Malaysian Journal of Economic Studies 51(2): 217-229, 2014 ISSN 1511-4554

Evaluation of the Relative Performance of RAS and Cross-Entropy Techniques for Updating Input-Output Tables of Malaysia

M. Yusof Saari[a]
Azman Hassan[b]
Muhammad Daaniyall Abd Rahman[c]
Azali Mohamed[d]
Universiti Putra Malaysia

Abstract: The growing interest in applied general equilibrium models for policy analyses increases the demand for up-to-date input-output tables. Constructing survey-based input-output tables for a recent year is costly and time-consuming. These constraints have led to the emergence of non-survey updating techniques. In this paper, the relative performance of two prominent non-survey techniques designed to update input-output tables, the RAS and cross-entropy, is compared. Results show that the cross-entropy technique is superior to the RAS regardless of the size of matrices. For both techniques, our analyses suggest that the accuracy of estimates improves with a high level of sectoral aggregation.

Key words: Cross-entropy, input-output, RAS, updating techniques
JEL classification: C67, D57, O21

1. Introduction

Recent decades have witnessed a rising demand for accurate and up-to-date input-output tables for Malaysia. This is mainly due to the growing demand for more recent and consistent economy-wide models to support policy analysis conducted by various planning agencies at national and regional levels. Input-output analysis provides policy makers and economic analysts with a powerful tool for policy simulation and impact analysis. It also serves the underlying data framework for applied general equilibrium models—social accounting matrix (SAM) and computable general equilibrium (CGE). However, the usefulness of the input-output tables and the associated SAM and CGE models is still hampered by the long delay with which it tends to appear. The input-output tables in Malaysia are usually published every five years, while national product and expenditure data are available annually, but with a lag. If an input-output table is completed with more than a 10-year lag compared to its vintage, then results from the input-output model may be biased due to the fact that the level of production technology, income and consumption patterns may have changed substantially within that period.

Estimating the input-output tables for a recent year has proven to be a challenging task for two main reasons. First, construction of input-output tables demands accurate

[a,b,c,d] Department of Economics, Faculty of Economics and Management, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor. [a]Email: *yusofsaari_upm@upm.edu.my* or *mysaari@gmail.com* (Corresponding author); [b]Email: *azmanhs@upm.edu.my;* [c]Email: *daaniyall@upm.edu.my*; [d]Email: *azali@upm.edu.my*

statistical data from various sources (such as industrial censuses, labour force surveys, government accounts, household surveys and balance of payments) along with well-trained personnel. Second, construction of input-output tables through censuses or surveys is prohibitively costly. This pursuit has led to the emergence of non-survey updating techniques. Non-survey techniques such as RAS (Dietzenbacher and Miller 2009; Miller and Blair 2009) and cross-entropy method (Golan *et al.* 1994; Robinson *et al.* 2001) are commonly applied and they do provide better estimation. In the updating process, these non-survey techniques try to minimise the difference between the estimated and the existing input-output tables. Hence, this poses an important question as to whether these two approaches differ significantly in practice and if not, it may not matter much which approach is used in most cases.

Our purpose in this paper is twofold. The first objective is to compare the relative performance of RAS and cross-entropy methods in updating Malaysian input-output tables for 2005—updated based on 2000 input-output table. The results are then compared with the actual 2005 input-output tables. The second objective is to analyse the degree to which aggregation may affect accuracy of the estimation. National products and expenditure data are often produced on a aggregated sectoral breakdown and thus updating input-output tables at an aggregated level may reduce estimation errors. Our contribution to the Malaysian literature is essentially an empirical issue. To the best of our knowledge, this is the first attempt to apply the cross-entropy method and analyse the superiority of the RAS and cross-entropy for updating input-output tables in Malaysia.

The next section reviews the literature on the development and application of RAS and the cross-entropy approaches for updating input-output tables with specific reference to the Malaysian case. Section 3 discusses the input-output modeling framework along with its main assumption of fixed input coefficients. Section 4 explains the RAS procedure, followed by a discussion of the cross-entropy approach. Section 5 describes data used to run the analysis while Section 6 presents the results of comparison between RAS and the cross-entropy method. Section 7 provides the concluding remarks.

2. Literature Review

RAS and cross-entropy are not the only formal non-survey methods for updating input-output, but they are representative of the methods that have been used in practice. Other methods such as Stone-Byron method of restricted least square [first developed by Stone (1977) and formalised by Byron (1978)], linear programming (Davis *et al.* 1977) and transaction proportional to value added (TPVA)(Jalili 2000) may also be applied. We chose only RAS and cross-entropy and compared their relative performances for two main reasons: (i) they are superior to other non-survey techniques and yield reliable estimates (Miller and Blair 2009) and (ii) they are cost-effective and flexible to incorporate additional information (the Stone-Byron method relies on restricted least squares).

For early work on updating input-output tables, one can refer to Stone (1961) and Stone and Brown (1962) who had developed a 'biproportional' procedure which was later known as RAS (the term RAS refers to the notation used by Stone in a series of papers on this topic). The practicality of the biproportional RAS technique has made it not only one of the most important tools for updating (Thissen and LÖfgren 1998) but also for balancing (Saari *et al.* 2014), regionalising (Jiang *et al.* 2011) and deflating (Dietzenbacher and Hoen 1998) input-output tables. The uniqueness of the RAS technique is that the updating procedures can be applied for both transaction matrix and coefficient matrix.

The cross-entropy is a non-linear technique that has been substantially used by Sherman Robinson and his associates in the International Food Policy Research Institute (IFPRI) group in compiling and balancing input-output tables and social accounting matrices. This method was first introduced to the input-output updating problem by Golan *et al.* (1994) and then was further formalised and extended by Robinson *et al.* (2001). Essentially, the cross-entropy method is formally similar to the generalised RAS method but nevertheless, there are some significant differences and additional complexities. First, the updating procedures are based on the coefficient matrix rather than the transaction matrix. Second, the procedures include the estimation of a set of error weights, which are part of the generation of error variables.

Several studies have examined the sensitivity and superiority of RAS and cross-entropy techniques compared to other techniques for different countries. For example, Davis *et al.* (1977) found RAS to be a superior technique compared to the linear programming technique for updating the input-output tables for the United States of America. For the Soviet Union's input-output tables, Jalili (2000) observed that RAS estimates coefficients that are identical to the actual coefficients, compared with LaGrangian optimisation method, transaction proportional to value added and NAÏVE method. To compare RAS and cross-entropy, Robinson *et al.* (2001) carried out a range of Monte Carlo experiments using Mozambique data and suggested the superiority of the cross-entropy over the RAS technique.

Although there are numerous studies on updating input-output tables, to the best of our knowledge, empirical evidence on application of RAS and cross-entropy for updating Malaysian input-output tables is simply not available. Though the most related discussion on using RAS technique within the context of Malaysia was deliberated by Saari and Rashid (2009), where they applied the technique to develop regional input-output tables for the state of Selangor, the authors did not compare other available techniques for sensitivity and better approximation. Therefore, due to limited references, this paper extends and provides novel contribution to the literature on alternative techniques in updating the input-output tables for Malaysia.

3. Structure of an Input-Output Table

Input-output analysis is concerned with the quantitative analysis of interdependencies among production sectors in an economy through consuming and producing output. It is represented in matrix form so that the flows of goods and services from one productive sector to another can be traced consistently. The formal properties for the simplified input-output accounting system are shown in Table 1.

The $(n \times n)$ matrix Z denotes the intermediate deliveries and each element of z_{ij} indicates the amount of commodity sector i used by sector j. The $(n \times k)$ vector of f represents final demand components (i.e. private consumption, c, investment, i.

Table 1. Simplified input-output table

	Intermediate demand							Final demand				Total
	S1	S2	S3	.	.	.	Sn	c	i	g	e	output
Sector 1 (S1)												
Sector 2 (S2)												
Sector 3 (S3)												
.				Z					f			x
.												
.												
Sector (Sn)												
Import				M								
Indirect tax				T								
Value added				V								
Total input				x'								

government consumption, *g* and export, *e*). Primary input components—the (1 x *n*) vector *m* gives the sectoral imports (1 x *n*), vector *v* shows value added and (1 x *n*) vector *t* is indirect tax for each sector. Summing across the columns, the total gross output, *x*, throughout the economy is found as[1]

$$x = Zi + f$$

The same value (because $x' = x$) can be found by summing across the rows

$$x' = i'Z + m + v + t$$

These are simply two alternative ways of summing all the elements in the table.

Based on input-output flows in Table 1, we can transform them into a model. The independencies among production activities can be shown based on the following material balance equation:

$$x = Ax + f \quad (1)$$

where *x* is the vector for gross output, *A* $(A = Z\hat{x}^{-1})$ is known as the technical coefficient or input-output coefficient and *f* is the vector for final demand. In the standard input-output model, Equation (1) can be transformed and solved in matrix notation as follows:

$$x = (I - A)^{-1} f = Lf \quad (2)$$

where *I* is the identity matrix, and $(I-A)^{-1}$ is known as the Leontief inverse matrix. Each element of the Leontief inverse matrix shows total output effects (both the direct and

[1] For clarity, matrices are indicated by bold, upright capital letters; vectors by bold, upright lower case letters, and scalar by italicised lower case letters. Vectors are columns by definition, so that row vectors are obtained by transposition, indicated by a prime (e.g. X'). A diagonal matrix with the elements of vector x on its main diagonal and all other entries equal to zero are indicated by a circumflex (e.g. $\hat{x}$). A summation vector is represented by *i*.

indirect effects) for any sector j to satisfy each unit of final demand. In this model formulation, quantity levels are assumed to be varied while prices are fixed.[2] To keep the prices fixed, assumptions of an excess capacity and unused resources have to co-exist. Consequently, expanding production is not hindered by supply constraints, such as limits for labour or imported inputs.

In Equation (2), level of output is exogenously determined by the exogenous final demand. Growth in final demand affects output through the fixed Leontief inverse matrix (because of fixed input-output coefficient), which is equivalent to

$$\Delta x=(I-A)^{-1}\Delta f=L\Delta f \quad (3)$$

It is clear that the input-output framework assumes constancy of multiplier coefficients (Leontief inverse matrix) over time. Linearity of the input-output model implies that composition of commodities (or products) used for production inputs are fixed and analyses are run with an absence of substitution possibilities among inputs.

For short-run impact analysis, the linearity assumption does not seem unreasonable. In medium- and long-term analysis, in particular, for projection of economic growth, linearity in the coefficients cannot be taken for granted. Over time, input-output coefficients could change due to any or a combination of three general causes: (i) changes in technology of production, (ii) changes in the relative prices of inputs, and (iii) changes in the product mix of a particular sector. For developing and transition economies that are characterised by rapid structural changes such as in China and Vietnam, input-output coefficients are likely to be unstable. For matured economies or developed countries, however, input-output coefficients tend to be stable.

In Malaysia, the input-output coefficients for some sectors may appear somewhat less stable. For example, we have used input-output coefficients for 2000 to estimate output in 2005 by taking final demand in 2005 as exogenous (all figures are in current prices). Results show that the 2000 input-output coefficients underestimated the 2005 output by 30 per cent. In view of the likely occurrence of significant coefficient changes over time, the lengthy construction periods for input-output tables highlight the need for effective updating procedures. The next section discusses the application of RAS and cross-entropy techniques for updating input-output coefficients.

4. Non-survey Updating Techniques

In updating input-output tables, the non-survey techniques attempt to minimise the difference between estimated and old tables (or matrices). The problem is essentially to find a new matrix which is in some sense 'close' to the existing matrix. In relation to the updating techniques, RAS and cross-entropy methods have proven to be both popular and operational in input-output construction (Round 2003). The relatively close analytical relationships between the most frequently used alternative methods for updating input-output tables suggest that if the required adjustments are relatively small then the differences between the methods are also likely to be small (Schneider and Zenios 1990).

[2] The dual model for the quantity model is a price model (also known as a cost-push model). The model is useful for analysis of price shocks, given prices may vary while quantities are assumed to be fixed. It is unnecessary to detail the price model because it also utilises the same fixed Leontief inverse matrix (Miller and Blair 2009).

4.1 RAS Approach

The RAS technique iteratively adjusts an old input-output matrix with new information on the row ($u, u_i = \Sigma_j z_{ij}$) and column ($w, w_j = \Sigma_i z_{ij}$) sums, but does not have new information on the intermediate deliveries (*Z*). The RAS technique generates an estimate from 3n pieces of information for the targeted year (i.e. in our case 2005). These are (i) total gross output ($x=x'$), (ii) total final demand (*f*), and (iii) total primary inputs—import (*m*), value added (*v*) and indirect tax (*t*). Once x, *f, m, v* and *t* are available, *u* and *w* can be simply obtained by the following accounting constraints: $u=x-f$ and $w=x'-(m+v+t)$. In this study, entries for *x, f, m, v* and *t* are taken directly from 2005 input-output tables.

With the new or targeted row and column sums, RAS generates a new matrix x^* for 2005 from the old matrix A (in our case 2000 input-output tables) by means of 'biproportional' row and column operations. That is,

$$x^* = \hat{r}A\hat{s} \tag{4}$$

where *r* and *s* are diagonal matrices with positive entries on the main diagonal. Elements of *r* and *s* also can be termed as 'scaling factor' that ensures the new row and column sums match the targeted row and column sums (see Saari and Zakariah 2006; Miller and Blair 2009 for detailed explanation of RAS). Results obtained from Equation (4) are compared with actual 2005 input-output table.

Equation (4) only deals with positive entries, but it also can be further extended to deal with a matrix that consists of positive as well as negative entries (for details see Junius and Oosterhaven 2003). Such a problem is usually observable in a social accounting matrix and in very special input-output tables. We consider such analysis as beyond the scope of this study. In addition to the intermediate flows, RAS can also iteratively update a new matrix A^* for 2005. That is,

$$A^* = \hat{r}A\hat{s} \tag{5}$$

Dietzenbacher and Miller (2009) show that updating transaction flows as in Equation (4) yields the same results as in updating the coefficient in Equation (5). This property is very attractive in practice since there are usually no reasons to favour one over the other. Within the set of commonly used updating procedures, this property is thus a distinctive feature of RAS.

4.2 Cross-entropy Approach

There are two main differences between cross-entropy and RAS. First, the cross-entropy minimisation procedures are based on the derivation of input-output coefficients *A* rather than transaction flows *x*. Second, cross-entropy requires two input-output coefficients (one for existing or old matrix and one for estimated matrix) to be minimised using the entropy approach.

The starting point for the cross-entropy approach is the information theory developed by Shannon (1948). Later, Theil (1967) introduced this approach into economics. Consider a set of *n* events of $E_1, E_2, \ldots, E_n$ with probabilities of $q_1, q_2, \ldots, q_n$ (prior probabilities). A message comes in implying that odds have changed, transforming the prior probabilities into posterior probabilities $p_1, p_2, \ldots, p_n$. Following Shannon, the 'information' received with the message is equal to $-\ln p_i$. However, each E_i has its own prior probability q_i, and the 'additional' information from p_i is given by

$$-\ln p_i/q_i = -[\ln p_i - \ln q_i] \quad (6)$$

Taking the expectation of the separate information values, we find that the expected information value of a message (or data) is

$$-I(\mathrm{p:q}) = -\sum_{i=1}^{n} p_i \ln p_i/q_i \quad (7)$$

where $I(p:q)$ is the Kullback and Leibler (1951) measure of the cross-entropy distance between two probability distributions.

The application of cross-entropy for updating input-output was introduced by Golan *et al.* (1994). Robinson *et al.* (2001) further formalised and extended the approach to deal with several circumstances in updating procedures. Cross-entropy formulation aims to find a new set of coefficient matrix $A^*=[a^*_{ij}]$ that minimises the entropy distance between estimated coefficient matrix $A=[a_{ij}]$, and the existing or prior coefficient matrix $A=[a_{ij}]$. The problem is:

$$a^*_{ij} = \min\left[\sum_i \sum_j a_{ij} \ln a_{ij}/\bar{a}_{ij}\right] \quad (8)$$

subject to

$$\sum_j a_{ij} x^*_{ij} = x^*_{ij}\,;\ \sum_j a_{ij} = 1\,;\ \ 0 \le a_{ij} \le 1 \quad (9)$$

The solution for (8) is obtained by setting up the Lagrangian and solving it. That is,

$$a_{ij} = \bar{a}_{ij}\exp(\lambda_i x^*_j)/\sum_i \sum_j \bar{a}_{ij}\exp(\lambda_i x^*_j) \quad (10)$$

where λ_i is the Lagrange multipliers associated with the information on row and column sums, and the denominator is a normalisation factor. The RAS algorithm can also be used to provide a solution for Equation (10) by adding a formal biproportional constraint to row and column sums.

We applied the cross-entropy approach in two different circumstances. First, to compare with the performance of RAS, we generated an estimated coefficient matrix A* for 2005 by minimising the entropy distance between the actual 2005 input-output coefficients (i.e. $\bar{\mathrm{A}}$) and the existing 2000 input-output coefficients (i.e. A), given the row (u) and column (w) sums as the constraints (i.e. similar constraints imposed in RAS). Results obtained from this estimation were then compared with the actual 2005 input-output tables.

The first situation implies that the initial estimated input-output coefficients for the targeted year are available. In most cases, initial estimated input-output coefficients are usually not available. Alternatively, we could generate 'synthetic' input-output coefficients (see Golan *et al.* 1994). This is our second case of updating and it presents an attractive feature of the cross-entropy approach. We may define this special case of application as generalised cross-entropy.

Let us define the synthetic input-output coefficient matrix as $\tilde{\mathrm{A}} = [\tilde{a}_{ij}]$. It is produced as follows: each element of the actual flows in the 2000 input-output table is multiplied by a random number generated from a normal distribution with specified mean and standard deviation (1,0.05). Each element of synthetic input-output coefficients ($\tilde{a}_{ij}$) then replaces the $\tilde{a}_{ij}$ in (8) – (10) and computations proceed accordingly.

5. Data

Two sets of input-output tables for 2000 and 2005 have been used to run the analysis (see Department of Statistics Malaysia 2005; 2010). Both tables classify production sectors according to the Malaysia Standard Industrial Classification (MSIC) (Department of Statistics Malaysia 2000). However, the sectoral breakdown for the two tables differs, in which the former has 94 sectors while the latter consists of 120 sectors. To make all matrices comparable, we harmonised both tables through aggregation of some sectors. As a result, the harmonised versions of the 2000 and 2005 input-output tables consisted of 75 sectors. To analyse the extent to which a different level of aggregation affects the estimation, we have further reduced the size of input-output tables for both years into 10 aggregate sectors. The full lists of sectors for 76-sector and 10-sector versions of input-output tables are available in the Appendix.

6. Results and Discussion

The purpose of this section is to compare the relative performances of RAS and cross-entropy techniques in updating input-output tables for 2005. For comparison purposes, we have set a problem in which both RAS and cross-entropy can be applied. That is, finding new input-output coefficients for the targeted year which in some sense is 'close' to the existing coefficients. For the RAS technique, the coefficients for the targeted year were estimated by adjusting the 2000 input-output coefficients by biproportional means given the row and column sums of the actual intermediate deliveries in 2005 as the minimisation constraints. For the cross-entropy technique, coefficients for targeted year were obtained by minimising the entropy distance between 2000 and 2005 (actual) input-output coefficients given the same constraints of row and column sums. We also extended the application of cross-entropy technique in the case where the 2005 input-output coefficients were assumed to be unavailable. Alternatively, we estimated the synthetic 2005 input-coefficients and minimised accordingly along with 2000 input-output coefficients (using the same row and column sums). Results from these three updating techniques were compared to the actual 2005 input-output coefficients.

In the input-output literature on non-survey techniques, several measures have been developed to evaluate the closeness of updating techniques. The most commonly used measures are the standardised total percentage error (STPE), mean absolute deviation (MAD), mean absolute percentage error (MAPE) and dissimilarity index (DI).[3] Among these measures, we have applied MAD and DI because STPE and MAPE are over-sensitive to small coefficient estimates and unable to deal with zero-value coefficients. To complement the MAD and DI measures, additional measures of closeness were utilised. These included Pearson correlation, Spearman rank correlation and logarithmic prediction error.[4] All the five measures are summarised in Table 2.

---

[3] The closeness measures used in this paper have also been used by Thissen and LÖfgren (1998) and Miller and Blair (2009).

[4] These three measures have also been used by Saari (2007).

Table 2. Measures of 'closeness' of updating techniques.

Measures	Formula	Notes
Mean absolute difference (MAD)	$\left(1/n^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\left\lvert a_{ij}^{*}-\bar{a}_{ij}\right\rvert$	Represents the average amount (whether positive or negative) by which an estimated coefficient differs from the actual coefficient. DI was developed to deal with zero-value coefficient.
Dissimilarity Index (DI)	$\left(1/n^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\left\lvert a_{ij}^{*}-\bar{a}_{ij}\right\rvert}{\left(a_{ij}^{*}+\bar{a}_{ij}\right)}$	
Pearson correlation (r)	$\frac{1}{n-1}\sum_{i=1}^{n}\frac{\left(a_{ij}^{*}-\bar{a}_{ij}^{*}\right)}{s^{*}}\frac{\left(\bar{a}_{ij}-\bar{\bar{a}}_{ij}\right)}{\bar{s}}$	Pearson correlation measures the strength of a linear association between two different input coefficients. Spearman correlation accesses how well the relationship between two input coefficients can be described as a monotonic function.
Spearman rank correlation	$1-\frac{6\sum_{i=1}^{n}d_i^2}{n^3-n}$	
Logarithmic prediction error (e)	$\left(1/n^2\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\ln\left\lvert\frac{\bar{a}_{ij}}{a_{ij}^{*}}\right\rvert$	Extended based on Tilanus (1966) to capture average deviation between an estimated coefficient and actual coefficient.

Notes: n = number of sectors, $\bar{a}_{ij}$ = actual 2005 coefficients, a_{ij}^{*} = estimated 2005 coefficients, d_i^2 = the difference between ranks of the sectors, S = sum of scores, $\bar{a}_{ij}^{s}$ = sample mean for estimated 2005 coefficients, $\bar{\bar{a}}_{ij}$ = sample mean for actual 2005 coefficients and S^{*}, $\bar{s}$ = standard deviation.

Table 3. Results of estimating performances

	n = 76			n = 10		
	RAS	CE	GCE	RAS	CE	GCE
Dissimilarity index (DI)	0.5829	0.4717	0.5829	0.4069	0.3369	0.3972
Mean absolute difference (MAD)	0.0049	0.0037	0.0049	0.0196	0.0109	0.0196
Pearson correlation	0.7740	0.8750	0.7740	0.8430	0.9540	0.8420
Spearman correlation	0.7370	0.7970	0.7370	0.8410	0.9050	0.8410
Logarithm prediction error (e)	0.5092	0.6039	0.5188	0.3508	0.3188	0.3490

Note: CE and GCE stand for cross-entropy and generalised cross-entropy, respectively.

It is important to note that the degree of closeness of the estimates was evaluated at average for all sectors. One may be interested in evaluating estimates for each single sector across updating techniques, so that stability of the input-output coefficient can be validated for each sector. However, our primary concern in this paper is on evaluating relative performance of updating techniques and we consider detailed sectoral analysis as beyond the scope of this paper.

The results of relative performance for the different updating techniques are presented in Table 3. For DI, MAD and logarithmic prediction error, the closer the value of the statistics to zero, the better the results obtained. This is contrary to the Pearson and Spearman correlation coefficients where better results are achieved when correlation coefficients are closer to one. The results of the estimates in Table 3 are straightforward.

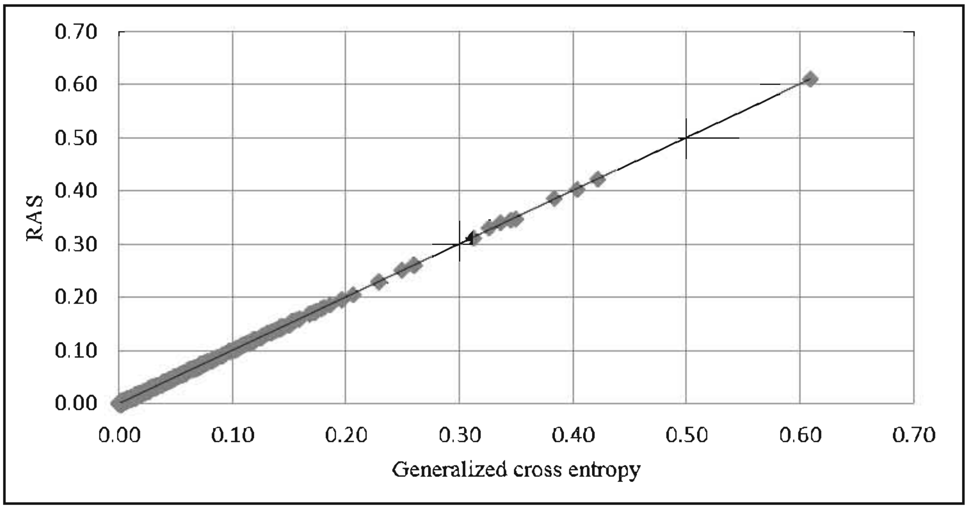

Figure 1. RAS versus generalised cross-entropy

Results at disaggregated sectors (n=76) clearly show that the cross-entropy outperforms the RAS significantly for all measures of closeness. Surprisingly, the generalised cross-entropy and RAS produced similar results. Figure 1 distributes the estimated input-output coefficients obtained from RAS (on Y-axis) against that of generalised cross-entropy (on X-axis). The solid 450 represents situations where two techniques give the same answer. It provides a clear indication that RAS and generalised cross-entropy techniques perform similarly in updating input-output coefficients.

A comparison between RAS and generalised cross-entropy provides an important implication for updating input-output coefficients. Practitioners are more familiar with the RAS technique but this technique requires 3n pieces of information—total output, final demand and primary input to be known in advance. The generalised cross-entropy approach only requires row and column sums of intermediate deliveries given that the synthetic input-output table can be generated from the randomly normal distribution. This provides an advantage to the generalised cross-entropy approach in the sense that it requires less information (thus time-efficient) but provides results that are similar to the RAS.

Following this, we analysed the extent to which the results for a highly aggregated input-output differ from that of disaggregated input-output. Results are given in the last three columns of Table 3. We have aggregated the input-output table into ten main sectors because the annual national account statistics are commonly available at these sectoral breakdowns (see Department of Statistics Malaysia, various years). All the closeness measures show an improvement of the estimation compared to the results of 76 sectors. This implies that size matters in updating procedures, where working with a highly aggregated input-output table reduces estimation error.

Input-output coefficients for some sectors may be more stable than others. For example, Saari (2007) showed that coefficients for the services and agriculture sectors were stable (i.e. fixed) between the periods 1983 and 1987. When the stabilised (fixed input) sectors that are large in size are aggregated with small-size and unstable coefficient sectors, the large sectors would offset the small sectors and thus play a dominant role in the estimation process. This aggregation effect can be clearly observed for the manufacturing sector. For the cross-entropy estimates, we found the DI measures for the manufacturing sector to have reduced from 0.5107 for 76 sectors to 0.1717 for 10 sectors. Of the total 47 manufacturing sectors, 24 sectors registered DI values below the average (i.e. 0.5107) and these sectors accounted for more than 57 per cent of the manufacturing output. Results suggest that updating input-output coefficients should start with a high level of sectoral aggregation.

The relative performances of the two updating techniques remain unchanged regardless of the size of matrices. That is, the results indicate a clear superiority of the cross-entropy technique over RAS. But RAS and generalised cross-entropy provide similar results.

7. Conclusions

The main constraint to use of the input-output and applied general equilibrium models (such as social accounting matrix and computable general equilibrium) is obtaining recent input-output tables. Survey-based input-output tables are costly, time-consuming and require well-trained personnel. This pursuit has led to the emergence of non-survey updating techniques. This paper analyses the relative performances of the two non-survey techniques, RAS and cross-entropy, which are mostly applied in practice. Three important implications are drawn from this study. First, the cross-entropy technique should be preferred for updating procedures. Results show that the cross-entropy technique outperforms the RAS technique regardless of the size of matrices. Second, the cross-entropy technique is more flexible in its demand for information and thus the recent input-output tables can be made available with a little delay. We have developed a generalised version of cross-entropy to deal with less information. Results show that this approach produces exactly the same results as obtained from the RAS technique. Third, updating input-output tables should start with a highly aggregated matrix rather than a disaggregated matrix because errors of estimation are reduced with a high level of aggregation.

References

Byron, R.P. 1978. The estimation of large social account matrices. *Journal of Royal Statistical Society, Series A* **141**: 359-367.

Davis, H.C., E.M. Lofting and J.A. Sathaye. 1977. A comparison of alternative methods of updating input-output coefficients. *Technological Forecasting and Social Change* **10**: 79-87.

Department of Statistics Malaysia. 2000. Malaysia Standard Industrial Classification. Putrajaya: Department of Statistics Malaysia.

_________ . 2005. Input-output tables Malaysia 2000. Putrajaya: Department of Statistics Malaysia.

_________ . 2010. Input-output tables Malaysia 2005. Putrajaya: Department of Statistics Malaysia.

__________ . Annual National Products and Expenditure Accounts. Putrajaya: Department of Statistics Malaysia. Various years.

Dietzenbacher, E. and A.R. Hoen. 1998. Deflation of input-output tables from the user's point of view: a heuristic approach. *Review of Income and Wealth* **44**: 111-122.

Dietzenbacher, E. and R.E. Miller. 2009. RAS-ing the transactions or the coefficients: it makes no difference. *Journal of Regional Science* **49:** 555-566.

Golan, A., G. Judge and S. Robinson. 1994. Recovering information from incomplete or partial multisectoral economic data. *The Review of Economics and Statistics* **76**: 541-549.

Jalili, A.R. 2000. Evaluating relative performances of four non-survey techniques of updating input-output coefficients. *Economics of Planning* **33**: 221-237.

Jiang, X., E. Dietzenbacher and B. Los. 2011. Improved estimation of regional input-output tables using cross-regional methods. *Regional Studies* **22**: 209-220.

Julius, T. and J. Oosterhaven. 2003. The solution of updating or regionalizing a matrix with both positive and negative entries. *Economic System Research* **15**: 87-96.

Kullback, S. and R.A. Leibler. 1951. On information and sufficiency. *Annals of Mathematical Statistics* **4**: 99-111.

Miller, R.E. and P.D. Blair. 2009. *Input-output Analysis: Foundations and Extensions.* Cambridge: Cambridge University Press.

Round, J.I. 2003. Constructing SAMs for developing policy analysis: lessons learned and challenges ahead. *Economic System Research* **15**: 161-182.

Robinson, S., A. Cattaneo and M. El-Said. 2001. Updating and estimating a social accounting matrix using cross-entropy methods. *Economic System Research* **13**: 47-64.

Saari, M.Y. and A.R. Zakariah. 2006. *Analisis dan Aplikasi Input-Output.* Kuala Lumpur: Dewan Bahasa dan Pustaka.

Saari, M.Y. 2007. Testing validity of the Leontief hypothesis by using alternative forecasting techniques. *Journal of International Economic Review* **1**: 139-152.

Saari, M.Y., E. Dietzenbacher and B. Los. 2014. Income distribution across ethnic groups in Malaysia: results from a new social accounting matrix. *Asian Economic Journal* forthcoming.

Saari, M.Y. and Z.A. Rashid. 2009. Pembangunan jadual input-output wilayah dan analisis ke atas struktur industri Selangor. *International Journal of Management Studies* **16**: 1-30.

Schneider, M.H. and S.A. Zenios. 1990. A comparative study of algorithms for matrix balancing. *Operations Research* **38**: 439-455.

Shannon, C.E. 1948. A mathematical theory of communication. *Bell System Technical Journal* **27**: 397-423.

Stone, R. 1961. Input-output and National Accounts. Paris: Organization for European Economic Cooperation.

Stone, R. 1977. *Social Accounting for Development Planning.* Cambridge: Cambridge University Press.

Stone, R. and A. Brown 1962. *A Computable Model of Economic Growth, A Programme for Growth.* London: Chapman and Hall.

Theil, H. 1967. *Economics and Information Theory.* Amsterdam: North-Holland.

Thissen, M. and H. LÖfgren. 1998. A new approach to SAM updating with an application to Egypt. *Environment and Planning A* **30**: 1991-2003.

Tilanus, C.B. 1966. *Input-output Experiments: the Netherlands, 1948-1961.* Rotterdam: Rotterdam University Press.

Appendix. Classification of sectors

76 sectors		10 sectors
Agriculture	Rubber Processing	Agriculture
Rubber primary products	Rubber products	Mining and quarrying
Oil Palm primary products	Plastic products	Manufacturing
Livestock etc.	Sheet Glass and Glass Products	Utilities
Forestry and logging	Clay products	Buildings & construction
Fishing	Cement, lime and plaster	Wholesale & retail trade, restaurant & hotel
Crude petrol, natural gas &coal	Other non-metal products	Transport and communication
Metal ore mining	Iron & steel	Finance, real estate and business
Stone, clay & sand quarrying	Basic precious and non-ferrous metals	Other services
Meat & meat production	Other metal products	Government services
Dairy products	Structural metal products	
Preserved fruits & vegetables	Industrial and Household Machinery	
Preserved seafood	Radio, TV & com. equipment	
Oils and fats	Elect. appliances and housewares	
Grain mills	Other electrical machinery	
Bakery products	Other manufacturing products	
Confectionery	Motor vehicles	
Other foods processing	Other transport equipment	
Animal feeds	Instruments and clocks	
Wine and spirits	Electricity and gas	
Soft drinks	Waterworks	
Tobacco Products	Buildings & constructions	
Yarn & cloth	Wholesale & retail trade	
Finishing of textiles	Hotels & restaurants	
Other textiles	Transport	
Wearing apparel	Communication	
Leather products	Banking services	
Footwear	Other Financial Institution	
Sawmill products	Insurance	
Other wood products	Real estate	
Paper products and furniture	Business services	
Printed products	Education	
Industrial chemicals	Health	
Other chemical products	Other private services	
Drugs and medicines	Recycling	
Soap and cleaning preparations	Public administration	
Petroleum Refinery	Defence and public order	
	Other public administration	