of both the sectoral classifications used in the contributed table and in the agricultural I-O tables; and it is the *finest* (that is, the most detailed) such classification. In this context, the FCA classification matches the contributed table sectoral classification in agriculture, and the agricultural I-O classification outside agriculture.

In the central step in the procedure, we rebalance the agricultural I-O table so that it is consistent with the contributed table at the finest common aggregation. That is, if one aggregates both the rebalanced agricultural I-O table and the contributed table to the FCA classification, the results are the same. Since the FCA classification has exactly as much agricultural detail as the contributed I-O table, this means that that if one aggregates the agricultural sectors in the rebalanced agricultural I-O table to the same sectoral classification as the contributed table, they agree with the contributed I-O table. In other words, the rebalanced agricultural I-O table is a disaggregation of the contributed I-O table in agriculture.

Following an entropy-theoretic approach, the rebalancing is undertaken using scaling factors. We define a *block scaling factor* for each *block* of cells in the agricultural table corresponding to a single cell in the FCA. The block may be 1×1, 1c*n*, *m*×1, or *m*×*n*, depending on whether the row and column containing the cell in the FCA are aggregated relative to the agricultural I-O classification. Suppose for example that the only disaggregation required is to separate “wool; ruminant meat” in the contributed table into two sectors, “wool” and “ruminant meat”. Then the agricultural I-O table contains a 2×2 block corresponding to intra-industry usage of “wool; ruminant meat” in the FCA table, a 2×1 block for each element of the“wool; ruminant meat” sales structure in the FCA table, and a 1×2 blocks for each element of the cost structure of the “wool; ruminant meat” industry in the FCA table.

We balance the table by matching each cell to a block and scaling it by its block scaling factor. We set the block scaling factors to achieve certain block target totals; we get the block target totals by aggregating the contributed table to the FCA. Thus after balancing, the agricultural I-O table is a disaggregation of the aggregated contributed table.

At the same time, we impose targets on individual row and column totals in the agricultural I-O table. Each row total represents total sales of some input: a domestically produced commodity, an imported commodity, or a primary factor. Aggregated to the FCA, these must agree with the contributed table; but within each aggregate commodity, we maintain the individual commodity shares of the original agricultural I-O table. So in the example previously given, if the ratio of total sales of wool to total sales of ruminant meat is 30:70 in the original agricultural I-O table, and total sales of "wool; ruminant meat" in the contributed table are $6 billion, then the total sales targets for "wool" and "ruminant meat" in the balanced table is 30 per cent of $6 billion, or $1.8 billion.

Just as we use block scaling targets to achieve the block target totals, we use row and and column scaling factors to achieve the row and column target totals. Thus each cell in the agricultural I-O table is scaled by three factors: one row-specific, one column-specific, and one block-specific. There are some further complications, relating to data categories with mixed sign data (changes in stocks, production taxes) and to avoiding singularities in the equation system (unacceptable in GEMPACK, liable to occur because of linear dependencies between the target totals), but in essence the balancing program just sets and applies the scaling factors.

It is debatable whether targeting the row and column totals is beneficial. The decision to apply them was made largely on grounds of conservatism. We targeted the row and column totals in the previous data base release; since the disaggregation procedure seemed to work reasonably well then, we thought it safer to follow a similar approach with the current release. But since rebalancing without row and column target totals has worked well outside agriculture in the current release (section 8.D.3), we may consider extending it to agriculture in future releases.

Table 8.D.1 Discrepancy in Industry Shares in Agricultural Factor Costs between Agricultural and Contributed I-O Tables

GTAP Region	Entropy Discrepancy
Panama	2.94
Myanmar	1.93
Hong Kong	1.57
Kyrgyzstan	1.44
Pakistan	1.35
Ethiopia	1.14
India	0.98
Armenia	0.92
Peru	0.74
Morocco	0.71
Malaysia	0.69
Philippines	0.67
Ecuador	0.61
Iran	0.58
Nicaragua	0.55
Tunisia	0.53
Indonesia	0.46
Paraguay	0.45
Senegal	0.43
Nigeria	0.43
Georgia	0.42
Laos	0.41
Viet Nam	0.40
Canada	0.39
Uganda	0.38
Bolivia	0.37
Azerbaijan	0.34
Egypt	0.26
Belarus	0.25
Sri Lanka	0.23
Russia	0.21
Kazakhstan	0.20
Mauritius	0.20
Chile	0.19
Costa Rica	0.19
China	0.16
Ukraine	0.16
Turkey	0.12
Thailand	0.11

Comparing the agricultural with the contributed I-O tables would be a very extensive project; we make only a beginning here. We compare industry shares in agricultural factor costs across the two tables. Table 8.D.1 uses the *entropy discrepancy* to indicate the degree of discrepancy between the two structures. To compare the two I-O tables we need to aggregate them to the FCA; therefore comparing results across countries may be misleading, since the smaller the number of agricultural sectors in the contributed table, the lower the discrepancy is likely to be. In the extreme, if the contributed table has just one agricultural sector, then the discrepancy is necessarily zero.

The regions with the largest discrepancies include small economies such as Panama and Hong Kong. Taking account of the absolute size of the agricultural sector, the most significant discrepancies are those for India, China, and Pakistan. Table 8.D.2 shows the factor cost shares for these regions.

While the differences between the two sources are generally moderate for China, for India and Pakistan they are considerable. We note in particular, that for India, the share of beverages and tobacco is much higher in the agricultural than in the contributed table, and for Pakistan, the share of “other food products” is much higher in the agricultural table and the share of livestock much higher in the contributed table. Features common to all three listed countries are that the agricultural tables show lower shares for livestock and higher shares for beverages and tobacco products than do the contributed tables.

While the differences between the two sources are generally moderate for China, for India and Pakistan they are considerable. We note in particular, that for India, the share of beverages and tobacco is much higher in the agricultural than in the contributed table, and for Pakistan, the share of “other food products” is much higher in the agricultural table and the share of livestock much higher in the contributed table. Features common to all three listed countries are that the agricultural tables show lower shares for livestock and higher shares for beverages and tobacco products than do the contributed tables.

8.D.3 Non-Agricultural Disaggregation

In the final stage we disaggregate the non-agricultural sectors. At the beginning of this stage, the contributed tables are fully disaggregated in agriculture, but in their original classification outside agriculture; at the end, they are fully disaggregated across all sectors. The procedure applies to 36 regions; of the 43 regions subject to disaggregation, all but seven require some disaggregation outside agriculture. The procedure is broadly similar to that for the agricultural disaggregation (section 8.D.2), but with some differences:

— The disaggregation is based on the representative I-O table (section 8.D.1), not a region-specific agricultural I-O table.

Table 8.D.2 Shares in Agricultural Factor Costs for Selected Countries: Comparison between I-O Tables (percent)

	Contributed Table	Agricultural Table
India		
Paddy rice	14.8	16.8
Wheat	8.2	3.0
Other grains	2.7	1.6
Vegetables and fruits	16.5	10.5
Oilseeds	11.5	4.5
Sugar cane and beet	5.4	2.0
Plant-based fibres	3.5	2.8
Other crops	10.5	3.1
Cattle and sheep	0.0	0.3
Other livestock; wool	6.9	0.6
Other food products	16.6	32.7
Other meat	0.0	0.3
Vegetable oils	0.9	0.9
Sugar	0.9	1.4
Beverages and tobacco	1.7	19.4
China		
Crops	58.0	56.9
Livestock	23.7	13.0
Meat and dairy products	1.2	4.9
Other food products	10.9	12.2
Beverages and tobacco products	6.2	12.9
Pakistan		
Paddy rice	2.6	3.8
Other cereals	6.9	5.4
Vegetables and fruits	13.2	7.1
Oilseeds	1.0	3.5
Sugar case and beet	3.6	2.0
Plant-based fibres	5.0	8.6
Other crops	8.9	0.4
Livestock	39.8	12.8
Other food products	6.7	32.3
Vegetable oils	0.8	0.4
Processed rice	2.3	2.7
Sugar	5.1	1.2
Beverages and tobacco products	4.2	19.8

— We don't need to construct an FCA sectoral classification (section 8.D.2). Since the representative table uses the full GTAP sectoral classification, the FCA is just the classification used in the incoming contributed table (i.e., fully disaggregated in agriculture but not more or less aggregated outside agriculture).

— In rebalancing the representative table to match the contributed table, we do not set row or column target totals for disaggregate sectors.

— Since the rebalanced representative table is already fully disaggregated, we do not need a mutual disaggregation step.

The decision not to target individual row or column totals in the non-agricultural disaggregation contrasts with the treatment of agriculture. For agriculture we could use region-specific agricultural I-O tables to set the target totals; for the non-agricultural sectors however we do not have a corresponding region-specific information source.

Although we do not target row or column totals, we still use column-specific scaling factors, at least for columns representing industries. We use them now to maintain sectoral balance, that is, to ensure that each industry's total costs is equal to its total sales. But we do not need row scaling factors anymore. Thus each cell is subject to two scaling factors: a block scaling factor, to ensure agreement with the aggregate table (as in section 8.D.2), and a column scaling factor, to ensure sectoral balance.

Under this approach, the shares of disaggregate sectors in total sales of an aggregate sector emerge from the rebalancing process. They are typically close to those in the representative table, but can differ for compositional reasons. Table 8.D.3 reports the cases where the change is greatest, as measured by the entropy discrepancy, not in total sales but in total domestic absorption, counting domestic product and imports together.

Of the 158 non-agricultural sectors disaggregated in 36 regions, table 8.D.3 shows the ten with the greatest entropy discrepancy. Of these, seven involve minerals, either fossil fuels only, or both fossil fuels and other minerals. These sectors figure prominently because they are usually sizable, many tables group them together, and their sales dispositions differ widely. Unlike the agricultural sector revisions reported in table 8.D.2 in section 8.D.2, these revisions need not indicate any data issues: those arose from inconsistencies between data sources, while these arise from the incorporation of region-specific information into a region-generic proxy data source.

In Uganda, for example, "other minerals" accounts for 72 per cent of total minerals absorption, compared to 19 per cent for the representative table. The main reason is that the construction accounts for 60 per cent of minerals absorption in Uganda, compared to 3 per cent in the representative table, and that industry's usage of minerals skews heavily toward "other minerals". For Costa Rica, similarly, "other minerals" accounts for 70 per cent of total minerals absorption; here not construction but "other manufacturing" accounts for an unusually high share of minerals usage, 79 per cent, compared to 0.3 per cent in the representative table, and "other manufacturing" usage of minerals likewise skews toward "other minerals". For these cases at least, the policy of not targeting row totals seems justified; we should not wish greatly to reduce the "other minerals" intensity of the Ugandan construction industry or the Costa Rican "other manufacturing" industry in order to meet such targets.

We note in passing that there appears to be an error in the aggregation mapping for Mauritius, with not "oil" but "gas" mapped to "oil_omn". The effects of this error on the final data base are reduced by our using IEA-based energy data (Chapter 11) to impose energy usage targets on the disaggregate tables (Chapter 15.A).

Table 8.D.3 Aggregation Shares in Absorption, Selected Commodities (per cent)

	Representative Table	Disaggregated Regional Table
Uganda, mine: 0.67		
Coal	16.9	5.9
Oil	29.3	15.9
Gas	34.8	6.6
Other minerals	19.0	71.6
Mauritius, oil_omn		
Gas	64.7	14.2
Other minerals	35.3	85.8
Costa Rica, cogm: 0.58		
Coal	16.9	8.1
Oil	29.3	9.1
Gas	34.8	12.8
Other minerals	19.0	70.0
Bolivia, silvczpe: 0.30		
Forestry	47.5	83.2
Fishing	52.5	16.8
Malaysia, col_omn: 0.21		
Coal	47.0	17.1
Other minerals	53.0	82.9
Azerbaijan, omf: 0.17		
Wood products	16.4	11.2
Machinery, eqpt	65.5	44.0
Other manufactures	18.1	44.8
Thailand, colomn: 0.16		
Coal	47.0	20.4
Other minerals	53.0	79.6
Guatemala, clolg: 0.16		
Coal	20.8	4.4
Oil	36.2	54.0
Gas	43.0	41.6
Russia, osg_ros: 0.14		
Recreatl services	27.0	7.5
Health, education	73.0	92.5
Kyrgyzstan,oilgas: 0.13		
Coal	20.8	44.6
Oil	36.2	22.7
Gas	43.0	32.7

Header rows show: *Country*, *sector*: *entropy discrepancy*.

References

Huff, K., McDougall, R. and Walmsley, T. 2000. "Contributing Input-Output Tables to the GTAP Data Base," GTAP Technical Paper No. 1, Center for Global Trade Analysis, Purdue University.

Liu, J. and McDougall, R. 1998. "Disaggregation of Input-Output Tables," Chapter 13 in McDougall, R., Elbehri, A., and Truong, T.P. *Global Trade, Assistance, and Protection: The GTAP 4 Data Base*. Center for Global Trade Analysis, Purdue University.