gpcema.gms

a general purpose cross-entropy matrix adjustment program (quick user guide)

André Lemelin[1]

December 2013

partnership for economic policy

pep

1 Centre INRS-UCS, Université du Québec, Montréal
andre_lemelin@ucs.inrs.ca

GPCEMA.GMS

A GENERAL PURPOSE CROSS-ENTROPY MATRIX ADJUSTMENT PROGRAM

This note is a quick user guide for the GAMS matrix adjustment program GPCEMA.gms. GPCEMA stands for « General Purpose Cross-Entropy Matrix Adjustment ». The matrix adjustment problem tackled in GPCEMA.gms is to adjust an existing matrix $\mathbf{T^0}$ (called the prior matrix) so that the adjusted matrix $\mathbf{T^*}$ (the posterior matrix) has row and column sums that conform to known marginal totals. The classic example of this problem is that of adjusting an outdated input-output table to current industry and commodity output data.

GPCEMA.gms is based on the minimum information-gain principle (also known as minimum cross-entropy), as it has been extended to deal with negative entries by Junius and Oosterhaven (2003). The standard cross-entropy minimization technique considers the structure of the prior matrix as a probability distribution. Since a probability cannot be negative, the contribution of Junius and Oosterhaven (2003) has been to propose a way of dealing with negative entries.

Cross-entropy is a measure of the one-way divergence[1] of the posterior probability distribution (the adjusted matrix), from the *a priori* distribution (the unadjusted matrix). The adjustment process consists in adjusting the matrix to its marginal totals in such a way as to minimize that divergence, which is readily interpreted in information theory as minimizing the quantity of extraneous information imposed upon the *a priori* matrix to fit the marginal totals. More elaborate presentations of the minimum information-gain principle and the cross-entropy adjustment method are found in Lemelin (2009 and 2011), where the reference list contains material for further reading. For an application, see Lemelin (2011).

The program automatically deals with situations where row sums, column sums, or both are constrained to zero. Consider, for example, a two-line world trade table with one column for each country, where the first line is the current account balance, and the second is the capital and financial account balance: in such a table, row sums and column sums are constrained to be zero (the current account balance is of equal magnitude to the capital and financial account balance and of opposite sign). Next, consider a closed system of balance sheets, such as a world table of country international investment positions (IIP). Here, row sums must be zero, because the worldwide sum of assets, minus liabilities of any given

[1] It is not a distance measure, because it is not symmetrical.

category must be zero; column sums are countries' net worth, or international investment position, which may be positive or negative[2].

Mathematical statement

GPCEMA.gms solves the following problem (Lemelin, 2010 and 2011):

$$\min_{\{z\}} H\left(\mathbf{T}^*\middle\|\mathbf{T}^0\right) = \frac{1}{d}\sum_{i,j}\left|t_{ij}^0\right| z_{ij}\ln\left(z_{ij}\right) + \ln\left(\frac{c}{d}\right) \tag{01}$$

subject to

$$\sum_j t_{ij}^0 z_{ij} = u_i \tag{02}$$

$$\sum_i t_{ij}^0 z_{ij} = v_j \tag{03}$$

where

$H\left(\mathbf{T}^*\middle\|\mathbf{T}^0\right)$	is the Kullback-Leibler cross-entropy measure of discrepancy of the posterior distribution derived from matrix $\mathbf{T^*}$ relative to the prior distribution derived from $\mathbf{T^0}$;
t_{ij}^0	are the elements of prior matrix $\mathbf{T^0}$;
z_{ij}	are non-negative proportional adjustment factors;
$t_{ij}^* = t_{ij}^0 z_{ij}$	are the elements of the adjusted posterior matrix $\mathbf{T^*}$;
u_i	are the exogenous row totals;
v_j	are the exogenous column totals;
$c = \sum_{i,j}\left\|t_{ij}^0\right\|$	is the overall sum of the elements of $\mathbf{T^0}$;

$d = \sum_{i,j}\left|t_{ij}^*\right| = \sum_{i,j}\left|t_{ij}^0\right| z_{ij}$ is the overall sum of the elements of $\mathbf{T^*}$.

Formulating the minimand in terms of absolute values $\left|t_{ij}^0\right|$ is Junius and Oosterhaven's "trick" to deal with negative entries. In fact, this approach separates the information retrieved from *a priori* matrix $\mathbf{T^0}$

---

2 This is the application in Lemelin (2011). The method is also used in Lemelin, Robichaud and Decaluwé (2013).

into two components: the relative magnitudes of cells, captured by the absolute values $\left|t_{ij}^{0}\right|$, and their sign (+ or –). The latter information is treated as certain and preserved in the adjustment procedure.

Other GAMS adjustment programs

Other matrix adjustment programs are offered to modellers. Two of them are briefly presented here.

PEP (Lemelin, Fofana and Cockburn, 2013)

In contrast to the one proposed here, the Sambal program originally developed by Fofana *et al*. in 2005 (revised edition, Lemelin *et al.*, 2013) is specifically designed for SAM balancing, taking advantage of the particular double-entry accounting structure of a SAM, where the total receipts of every account (SAM row) must be equal to its total payments (SAM column). The problem Sambal solves is equivalent to

$$\min_{\{t^*\}} H'\left(\mathbf{T}^* \middle\| \mathbf{T}^0\right) = \sum_i \sum_j t_{ij}^* \ln \frac{t_{ij}^*}{t_{ij}^0} \tag{04}$$

subject to

$$\sum_{i,j} t_{ij}^* = \sum_{i,j} t_{ij}^0 \tag{05}$$

$$\sum_i t_{ji}^* = \sum_i t_{ij}^* \tag{06}$$

Additional information, such as known cell values or row or column totals, may be taken into account as supplementary constraints. While the standard entropy method does not allow negative values, Sambal solves this problem by transposing these values to their counterpart cells before balancing the SAM. Indeed, as the SAM represents flows from one account to another, a negative flow from account A to account B is equivalent, in terms of accounting, to a positive flow of equal magnitude from account B to account A. The reverse transposition is performed after the SAM has been balanced, to restore negative values to their original positions. It should be noted however, that the reverse transposition eliminates one of any pair of cross-flows of opposite signs: if the original matrix has both a negative flow from A to B and a positive flow from B to A, the smaller one of the two in absolute value will be set to zero in the balanced matrix.

Finally, Sambal offers a choice of minimands. The default minimand is the entropy-based one in (04). But the user may choose to use a quadratic minimand, labeled “OLS” for Ordinary Least Squares:

$$\min_{\{t^*\}} \sum_i \sum_j \left(\frac{t_{ij}^* - t_{ij}^0}{t_{ij}^0} \right)^2 \tag{07}$$

IFPRI (Robinson, Cattaneo and El-Said, 2001)

Robinson *et al*. (2001) have developed a cross-entropy SAM estimation method to deal with a situation in which initial data are inconsistent and measured with error[3]. In that respect, their approach is far more elaborate than the one presented here. On the other hand – and therein lies a second difference with the program presented here –, they deal with negative entries in the same way as Lemelin, Fofana and Cockburn (2013), by netting them out using flow transposition.

But there is another major difference between the approach proposed by Robinson *et al.* and the ones implemented here and by Lemelin *et al* (2013). This is most easily described in the case where there is no measurement error, and marginal totals are known with certainty.

Robinson *et al*. (2001, p. 56) apply the cross-entropy formula to the coefficient matrix. Ignoring measurement errors, the problem they solve is

$$\min_{\{a^*\}} H'' = \sum_i \sum_j a_{ij}^* \ln \frac{a_{ij}^*}{a_{ij}^0} \tag{08}$$

subject to

$$\sum_i a_{ij}^* = 1 \tag{09}$$

$$\sum_j a_{ij}^* v_j = u_i \tag{10}$$

where

$a_{ij}^0 = t_{ij}^0 \Big/ \sum_h t_{hj}^0$	are the SAM coefficients derived from the prior matrix $\mathbf{T^0}$;
a_{ij}^*	are the adjusted SAM coefficients.

3 Two versions of the GAMS code are available. The first uses mixed complementarity programming (MCP), and the second non-linear programming (NLP). Both can be copied (not downloaded) from: http://www.gams.com/modlib/libhtml/cesam.htm, and http://www.gams.com/modlib/libhtml/cesam2.htm respectively.

The adjusted SAM, $\mathbf{T^*}$, is then computed as

$$t_{ij}^* = a_{ij}^* v_j \tag{11}$$

Note that, in view of (09), each column of the coefficient matrix may be formally viewed as a probability distribution, so that the minimand in (08) is a sum of several Kullback-Leibler cross-entropy measures, one per column. Now, given known row and column sums, adjusting the SAM coefficients as in the IFPRI method, rather than the transactions matrix, is tantamount to assigning weights to the terms in the right-hand side of (01) above; these weights are inversely proportional to the new column sums.[4]

Description of the GAMS program

The GAMS program is abundantly commented, and the user is invited to read the comments carefully.

Data is read into the program from an Excel file. The interior values of the matrix are the elements of prior matrix $\mathbf{T^0}$. The last column in the table is the vector of exogenous row totals $\mathbf{u}$, and the last line is the vector of exogenous column totals $\mathbf{v}$.

The problem to be solved is defined as a set of equations which include a target function to be minimized, row and column sum constraints, and, depending on the problem at hand, other constraints described below. The target function is the Kullback-Leibler cross-entropy formula (01).

The program reads the data and determines to which case the problem belongs:

- Case 1: at least one row total u_i and one column total v_j are nonzero;
- Case 2: all column totals, but not all row totals are zero;
- Case 3: all row totals, but not all column totals are zero;
- Case 4: all row totals and all column totals are zero.

A case 2 or case 3 problem is then converted into a case 4 problem. In a case 2 problem, a new column is added to the prior matrix: if $\mathbf{T^0}$ has m columns, an $(m+1)^{\text{th}}$ column is added with $t_{i,m+1}^0 = -u_i$. Then, row totals are changed to $u_i = 0$, and the constraints $z_{i,m+1} = 1$ are added to the set of equations. In a case 3 problem, a new row is added to the prior matrix: if $\mathbf{T^0}$ has n rows, an $(n+1)^{\text{th}}$ row is added with $t_{n+1,j}^0 = -v_j$. Then, column totals are changed to $v_j = 0$, and the constraints $z_{n+1,j} = 1$ are added to the set of equations.

---

4 Fore a more elaborate discussion and a mathematical demonstration, see McDougall (1999) or, alternatively, Lemelin *et al*. (2013).

Next, if the sum of column totals and the sum of row totals are both nonzero (case 1), the program adjusts column totals proportionately so that their sum equals the sum of row totals (it could be the other way around), in order to eliminate any rounding error discrepancy and ensure that the feasibility constraint is fulfilled exactly. In the event of a null sum of column totals and a nonzero sum of row totals, or vice-versa, the problem is infeasible and the program exits with a warning.

In a « pure » case 4 problem (not converted from case 2 or 3), since the solution is defined only up to a multiplicative factor, the constraint $\sum_{i,j} \left| t_{ij}^0 \right| z_{ij} = \sum_{i,j} \left| t_{ij}^0 \right|$ is added to the set of equations as a normalization rule.

The program uses the CONOPT GAMS solver to find the solution to the problem and writes the results to a GDX file, which is then converted to an Excel file.

Files

The *GPCEMA.zip* file contains the GAMS program file *GPCEMA.gms*. Several input Excel files are included with the program as examples. The default example is *Input.xls*, a standard adjustment problem. *Zero_col_tot.xls* is an example of a case 2 problem, while *Zero_line_tot.xls* is an example of a case 3 problem. Lastly, *Zero_margins.xls* is an example of a case 4 problem. Finally, the file called *Input.gms* is a text file that may be used as an alternate method of reading the problem data.

References

Junius, T. and J. Oosterhaven (2003), « The solution of updating or regionalizing a matrix with both positive and negative entries », *Economic Systems Research*, 15, pp. 87-96.

Lemelin, André (2009), « A GRAS variant solving for minimum information loss », *Economic Systems Research*, 21(4), p. 399-408.

Lemelin, André (2011) *Balancing international financial data: Fun and games with GAMS and cross-entropy*, Poverty & Economic Policy Research Network (PEP), MPIA Working Paper 2011-07. http://www.pep-net.org/publications/working-papers/

Lemelin, André, Ismaël Fofana and John Cockburn (2013) *Balancing a Social Accounting Matrix: Theory and application (revised edition)*, Partnership for Economic Policy (PEP), Université Laval, Québec, 24 p.
http://www.pep-net.org/programs/mpia/training-material/

Lemelin, André, Véronique Robichaud and Bernard Decaluwé (2013) « Endogenous current account balances in a world CGE model with international financial assets », *Economic Modelling*, 32 (May), 146-160.
http://dx.doi.org/10.1016/j.econmod.2013.01.046

McDougall, R. A. (1999) *Entropy Theory and RAS are Friends*, GTAP Working Paper No. 03, GTAP, Purdue, Indiana.
https://www.gtap.agecon.purdue.edu/resources/download/10.pdf.

Robinson, S., A. Cattaneo and M. El-Said (2001) « Updating and estimating a Social Accounting Matrix using Cross-Entropy methods », *Economic Systems Research*, 13, pp. 47-64.