Chapter 8.B

Food and Agricultural Data Base: Local Modifications

Robert A. McDougall

8.B.1 Overview

Some local modifications were made to the agricultural and food products (*AFP*) dataset described in chapter 8.A. For reasons relating to the overall data base construction system, we generate input-output (I-O) tables not just for the countries that need them immediately but for all countries. For convenience in the subsequent I-O table disaggregation (chapter 8.D), we fill in the I-O structure to match more closely the structure of the contributed tables. And to avoid strange I-O structures for some industries in some countries, we adopt a new procedure for balancing the I-O tables (*cf*. section 8.A.2.4). Of these modifications, the last is the most substantial.

For the GTAP 6[1] Data Base, we use the AFP dataset to disaggregate AFP sectors in thirteen countries (chapter 8.D). But this may change in later point releases (that is, in release 6.1, *etc*.), as indeed it changed during the preparation of GTAP 6 itself. To avoid recreating the AFP dataset each time we need AFP disaggregation for a new country, we generate at the outset I-O tables not just for the countries needed for GTAP 6 but for all standard GTAP countries.

The AFP I-O dataset is in effect a collection of single-country I-O tables, with a sectoral classification heavily oriented towards agriculture. More specifically, the commodity classification distinguishes the 20 AFP sectors and a residual *other* commodity; the sectoral classification distinguishes the 20 AFP sectors, three AFP-using industries (textiles, leather products, and chemical, rubber, and plastic products), and a residual *other* industry. The initial version has certain special features, reflecting the nature of its sources:

— It distinguishes between usage of domestic products and imports for the 20 AFP commodities, but not for the residual commodity.

— It contains no data for usage of non-AFP inputs by non-AFP industries. In particular, it contains no data for usage of the residual commodity by the four non-AFP industries, for employment of primary factors by the non-AFP industries, or for final demand for the residual commodity.

These special features reflect the nature of the source data on which the AFP tables are based. But the agricultural I-O disaggregation procedure, where the AFP data are used, works by

[1] Editor's note: Although this chapter illustrates these modifications for GTAP 6 Data Base, the procedure remains unchanged for version 8 Data Base as well.

adjusting the AFP I-O table for each country to match the more aggregate data in the contributed table. It is therefore convenient for the disaggregation procedure for the AFP I-O table to match the contributed table in structure, and to present non-zeros in all appropriate cells. So, for convenience, we separate residual commodity usage into domestic product and import components, and fill in the empty non-AFP cells. Since in the end the non-AFP cells are governed by the contributed table, the values used in filling in these cells in the AFP dataset are not very consequential; it suffices that they be of more or less the right order of magnitude.

Another special characteristic of the AFP dataset is that the constituent I-O tables are notably sparse. This is partly by design — most AFP commodities are assumed to have only a few uses — and partly an effect of the estimation procedure — the Tobit procedure yields many zero estimates. But since sparse tables can be difficult or impossible to balance, it is expedient to replace some of the zeros with small non-zero values. This we do for the zeros resulting from the Tobit estimation procedure, not for the zeros imposed by design.

The most substantial adjustment pertains to the balancing of the I-O tables. As described in chapter 8.A, the AFP dataset is constructed in two stages: first, money values of production (and certain other variables) are estimated for each AFP commodity and country; then, the AFP I-O tables are estimated consistent with those totals. In the second stage, total costs and total sales of each AFP sector in each country must agree with the production estimates obtained in the first stage. To ensure this, the second stage itself has two substages. First, money values of usage are estimated cell by cell; then, these initial estimates are balanced against the production totals. The balance conditions apply both to the row totals (total sales) and the column totals (total costs) of the tables.

In the original AFP dataset, this is done in the simplest possible way. In cell-by-cell estimation, only primary factor employment and AFP commodity usage are estimated; usage of the residual commodity is ignored. After completion of the cell-by-cell estimation, the AFP commodity rows are rescaled to match total sales to total production. Residual commodity usage is then set to match total costs to total production. While this approach has the advantage of simplicity, it has for our purposes a critical disadvantage in that it yields many negative values for usage of the residual commodity. These unfortunately are unusable in our disaggregation procedure. Accordingly, we adopt a new approach to the balancing: we generate independent estimates of residual commodity usage, and then, employing a more complex RAS-like procedure, rebalance simultaneously both the rows and the columns against the production totals.

Unlike the estimates for non-AFP usage by non-AFP industries, the estimates for non-AFP usage by AFP industries are consequential, since they can strongly affect the cost structures of the AFP industries in the disaggregated tables.

Table 8.B.1 summarizes the overall data generation story. Blocks marked “0” are filled with zeros by design; those marked “1” use the cell-by-cell estimates from the initial construction (and these are the largest and most important part of the dataset); those marked “2” are re-estimated during local adjustment; those marked “3” are filled in for the first time during the local adjustment (though the exact values are not consequential). We note however that the new balance procedure can affect all non-zero cells in the dataset, including those in the “1” block.

Table 8.B.1 Handling of Data Cells in AFP Data Construction

	AFP	AFP-Using	Residual	Final Demand
Domestic AFP	1	1	0	1
Domestic Residual	2	3	3	3
Imported AFP	1	1	0	1
Imported Residual	2	3	3	3
Final Demand	1	3	3	0

0 : Zero by design.
1 : Estimated in the initial construction.
2 : Estimated in the initial construction, for domestic product and imports jointly: re-re-estimated in the local adjustments for domestic product and imports separately.
3 : Added in the local adjustments.

Altogether, the local adjustments constitute a new front end to the AFP dataset construction process:

1. Estimate usage of the residual commodity by AFP industries.
2. Separate residual commodity usage into usage of domestic products and usage of imports. Estimate usage of non-AFP inputs in non-AFP production and in final demand.
3. Adjust the dataset to replace zero estimates from the Tobit estimator with small non-zeros.
4. Balance the I-O tables against the production and import estimates obtained in the first stage of the initial construction.

The most important steps are (1) and (4); (2) and (3) may be considered minor technical adjustments. In the remainder of this chapter, we discuss in more detail all these steps but mostly steps (1) and (4).

8.B.2 Estimating Usage of the Residual Commodity by AFP Industries

To estimate non-AFP intermediate usage by AFP industries, we first estimate the share of non-AFP intermediate usage in total input costs, then apply that share to the agricultural and food data base estimate of AFP production. In the spirit of the AFP dataset construction, rather than simply apply a global average share, we estimate a share equation econometrically, using per capita GDP as an explanatory variable. For data, we use a development version of the GTAP Data Base, together with the population and per capita GDP data described in chapter 6.

Our leading concern is to keep the estimated shares within reasonable bounds, even for extreme values of the explanatory variable. To this end, we use a special form for the dependent variable:

$$f(S) = \log \frac{S}{\sqrt{S(1-S)}} = \frac{1}{2}\left[\log S - \log(1-S)\right] \tag{1}$$

where *S* denotes the share of non-AFP intermediate usage in total input cost. We note that the function *f(S)* is increasing in *S*, that *f(S)* decreases without limit as *S* approaches 0 from above, and that *f(S)* increases without limit as *S* approaches 1 from below. The advantage of this form is that all possible values of *f(S)* correspond to feasible *S*, that is, to shares lying between zero and one.

With this transformed dependent variable, we apply a supply linear regression. For the independent variables we use a constant and the logarithm of per capita GDP. We estimate separately for each AFP industry.

We find that, in general, the share increases with increasing per capita GDP. We observe:

— The regressions are polarized between very high significance for some industries and very low significance for others. Only one industry, wool, shows a moderate confidence level, between 95 and 99 per cent.

— The regressions are significant for all primary products and insignificant for most food products.

— In all significant regressions, the coefficient on the GDP variable is positive.

We cannot draw inferences about the real world directly from these results, since the data include synthetic I-O tables for composite regions, constructed on the assumption that I-O structure and per capita GDP are linked. Nevertheless, the results do show that a strikingly strong positive relation exists between per capita GDP and the cost share of the residual input in agriculture in the GTAP Data Base. In future, it might be desirable to revise the estimation using the contributed tables rather than a previous finished data base.

Table 8.B.2 Estimation of Cost Share for Residual Input

Commodity	Estimated Coefficient for GDP	*F* statistic	Significance[1]
Paddy rice	0.057	14.2	2
Wheat	0.098	27.2	2
Other grains	0.116	33.8	2
Vegetables and fruits	0.101	23.3	2
Oilseeds	0.129	38.4	2
Sugar cane, beet	0.096	15.6	2
Plant-based fibers	0.072	7.6	2
Other crops	0.075	16.1	2
Cattle, sheep	0.066	9.2	2
Other livestock	0.174	82.7	2
Raw milk	0.133	31	2
Wool	0.117	4.7	1
Cattle, sheep meat	-0.056	3.6	0
Other meat	0.09	13.1	2
Vegetable oils	0.027	2.3	0
Dairy products	0.015	0.5	0
Processed rice	0.011	0.1	0
Sugar	0.015	0.8	0
Other food products	0.119	49.3	2
Beverages, tobacco	0.087	58	2

[1]Significance indicators are:
0 not significant at 95 per cent confidence level
1 significant at 95 per cent confidence level
2 significant at 99 per cent confidence level

Combining our industry-specific coefficient estimates with country-specific per capita GDP data, we estimate the share of intermediate usage of the non-AFP commodity in total input costs of each AFP industry. Applying these to the agriculture and food data base production estimates, we obtain estimates of the money value of intermediate usage of the non-AFP commodity for each AFP industry in each GTAP country.

8.B.3 Filling Out the I-O Tables

As noted above (section 8.B.1), it is convenient for the I-O table disaggregation (chapter 8.D) that the AFP tables should be complete, in the sense that they should contain non-zero estimates for all cells corresponding to non-zero cells in the fully disaggregated tables. To this end, we perform several minor tasks:

— estimate usage of the residual commodity and the individual primary factors by non-AFP industries;

— split the residual commodity into imports and domestic product, and

— estimate final domestic consumption and imports of the residual commodity.

The results obtained for the split of residual commodity usage by AFP industries are somewhat important; if we arbitrarily vary import shares across industries, this will carry through to the disaggregate tables. For the other cells, the exact results are not critical; all that is required is that, for computational reasons, the estimates should be of the right sign and a reasonable order of magnitude.

For those other cells, we first estimate usage of the residual commodity and employment of each primary factor by each AFP-using industry. We can now calculate total factor usage by the residual industry, as economy-wide factor usage less total factor usage by AFP and AFP-using industries. Unfortunately, in some cases, this yields a negative estimate for some countries, including two whose contributed tables require agricultural disaggregation, Vietnam and Uganda.

To avoid these negatives, we proceed as follows: We obtain an econometric estimating equation for some transformation of primary factor usage by the residual industry; we determine not only the mean but the standard deviation of the dependent variable; we set the mean less three standard deviations as the minimum acceptable value for the dependent variable (the transformation applied is such that this minimum always corresponds to a positive level of primary factor usage); we adjust factor usage so as never to fall below this minimum. This entails adjusting not only the countries for which the initial residual calculation was negative, but some for which it was positive but unexpectedly small. In all, we adjust estimates for 63 out of 226 countries: Vietnam and Uganda, eight other countries with initial negatives not subject to I-O table disaggregation, and 53 countries with positive but unexpectedly small initial estimates (including Bulgaria and New Zealand, but no countries subject to I-O table disaggregation). These results may have some diagnostic value, as indicating possible overestimation of AFP factor employment in those countries (they may of course also indicate errors in our econometric estimates of residual factor employment).

Further steps include distributing factor employment by the residual industry across individual primary factors, and ensuring positive estimates for final domestic consumption and exports of the residual commodity.

To estimate the share of imports in usage of the residual commodity in each AFP industry, we again use the transformation of equation (1) to the dependent variable. For independent variables, we take the logarithm of per capita GDP and a transformed macro import share variable. We estimate positive coefficients for the macro import share variable for all industries, and positive coefficients for the per capita GDP variable for most industries. Cross-industry variation in the coefficients will lead to some cross-industry variation in the import shares in the disaggregated tables; the estimated coefficients are such that the cross-industry variation remains within reasonable limits. The usefulness of the resultant information content is a question for further research.

To replace zero estimates from the Tobit estimators by small positives, we set matrix shares in each table to a weighted average of the initial matrix shares and the cross-country average matrix shares, assigning to the initial matrix shares a weight close to but strictly less than one, and to the average matrix shares a weight close to but strictly greater than zero.

8.B.4 Rebalancing the I-O Tables

Because we have estimated usage of the residual commodity independently, rather than calculating it as a residual, we have a somewhat complex rebalancing to perform. We must adjust the data so that, for each AFP sector, total sales and total costs agree with Everett Peterson's target totals (chapter 8.A); and we must also ensure that total sales of the residual commodities match total costs summed over the four non-AFP industries (three AFP-using industries, and the residual industry). This we achieve using an entropy-theoretic RAS-like rebalancing procedure.

Table 8.B.3 compares initial estimates with target totals. We order the results by a measure of the seriousness of the discrepancy, and show just the cases where the discrepancies are greatest. We restrict the comparison to countries that actually undergo agricultural disaggregation for this GTAP release. We cover three kinds of totals: total imports by commodity and country, total sales of domestic product by commodity and country, and total costs by industry and country.

Some observations stand out:

- While many of the estimates (among those not reported here!) are impressively accurate, the inaccurate ones can be very inaccurate indeed.
- All the greatest discrepancies occur for China and India (India is listed in chapter 8.A as one of the countries for which fully disaggregated I-O data were available; in the course of preparation of the data base, however, a decision was taken to discard some of the sectoral detail from the original contributed table).
- In general, imports fare worse than domestic products.
- In all the greatest discrepancies, the initial level exceeds the target rather than *vice versa*.

Further investigation shows that the discrepancies in totals are driven by discrepancies for certain individual cells. For example, as shown in the first entry in the table, the initial estimate for vegetable and fruit imports into China is too high; this is driven by high estimates for usage of imported vegetables and fruits in certain industries, for example, "other livestock"; the same cell helps drive the second discrepancy noted in the table, excess total costs in the China's "other livestock" industry.

We note that there is no danger of a huge usage of vegetable and fruits imports by the China's "other livestock" industry occurring in the final GTAP Data Base. First, we rebalance here against the import total. Second, in the I-O disaggregation, the disaggregate number is constrained by the corresponding aggregate number in the contributed I-O table, in this case, the total value of usage of imported crops by the livestock industry. It is likely, however, that of the available total, too great a share is allocated to vegetable and fruit usage by the "other livestock" industry.

Table 8.B.3 Comparison of Initial AFP I-O Table Estimates with Target Totals (US$ Billion)

Commodity	Country	Variable	Initial	Target
Vegetables and fruits	China	imports	1951	0
Other livestock	China	costs	2502	80
Oilseeds	China	imports	670	0
Oilseeds	India	imports	364	0
Milk	India	costs	563	15
Wheat	China	imports	206	2
Vegetables and fruits	India	imports	143	0
Cattle and sheep	China	costs	212	7
Wheat	India	imports	103	0
Wool	China	costs	96	3
Other crops	India	imports	51	0
Milk	China	costs	69	2
Other livestock	India	costs	61	1
Cattle and sheep	India	costs	43	1
Other food products	China	sales	237	126
Beverages and tobacco	China	sales	196	106

It may be noted that the results reported in table 8.B.3 pertain entirely to less developed economies, and, more specifically, to China and India. Beyond the top sixteen entries reported in that table, the first occurrence of a country either than China and India is for Indonesia, in eighteenth place. This may suggest that some special problems arise for developing countries in general, or for China and India in particular. But this conclusion would be hasty. First, it so happens that in GTAP 6, the only countries undergoing agricultural disaggregation are less developed economies; hence no other countries can appear in the table. Furthermore, our measure of seriousness of discrepancy takes account of the absolute magnitudes involved, and these of course are much greater for China and India than for other developing economies.