FROM IDENTITIES TO NORMAL EQUATIONS: AN EASY APPROACH TO LEAST SQUARES

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

A paper I presented six years ago at a meeting sponsored by the New York Academy of Sciences sought "to exhibit the classical method of least squares without recourse to the conventional summary normal equations" [1]. It emphasized two procedures that explicitly introduce the n unknown residuals into the n observation equations. According to one of these procedures, the rectangular n x m observation matrix is expanded by a simple rule into a much larger invertible square matrix. A supermatrix system equivalent to conventional normal equations is immediately obtained, and it becomes possible to delegate all arithmetic processing to computer specialists. The second procedure is to set up "normal identities" in an obvious manner and then to eliminate certain summary terms that contain residuals and that an adjustment process might reasonably be expected to reduce to zero. The result is a conventional system of normal equations or something very similar to it -- plus some footnote information on the minimum value of the sum of squared residuals. Having already explored my supermatrix approach in some degree [2], I return at this time to my normal-identity approach.

Let us consider the familiar least-squares case of fitting the line $y = a + bx$ with only the y_i subject to error. Here, the number of unknown constants is $m = 2$ and the number of observations is n. For simplicity, we assume that the observations have equal weight.

Actually, each observation equation is an identity containing an additional term, the variable residual r_i. Explicitly introducing this term, we write $y \equiv a + bx_i + r_i$. Since there are n observations, we have n residuals.

We now proceed to develop normal identities from the observation identities. We multiply each observation identity by the coefficient of a (i.e., by 1) and sum to obtain the first normal identity. Next, we multiply each observation identity by the coefficient of b (i.e., by x_i) and sum to obtain the second normal identity. Finally, we multiply each observation identity by r_i and sum to obtain the third normal identity.

The resulting system looks like this:

$$\begin{aligned} \Sigma y &\equiv na + b\Sigma x + \Sigma r \\ \Sigma xy &\equiv a\Sigma x + b\Sigma x^2 + \Sigma xr \\ \Sigma ry &\equiv a\Sigma r + b\Sigma xr + \Sigma r^2 \quad . \end{aligned}$$

The determinant of the right-hand side is axisymmetric. The unknowns include a, b, and the terms in r.

How may this system be solved? One obvious scheme is to assume $\Sigma r = 0$ and $\Sigma xr = 0$ and restrict attention to the first two lines, since only two unknowns, a and b, really need to be found. The third line is redundant for solution, but it states a consistency condition, a necessary implication of the adjustment process.

Our two assumptions, it might be noted, tell us exactly the same thing that the two normal equations do. The assumptions tell us with reference to residuals what the normal equations tell us with reference to the unknowns of primary interest.

The third line, which simply states that $\Sigma ry = \Sigma r^2$, is a mathematical footnote. It tells us what our assumptions mean with respect to the minimum value of the sum of squared residuals. Indeed, it tells us what we mean by "least squares" in this case. If this information is deemed exceptionable, if this implication offends common sense or some prior principle, the process should be reconsidered or different sums should be eliminated.

The normal identity approach makes it clear that the simple condition $\Sigma r = 0$ arises in least-squares adjustment only when a free constant (such as a) exists. For any other linear model (e.g., the one used by Gauss in his Theoria Motus to illustrate the adjustment process), the residuals always appear with unequal weights in the normal identities. Hence, in such instances, only sums of weighted residuals may be set equal to zero -- even though the observations themselves are unweighted. Furthermore, in the absence of a free constant, a fitted line cannot pass through the point of unweighted means of the observed values.

*The author's views do not necessarily represent positions of The W. E. Upjohn Institute for Employment Research.

Much more complicated cases of curve-fitting may also be investigated with the aid of normal identities. Let us consider briefly the case in which both variables, the y_i and the x_i, are subject to error. This problem has attracted the attention of many statisticians over a long span of time. A paper published in 1959 provides an impressive bibliography -- 53 items appropriately extending from A (Adcock, 1878) to Z (Zucker, 1947). This bibliography, however, must still be far from exhaustive for the period covered [3].

We again start with $y_i = a + b_i$ and do not weight the observations. Inserting s_i for the residual corresponding to y_i and inserting t_i for the residual corresponding to x_i, we obtain $y_i + s_i \equiv a + b(x_i + t_i) \equiv a + bx_i + bt_i$ as the prototype observation identity.

This time, we obtain five normal identities as we subject the observation identities to multiplication, in turn, by 1, x_i, y_i, s_i, and t_i and sum the results. The whole system looks like this:

$$\begin{aligned}
\Sigma y + \Sigma s &\equiv na + b\Sigma x + b\Sigma t \\
\Sigma xy + \Sigma xs &\equiv a\Sigma x + b\Sigma x^2 + b\Sigma xt \\
\Sigma y^2 + \Sigma ys &\equiv a\Sigma y + b\Sigma yx + b\Sigma yt \\
\Sigma sy + \Sigma s^2 &\equiv a\Sigma s + b\Sigma sx + b\Sigma st \\
\Sigma ty + \Sigma ts &\equiv a\Sigma t + b\Sigma tx + b\Sigma t^2 \quad .
\end{aligned}$$

What assumptions could we reasonably make in order to solve this system? The sums of unweighted residuals presumably should be made equal to zero: $\Sigma s = \Sigma t = 0$. We may also suppose the independence of: (1) the two sets of residuals and (2) the observed values of one variable and the residuals associated with the other. Thus, we also assume $\Sigma xs = \Sigma yt = \Sigma st = 0$.

After making these simplifications, we are left with this pattern of equations:

$$\begin{aligned}
\Sigma y &= na + b\Sigma x \\
\Sigma xy &= a\Sigma x + b\Sigma x^2 + b\Sigma xt \\
\Sigma y^2 + \Sigma ys &= a\Sigma y + b\Sigma yx \\
\Sigma sy + \Sigma s^2 &= 0 \\
0 &= b\Sigma xt + b\Sigma t^2
\end{aligned}$$

The fourth and fifth lines tell us what the adjustment process, if it is accepted, means for the values of Σs^2 and Σt^2. Unless some relationship between these sums of squared residuals is posited, we cannot solve the system, since we still have too many unknowns.

If we assume that $\Sigma s^2 = k\Sigma t^2$, we arrive at the quadratic equation that is often shown in the literature as the key to complete solution. This additional assumption entails $\Sigma sy = \Sigma xt$. Substituting in the first three summary equations and simplifying, we obtain this expression:

$$b^2 M_{xy} + b(kM_x - M_y) - kM_{xy} = 0,$$

where M_{xy}, M_x, and M_y refer to the moments appearing in the familiar formulas for the correlation coefficient and for the variances of x and y. The quadratic expression may be solved readily for b; and, giving different values to the parameter k, we obtain various special cases of interest [4]. If we set k = 1, we have the well-known case of orthogonal regression, which is usually discussed in terms of polar coordinates and solved with respect to the tangent of an angle [5].

Obviously, the normal-identity approach is versatile, and it should have both pedagogic and theoretical interest. It seems to treat the adjustment process as a deterministic, rather than as a probabilistic, one; but a transition from "mathematics" to "statistics" is made via the assumptions. Since the assumptions refer to summary terms involving residuals, a range of choices may be explored advantageously when more than one variable is subject to error. The "errors-in-variables" model, moreover, is nowadays contrasted in econometrics with the "errors-in-equations" model, and the normal-identity tool ought to be useful in the investigations pursued [6].

REFERENCES

[1] I. H. Siegel, "Least Squares 'Without Normal Equations'," _Transactions of the New York Academy of Sciences_, February 1962, pp. 362-371.

[2] See these papers by I. H. Siegel: "Simplified Least-Squares Approach for an Age of Computers," 1966 Social Statistics Section Proceedings of the American Statistical Association, pp. 398-400; "Deferment of Computation in the Method of Least Squares," Mathematics of Computation, April 1965, pp. 329-331; and "Least Squares with Less Effort," 1964 Business and Economic Statistics Section Proceedings of the American Statistical Association, pp. 284-285.

[3] The bibliography appears at the end of a paper by Albert Madansky, "The Fitting of Straight Lines When Both Variables Are Subject to Error," Journal of the American Statistical Association, March 1959, pp. 172-205. It is called "a complete survey of the literature" by E. Malinvaud (Statistical Methods of Econometrics, Rand McNally, Chicago, 1966, p. 362), but many relevant items are not included, such as: Mansfield Merriman, A Text-Book in the Method of Least Squares, 8th ed., Wiley, New York, 1911, pp. 127-128, 216-217; a paper by Merriman in Report of the Superintendent of the U.S. Coast and Geodetic Survey, Fiscal Year 1890, pp. 687-690; Henry Schultz's "Comments" on R. G. D. Allen's paper in Economica, May 1939, pp. 202-204; L. J. Reed, "Fitting Straight Lines," Metron, No. 3, 1921, pp. 54ff.; and P. J. Dwyer and M. S. Macphail, "Symbolic Matrix Derivatives," Annals of Mathematical Statistics, December 1948, especially pp. 531-532.

[4] See, for example, J. Johnston, Econometric Methods, McGraw-Hill, New York, 1960, pp. 150-162, as well as Allen's Economica paper, pp. 191-201.

[5] See, for example, Madansky, loc. cit., p. 202; Y. V. Linnik, Method of Least Squares and Principles of the Theory of Observation, Pergamon, New York, 1961, pp. 11-13, 320-322; and Reed, loc. cit.

[6] Malinvaud, op. cit., Chapter 10; and Johnston, op. cit., Chapter 6.