GRAS and RAS versus Minimising Absolute and Squared Differences in Coefficients

JAN OOSTERHAVEN & BERTUS G. TALSMA

ABSTRACT Jackson & Murray (2004) claim that their sign-preserving minimisation of squared differences in coefficients produces a smaller information loss in updating IO transaction matrices than the iterative GRAS algorithm of Junius & Oosterhaven (2003). Here we sort out differences in measures from calculation errors, and show that the information loss needs to be measured with absolute terms when increasing and decreasing cell values occur together. The new and improved numerical results show that GRAS outperforms both sign-preserving alternatives in all but one comparison of lesser importance. They furthermore show that minimising absolute differences consistently outperforms minimising squared differences. Finally and most surprisingly, they show that the classic solution of using RAS with the negative cells excluded outperforms GRAS when IO coefficients and multipliers are compared, and outperforms both sign-preserving alternatives in all but one comparison.

KEYWORDS: RAS; biproportional; updating; input-output; information loss

Jan Oosterhaven, Faculty of Economics, University of Groningen, Postbus 800, 9700 AV Groningen, the Netherlands. E-mail: j.oosterhaven@eco.rug.nl. Bertus G. Talsma, Faculty of Economics, University of Groningen, Postbus 800, 9700 AV Groningen, the Netherlands. E-mail: b.g.talsma@student.rug.nl.

1. Introduction

In the *ESR* special issue on biproportional techniques in input-output analysis (Lahr & de Mesnard, 2004), Jackson & Murray (2004) (abbreviated to JM) compare four linear and three non-linear programming alternatives with the well-known RAS iterative biproportional scaling algorithm (Stone, 1961). All eight alternatives solve the problem of finding the unknown cells a new matrix with given row and column sums, such that it is as close as possible to an old matrix with the same dimensions. The only difference between the eight methods is the goal function (i.e. the measure for the distance between the two matrices) that is minimized. The only restriction on the problem definition is that all cells need to be semi-positive.

JM furthermore develop and test two programming alternatives that are able to also deal with negative cells and totals, and compare their results with those of Junius & Oosterhaven (2003) (abbreviated to JO) who developed a generalization of RAS (GRAS) with the same property. The only restriction on the problem formulation of these three alternatives is that the sign of the cells of the old matrix are preserved in the new matrix. This restriction is a mild one and is often desired, as most IO transactions have an economic content that precludes a change in sign. Prices, quantities and taxes all need to stay semi-positive, and subsidies need to stay semi-negative. Only net exports and changes in stocks may change signs theoretically, although changes in comparative advantage remain unlikely.

In this article we only discuss and extend on this last contribution. First, JM find that their sign preserving squared differences (SPSD) programme, applied to the JO prototype IO table, produces a smaller information loss than GRAS. Second, JM find that their sign preserving absolute differences (SPAD) programme produces an information loss that is only little larger than GRAS, which is considered “encouraging” (p.147) as it involves a simple linear programme as opposed to the other two, more convoluted non-linear methods. The first conclusion, however, must be wrong since JO mathematically prove that the solution of the iterative GRAS algorithm minimises the information loss of the comparably formulated updating programme. Hence, either different information loss measures are compared or calculation errors are made. Consequently, the second conclusion needs to be modified too, be it only numerically.

First, we will root out differences in approaches from true errors. Doing so, we will discover that the standard way to measure information loss only produces sensible results when solely increasing IO values are compared. When IO coefficients are compared, absolute values need to be taken as about half of the coefficients will decrease in size. Second, although the information loss measure has a strong theoretical foundation (Shannon, 1948;

Kullback & Liebler, 1951), the stability of IO multipliers is analysed too, as we consider it equally important. Besides, adding a comparison of IO coefficients and IO multipliers is also a question of fairness, as those comparisons do not per definition favour GRAS over SPAD and SPSD. Finally, the differences in IO values, coefficients and multipliers will also be measured for the traditional way to solve the problem of updating matrices with negative cells, namely to first exclude the negatives, to then RAS the semi-positives, and to then return the negatives. We will abbreviate this traditional solution as RAS^{ex-} (see Junius & Oosterhaven, 2003, for mathematical details). The last section will summarize the partly surprising results.

2. Information loss in IO values and coefficients

Our point of departure is Table 6 from JM with its IO coefficient matrices and its disputed information loss (IL) values, and Tables 1 and 2 from JO with its prototype old IO table and the new row and column totals. When the IO coefficients were used to reconstruct and check the updated GRAS, SPAD and SPSD cell values and IL values, the following conclusions had to be drawn:

1. The GRAS coefficients were calculated by dividing the cell values of the updated table by total use instead of by total input (see the Appendix for the updated IO tables).
2. The SPAD and SPSD coefficients were correctly calculated by the dividing the updated cell values by the column totals of the whole IO table, i.e. by total input.
3. The OLD coefficients were calculated by dividing the old cells values by the new total use instead of by the old total input (see the Appendix for the old IO table).
4. The information loss in JM was calculated by means of IL = $\Sigma_{ij}\ x_{ij}\ \ln(x_{ij}/a_{ij})$, with x_{ij} representing the updated coefficients and a_{ij} representing the old coefficients, whereas the information loss in JO was calculated by means of IL = $\Sigma_{ij}\ |x_{ij}|\ \ln(x_{ij}/a_{ij})$, with x_{ij} representing the updated cell values and a_{ij} representing the old cell values.

Point 1 and 3 will need to be corrected, as they represent two simple mistakes. Point 4 requires some more discussion as it represents a difference in approach.

First, in view of the JM preference to formulate their updating problems in terms of IO coefficients instead of transactions, comparing coefficients is their natural choice that will tend to favour SPAD and SPSD as these methods minimise the absolute and squared differences in coefficients instead of in transactions. RAS^{ex-} and GRAS, however, minimise the information loss in IO transactions, while GRAS per definition produces the best results when transactions are compared.[1] To reach a balanced view, Table 1 shows IL results defined on transactions as well as on coefficients.

Second, JM use the IL formula in its original form, which was appropriate for comparing positive values. However, this formula is no longer appropriate when positive and negative values are compared simultaneously, even when the signs are preserved.[2] First and most importantly, the first term x_{ij} represents the weighing of the flattened relative difference of the second term $\ln(x_{ij}/a_{ij})$, and weights always need to be positive. Second, the IL measure with absolute weights also is the distance measure that is minimized in deriving the GRAS solution. Third, SPAD and SPSD minimise the sum of absolute and squared differences, which also does not allow for compensating positive and negative errors. Thus, we will use the information loss formula with absolute weights.

The numerical results of the above decisions are presented in the first part of Table 1. The information loss in transaction values when using GRAS is almost three times as small as that of using SPAD or SPSD, which seems to indicate a clear case of superior performance. Surprisingly, GRAS is only a little better than the traditional solution of using RAS exclusive of the negative cells in the old table. Looking at the information loss in coefficient values, Table 1 seems to show the expected superiority of SPAD and SPSD. However, it also shows unexpected negative signs for both GRAS and RAS^{ex-}, despite taking absolute values for the weights in the IL formula.

Table 1. Performance of the updating alternatives in terms of Information Loss

Method	RAS^{ex-}	GRAS	SPAD	SPSD
*Comparing all 3x4 cell values with positive weights in the IL formula**				
Transactions	7.90	6.79	19.69	19.86
Coefficients	-1.66	-1.43	0.24	0.23
Comparing all 3x4 cell values with positive full terms in the IL formula				
Transactions	10.45	11.22	19.69	21.33
Coefficients	1.87	1.70	0.24	0.42
Comparing only the 2x3 sub-matrix with intermediate and consumption deliveries				
Transactions	9.17	9.00	11.55	11.51
Coefficients	0.13	0.20	0.21	0.26

* This is the only comparison in which GRAS per definition performs best when transactions are compared, but the IL formula used here is not satisfactory as is shown by the negative values on the second row.

An inspection of the IL details reveals the reason for the negative outcomes. When the argument of the logarithm (x_{ij}/a_{ij}) becomes smaller than one its logarithm becomes negative, and positive errors are numerically compensated by negative errors, wrongly suggesting a smaller information loss. This unwarranted property went unnoticed up till now. The obvious reason is that the updated IO table almost always represented a larger economy than the old IO table. Consequently, the few negative terms were usually small and remained unnoticed. Following JM, however, we now also compare IO coefficient matrices and these do not grow.

In fact, one may expect a more or less comparable number of growing and declining coefficients. The solution to the problem is simple. In all cases, one should always measure the information loss with positive terms: IL = $\Sigma_{ij}\ |x_{ij} \ln(x_{ij}/a_{ij})|$.

The second part of Table 1 shows the result from using the correct IL formula for comparing matrices with both growing and shrinking cells. When IL is measured in transaction space, RAS^{ex-} and GRAS produce an information loss that is almost two times as small as that of SPAD and SPSD. Surprisingly, the old approach does a little better than GRAS, and minimising absolute differences does a little better than SPSD. When IL is measured in coefficient space, SPAD and SPSD do considerably better than RAS^{ex-} and GRAS, which also was expected. Now, as expected, GRAS does do better than RAS^{ex-}, and again SPAD does better than SPSD. In this case, the latter difference is quite large (0.24 versus 0.42) and opposite of that originally found by JM. Obviously, minimising squared differences puts too large a weight on minimising the larger differences that tend to occur in the smaller coefficients.

Instead of comparing all the updated coefficients, one might argue that it is more important to only compare the coefficients that may actually be used in model building. Thus, the final part of Table 1 shows the result of excluding the ‘net taxes’ row and ‘net export’ column from the IL formula. As expected, GRAS outperforms the other methods when the information loss in transactions is considered, but the differences are small. When the IL in coefficients is considered, SPSD does unexpectedly worse than the other methods, and RAS^{ex-} does surprisingly better than the other ones.

3. Differences in IO multipliers

The interesting last question is which of the four methods performs best when IO multipliers are compared. Table 2 shows the changes in the column sums of the Leontief-inverse when it is updated from the old level. The pattern of the changes in the important coefficients, analysed in the last row of Table 1, is strengthened in Table 2. Each sector’s output multiplier is considerably more stable when RAS is applied with the old trick of excluding the negative cells. GRAS is a little better than SPAD, while SPSD produces by far the most unstable output multipliers.

Table 2. Performance of the updating alternative in terms of IO output multipliers

IO table	OLD level	RAS^{ex-} change	GRAS change	SPAD change	SPSD change
Goods sector	2.510	-0.162	-0.458	+0.545	+1.083
Services sector	1.776	-0.052	-0.135	+0.149	+0.372

An inspection of the values of the old and the updated IO tables (see the Appendix) explains part of the difference in the results. The row and column that contain the negative cells in the old table both have totals that are decreasing (the row sum of ‘net taxes’ even changes sign), whereas all other rows and columns totals are increasing. This puts an extra strain on the treatment of the two negative cells. In this specific case, GRAS decreases them, RAS^{ex-} leaves them unchanged, and SPAD and SPSD increase them. Consequently, the positive cells in the GRAS table need to increase least, whereas the positive cells in the SPAD and SPSD tables need to increase most. This leads to a downward deviation of the GRAS multipliers, and an upward deviation of the SPAD and SPSD multipliers, with the RAS^{ex-} multipliers nicely in the middle with a relatively small downward deviation.

In this extreme numerical case, the old trick works best, but that may well depend upon the specific pattern of the changes in the row and column totals. Further empirical testing is needed before more final conclusions may be drawn as regards the stability of the output multipliers with different updating alternatives.

4. Conclusion

However, some other conclusions are possible. First it was found that, especially when one compares matrices with positive and negative cells, but more generally when one compares matrices with some cells growing and some shrinking, absolute values need to be taken in calculating the aggregate information loss (IL). Consequently, the conclusions of both JO and JM need to be modified.

With the correct definition of IL, the generalisation of RAS by JO no longer produces the smallest loss of information per definition. Nevertheless, when GRAS is compared with minimising absolute and squared differences it outperforms both, except when the information loss over all (important and unimportant) coefficients is compared. Furthermore, with the correct definition of IL and the numerical mistakes removed, it appears that minimising absolute differences (SPAD) consistently outperforms minimising squared differences (SPSD).

Most surprising, however, is the strong performance of the old trick of excluding the negatives and simply applying RAS to the rest of the IO table. In the more important comparisons the old trick outperforms GRAS, and in all but one unimportant comparison it outperforms minimising both absolute and squared differences in coefficients.

Notes

[1] JM, following Okuyama *et al*. (2002), suggest that RAS defined on coefficients produces a different outcome than RAS defined on transactions. This suggestion is incorrect. A careful inspection of JM's own description of RAS in terms of coefficients (p.137) simply reveals its mathematical equivalence with RAS in terms of transactions.

[2] In all cases the cells with zeros in the old table of course need to be excluded from the IL measure, as the argument of the logarithm is not defined for those cells.

References

Jackson, R. W. & Murray, A. T. (2004) Alternative Input-Output Matrix Updating Formulations, *Economic System Research*, 16, pp. 135-148

Junius, T. & Oosterhaven, J. (2003) The Solution of Updating or Regionalizing a Matrix with both Positive and Negative Entries, *Economic System Research*, 15, pp. 87-96

Kullback S. & Liebler, R. (1951) On information and sufficiency, *Annals of Mathematical Statistics*, 22, pp. 79-86.

Lahr, M.L. & de Mesnard, L. (2004) Biproportional Techniques in Input-Output Analysis: Table Updating and Structural Analysis, *Economic Systems Research*, 16, pp. 115-134.

Okuyama,Y., Hewings, G. J.D., Sonis, M. & Israilevich, P. (2002) An econometric analysis of biproprortional properties, *Journal of Regional Science*, 42, pp. 361-388.

Shannon, C. E. (1948) A mathematical theory of communication, *Bell System Technical Bulletin Journal*, 27, pp. 623-659.

Stone, R. A. (1961) *Input-Output Accounts and National Accounts* (Paris, Organization for European Economic Cooperation).

Appendix. The old and the updated input-output tables

In the tables below, the two-digit numbers are estimated, whereas the others are pre-determined. Because of separate rounding-off, the two-digit numbers do not necessary add-up to the required totals.

OLD table	Goods	Services	Consumption	Net exports	Total output
Goods	7	3	5	-3	12
Services	2	9	8	1	20
Net taxes	-2	0	2	1	1
Total use	7	12	15	-1	33
Value added	5	8	0	0	13
Total input	12	20	15	-1	46

RAS^{ex-} update	Goods	Services	Consumption	Net exports	Total output
Goods	8.43	3.57	6.00	-3	15
Services	2.57	11.43	10.23	0.77	25
Net taxes	-2	0	0.77	0.23	-1
Total use	9	15	17	-2	39
Value added	6	10	0	0	16
Total input	15	25	17	-2	55

GRAS update	Goods	Services	Consumption	Net exports	Total output
Goods	7.89	3.42	5.92	-2.24	15
Services	2.54	11.58	10.68	0.20	25
Net taxes	-1.43	0	0.40	0.03	-1
Total use	9	15	17	-2	39

SPAD update	Goods	Services	Consumption	Net exports	Total output
Goods	9.50	3.75	7.75	-6.00	15
Services	2.51	11.25	9.25	2.00	25
Net taxes	-3.00	0.00	0.00	2.00	-1
Total use	9	15	17	-2	39

SPSD update	Goods	Services	Consumption	Net exports	Total output
Goods	9.72	4.47	6.80	-5.98	15
Services	2.94	10.53	9.52	2.01	25
Net taxes	-3.66	0.00	0.68	1.98	-1
Total use	9	15	17	-2	39