Using Entropy to compare IO tables

By

Terrie L. Walmsley

And

Robert McDougall

GTAP Research Memorandum No. 9

Using Entropy to Compare IO tables

By
Terrie L. Walmsley and Robert McDougall

1. Introduction

The importance of obtaining quality I-O tables for the GTAP Data Base[1] has prompted the need for a method to compare contributed I-O tables. Comparisons are useful for a number of reasons.

a) Time comparisons can be used to show how an economy has changed over time.

b) Country comparisons may be used to show differences in cost or use structures between countries.

c) Comparisons may also be used to highlight errors in construction procedures or mappings and therefore assist us in improving the quality of the data.

In this document we look out how entropy can be used to compare I-O tables. The paper is intended to be a practical guide to users who wish to view, examine and use the results from the entropy methods used in the GTAP data Base I-O checking procedure. Section 1 outlines the entropy method used to compare IO tables. Section 2 provides a step by step guide to users on how to view and interpret the results of the entropy programs and section 3 looks at some examples of the entropy program results.

2. The Entropy Method

The entropy method calculates the difference between the shares in the I-O table as an equal weighted average of each share multiplied by the natural log of the ratio of that share to the equivalent share in the comparison I-O table. In this case the example is input or cost shares, although other shares can be examined in a similar way.

$$E_{i,u} = 0.5\left[S^{*A}{}_{i,u}\left[LOG_e(S^{*A}{}_{i,u}/S^{*B}{}_{i,u})\right]\right] + 0.5\left[S^{*B}{}_{i,u}\left[LOG_e(S^{*B}{}_{i,u}/S^{*A}{}_{i,u})\right]\right]$$

where: $E_{i,u}$ is the entropy measure of the difference between $S^{*A}{}_{i,u}$ and $S^{*B}{}_{i,u}$.

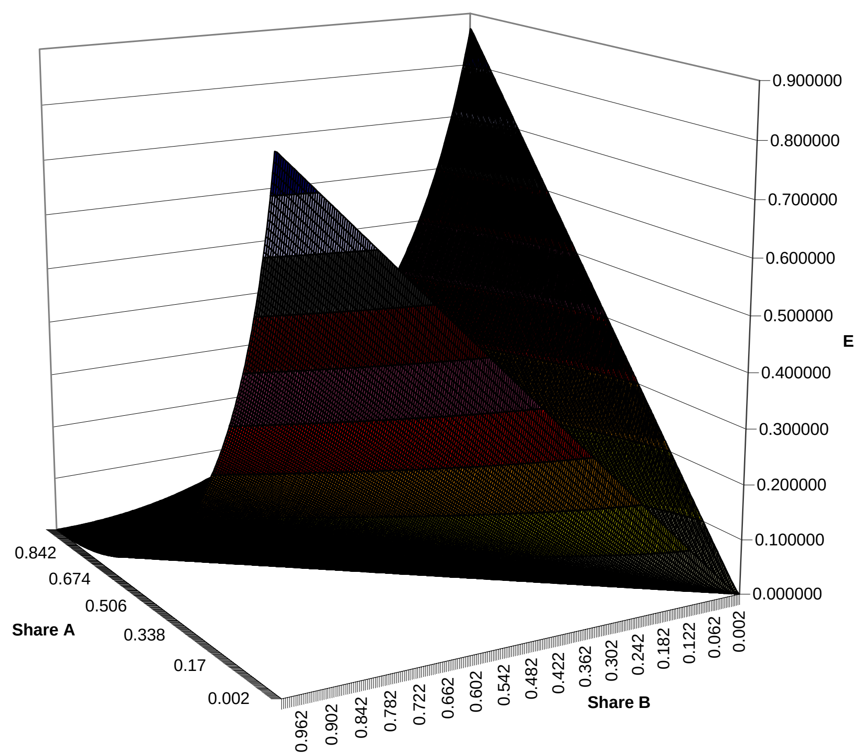
[1] See Huff, McDougall and Walmsley (2001) on how to contribute an I-O table to the GTAP Data Base.

$S^{*A}_{i,u}$ is the adjusted share of input i in used in use u in I-O table A[2].

$S^{*B}_{i,u}$ is the adjusted share of input i used in use u in I-O table B.

The results of this method are shown in Figure 1. Along the diagonal where shares in the two I-O tables are equal $E_{i,u}$ is zero. The further one moves away from the diagonal, i.e. the further apart the shares of the two I-O tables the larger is $E_{i,u}$. Where shares are small $E_{i,u}$ rises linearly, however if shares are large $E_{i,u}$ rises at an increasing rate.

Figure 1: Relationship between $S^{*A}_{i,u}$, $S^{*B}_{i,u}$ and $E_{i,u}$ [3]



An adjustment is made to both the input shares in I-O tables A and B. The adjustment made to the shares is given by:

---

[2] Note that the I-O data is first aggregated. No distinction is made between domestic and imported intermediate inputs. The resulting shares are input i (domestic and imported) used in use u. This allows the program to check differences in cost structures and not differences in each country's comparative advantage.

[3] Assuming minimum and maximum shares are to three decimal places only. Further refinement would lead to much larger tails.

$$S^{A}{}_{i,u}{}^{*} = \left[S^{A}{}_{i,u}\left[1 - TINY\right]\right] + \left[S^{B}{}_{i,u}\left[TINY\right]\right]$$

$$S^{B}{}_{i,u}{}^{*} = \left[S^{B}{}_{i,u}\left[1 - TINY\right]\right] + \left[S^{A}{}_{i,u}\left[TINY\right]\right]$$

where: $S^{A}{}_{i,u}$ is the share of input i in used in use u in I-O table A.

$S^{B}{}_{i,u}$ is the share of input i in used in use u in I-O table B.

TINY is a small number (= 0.1).

The purpose of this adjustment is twofold. Firstly, it ensures that zero shares are taken into account; and secondly, it affects the value of $E_{i,u}$ and the rate at which it increases with large shares. By manipulating TINY one can alter how differences between shares are perceived. In general the difference between two large shares is considered to be less important than the same absolute difference between two smaller shares. Lowering TINY reduces the rate at which $E_{i,u}$ increases for large shares and therefore gives less prominence to differences in large share. Table 1 can be used to illustrate the impact of a change in TINY. Table 1 shows the difference (in terms of $E_{i,u}$) between $S^{*A}{}_{i,u}$ = 0% and $S^{*B}{}_{i,u}$ = 5%. When TINY = 0.1, the equivalent $E_{i,u}$ is obtained from the difference between $S^{*A}{}_{i,u}$ = 100% and $S^{*B}{}_{i,u}$ = 66.2%. Lowering TINY to 0.0001, raises the value of $E_{i,u}$ to 0.22 and changes the equivalent difference to $S^{*A}{}_{i,u}$ = 100% and $S^{*B}{}_{i,u}$ 44.2%. Since differences in large shares should not be completely discounted TINY was set equal to 0.1 in Figure 1 and in the examples below.

Table 1: Impact of altering TINY

TINY	$S^{*A}{}_{i,u}$	$S^{*B}{}_{i,u}$	$E_{i,u}$
0.1	0%	5%	0.043944
0.1	100%	66.2%	0.043228
0.0001	0%	5%	0.229715
0.0001	100%	44.2%	0.227692

3. An Example

The entropy check program is run using the GEMPack[4] software suite developed by Kenneth Pearson at the Centre of Policy Studies, Monash University. The results are then converted to a gms file using SEEHar's HAR to GAMS facility. The list produced in the gms file can then be taken to excel and ordered. The resulting excel worksheet contains the following information:

Table 2: Output from the Entropy program examining differences in input shares

Row	col	ENT	IO shr	CMP shr
Pdr	pcr	0.631387	0.026656	0.836682
Capital	pdr	0.405307	0.050451	0.6541
c_b	sgr	0.363641	0.569699	0.03992
Gas	gdt	0.308297	1E-06	0.350781
Capital	v_f	0.283504	0.063876	0.540802
Ctl	cmt	0.281263	0.00157	0.327773
Ofd	oap	0.254115	0.00841	0.327318
Land	c_b	0.250721	0.285834	0.000113
Capital	osd	0.240764	0.061199	0.477843
Labor	ctl	0.23586	1E-06	0.268362
v_f	ctl	0.228978	0.33102	0.017001
Land	pfb	0.22729	0.25926	0.00013
capital	gro	0.223822	0.064815	0.464691
land	v_f	0.223026	0.253911	3.1E-05
land	ocr	0.222904	0.253911	5.84E-05
land	osd	0.213697	0.24327	2.57E-05
oap	omt	0.20876	1E-06	0.237528
land	gro	0.197597	0.225831	0.000201
capital	wht	0.191431	0.050838	0.38563

Where: row is the input i;

col is the use u;

ENT is the value of $E_{i,u}$ for the input shares.

cmpshr are the $S^{*A}{}_{i,u}$ **,** the adjusted input shares of the representative table created during the GTAP data base construction process. The representative table is a weighted average of a number of regional I-O tables.

ioshr is $S^{*B}{}_{i,u}$ **,** the adjusted input share of the regional I-O table being

[4] http://www.monash.edu.au/policy/gempack.htm

investigated.

As stated above the greater the difference between shares in the two I-O tables the greater the value of $E_{i,u}$. The list is ordered so that higher $E_{i,u}$ appear at the top of the table. To check an I-O table you should therefore work from the top of the list to the bottom, examining each one in turn until the differences are negligible. To understand the table we interpret the first row in Table 2. It should be read as follows: Paddy rice (pdr) used in processed rice (pcr) on average (representative table) accounts for 84% of the costs of processed rice, however in the I-O table contributed, paddy rice (both imported and domestic) represents only 2.6% of costs.

Let us now turn to a specific example of how this type of analysis could be used to improve the quality of the I-O tables.

Table 3: An extract from Table 2

row	col	ENT	IO shr	CMP shr
ctl	cmt	0.281263	0.00157	0.327773
oap	omt	0.20876	1E-06	0.237528

The entropy program is used to compare a contributed I-O table to a representative table. The purpose of this is to identify problems in the contributed I-O. This comparison is useful because in general we would not expect input shares to differ significantly across countries. Inputs into production should not change significantly simply because they are produced in another country, for example fish is unlikely to be an input into the wearing apparel industry in any country, while textiles will be.

Table 3 depicts the entropy results for the input shares into the production of cattle meat and other meat in a GTAP contributed I-O table. The results show that cattle are not the primary input into cattle meat and other animal products are not the primary input into other meat in the contributed I-O table, while the representative table shows that both these shares should be significant. This is also what we would expect given what we know about the production of beef, pork and chicken.

Upon further investigation it may be found that there is an incorrect mapping between the GTAP sectors and the source I-O table, which led to an overlap between cattle and other animals. Alternatively cattle and other meats may have been one category in the original I-O data which were then split without taking into consideration the specific cost structures of these commodities.

In some cases however high entropy values may not be the result of mistakes but peculiarities in the cost structures of the country involved. Understanding these peculiarities can significantly assist us in making decisions which could significantly improve the quality of the data.

4. Conclusion

The entropy method provides a useful method for quantifying differences in I-O tables which we are currently using to assist us in improving the quality of the GTAP data base. This methodology could also be used to improve our understanding of the countries we are examining and allow us to examine how those economies change over time.

5. References

Huff K., R. McDougall and T. L. Walmsley (2001): Contributing Input-Output Tables to the GTAP Data Base, GTAP Technical Paper No. 1, Center for Global Trade Analysis, Purdue University.