INSTANT SUPERMATRIX SETUPS FOR MULTIVARIATE, ITERATIVE, AND MULTI-EQUATION MODELS

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

Aim and Background

My purpose is to illustrate the usability of data supermatrices in more complex cases of least-squares estimation than I have been treating earlier. My motivation derives in part from inquiries occasioned by a paper that I presented at the August 1968 meeting of the American Statistical Association [1] and by another that appeared in the December 1968 American Statistician [2]. An additional stimulus has been provided by the development of a new computer program for regression analysis that takes account of the special features of least-squares supermatrices. The availability of such a program, discussed in another paper presented at this meeting, should encourage the practical utilization of the supermatrix approach.

In the classical least-squares cases, which involve unweighted or weighted observations (with no auxiliary conditions on the unknown parameters), the supermatrix method permits instant organization of the data, without any prior processing at all, into the algebraic equivalents of normal equations. If, in the weighted case, some prior arithmetic processing is allowed, an alternative supermatrix system may be written that is more compact but still algebraically equivalent. As a rule, there is a tradeoff between the acceptable amount of prior computation and the acceptable size of the supermatrix system.

It was for the relatively simple classical cases that the supermatrix approach was originally developed. In my earliest gropings toward this method (see my 1941 papers in the Journal of the American Statistical Association [3], [4]), I used a weighted least-squares model and rectangular data matrices in studying the relationship between aggregative index numbers that embody different sets of weights.

I have already demonstrated the extensibility of supermatrix design beyond the simplest and most familiar situations. Thus, all or most prior arithmetic operations are avoidable (a) in the fitting of a straight line when all observations are subject to error (othogonal regression represents a special subcase)[2] and (b) in least-squares adjustment when the unknown constants are subject to side constraints [5].

Now, I consider additional, more difficult, instances. If these instances may be said to exemplify a common theme, it is "multiplicity." Thus, I say something below about the estimation of parameters for a function with multiple variables subject to error. I also offer two illustrations of multiple-step approximation; one is iterative in the sense of ultimate approach to least-squares estimation, and the other is iterative in the sense that the original linear model undergoes refinement by the addition of parameters and observations. Finally, I make some remarks on models involving multiple equations subject to disturbance.

Before proceeding with the discussion of new cases, I want to comment further on least-squares supermatrix systems. All of these have the general form $Fg = h$. Here, F is a square design matrix -- really, a supermatrix -- containing unprocessed observations on one or more regressors and providing coefficients for individual error terms. These coefficients state assumptions regarding the combination of errors. The design matrix may also include, if the problem requires, assumptions relating to additional Lagrangian terms. The unknown constants, the (usually) unknown observational errors (really, "residuals"), and the unknown Lagrangian multipliers (if required) are incorporated in a supervector, g. Another supervector, h, (usually) shows the observations on the dependent variable and the aggregate or summary conditions imposed on the errors and on the Lagrangian terms (the aggregates typically are zero).

F, g, and h have well-defined structures. That is, they are partitionable in rather obvious ways. The elements they contain are organized in characteristic "packages," according to the nature of the adjustment problem to be solved. Since these packages,

*The author's views should not be ascribed to the Upjohn Institute.

moreover, enter discernibly into simultaneous submatrix equations, the supermatrix system may be conceived as the result of a "stacking" procedure [5].

One of the ways of exhibiting the structure of the supermatrix system is to cast it into this form:

$$\begin{bmatrix} O & Q \\ P & I \end{bmatrix} \cdot \begin{bmatrix} d \\ r \end{bmatrix} = \begin{bmatrix} O \\ Y \end{bmatrix}$$

$$(F \quad \cdot \quad g \quad = \quad h) \ .$$

Here, O is a block of zeros; I stands for one or more (diagonal) identity matrices; P shows the observed values of the independent variables or regressors; Q gives the coefficients of the individual error terms (residuals); d represents a subvector of unknown parameters; r is the subvector of unknown errors; and Y typically refers to observed (sometimes, assumed or computed) values of the dependent variable. When the observations on the dependent and independent variables are subject to error, the elements of r may be composite. In general, $Qr = O$ amounts to a simple direct statement of the normal equations in terms of the errors (i.e., residuals); and $Pd + Ir = Y$ (strictly, an identity rather than an equation) sets out the observed data in the form of the function, the prototype relationships, that is to be estimated.

In my first paragraph, I mentioned that another paper being presented at this ASA meeting deals with the programming of supermatrix systems (in FORTRAN) for efficient computer solution. My collaborator is Mr. Mac Shaibe, of the U.S. Bureau of Labor Statistics. As a result of his efforts, I expect supermatrix regression, which is not just a curious technical toy, to become more widely recognized as a potentially useful tool.

Multiple Variables Subject to Error

The multivariate case in which all observations are subject to error may be handled in (a) exactly the same way that I have treated the straight line with x's and y's subject to error [2] or (b) an alternative way that leads to a still larger supermatrix system. This larger system can be set up instantly with virtually no prior arithmetic processing. Some awkwardness may arise in actual computation, however, since the design matrix itself includes unknown parameters!

To work out the required supermatrix pattern, I have made use of what I call "normal identities" [6]. These identities, derived from the observation equations in a simple systematic manner, are reducible to normal equations when the appropriate error-affected aggregates are set equal to zero. Since not all of the error-affected aggregates in the identities can reasonably be assumed to vanish, however, we are left with more unknowns than equations and accordingly have to make additional assumptions to connect the unannihilated aggregates. In any event, the suppression and interconnection of aggregates are not matters of arbitrary choice; for the object is to obtain by the supermatrix method the same result that is given by the relevant normal equations. From a scrutiny of normal identities, we learn how to treat error aggregates for the derivation of the correct normal equations; and these normal equations have to be examined in turn for clues to their "explosion" into supermatrix equivalents.

For simplicity of exposition, I consider a function of the form $y = a + bx + cz$, where a, b, and c are the unknown parameters and where the observed values of y, x, and z are subject to error. Thus, I really start with equations (identities) of the form

$$y_i + s_i = a + bx_i + bt_i + cz_i + cu_i$$

($i = 1, \ldots, n$), where s_i, bt_i and cu_i represent unknown error terms that are to be "purged" from the observations. I call attention to the appearance of b and c as separate unknowns in the subvector d and also in combination with the error terms t_i and u_i, respectively, in the subvector r.

In addition to a block of zeros ($3n + 3$, 3) in the "northwest" corner of the design matrix, I obtain these packages of elements for the supermatrix system:

$$Q = \begin{bmatrix} 1 \ldots & 1 \ldots & 1 \ldots \\ y_1 \ldots & x_1 \ldots & z_1 \ldots \\ z_1 \ldots & y_1 \ldots & x_1 \ldots \\ kx_1 & & by_1 \\ \ddots & & \ddots \\ & mz_1 & cy_1 \\ & \ddots & \ddots \end{bmatrix},$$

$$[P \; I] = \begin{bmatrix} 1 & x_i & z_i & 1 & & 1 & & 1 & \\ \vdots & \vdots & \vdots & & \ddots & & \ddots & & \ddots \end{bmatrix},$$

$$\begin{bmatrix} d \\ r \end{bmatrix} = [a \; b \; c \quad bt_1 \ldots \quad cu_1 \ldots \quad -s_1 \ldots]', \text{ and}$$

$$\begin{bmatrix} 0 \\ Y \end{bmatrix} = [0 \; 0 \; 0 \quad 0 \ldots \quad 0 \ldots \quad y_1 \ldots]'.$$

Dots indicate an extension to n terms. The symbols k and m appearing in Q represent constants to which values may be assigned according to assumptions desired with regard to relative variances. Thus, having only four normal equations but three parameters, a, b, and c, and three unknown and nonvanishing error aggregates, Σsy, $b\Sigma tx$, and $c\Sigma uz$, I interconnect the sums (which really represent variances) in this way: $\Sigma sy = k\Sigma tx = m\Sigma uz$. Notice that the terms kx_i and mz_i are the only ones in the design supermatrix inviting or requiring any prior computation.

Two Iterative Cases

In my first example of sequential approximation, I want to estimate the parameters of a straight line, but I choose to guess initially at the least-squares y's corresponding to the given x's. I could also introduce a second or third guess if I choose to do so; the supermatrix system is not overtaxed thereby. What is the supermatrix setup that assures "closure," that makes up for my bad guesses and yields the correct least-squares estimates of the parameters anyway?

For simplicity, I assume only one guessing round. I set up the supermatrix system for the observations $y_i = a + bx_i + e_i$ $(i = 1, \ldots, n)$ in this manner:

$$Q = \begin{bmatrix} 1 & & 1 & & 1 & \ldots \\ & 1 & & 1 & & \ldots \\ x_1 & & x_2 & & x_3 & \ldots \\ & x_1 & & x_2 & & \ldots \end{bmatrix},$$

$$[P \; I] = \begin{bmatrix} 1 & & x_1 & & 1 & & & & \\ & 1 & & x_1 & & 1 & & & \\ 1 & & x_2 & & & & 1 & & \\ & 1 & & x_2 & & & & 1 & \\ 1 & & x_3 & & & & & & 1 \\ \vdots & \vdots & \vdots & \vdots & & & & & \ddots \end{bmatrix},$$

$$\begin{bmatrix} d \\ r \end{bmatrix} = [a'a'' \quad b'b'' \quad e_1' e_1'' e_2' e_2'' \ldots]',$$

$$\begin{bmatrix} 0 \\ Y \end{bmatrix} = [0 \; 0 \quad 0 \; 0 \quad y_1' e_1' y_2' e_2' \ldots]'.$$

Here the y_i' represent guesses; and the $e_i' = y_i - y_i'$ represent not only a set of error terms in supervector g but also a set of residual ordinates to be included in the supervector h for the assurance of "closure"! The errors, e_i', are actually <u>known</u> in this case. The observed y_i do <u>not</u> enter as such into h. The increments a" and b" are readily computable, together with the a' and b' corresponding to the guessing round, in a single pass.

A more striking pattern may be obtained by a rearrangement of terms in the design supermatrix and in the two supervectors, as in Figure 1.

$$\left[\begin{array}{c|c}
\begin{array}{c|c} 0 & \begin{array}{c} 1 \cdots \\ x_1 \cdots \end{array} \\ \hline \begin{array}{cc} 1 & x_1 \\ \vdots & \vdots \end{array} & \begin{array}{c} 1 \quad \\ \ddots \end{array} \end{array} & 0 \\ \hline
0 & \begin{array}{c|c} 0 & \begin{array}{c} 1 \cdots \\ x_1 \cdots \end{array} \\ \hline \begin{array}{cc} 1 & x_1 \\ \vdots & \vdots \end{array} & \begin{array}{c} 1 \quad \\ \ddots \end{array} \end{array}
\end{array}\right] \cdot \left[\begin{array}{c} a' \\ b' \\ e_1' \\ \vdots \\ a'' \\ b'' \\ e_1'' \\ \vdots \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ y' \\ \vdots \\ 0 \\ 0 \\ e_1' \\ \vdots \end{array}\right]$$

Fig. 1

Notice that the leftmost expression is a "supermatrix of supermatrices," replicating the F required for classical unweighted least squares. This form is of interest for the next example and for our brief remarks on multi-equation models.

My second example shows the instant supermatrix setup for "stepwise" regression, a concept due to Goldberger and others [7], [8]. A linear relationship is estimated progressively for one set of error-affected y's from two or more sets of x's. The procedure is usually presented for only two sets of x's but additional sets can be accommodated easily within the supermatrix framework.

For two steps, the supermatrix system appears as in Figure 2.

$$\left[\begin{array}{c|c}
\begin{array}{c|c} 0 & \begin{array}{c} 1 \cdots \\ x_1 \cdots \end{array} \\ \hline \begin{array}{cc} 1 & x_1 \\ \vdots & \vdots \end{array} & \begin{array}{c} 1 \quad \\ \ddots \end{array} \end{array} & \\ \hline
 & \begin{array}{c|c} 0 & \begin{array}{c} 1 \cdots \\ x_1^* \cdots \end{array} \\ \hline \begin{array}{cc} 1 & x_1^* \\ \vdots & \vdots \end{array} & \begin{array}{c} 1 \quad \\ \ddots \end{array} \end{array}
\end{array}\right] \cdot \left[\begin{array}{c} a \\ b \\ e_1 \\ \vdots \\ a^* \\ b^* \\ e_1^* \\ \vdots \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ y_1 \\ \vdots \\ 0 \\ 0 \\ e_1 \\ \vdots \end{array}\right]$$

Fig. 2

Here, the y_i are observed ordinates and the e_i are residuals -- the differences between the observed ordinates and the computed estimates derived from the first regression step. The second step involves, in effect, the regression of the first residual ordinates on the x_i^*, a second set of observations. Since the e_i are not known in advance in this case (i.e., they result from computation, not from guessing), the solution of the supermatrix system cannot be accomplished in one pass.

Multi-equation Models

More ambitious applications of instant supermatrix design are suggested by a glance at texts exhibiting: the generalized least-squares method of Aitken, the three-stage approach to least-squares estimation of Zellner and Theil, and the discrete-time linear engineering formulations that accommodate "transitions" with "noise" from an initial state to subsequent ones [9], [7], [8]. In these instances, which involve the simultaneous estimation of all the parameters in a multi-equation model, the supermatrix setup is introducible in the last phases. The column of disturbances is absorbed into the supermatrix equation $Fg = h$. The elements of F and h would commonly have undergone substantial processing rather than represent raw data.

The two examples discussed in the preceding section also suggest analogues in which multiple equations are treated simultaneously. In these analogues, the errors arising in the first round or step are not introducible as data in successive ones.

Another possible application of the supermatrix approach is to instances in which a full linear system is to be estimated from observations made on fragmentary linear "shards." All the y_i may, for example, represent intelligence quotients or performance scores of a given kind. Associated with a subset of the y_i are ratings in one or more "predictive" tests; associated with other subsets of y (they may differ in number of elements) are ratings in still other tests or in various combinations of tests. A supermatrix system can be designed immediately for the derivation of the single least-squares equation accommodating all of the data simultaneously. This equation may be useful in its own right or for comparison with the regression estimates corresponding to y_i subsets.

Finally, it appears feasible to estimate by supermatrix methods the unknowns of a multi-equation model that consists of (a) accounting identities presumed to be exact and (b) inexact regression relationships. The latter equations may, of course, include highly processed "observations." They may express each of several endogenous variables in terms of exogenous variables, other endogenous variables, or both [10].

REFERENCES

[1] Siegel, I.H., "The Supermatrix Approach to Least-Squares Adjustment," The American Statistician, December 1968, 22-23.

[2] ____________, "Instant Matrix Design: Straight Line With Errors in Both Variables," Proceedings of the Social Statistics Section, American Statistical Association, 1968, 247.

[3] ____________, "Difference Between Paasche and Laspeyres Index-Number Formulas," Journal of the American Statistical Association, September 1941, 343-350.

[4] ____________, "Further Notes on Index-Number Differences," ibid., December 1941, 519-524.

[5] ____________, "Simplified Least-Squares Approach for an Age of Computers," Proceedings of the Social Statistics Section, American Statistical Association, 1966, 398-400.

[6] ____________, "From Identities to Normal Equations: An Easy Approach to Least-Squares," Proceedings of the Social Statistics Section, American Statistical Association, 1967, 354-356.

[7] Goldberger, A.S., Econometric Theory. New York: Wiley, 1964.

[8] Deutsch, R., Estimation Theory. Englewood Cliffs: Prentice-Hall, 1965.

[9] Christ, C.F., Econometric Models and Methods. New York: Wiley, 1966.

[10] Fox, K.A., Sengupta, J.K. and Thorbecke, E., The Theory of Quantitative Economic Policy. Chicago: Rand McNally, 1966.

POSTSCRIPT

Since it is my nature to abhor a vacuum, I shall use some of the abundant space that remains on this page for a few additional remarks on sequential estimation.

First, the example I used for stepwise regression was deliberately patterned on the example preceding it and hence may not at first resemble the illustrations usually given. The supermatrix approach is adaptable, of course, to instances in which the second set of observed independent variables does not, say, include a column of 1's.

I wish to point out also that the classical cases of least-squares adjustment may readily be restated to involve sequential procedures. Indeed, the example that starts with a guess suggests one of the many routes that might be taken toward enlargement of the standard design supermatrices without alteration of the final results.

Another route is suggested by a theorem published by Jacobi in 1841 (it is shown in Whittaker and Robinson, Calculus of Observations). This theorem, which may be translated into supermatrix form at once, says that the unweighted least-squares straight line is equivalent to a certain weighted average of the ${}^{n}C_{2}$ lines derivable from n given values of x and y.

Still another way is to introduce pairs of mathematically redundant unit submatrices into the design supermatrix for the unweighted case. These diagonal matrices are like "idlers" or "gears." The first pair would link the residuals to a set of Lagrangian coefficients; the second pair would link these coefficients to another set of multipliers; and so on.

The weighted case is more interesting. Here, too, pairs of extra diagonal matrices may be inserted into the design supermatrix, one diagonal showing the weights and the other showing 1's. If two or more tiers are introduced, the weights included in each may be raised to a fractional power; but the introduction of tiers must cease when the sum of the fractional powers reaches unity.

I conclude with the comment that an unweighted least-squares system may also be translated into a larger weighted one in which Q does not represent the transpose of P. In this setup, a tier containing a compensatory matrix of pseudo-weights and a diagonal unit matrix is included in the design supermatrix. The matrix of pseudo-weights, however, is not diagonal and not unique, and it requires some computation.