

August 22, 2003

Notes on Balancing Matrices with Uncertain Totals

Robert Michael Field

The gross output data necessary for an input-output table are frequently unavailable or of uncertain reliability. Rather than force the rows and columns of a table to add to estimated or uncertain totals, the algorithm presented in this paper (which is called *gvras* in *G7+*) allows some change in the gross outputs during the balancing process.

In our work on the *MuDan* model for the Chinese economy, we have been faced with a great deal of uncertainty about the gross outputs. Gross output data are not available for some sectors, and for others we do not have great confidence in their accuracy. It became obvious during the process of making the 1997 input-output table consistent with the reported national income data that allowing some change in the gross outputs introduced less distortion in the balanced table than using the gross outputs as reported.

AN EXAMPLE

The algorithm is illustrated using a 5-sector version of the Chinese table. The data available for control totals, the nature of the algorithm, and the results of its application to the table are discussed, below.

THE DATA

Figure 1 shows the 5-sector table. The table is in the usual format, and the only subtotals are the row and column for intermediate inputs.

Figure 2 shows the gross value data available for use as control totals and compares them with the data from the I-O table. For agriculture, the reported gross value is accepted. For sectors mining, manufacturing and power, total gross values are not available; however, gross values are available for independent accounting enterprises. In addition, the gross value of industry as a whole (that is, the sum of gross values for mining, manufacturing and power) is available, and the gross value of other enterprises can be derived as a residual. For construction and tertiary industry, no gross value data are available.

Figure 3 shows the final demand data available for use as control totals and compares them with the data from the I-O table. There is no reported total for the error column, but it can be derived as the difference between GDP and GDE.

Figure 4 shows the value added data that can be used as preconditions in the process of balancing the matrix and compares them with the data from the I-O table. Preconditions, of course, are known values for part or all of a cell that can be removed from the cell and from the corresponding row and column totals before the matrix is balanced and restored afterwards – thus preventing the balancing process from changing values that are known. In this example, value added is known for all of agriculture, for part of mining, manufacturing and power, and for all of construction and tertiary industry.

THE ALGORITHM

In order to allow some or all of the gross outputs to change, the algorithm works by balancing the entire table rather than just the A matrix. This requires first that:

(a) a row (between value added and gross output) and a column (between final demand and gross output) be opened for each group of sectors that has a known total, and a single row and column be opened for those sectors that do not have a known total.

(b) the portion of gross output that is considered uncertain – which can be all or part or none – be subtracted from the total gross output and from the empty cell in the row (and column) that has been inserted, and

and second, that the scaling procedure be able to handle sets that include negative numbers and sets that sum to zero.

Figure 5 shows the modified version of the I-O table that is to be balanced. The row and column for intermediate input have been removed, total value added has been moved across to the column for gross output, and the various categories of final demand has been moved down to the row for gross output. In addition, two rows and two columns have been inserted for that portion of gross output that will be allowed to change during the process of balancing the matrix. The first inserted row and the first inserted column are for those sectors that have a known sum of the portion of gross output; the second row and second columns are for those sectors that do not have a known sum.

Figure 6 shows the table with the uncertain gross outputs moved inside the table. The details are as follows:

- For agriculture, the entire gross output (23,788) is known, so there is no uncertain gross output and nothing is moved inside the table.
- For mining, manufacturing and power, the gross values of independent accounting enterprises (4,681; 59,985; and 3,686) are retained in the gross outputs column and the differences between total gross outputs from the table and the gross values of independent accounting enterprises (which have negative values) are inserted in the rows and column for uncertain gross outputs with a known total.

- For construction and tertiary industry, no part of the gross output is known, so the entire amount is moved inside the table (59,823).

Figure 7 gives the details of the gvras command.

Figure 8 gives the details of the (optional) preconditions file. Each line in the file (except comments, which begin with #) imposes a single precondition that contains four elements, as follows:

- The precondition command “zero” or “part”. Zero indicates that exact value of a cell in the table is known; and part that a portion of the value is known.
- The row of the cell.
- The columns of the cell
- The name of the bank or vam file in which the known value is found.

Prior to executing the gvras command, both the row total and column total are reduced by the value from the bank or vam file. If the precondition command is “part”, the value in the cell is also reduced by the value from the bank or vam file; or if the command is “zero”, the value in the cell is set to zero. After the gvras command is executed, the known value is restored to the cell, the row total and the column total.

The table is now ready to balance by the following iterative procedure:

- Scale the columns.
- Scale the row(s) for uncertain gross outputs with a known total.
- Pivot the row(s) for uncertain gross output to the columns for uncertain gross output. This step is necessary to insure that the row sums will equal the column sums sector by sector when the balancing is complete and the extra rows and columns have been removed.
- Scale the rows.
- Scale the columns(s) for uncertain gross outputs with a known total.
- Pivot the columns(s) for uncertain gross output to the rows for uncertain gross output.

THE RESULTS

Figure 9 shows the portion of the G7+ output file that is generated by the gvras command. It first reports that the reconditions are being removed and the year for which the table is being balanced. It then lists, iteration by iteration, the largest column and largest row scale factor. And finally it reports whether or not the algorithm converged, that the preconditions have been restored and that the preconditions file has been closed.

A more detailed version of the process of convergence is shown in the Appendix Table. In the first iteration, it gives the scale factor for every row and column; in the second through the fifth iteration, it gives every scale factor that is greater than five

percent, or the largest row and the largest scale factor if none is greater than five percent. Finally it gives the largest row and the largest scale factor for every fifth iteration. This additional detail sometimes reveals problems that are not shown in the data presented in the G7+ output file.

Figure 10 compares the matrix before and after balancing with the `gyras` command, and gives the difference and the difference in percent. The first thing that strikes the reader is the very large size of the differences in the error column. These differences, however, are not the result of some aberration in the algorithm but the result of the difference between GDP and GDE in the table (-1,042) and in the national accounts (-432). Similarly, the difference in the imports column is relatively large – but once again, it is the result of the difference between imports in the table and in the national accounts. The other differences are all small.

Appendix Table

gvras.log -- Proportional scale factors for balancing
matrices with uncertain totals

iter[1], col[1], cscale = -0.036092
iter[1], col[2], cscale = -0.000000
iter[1], col[3], cscale = 0.000000
iter[1], col[4], cscale = 0.000000
iter[1], col[5], cscale = 0.014918

iter[1], col[6], cscale = -0.020772
iter[1], col[7], cscale = -0.000001
iter[1], col[8], cscale = 0.038547
iter[1], col[9], cscale = -0.122629
iter[1], col[10], cscale = 0.585610

iter[1], row[7], GVRSCALE = 0.034277

iter[1], row[1], rscale = -0.023895
iter[1], row[2], rscale = -0.006352
iter[1], row[3], rscale = -0.003227
iter[1], row[4], rscale = -0.021860
iter[1], row[5], rscale = -0.007857

iter[1], row[6], rscale = 0.046065

iter[1], col[11], GVCSCALE = 0.003651

iter[2], col[2], cscalemax = -0.015398

iter[2], row[6], rscalemax = 0.012667

iter[3], col[2], cscalemax = -0.007913

iter[3], row[6], rscalemax = 0.005104

iter[4], col[2], cscalemax = -0.003507

iter[4], row[6], rscalemax = 0.002249

iter[5], col[2], cscalemax = -0.001572

iter[5], row[6], rscalemax = 0.001035

iter[6], col[2], cscalemax = -0.000724

iter[6], row[6], rscalemax = 0.000486

iter[7], col[2], cscalemax = -0.000340

iter[7], row[6], rscalemax = 0.000231

iter[8], col[2], cscalemax = -0.000161

iter[8], row[6], rscalemax = 0.000110

iter[9], col[2], cscalemax = -0.000077

iter[9], row[6], rscalemax = 0.000053

iter[10], col[2], cscalemax = -0.000037

iter[10], row[6], rscalemax = 0.000025

iter[11], col[2], cscalemax = -0.000018

iter[11], row[6], rscalemax = 0.000012

iter[12], col[2], cscalemax = -0.000008

iter[12], row[6], rscalemax = 0.000006

iter[13], col[2], cscalemax = -0.000004

iter[13], row[6], rscalemax = 0.000003

iter[14], col[2], cscalemax = -0.000002

iter[14], row[6], rscalemax = 0.000001

iter[15], col[2], cscalemax = -0.000001

iter[15], row[6], rscalemax = 0.000001