SIMPLIFIED LEAST-SQUARES APPROACH FOR AN AGE OF COMPUTERS

Irving H. Siegel
The W. E. Upjohn Institute for Employment Research*

A Supermatrix System with Unprocessed Elements

The increasing availability of electronic data-processing services greatly enhances the practicability of an approach to least-squares adjustment developed some years ago by the present author. The straightforwardness of programming and of computer solution was demonstrated in tests conducted for the author by Daniel E. Cowgill in 1963-64 at the Computer Sciences Center of Research Analysis Corporation. The treatment of weighted and unweighted observations by the author's method and the design, in addition, of certain related adjustment procedures that do not employ the least-squares principle were outlined in a paper presented at the 1964 meeting of the American Statistical Association.1/ Below, reference is made again to the least-squares cases considered earlier, and the approach is shown to be also applicable to situations involving linear restrictions.

The author's concept should appeal to statisticians, who need only to organize the data in a certain manner that permits the delegation of literally all the mathematical operations to computer specialists. The method features instant design, without any arithmetic processing of the data whatsoever, of an algebraic equivalent of a set of conventional "normal equations." This equivalent system is actually a correct supermatrix equation constructed by mere arrangement of the original information "packages" with suitable "dunnage". The proper way to organize the data is easy for the statistician to learn and to teach his subordinates.

Data-processing specialists can assume the complete burden of programming and calculation. The traditional intermediate step of computing normal equations is avoidable altogether--or it can be incorporated, if desired, in the computing routine. The unprocessed data, even when voluminous, may be stored intact on tape; and, in addition, it is a simple matter to add, eliminate, or amend the observations, weights, and restraints.

The essence of the author's approach is to expand, rather than to compress, the original oblong matrix of the supposedly error-free observational data into an invertible square matrix. In the conventional approach, it will be recalled, the oblong is compressed by suitable operations into a number of "normal equations" equal to the number of unknown constants. Our alternative requires the construction instead of a design "supermatrix", which we shall call an S-matrix,2/ by the introduction of extra unprocessed information--e.g., conditions imposed on error terms, equal (i.e., unit) weights or unequal weights, and linear restraints. By such additions, the number of variables in the S-system as a whole is considerably enlarged. Explicit solutions are not needed for the extra variables, although they may as a rule be obtained routinely with modern computing equipment. The conditions imposed on the errors, incidentally, are restatements of the conditions expressed in the normal equations derived in the conventional approach.

Transposes of original data matrices play an important role in the formation of the design S-matrix.

---

* The author's views should not be ascribed to The W. E. Upjohn Institute for Employment Research.

1/ I. H. Siegel, "Least Squares with Less Effort", 1964 Business and Economic Statistics Section Proceedings of the American Statistical Association, pp. 284-85. See also two other pertinent papers by I. H. Siegel: "Least Squares 'Without Normal Equations'," Transactions of N.Y. Academy of Sciences, February 1962, pp. 362-371; and "Deferment of Computation in the Method of Least Squares", Mathematics of Computation, April 1965, pp. 329-331.

2/ At Research Analysis Corporation, "S-matrix" stood, not only for "supermatrix", but also for "swollen matrix" and, still better, for "Siegel matrix".

Specifically, the package of conditions imposed on the errors is a transpose of the matrix of the observed error-free coefficients. Hence, any puzzlement raised by the preceding paragraph with respect to the sources of additional needed information is groundless. Indeed, the method of least squares has, in addition to its other recognized merits, an unadvertised distinction of requiring "minimum information" for adjustment, inasmuch as it makes double use, through the transpose, of original observational data.

Our approach calls for the "stacking" of submatrix equations into the S-system. The oblong of error-free coefficients, its transpose, and a diagonal unit of weights are basic ingredients of all S-systems involved in least-squares adjustment. Another essential ingredient of the design S-matrix is the zero block--useful for padding and for spacing the other packages.

Organizing the Data: Unweighted, Weighted, Constrained Cases

The observational equations to be adjusted may be written in matrix form as:

$$Xb + e = Y.$$

Here X refers to n sets of observations on error-free coefficients of the k variables $b(n>k)$, Y is a column of n observed values, and e is a column of n unknown error terms to be purged from the elements of Y. This matrix equation becomes a submatrix equation in the S-system, comprising one tier.

Looking at the above matrix equation as though it were an ordinary linear equation in two unknowns, b and e, we may ask: What additional equations would yield a unique solution for these unknowns? The simplest of all possible companion equations is $X'e = 0$, where X' is the transpose of X. Using this equation as another tier, we have the S-system:

$$\begin{bmatrix} I & X \\ X' & 0 \end{bmatrix} \cdot \begin{bmatrix} e \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \end{bmatrix},$$

where I is a unit matrix and each 0 represents a zero block of appropriate dimensions. It may easily be verified that the two submatrix equations embedded in this S-system are equivalent to the unweighted least-squares matrix normal equation--i.e., equivalent to $X'Xb = X'Y$. It should also be observed that none of the elements in either tier has been processed arithmetically, yet the supermatrix system corresponds to the conventional normal equations. The system is ready for computer processing, being a solvable matrix equation rather than just a turgid array of information.

When the n observations are weighted, we have three tiers. We must introduce a diagonal n x n weight matrix and a set of n arbitrary unknown variables, λ, if all arithmetic preprocessing of the elements is to be avoided. The following S-system may readily be proved equivalent to the weighted matrix normal equation, $X'WXb = X'WY$:

$$\begin{bmatrix} 0 & I & X \\ I & W & 0 \\ X' & 0 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda \\ e \\ b \end{bmatrix} = \begin{bmatrix} Y \\ 0 \\ 0 \end{bmatrix}.$$

Constrained least-squares adjustment may also be handled by the stacking of submatrix equations containing no preprocessed elements. For example, the famous Gaussian "method of correlatives", commonly illustrated for angle or line measurements in texts treating conditioned observations,3/ is readily managed in a two-tier or three-tier construction. When three tiers are used, each represents an equation in one or more of three sets of variables--the unknown coefficients that we wish to determine, the residuals, and the "correlatives" (undetermined Lagrange multipliers). When the observation equations are written in non-matrix form, they are generally more numerous than the number of unknown constants (e.g., the angles of a triangle), while the conditions to be satisfied (e.g., the sum of the three angles equals 180°) are always fewer.

In modern treatments of constrained least squares, m linear restrictions of the form

$$Z = Rb$$

3/ See, for example, the classical discussions of Merriman and Chauvenet or the more recent works of Linnik and Arley-Buch.

are usually introduced, 4/ where Z is a known column vector of m elements, $n \geq k > m$, and R is a known matrix of order m x k. Arbitrary variables, ϕ, are needed to express the conditions implicitly imposed on R', the transpose of R; they are m in number. The S-system becomes:

$$\begin{bmatrix} R' & X' & 0 \\ 0 & I & X \\ 0 & 0 & R \end{bmatrix} \cdot \begin{bmatrix} \phi \\ e \\ b \end{bmatrix} = \begin{bmatrix} 0 \\ Y \\ Z \end{bmatrix}$$

None of the data packages contains processed elements. When the first two submatrix equations are combined, they become the more recognizable matrix normal equation, $X'Xb = X'Y + R'\phi$, which remains subject to the restriction stated in the third submatrix equation, $Z = Rb$.

4/ See, for example, A. S. Goldberger, Econometric Theory, Wiley, New York, 1964, pp. 256-257; and R. L. Plackett, Regression Analysis, Oxford, 1960, pp. 52-53.