# A Cross-Country Examination of Entropy Errors
## Report prepared for the 2006 GTAP Board Meeting

**Tasneem Mirza and Terrie Walmsley**

This report examines the cost shares from 39 country I-O tables and compares them to an average share obtained from a representative table[1] using an entropy measure developed in Walmsley and McDougall (2004).

The entropy method attaches a number which measures the degree of inconsistency between the country specific I-O table share (ioshr) and the representative table share (cmpshr). This number can then be used to highlight any discrepancies which are considered 'too large'. For our purposes an entropy value greater than 0.05 indicates that the entropy level is "too high" and hence the I-O table share is significantly different from the share in the representative table. The entropy value itself has no meaning. If the representative table had a share of 50%, then an entropy value of greater than 0.05 would arise if the share in the country's I-O table was less than 26% or greater than 82%. The range shortens as the share in the representative table declines. That is the difference between 75% and 80% is considered smaller than the difference between 0% and 5%. The table below shows the bounds for other shares and full details are provided in Walmsley and McDougall (2004).

| Representative Table share | Maximum | Minimum |
|---|---|---|
| 0.1% | 0.014% | 0.27% |
| 10% | 27% | 1.4% |
| 25% | 49% | 9% |
| 50% | 82% | 26% |
| 75% | 100% | 45% |

Large differences or entropy values may arise as a result of:

a) a problem with the I-O table for the USA;
b) a problem with the representative table;
c) a problem with another I-O table which is affecting the share in the representative table; or
d) valid peculiarities particular to that country.

Table 1 depicts the input-output pairs of industries for which more than 20 countries (out of a total of 39 that we investigated) have shares which differ significantly from the share in the representative table[2]. The column 'Number of errors' indicates the number of countries that have errors for a particular pair of input-output industries.

[1] The representative table is an average I-O table for a typical country. It is found by essentially a weighted sum of the I-O tables of 34 countries. out of 39 countries that are investigated. Countries that frequently caused biases in the cmpshr values due to very high (or low) ioshr have been left out. The following countries have not been used to create the representative table: Russian Federation, Korea, Slovakia, Poland and Bulgaria.

[2] A header array file containing all the results is available from the Center for Global Trade Analysis.

**Table 1: Most frequently occurring unusual input-output shares**

| | Input Sector | Output Sector | Input cost share in Representative Table (cmpshr) | Reasonableness of Representative table share | Input cost share in IO table (ioshr) | Number of errors |
|---|---|---|---|---|---|---|
| 1 | 8 ocr | 11 rmk | 8.7% | Low | Errors are mostly due to low ioshr, approx. close to 0%. 7 countries have ioshrs between 15-25%; Bulgaria and Japan are outliers with high ioshrs of 39% and 29%. | 20 |
| 2 | 10 oap | 19 cmt | 28.1% | Low | Most countries that cause errors have low ioshr ranging between 0-2%. A few countries have high ioshr namely, Colombia (18.8%), Japan (49.1%) and Korea (38.6%). | 35 |
| 3 | 11 rmk | 19 cmt | 9.0% | Low | Almost all ioshr are below 1% and a few are below 3%. Colombia is the only country with a high ioshr of 15.0% which biases the cmpshr upwards and generates errors for low ioshr countries. | 31 |
| 4 | 9 ctl | 20 omt | 15.3% | Low | All error-causing countries have ioshr that are less than 1%. A few countries have high ioshr, namely, Australia (24.1%), Botswana (23.2%), Colombia (23.8%) and Zimbabwe (56.8%). | 33 |
| 5 | 10 oap | 20 omt | 17.8% | High | Unlike the above cases, errors are mostly caused by countries that have very large ioshr. Ioshr range between 20-50% for most countries. About 8 countries have ioshr less than 10% which biases the cmpshr downwards. | 21 |
| 6 | 11 rmk | 20 omt | 9.3% | Low | Most ioshr are close to 0%. Colombia is an outlier with an ioshr of 14.5% | 36 |
| 7 | 8 ocr | 21 vol | 6.7% | Low | Almost all ioshr are very close to 0% and a few are about 1%. Japan and Zimbabwe are the only countries with high ioshr of 20.7% and 6.5%, respectively. | 23 |
| 8 | 1 pdr | 23 pcr | 64.6% | High | About one-third of the countries have high ioshr ranging from | 29 |

| | | | | | | |
|---|---|---|---|---|---|---|
| | | | | | 20-99%. The rest of the ioshr are less than 10% with quite a few less than 1%. | |
| 9 | 6 c_b | 24 sgr | 7.6% | High | Errors are mostly generated by countries with large ioshrs. A few countries have small ioshr (11 countries have ioshrs less than 10%) which biases the cmpshr downwards. | 22 |
| 10 | 47 trd | 26 b_t | 16.2% | Low | Ioshr are less than 10% for most countries. Bangladesh (24.2%) and Colombia (34.4%) are outliers. | 22 |
| 11 | 16 oil | 32 p_c | 41.9% | High/Various | About one-third of the countries have ioshr between 20-60%. For the rest of the countries the ioshr vary widely between 0-20%. | 25 |
| 12 | 35 i_s | 35 i_s | 46.4% | Various | About half the countries have larger ioshr ranging from 20-60%. Errors are caused by countries with ioshr less than 20% | 23 |
| 13 | 38 mvh | 38 mvh | 44.7% | Various | A few countries have large ioshr namely, Belgium (57.8%), Japan (47.1%), Mozambique (67.0%), Slovakia (64.1%) and Slovenia (54.1%). For other countries it varies widely between 0-40%. | 23 |
| 14 | 17 gas | 44 gdt | 9.3% | High/Various | The ioshr have a large variance. Errors are generated for countries with too low (14) or too high (8 countries) ioshrs relative to the cmpshr'. | 22 |
| 15 | 49 wtp | 49 wtp | 22.6% | Various | Most countries have small ioshr (less than 10%). A few countries have high ioshr: Australia (24.0%), Bulgaria (21.3%), Germany (36.1%) and Japan (26.3%) | 30 |
| 16 | 48 otp | 50 atp | 28.5% | Low | Ioshrs range from 2-35% with quite a lot of variability. 3 countries have ioshr greater than 30% (Greece, Portugal and Japan). | 24 |
| 17 | 48 otp | 18 omn | 18.7% | Low | Most countries are between 0 and 10%. Japan has a high share (29%) which raises the cmpshr share. | 29 |

Here we reflect on each of the other "unusual shares". Rows 1-4, 6 and 7 appear to be the result of a small number of countries altering the share in the representative table adversely. In these cases most of the countries are within what we would consider a normal range. For example most countries do not use other animal products to produce cattle meat, however there are some important countries which have significant errors – Japan, Colombia and Korea – which have raised the share in the representative table and hence make many of the other countries look unusual when they are not. While only a few countries show unusual shares for each input-use combination, taken together with row 5 (oap used in omt) these rows reveal an overall problem with the disaggregation of livestock, meat and raw milk which affects a large number of countries: Australia, Bangladesh, Botswana, Bulgaria, Colombia, Estonia, Japan, Korea, Latvia, Malta, Romania, Slovenia, Tanzania, the United Kingdom and Zimbabwe. This is discussed further below.

Rows 8 and 9 reflect problems with the disaggregation of commodities into raw and processed goods, which is of particular concern in the developing economies. The use of paddy rice in processed rice for two thirds of the 29 countries was too low. Many I-O tables do not distinguish between processed and paddy rice, hence naive splits result in paddy rice being sold directly to consumption. Sugar beet used in sugar is also too low in 11 of the 39 I-O tables (Malta, Romania, Slovenia, Mozambique, Malawi, Korea, Estonia, Czech Republic, Colombia, Botswana and Bulgaria).

Rows 11, 12 and 14 relate to energy commodities. The share of oil used in petroleum and coal products (row 11) is expected to be high, although some countries that have no oil refineries but do undertake other petroleum and oil activity (e.g., coke ovens) may have low shares. So low oil shares can't be ruled out a priori. In the case of ferrous metals used in ferrous metals (row 12) intra-industry usage may also be sensitive to industry organization. In the case of gas used in gas distribution again we would expect the share to be high. However in some countries gas distribution may represent not gas transmission but other activities (e.g. steam and hot water supply), so again low gas shares cannot be ruled out.

Row 13 (motor vehicles used in motor vehicles) could be the result of the fact that motor vehicles are made up of both parts and final vehicles. Countries which produce final vehicles are expected to use more motor vehicles parts in their construction than countries that produce parts. Hence we would expect large differences in shares across countries.

Similarly Row 15, the share of water transport used in water transport may also differ across countries depending on the types of water ways (e.g. countries with large coastlines seem to have higher shares).

Row 16, other transport in air transport appears to depend on whether the country is developed or developing. Developing countries have low shares, while developed countries have higher shares, possible reflecting the fact that they have larger airports with more traffic.

Some input industries have entropy errors more frequently than other input industries. From our dataset the following input industries have shown more than 100 entropy errors across all output industries and all countries.

| | Input Sector | Countries with largest number of errors across all output sectors | Number of errors |
|---|---|---|---|
| 1 | 47 trd | Out of 39, 13 countries have more than 10 errors. Bangladesh (23), Czech Republic (16), Greece (15) and Romania (15) have the largest errors, respectively. | 253 |
| 2 | 48 otp | Errors range between 0-6 per country and are evenly distributed among countries. Majority of errors come from the two uses listed above (atp and omn). | 103 |
| 3 | 52 ofi | All countries have less than 6 errors, except, Greece (19) and Portugal (28). | 114 |
| 4 | 54 obs | 7 countries have more than 10 errors. The number of errors per country varies largely, ranging from 0 to at most 15 errors. Uses where most of these errors occur are oil (14), gas (11), ely (11), otp (19), ofi (15), isr (19), obs (11). | 189 |
| 5 | 55 ros | Most countries have no errors and a few countries have errors between 1-4. Belgium (22), Germany (41) and Sweden (20) have large number of errors (which comprises of about 72% of the total number of errors) for the ros (Recreational and other services) input industry | 115 |

In the case of row 1, beverages and tobacco is the use is the use with the most errors (22, row 10 Table 1), but other food processing sectors also have about 10 errors. Some countries appear to consistently have shares higher than the representative table (Bangladesh) while others are consistently below (Czech Republic, Greece and Romania) indicating differences in trade in those countries.

In row 4, other business services certain developed economies (France and Japan) seem to be raising the shares in the representative table.

The unusual shares in row 2 were discussed in Table 1 and those in Rows 3 and 5 are primarily the result of the few countries listed.

The following lists the output industries for which more than 100 entropy errors were observed across all input industries and all countries. Both these problems were discussed above.

| Ouput Sector | Countries with largest errors across all input sectors | Number of errors |
|---|---|---|
| 19 cmt | Errors range between 0-4 per country and are evenly distributed among countries. | 111 |
| 20 omt | Errors range between 0-6 per country and are evenly distributed among countries. | 144 |

A number of conclusions are drawn below:

It is always difficult to decide which countries should be included in the representative table. Even though a contributor may contribute a table with all sectors, the shares in that I-O table may differ considerably from the 'norm' for any number of both valid and invalid reasons; and hence will affect the shares in the representative table. The representative table is used to a) disaggregate non-agricultural sectors in country I-O tables which do not have sufficient non-agricultural disaggregation and b) in the checking

procedures prior to incorporation into the GTAP Data Base. In the future the representative table may also be used to help fill in missing data for countries with very limited data (see Method of constructing IO tables from very limited data for GTAP purposes by Mark Horridge in supplementary materials). While the agricultural shares in the representative table are not used in the construction process itself, if in the future the representative table becomes more widely used we should consider using FAO data to improve the agricultural shares in the representative table. Another idea which has been suggested is the creation of two representative tables: one for developed economies and another for developing.

A number of problems arose with the cost shares in agriculture. These problems are related to the fact that the data from many developing countries do not disaggregate commodities into raw and processed goods – one reason for this that often products do not get processed at the commercial level but at the domestic household level. As a result, if disaggregation is done in a naïve way, without additional information, then the cost shares are likely to be dubious. Previously we preferred that contributors did as much disaggregation as possible. Now however, following a report by Troy Podbury on the GTAP 6 Data Base, we ask that a contributor only disaggregate agriculture if they have additional information available on the cost structures of the industries. Then we use FAO data to do this agricultural disaggregation and hopefully avoid many of these problems. We also scrutinize this aspect of the contributed I-O tables more closely than was done previously. Three examples of this problem can be seen in the results of this report:

1. Cattle (ctl) and other animal products (oap), and cattle meat (cmt), other meat (omt) and raw milk (rmk) are often aggregated in the original country I-O data and in some cases have been disaggregated by the contributor. Australia, Bangladesh, Botswana, Bulgaria, Colombia, Estonia, Japan, Korea, Latvia, Malta, Romania, Slovenia, Tanzania, the United Kingdom and Zimbabwe have one or more strange shares amongst these products.
2. The use of paddy rice in processed rice for two thirds of the 29 countries was too low. Many I-O tables do not distinguish between processed and paddy rice, hence naive splits result in paddy rice being sold directly to consumption.
3. Sugar beet used in sugar is also too low in 11 of the 39 I-O tables (Malta, Romania, Slovenia, Mozambique, Malawi, Korea, Estonia, Czech Republic, Colombia, Botswana and Bulgaria).

**References**

Walmsley, T. L., and R. McDougall, 2004, Using Entropy to Compare IO tables Center for Global Trade, Analysis https://www.gtap.agecon.purdue.edu/access_member/resources/res_display.asp?RecordID=1676

Podbury, T. “Suggestions for improving the basic structure of GTAP Data Base 6”, mimeo.