1987:13

ISSN 0347-2809

PREPRINT

GENERALIZED LEAST SQUARES ESTIMATION
WITH ESTIMATED COVARIANCE MATRIX

VLADIMIR KHOKHLOV

STURE HOLM

NO 1987-13/ISSN 0347-2809

DEPARTMENT OF MATHEMATICS
CHALMERS UNIVERSITY OF TECHNOLOGY
THE UNIVERSITY OF GÖTEBORG

GENERALIZED LEAST SQUARES ESTIMATION WITH ESTIMATED COVARIANCE MATRIX

VLADIMIR KHOKHLOV

STURE HOLM

NO 1987-13/ISSN 0347-2809

1987

MATHEMATICAL STATISTICS
DEPARTMENT OF MATHEMATICS
CHALMERS UNIVERSITY OF TECHNOLOGY
AND THE UNIVERSITY OF GÖTEBORG
S-412 96 GÖTEBORG, SWEDEN

1. Introduction.

A classical problem in regression analysis, often met in applied work, is that of correlated errors. If the error covariance matrix is unknown the classical unweighted least squares method is commonly used. But deleting the fact that the observations are correlated leads to loss of efficiency. There are no suitable general methods available to get efficient estimates for this case.

The best considering of different problems in the estimation and inference on unknown parameters in a linear model under various assumptions on the error term is given by C. Radhakrishna Rao (1967). He examined four general cases of structures of the dispersion matrix Σ.

In the first case the matrix Σ is an unknown arbitrary positive definite matrix. In the three other cases he used the matrix Σ of the form

$$\Sigma = X\Gamma X^T + Z\theta Z + \sigma^2 I,$$

where matrices Γ, Z, θ are known or unknown. He showed the efficiency of the estimators when the dispersion matrix is given in every case.

In our work we try to receive improved estimators for the unknown parameters of the regression model by using estimated covariance matrix of another structure.

2. Estimates of parameters in the model

The model of regression analysis is

$$\mathcal{E}(Y) = XB \qquad (1)$$

where Y is a $(N\times k)$-matrix of k variables and a number of observations equals to N, X is a $(N\times n)$-matrix and B is a $(n\times k)$-matrix of regression coofficients.

Let the covariance matrix of errors be

$$\mathcal{E}\{(Y-XB)(Y-XB^T)\} = \Sigma$$

Suppose that Σ has full rank, that it is positive definite and unknown.

In classical scheme of estimation we have to find an estimate of B by a minimization, i.e.

$$\hat{B}_{LS} = \arg \inf_B (tr\{(Y-XB)^T(Y-XB)\}).$$

Let us use a decomposition of the matrix Σ of the form

$$\Sigma = Q^T LQ$$

where L is diagonal.

In other words

$$\Sigma^{-1} = G^T G$$

where

$$G = L^{-\frac{1}{2}}Q.$$

It is now possible to change variables by

$$W = GY$$

$$V = GX$$

Thus

$$\mathcal{E}\{(W-VB)(W-VB)^T\} = \sigma^2 I$$

for some $\sigma^2 > 0$.

Hence for a model

$$\mathcal{E}\{W\} = VB$$

the observations became independent. And an estimate of B can be found as

$$\hat{B} = \arg \inf_B (tr\{(W-VB)^T(W-VB)\}).$$

On the other side

$$\hat{B} = \arg \inf_B (tr\{(Y-XB)^T G^T G(Y-XB)\})$$

$$= \arg \inf_B (tr\{(Y-Y-XB)^T \Sigma^{-1}(Y-XB)\})$$

The main idea of our work is to consider the matrices Σ with decompositions

$$\Sigma = AA^T + D^2 \qquad (2)$$

where A is a $(N \times m)$-matrix, and D^2 is a diagonal $(N \times N)$-matrix. But expression (2) is a main concept in factor analysis. The matrix A is the matrix of factor loadings, and D^2 is a diagonal matrix of uniquenesses.

If we have some estimate of Σ, it will be possible to get a maximum likelihood estimate of A and accordingly of Σ.

There exists an iterative procedure defined more precisely than in Lawley and Maxwell (1971). They got a non-convergent procedure of maximum likelihood estimate. A convergent one was offered in Khokhlov (1983). The estimate of A can be obtained by the method of consistent iterations

$$\hat{A}^{iT1} = S_0 \hat{D}_i^{-2} \hat{A}_i (\hat{A}_i^T \hat{D}_i^{-2} \hat{A}_i + I)^{-1}, \quad i = 0,1,\ldots \qquad (3)$$

where $S_0 = [(N-1)/N]S$, and S is a sample estimate of Σ, and by Newton-Kantozovich method which converges more rapidly

$$\hat{A}_{i+1} = (I - \hat{\Sigma}_i^{-1} S_0)^{-1} (S_0 \hat{\Sigma}_i^{-1} - \hat{\Sigma}_i^{-1} S_0) \hat{A}_i, \quad i = 0,1,\ldots \qquad (4)$$

where $\hat{\Sigma}_i = \hat{A}_i \hat{A}_i^T + \hat{D}_i^2$

If we get a simple estimate S by

$$S = (Y - X\hat{B}_{LS})(Y - X\hat{B}_{LS})^T$$

than S has not full rank. A full rank matrix S can be obtained in two ways. The fast way is by using Tikhonov regularizator r (Tikhonov 1978)

$$S = (Y - X\hat{B}_{LS})(Y - X\hat{B}_{LS})^T + rI$$

where r can be received by different methods. One of them is

$$r = \lambda_{max}\{(Y - X\hat{B}_{Ls})(Y - X\hat{B}_{Ls})\}$$

$\lambda_{max}(R)$ is the largest eigenvalue of a matrix R.

The second way is to substitute a matrix L in the decomposition

$$EE^T = Q^T LQ$$

by a diagonal matrix J which contains elements of a diagonal of the matrix EE^T such that

$$j_{11} \geq j_{22} \geq \ldots \geq j_{NN} > 0.$$

The matrix E is

$$E = Y - X\hat{B}_{LS}.$$

So in this case we have the matrix

$$S = Q^T JQ \tag{5}$$

of full rank. Morover

$$\sum_i J_{ii} = tr(S) = tr(EE^T)$$

which corresponds to one of the most important properties of eigenvalues

$$\sum_i \lambda_i(R) = tr(R) \ .$$

And because the determinant of EE^T equals zero we can consider that

$$det(S) = \Pi j_{ii}$$

which is another property of eigenvalues.

The matrix S obtained in the first way does not have these properties.

Hence the estimate of the matrix of errors' covariances is the matrix (5) where Q is a matrix eigenvector of EE^T and J is a full rank matrix of regulated diagonal elements of EE^T.

Then the estimate of matrix B in the model (1) is the solution of the minimization

$$\hat{B} = \arg \inf_B (tr(Y-XB)^T \hat{\Sigma}^{-1} (Y-XB))$$

where Σ is the matrix (2), the maximum likelihood matrix of errors' covariances, received by using the procedure (3) or (4).

So

$$\hat{B} = (X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}Y \ . \qquad (6)$$

The form of the estimate is the well known Gauss-Markov estimate.

3. Another type of estimator

The model (1) can be modified into another form by

$$Y = XB + E$$

where $\mathcal{E}\{E\} = 0$

and $\text{cov}\{E\} = AA^T + D^2$.

That means that

$$Y = XB + AZ + u \qquad (7)$$

where $\mathcal{E}\{Z\} = \mathcal{E}\{u\} = 0$

$$\text{cov}\{u\} = D^2$$

$$\text{cov}\{Z\} = I \quad \text{and} \quad \mathcal{E}\{Zu^T\} = 0.$$

So for the new model (7)

$$\mathcal{E}\{(Y-XB)(Y-XB)^T\} = \mathcal{E}(AZ+u)(AZ+u)^T\} = AA^T + D^2$$

and the model (7) is equivalent to the model (1).

Hence the estimate of B has to be found as

$$\hat{B} = \arg\inf(\text{tr}\{(Y-XB-AZ)^TD^{-2}(Y-XB-AZ)\} + \text{tr}\{ZZ^T-I\}).$$

Here the term $\text{tr}\{ZZ^T-I\}$ is needed because we must estimate not only B but also Z, and the matrix Z has to have a restriction

$$ZZ^T = I.$$

The equation (7) is solved by finding the first derivatives of a function

$$f = \text{tr}\{(Y-XB-AZ)^TD^{-2}(Y-XB-AZ)\} + \text{tr}\{ZZ^T-I\}$$

with respect to the matrices Z and B.

After this we will get the estimate from the equations

$$\partial f/\partial Z = 0; \quad \text{and} \quad \partial f/\partial B = 0.$$

Here

$$\frac{\partial f}{\partial Z} = \text{vec}(-A^TD^{-2}(Y-XB-AZ)+Z)$$

and $\frac{\partial f}{\partial B} = \mathrm{vec}(X^TD^{-2}(Y-XB-AZ))$.

Hence $-A^TD^{-2}Y + A^TD^{-2}XB + A^TD^{-2}AZ + Z = 0$ (8)

$$X^TD^{-2}Y - X^TD^{-2}XB - X^TD^{-2}AZ = 0 \qquad (9)$$

We receive the expression for Z from (8)

$$Z = (A^TD^{-2}A^T+I)^{-1}\ A^TD^{-2}(Y-XB) \qquad (10)$$

After this we substitute Z in (9) and find the estimate B

$$\hat{B} = (X^T[D^{-2}-D^{-2}A(A^TD^{-2}A+I)^{-1}A^TD^{-2}]X)^{-1}$$
$$\cdot X^T[D^{-2}-D^{-2}A(A^TD^{-2}A+I)^{-1}A^TD^{-2}]Y.$$

But a matrix

$$[D^{-2}-D^{-2}A(A^TD^{-2}A+I)^{-1}A^TD^{-2}] \qquad (11)$$

is simply the matrix $[AA^T+D^2]^{-1}$.

It is easy to prove this. Multiplying (11) by $AA^T + D^2$ gives as a result the matrix

$$[D^{-2}-D^{-2}A(A^TD^{-2}A+I)^{-1}A^TD^{-2}][AA^T+D^2]$$
$$= D^{-2}AA^T-D^{-2}A(A^TD^{-2}A+I)^{-1}A^TD^{-2}AA^T+D^{-2}D^2$$
$$-D^{-2}A(A^TD^{-2}A+I)^{-1}A^TD^{-2}D^2$$
$$= D^{-2}AA^T + I - D^{-2}A(A^TD^{-2}A+I)^{-1}(A^TD^{-2}A+I)A^T = I.$$

Thus the estimate $\hat{B}$ is the same as (6).

4. Properties of the estimates

The estimate (6) is unbiased since

$$\mathcal{E}(\hat{B}) = \mathcal{E}((X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}Y)$$
$$= (X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}\mathcal{E}(Y)$$
$$= B.$$

Further the estimate (6) has covariance matrix

$$\mathrm{cov}(\hat{B}) = \mathcal{E}((\hat{B}-B)(\hat{B}-B)^T) = (X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1} =$$
$$= \mathcal{E}((Y-XB)(Y-XB)^T)\hat{\Sigma}^{-1}X(X^T\hat{\Sigma}^{-1}X)^{-1}$$
$$= (X^T\hat{\Sigma}^{-1}X)^{-1}X^T\hat{\Sigma}^{-1}\Sigma\hat{\Sigma}^{-1}X(X^T\hat{\Sigma}^{-1}X)^{-1} \qquad (12)$$

The ordinary least squares estimate is

$$\hat{B}_{LS} = (X^TX)^{-1}X^TY$$

with covariance matrix

$$\mathrm{cov}(\hat{B}_{Ls}) = \mathcal{E}((\hat{B}_{Ls}-B)(\hat{B}_{Ls}-B)^T)$$
$$= (X^TX)^{-1}X^T\Sigma X(X^TX)^1. \qquad (13)$$

Let us compare (12) and (13). Then we have to use the lemma 2b (Rao, 1967). If X is a $(N\times n)$-matrix of rank n, let Z be a $(N\times(N-n))$-matrix of rank $(N-n)$ such that $X^TZ = 0$. Then

$$(X^TS^{-1}X)^{-1}X^TS^{-1} = (X^TX)^{-1}X^T - (X^TX)^{-1}X^TSZ$$
$$\cdot(Z^TSZ)^{-1}Z^T \qquad (14)$$

where S is any $(N\times N)$-positive definite matrix. It is easy to see that the neat expression is true for the same matrices

$$(Z^TSZ)^{-1}Z^TS = (Z^TZ)^{-1}Z^T -$$
$(Z^TZ)^{-1}Z^TS^{-1}X(X^TS^{-1}X)^{-1}X^T$Substituting (14) into (12) gives

$$\mathrm{cov}(\hat{B}) = (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1} - (X^TX)^{-1}X^T\Sigma Z$$
$$\cdot(Z^T\hat{\Sigma}Z)^{-1}Z^T\hat{\Sigma}X(X^TX)^{-1} - (X^TX)^{-1}X^T\hat{\Sigma}Z(Z^T\hat{\Sigma}Z)^{-1}$$
$$\cdot Z^T\Sigma[I-Z(Z^TSZ)^{-1}Z^TS]X(X^TX)^{-1}. \qquad (16)$$

If the matrices X and Z are as in the lemma then the unit matrix within the square brackets of (16) equals (see Rao, 1967)

$$I = Z(Z^TZ)^{-1}Z^T + X(X^TX)^{-1}X^T.$$

Using (15) for the right part of the expression within the square brackets, we finally get

$$\mathrm{cov}(\hat{B}) = (X^TX)^{-1}X^T\Sigma X(X^TX)^{-1} - (X^TX)^{-1}X^T\Sigma Z$$
$$\cdot(Z^T\hat{\Sigma}Z)^{-1}Z^T\hat{\Sigma}X(X(X^TX)^{-1} - (X^TX)^{-1}X^T\hat{\Sigma}Z(Z^T\hat{\Sigma}Z)^{-1}$$
$$\cdot Z^T\Sigma X(X^TX)^{-1} - (X^TX)^{-1}X^T\hat{\Sigma}Z(Z^TZ)^{-1}Z^T\hat{\Sigma}^{-1}X\cdot(X^T\hat{\Sigma}^{-1}X)^{-1}. (17)$$

Hence because $\hat{\Sigma}$ and Σ are positive definite the matrix

$$\mathrm{cov}(\hat{B}_{Ls}) - \mathrm{cov}(\hat{B}) = (X^TX)^{-1}X^T\Sigma Z(Z^T\hat{\Sigma}Z)^{-1}$$
$$\cdot Z^T\hat{\Sigma}X(X^TX)^{-1} + (X^TX)^{-1}X^T\hat{\Sigma}Z(Z^T\hat{\Sigma}Z)^{-1}Z^T\Sigma X(X^TX)^{-1}$$
$$+ (X^TX)^{-1}X^T\hat{\Sigma}Z(Z^TZ)^{-1}Z^T\hat{\Sigma}^{-1}X(X^T\hat{\Sigma}^{-1}X)^{-1} \qquad (18)$$

is positive definite too. This means that the estimate (6) is more efficient than the ordinary least squares estimate.But if $\Sigma = \sigma^2I$ and $\hat{\Sigma} = \hat{\sigma}^2I$, then (18) equals zero.

<u>References</u>

Khokhlov, V.V., The maximum likelihood method in factor analysis. Preprint of Ukr NIINTI, Kiev, 1983 (in Russian).

Rao, C.R., Least squares theory using an estimated despersion matrix and its application to measurement of signals. The Fifth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley-Los Angeles, vol. I(1967), pp. 355-372.

Tikhonov, N.I., The regularization theory, Moscow-Nauka, 1978 (in Russian).