Linear Least Squares for Correlated Data

Edwin B. Dean
NASA Langley Research Center
Mail Stop 430
Hampton VA 23665
804 865 4894

Presented at the
Tenth Annual International Conference of the
International Society of Parametric Analysts
July 25-27, 1988
Brighton, England

Introduction

Throughout the literature authors have consistently discussed the suspicion that regression results were less than satisfactory when the independent variables were correlated. Camm, Gulledge, and Womer [1], and Womer and Marcotte [2] provide excellent applied examples of these concerns. Many authors have obtained partial solutions for this problem as discussed by Womer and Marcotte [2] and Wonnacott and Wonnacott [3], which result in generalized least squares algorithms to solve restrictive cases.

This paper presents a simple but relatively general multivariate method for obtaining linear least squares coefficients which are free of the statistical distortion created by correlated independent variables.

The Method

The multivariate linear least squares problem is stated as

minimize $\sum_k \varepsilon_k^2$

with respect to $\bar{a}$ and ε_k

subject to $\varepsilon_k = y_k - \bar{x}_k^T \bar{a}$ for k = 1 ... n,

where y_k is the kth measurement of the dependent variable, $\bar{a}$ is the vector of m desired parameters a_i, $\bar{x}_k$ is the kth measurement of the vector

$$\bar{x} = \begin{pmatrix} x_0 \\ \dots \\ x_m \end{pmatrix}$$

of the independent variables x_i with $x_{0k} = 1$,

$$\bar{x}_k^T = \begin{pmatrix} x_{k0} \dots x_{km} \end{pmatrix}$$

is the transpose of $\bar{x}_k$, and ε_k is the kth value of the error ε.

The problem is called linear since the model

$$\varepsilon = y - \bar{x}^T \bar{a}$$

is linear in the parameter vector $\bar{a}$. Setting $x_{0k} = 1$ provides a constant term a_0 in the model.

Note that the linear least squares problem exists as an optimization problem independent of statistics. It also has a statistical interpretation when either the independent variables x_k or the error ε_k are considered as random variables. In this context the error is often called the residual.

Traditional econometric methods apply statistics to the analysis of the residuals. The foundations for this paper are based on a second approach found in Fukunaga [4]. He assumes that the data observations themselves are random variables. The independent variables in aerospace parametrics are typically observations of

possible projects as measured by system requirement, system performance, system design, or engineering process metrics. Since the finished project is only one of the many possible projects which could have been completed [5,6], the data observations are themselves drawn from the distributions of the measures of the possible projects. Thus Fukunaga's paradigm applies.

In classical linear least squares as described by Draper and Smith [7] we let $\bar{y}$ be the vector of the n dependent variable observations y_k, $\bar{\varepsilon}$ be the vector of the n errors ε_k, and

$$X = \begin{pmatrix} \bar{x}_1^T \\ . \\ \bar{x}_n^T \end{pmatrix}$$

be the vector of n independent variable observation transpose vectors $\bar{x}_k^T$, the problem may be restated in vector form as

$$\text{minimize} \qquad \bar{\varepsilon}^T \bar{\varepsilon} = (\bar{y} - X \bar{a})^T (\bar{y} - X \bar{a}).$$

Setting the partial derivatives with respect to $\bar{a}$ to zero we have the normal equations

$$X^T X \tilde{a} = X^T \bar{y}$$

where $\tilde{a}$ is the estimate of $\bar{a}$.

Noting that $X^T X$ is a m by m matrix, which with at least m distinct data points should have full rank, we assume that $X^T X$ may be inverted to obtain

$$\tilde{a} = (X^T X)^{-1} X^T \bar{y}.$$

It is important to note that, when the x_k are not considered random variables, the components a_i are linear functions of the random dependent observations y_k. Under this interpretation, $\tilde{a}$ provides a statistically unbiased estimate of $\bar{a}$ which has the minimum

variance of all linear unbiased estimators of $\bar{a}$, irrespective of distribution properties of the errors.

However, when the $\bar{x}$ are considered random variables, the components a_i are no longer linear functions of random variables. Following Papoulis [8] we use the property of conditional probability distributions that for independent random variables z_i

$$E\{g(\bar{z}) \mid \bar{z}\} = g(\bar{z})$$

to obtain

$$E\{\tilde{a} \mid \bar{x}, \bar{\varepsilon}\} = E\{(X^T X)^{-1} X^T \bar{y} \mid \bar{x}, \bar{\varepsilon}\} = (X^T X)^{-1} X^T \bar{y} = \bar{a}$$

since $\bar{y}$ is a function of $\bar{x}$ and $\bar{\varepsilon}$ Thus if the random variables $\bar{x}$ and the errors $\bar{\varepsilon}$ are statistically independent $\tilde{a}$ is still an unbiased estimate of a $^{-}$.

Unfortunately, statistical independence is a strong condition in practice since many of the metrics used in cost analysis are highly correlated. Thus it is desired to obtain coefficients a_i which are free of the statistical distortion caused by performing a linear least squares fit when there is correlation between the x_i.

A general method follows through which the coefficients a_i may be found by first transforming the x_i into a new set of random variables z_i, performing the least squares fit on the z_i, and then transforming the coefficients b_i found by the least squares process to obtain the desired coefficients a_i.

Following Fukunaga [4] an uncorrelated and if normally distributed a statistically independent set of random variables may be obtained using the eigenvalues and eigenvectors of the covariance matrix

$$H_x = E\{(\bar{x}-\mu_x)(x^{-}-\mu_x)^T\}$$

$$= E\{(x_i-\mu_{x_i})(x_j-\mu_{x_j})\}$$

$$= E \begin{pmatrix} (x_1-\mu_{x_1})(x_1-\mu_{x_1}) & \dots & (x_1-\mu_{x_1})(x_m-\mu_{x_m}) \\ \dots & \dots & \dots \\ (x_m-\mu_{x_m})(x_1-\mu_{x_1}) & \dots & (x_m-\mu_{x_m})(x_m-\mu_{x_m}) \end{pmatrix}$$

The eigenvectors of the covariance matrix H_x are directions of statistically independent random variables and the eigenvalues are the associated variances. The eigenvectors $\bar{\omega}_i$ are called principal components in statistical jargon. Figure 6.1 illustrates the principal component axes formed in two dimensions by correlated variables x_1 and x_2. The vector $\bar{\omega}_1$ indicates the direction of maximum variance ${\sigma_1}^2$. The vector $\bar{\omega}_2$ indicates the direction of minimum variance ${\sigma_2}^2$. Since the principal component vectors have unit magnitude by definition, the standard deviations may be represented as standard deviation vectors with origin translated to sample mean. Thus

$\bar{\sigma}_1 = \sigma_1 \bar{\omega}_1$ and

$\bar{\sigma}_2 = \sigma_2 \bar{\omega}_2$.

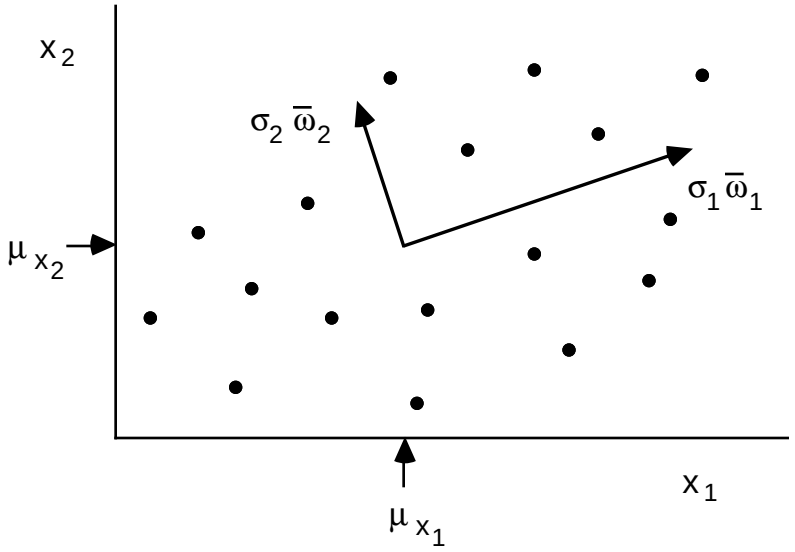

Figure 1
Principal Component Axes and Standard Deviation Vectors

When the distributions of the x_i are normal, the standard deviation vectors multiplied by the same constant ß form the axes of an

ellipse representing a ß standard deviation equal probability contour.

Let $\lambda_{ii} = \sigma_i^2$ be the ith eigenvalue of H_x and $\bar{\omega}_i$ be the eigenvector of H_x associated with λ_{ii}. Then there exists a diagonal matrix Λ with the λ_{ii} as diagonal components, a matrix Ω with columns $\bar{\omega}_i$, and the transpose Ω^T of Ω such that

$$H_x \Omega = \Omega \Lambda \quad \text{and} \quad \Omega \Omega^T = \Omega^T \Omega = I.$$

Thus H_x may be expressed as

$$H_x = \Omega \Lambda \Omega^T.$$

Form new random variables z_i by the rule

$$\bar{z} = A \bar{x}.$$

The covariance matrix H_z of the z_i has the form

$$H_z = A H_x A^T.$$

The square root matrix $M^{1/2}$ is defined by the property

$$M^{1/2} M^{1/2} = M$$

with inverse $M^{-1/2}$.

Letting

$$A = \Lambda^{-1/2} \Omega^T$$

then

$$H_z = \Lambda^{-1/2} \Omega^T H_x \Omega \Lambda^{-1/2} = \Lambda^{-1/2} \Omega^T \Omega \Lambda \Omega^T \Omega \Lambda^{-1/2} = I.$$

The transform Ω^T rotates the axes to coincide with the direction of the principal components. The transform $\Lambda^{-1/2}$ divides the new basis

vector magnitudes σ_i by σ_i to provide unit basis vector magnitudes coinciding exactly with the principal components.

Note that the standard deviation vectors of Figure 1 have equal magnitude when the variances are equal. Thus for a normal joint uncorrelated distribution the equal probability contours form a circle. This transformation is called a whitening transform since the transformed variables are representative of white (totally uncorrelated normal) noise.

The random variables z_i resulting from the transform are not only uncorrelated since the covariance matrix is diagonal, but are also statistically independent if the x_i are normally distributed.

The problem to be solved now is

minimize $\sum_k \varepsilon_k^2$

with respect to $\bar{b}$ and the ε_k

subject to $\varepsilon_k = y_k - \bar{z}_k^T \bar{b}$ for k = 1 ... n.

The composite transformation provides random variables

$$\bar{z} = \Lambda^{-1/2} \Omega^T \bar{x}$$

and restates the problem as

minimize $\sum_k \varepsilon_k^2$

with respect to $\bar{b}$ and the ε_k

subject to: $y_k - (\Lambda^{-1/2} \Omega^T \bar{x}_k)^T \bar{b} = \varepsilon_k$ for k = 1 ... n,

By setting

$$\bar{a} = \Omega \Lambda^{-1/2} \bar{b}$$

we have a solution in terms of the untransformed data.

Noting that

$$Z = X\,\Omega\,\Lambda^{-1/2}$$

we have

$$\tilde{a} = \Omega\,\Lambda^{-1/2}\,\tilde{b}$$

$$= \Omega\,\Lambda^{-1/2}\,(Z^T Z)^{-1}\,Z^T\,\bar{y}$$

$$= \Omega\,\Lambda^{-1/2}\,((\Lambda^{-1/2}\,\Omega^T X^T)\,(X\,\Omega\,\Lambda^{-1/2}))^{-1}\,(\Lambda^{-1/2}\,\Omega^T X^T)\quad \bar{y}$$

$$\tilde{a} = \Omega\,\Lambda^{-1/2}\,(\Lambda^{1/2}\;\Omega^T (X^T X)^{-1}\,\Omega\,\Lambda^{1/2})\,\Lambda^{-1/2}\,\Omega^T X^T\;\bar{y}$$

$$= \Omega\,(\Lambda^{-1/2}\,\Lambda^{1/2})\,\Omega^T (X^T X)^{-1}\,\Omega\,(\Lambda^{1/2}\;\Lambda^{-1/2})\,\Omega^T X^T\;\bar{y}$$

$$= (\Omega\,\Omega^T)\,(X^T X)^{-1}\,(\Omega\,\Omega^T)\,X^T\;\bar{y}$$

$$= (X^T X)^{-1}\,X^T\,\bar{y}.$$

Thus the parameter $\bar{a}$ is a solution to the original problem.

A simple computer algorithm for implementing this method is as follows.

Input the data vectors $\bar{x}_k$. Calculate the covariance matrix H_x. Find the eigenvalues and eigenvectors of H_x. Generate the transform matrix

$$A = \Lambda^{-1/2}\,\Omega^T.$$

Transform each data vector $\bar{x}_k$ by A to obtain $\bar{z}_k$. Apply conventional linear least squares to the data $(y_k, \bar{z}_k)$ to obtain the coefficients b_i. Use

$$\tilde{a} = A^T \tilde{b} = \Omega \Lambda^{-1/2} \tilde{b}$$

to transform the coefficients b_i by A^T to obtain the desired coefficients a_i.

Note that $\tilde{a}$ is an unbiased estimate of $\bar{a}$ since $\tilde{a}$ is a linear transform of $\tilde{b}$ which is an unbiased estimate of $\bar{b}$.

Wilkinson [9] demonstrates many methods for finding eigenvalues and eigenvectors. An efficient algorithm based on Hildebrand [10] coded in Pascal may be found in Flanders [11]. Multivariate least squares algorithms in Pascal may be found in Miller [12] which may be combined with the eigenvalues/eigenvector procedures to generate your own custom software for implementing this technique.

Conclusion

Assuming that the independent variables x_i are random variables, which is representative of most parametric cost applications, the linear least squares coefficients a_i obtained by the method in this paper are free of the statistical distortion from a linear least squares fit over the correlated independent variables x_i.

The result is that the analyst can use this technique without concern for colinearity or correlation of the independent variables.

Although, only one transformation has been discussed in this paper, there exists a class of transformations which will yield the same freedom from statistical colinearity distortion.

Finally the analyst should also note that they may perform any least squares technique they normally use, such as stepwise regression, to obtain the b_i with the data $(y_k, \bar{z}_k)$. The a_i are still found from the transformation

$$\bar{a} = \Omega \Lambda^{-1/2} \bar{b}.$$

References

[1] Camm, J. D., Gulledge, T. R., and Womer, N. K., "Production Rate and Contractor Behavior," *The Journal of Cost Analysis*, Volume 5, Number 1, Summer 1987.

[2] Womer, N. K., and Marcotte, R. C. J., "Airframe Cost Estimation Using and Error Components Model", *The Journal of Cost Analysis*, Volume 3, Spring 1983.

[3] Wonnacott, R. J., and Wonnacott, T. H., Econometrics, John Wiley and Sons, New York NY, 1970.

[4] Fukunaga, K., Introduction to Statistical Pattern Recognition, Academic Press, New York NY, 1972.

[5] Dean, E. B., Wood, D. A., Moore, A. A., and Bogart, E. H., "Cost Risk Analysis Based on Perception of the Engineering Process," presented at the Eighth Annual Conference of the International Society of Parametric Analysts, Kansas City MO, May 12-16, 1986.

[6] Mazzini, R. A., "The Evolutionary Theory of Cost Management," *Transactions of the 30th Annual Meeting of the American Association of Cost Engineers*, Chicago IL, 1986.

[7] Draper, N. R., and Smith, H., Applied Regression Analysis, 2nd Ed., John Wiley and Sons, New York NY, 1981.

[8] Papoulis, A., Probability, Random Variables, and Stochastic Processes, McGraw-Hill Book Company, New York NY, 1965.

[9] Wilkinson, J. H., The Algebraic Eigenvalue Problem, Clarendon Press, Oxford, England, 1965.

[1 0] Hildebrand, F. B., Introduction to Numerical Analysis, McGraw-Hill Book Company, 1956.

[1 1] Flanders, H., Scientific Pascal, Reston Publishing Company, Reston VA, 1984.

[1 2] Miller, A. R., Turbo Pascal Programs for Scientists & Engineers, SYBEX®, San Francisco CA, 1987.