On the Calculation of Minimum Variance Estimators for Unobservable Dependent Variables

by

I.D. Coope
Department of Mathematics and Statistics,
University of Canterbury, Christchurch, New Zealand

No. 99 August, 1993

Abstract – The determination of minimum variance estimators in an unuaual context is considered. The problem arises from an attempt to perform a regression with an unobservable dependent variable. The required minimum variance estimator is shown to satisfy a linear system of equations where the coefficient matrix has a simple structure. Uniqueness of the estimator is established by determining necessary and sufficient conditions on the data which guarantee positive definiteness of this coefficient matrix. Numerical aspects of the method of computation are also briefly explored.

Keywords: Minimum variance estimators, regression, unobservable dependency, Sylvester's law of inertia.

Accepted for publication in Computational Optimization & Applications.

1 Introduction

In this article the problem under consideration is that of finding optimum values for the variables $x_1, x_2, \ldots, x_k$, to minimize the function $S(X)$, where

$$S(X) = \text{trace}[(R + PX)K(R + PX)^T], \tag{1.1}$$

and X denotes the diagonal matrix with non-zero elements $X_{jj} = x_j$. The problem arises from an attempt to perform a regression with an *unobservable* dependent variable. Suppose that k independent *estimates* of an unobservable data series are given by $\mathbf{r}_j + \mathbf{p}_j x_j$, where the n-dimensional vectors, $\mathbf{r}_j$ and $\mathbf{p}_j$, $j = 1, 2, \ldots, k$, are columns of the $n \times k$ data matrices R and P respectively and $x_1, \ldots, x_k$ are scalars to be determined. The sum of squared deviations of these estimates about observed values cannot be minimized because there are no observations. If instead the variance (or more correctly the trace of the sample variance/covariance matrix) of the k estimates is minimized then the resulting function to be minimized can be defined by (1.1), where K is the (positive semi-definite) matrix

$$K = kI - \text{ones}(k, k), \tag{1.2}$$

and $\text{ones}(k, k)$ denotes the $k \times k$ matrix with all its elements unity.

The statistical merits of the approach outlined above are not considered here but the linear algebra underlying the minimization problem, the conditions under which a unique solution can be expected and how this might be calculated effectively, are investigated.

First it is noted that $S(X)$ is a quadratic function of the parameters $x_1, x_2, \ldots, x_k$, and hence a necessary condition for the variance to be minimized is that $\partial S/\partial x_j = 0$, $j = 1, 2, \ldots, k$. Differentiating (1.1) gives (see, for example [2]).

$$\frac{\partial S}{\partial X_{ij}} = \left[2P^T RK + 2P^T PXK\right]_{ij}. \tag{1.3}$$

Because X is diagonal, only the derivatives where $i = j$ are of interest. Therefore, the required stationary point satisfies

$$\text{diag}(P^T RK) = -\,\text{diag}(P^T PXK), \tag{1.4}$$

which is a system of k linear equations in the components of the vector $\mathbf{x} = X\mathbf{e}$, where $\mathbf{e} = (1, 1, \ldots, 1)^T$ is the k-dimensional vector whose components are all one.

Exploiting the special structure (1.2) of K enables equation (1.4) to be reduced to the simple matrix/vector equation

$$\hat{A}\mathbf{x} = -\hat{B}\mathbf{e}, \tag{1.5}$$

where

$$A = P^T P, \qquad \hat{A} = k\,\mathrm{diag}(A) - A; \tag{1.6}$$

$$B = P^T R, \qquad \hat{B} = k\,\mathrm{diag}(B) - B. \tag{1.7}$$

2 Uniqueness of the Minimum Variance Estimator

Equation (1.5) will provide a unique solution for $\mathbf{x}$ provided that the matrix $\hat{A}$ is non-singular. This is easily established by making repeated use of the following property of a congruence transformation (see, for example, Strang[3]).

Sylvester's Law of Inertia. If A and C are $n \times n$ real matrices, with A symmetric and C non-singular, then the matrix $C^T AC$ has the same number of positive eigenvalues as A, the same number of negative eigenvalues, and the same number of zero eigenvalues.

Thus a congruence transformation preserves the signs of the eigenvalues.

Theorem 2.1. Let A be a real $k \times k$ symmetric positive semi-definite matrix. Then the matrix

$$\hat{A} = k\,\mathrm{diag}(A) - A, \tag{2.1}$$

is positive definite if, and only if, A has positive diagonal elements and $\mathrm{rank}(A) > 1$.

Proof: First note that if A has a zero diagonal element, then so does the matrix $\hat{A}$, which cannot therefore be positive definite.

Alternatively, if A has positive diagonal elements define

$$C = \text{diag}\left[\frac{1}{\sqrt{a_{11}}}, \ldots, \frac{1}{\sqrt{a_{kk}}}\right]$$

to be the diagonal matrix whose elements are the reciprocal square roots of the diagonal elements of A. Then it is sufficient to show that $C^T\hat{A}C$ has positive eigenvalues, because this matrix is clearly congruent to $\hat{A}$. But

$$C^T\hat{A}C = kI - C^TAC, \tag{2.2}$$

so if $\lambda_1, \lambda_2, \ldots, \lambda_k$, are the eigenvalues of C^TAC, then the corresponding eigenvalues of $C^T\hat{A}C$ are:

$$\mu_j = k - \lambda_j, \; j=1,2,\ldots,k. \tag{2.3}$$

However, C^TAC is congruent to A and therefore has non-negative eigenvalues. Also $[C^TAC]_{ii}=1, \; i=1,2,\ldots,k$, therefore

$$\text{trace}(C^TAC) = k = \sum_1^k \lambda_j.$$

Thus $0 \leq \lambda_j \leq k, \; j=1,2,\ldots,k$ and $\lambda_j = k$ for some j if and only if all other eigenvalues are zero, that is if, and only if, $\text{rank}(A) < 2$. Clearly, $\mu_j > 0, \; j=1,2,\ldots,k$, if, and only if, $\text{rank}(A) > 1$. □

Corollary 2.1. Problem (1.1) has a unique solution if, and only if, the matrix P has no zero columns, and at least two linearly independent columns.

Proof: This follows directly from the definition of A in equation (1.5) and Theorem (2.1). □

It should be clear that it would be highly unusual in any practical case for a unique solution to not exist. In fact, even in such (pathological) cases $\hat{A}$ is positive semi-definite and equation (2.1) is consistent (and therefore infinitely many solutions exist) but this is not considered further here since this situation should never occur in practice.

3 Numerical Considerations

The obvious approach to solving the system (2.1) is to first form the vector $\hat{B}\mathbf{e}$, without forming $\hat{B}$, using the special form (1.5) of $\hat{B}$. This requires only $k(2n+1)$ multiplications and additions. Next the matrix $\hat{A}$ is formed at a cost of k^2n multiplications and a similar number of additions. Then the Choleski factors of $\hat{A}$ would be calculated, requiring about $\frac{1}{6}k^3$ multiplications/additions and the solution of (2.1) completed by forward and back substitution with two triangular coefficient matrices requiring a further k^2 multiplications/additions. Thus the dominant cost by this method is

$$k^2n + \frac{1}{6}k^3$$

multiplications/additions, which is also the dominant cost of computing a solution to the more familiar linear least squares problem via "normal equations" of the form:

$$P^TP\mathbf{x} = P^T\mathbf{b}. \tag{3.1}$$

However, for greater numerical stability, forming the normal equations (3.1) in the linear least squares case is sometimes avoided because the matrix P^TP can be very ill-conditioned. More precisely,

$$\text{cond}(P^TP) = \text{cond}(P)^2,$$

where $\text{cond}(P)$ denotes the spectral condition number of the matrix P, which is defined as the ratio of the largest to the smallest singular value of the matrix P. Instead of forming the matrix P^TP, the QR-factorisation, or singular value decomposition of the matrix P is sometimes recommended for greater numerical stability even though each of these approaches requires more computational effort (see [1], for example).

Unfortunately the QR-factors of P are not helpful in solving the system of equations (2.1), so it is relevant to ask if this system can also be prone to the numerical instability that can effect the normal equations (3.1) when P is ill-conditioned. Fortunately, it turns out that the system (2.1) is much less susceptible to ill-conditioning as the following example illustrates.

Let P be the $n \times k$ segment of the Hilbert matrix whose elements are $P_{ij} = 1/(i+j-1)$, $i = 1, 2, \ldots, n; \quad j = 1, 2, \ldots, k$. This matrix is notoriously ill-conditioned, even for quite moderate values of n, k, as Table 3.1 shows. However, the condition number of the matrix $\hat{A}$, grows much more slowly with n and k as can also be seen from Table 3.1. In this case, the slow growth of condition number with n and k would allow quite large systems to be solved accurately without having to resort to orthogonal factorisations (either QR or SVD).

		Condition Number		
n	k	P	$P^T P$	$\hat{A}$
7	3	2.11e+02	4.46e+04	1.47e+02
7	5	1.66e+05	2.77e+10	2.31e+02
9	3	1.83e+02	3.37e+04	1.30e+02
9	5	1.01e+05	1.02e+10	1.97e+02
9	7	1.24e+08	1.57e+16	2.91e+02

Table 3.1. Spectral Condition Numbers.

Of course, it is still possible to construct a matrix, P, which gives rise to a poorly conditioned matrix $\hat{A}$; indeed as Theorem 2.1 indicates, this could be achieved most easily either by choosing P to be arbitrarily close to a rank-1 matrix or by making one column of P arbitrarily small in norm. Hopefully, such situations would be unlikely to occur in practical cases.

Acknowledgement

The problem discussed in this paper arose in an economic setting and was introduced to the author by K. Wignall[4] of the Economics Department, University of Canterbury.

References

[1] G.H. Golub, and C.F. van Loan, *Matrix Computations*, 2nd Edition, Johns Hopkins University Press, Baltimore, (Johns Hopkins series in the mathematical sciences, 3), 1989.

[2] J.R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*, Wiley, New York, 1988.

[3] G. Strang, *Linear Algebra and its Applications*, 3rd Edition, Harcourt, Brace, Jovanavich , San Diego, 1988.

[4] K. Wignall, Private Communication.

Department of Mathermatics,

University of Canterbury

Christchurch

NEW ZEALAND