Wuhan University Journal of Natural Sciences Vol. 5 No. 4 2000,400~405

Article ID: 1007-1202(2000)04-0400-06

Selection of the Linear Regression Model According to the Parameter Estimation

Sun Dao-de

Department of Computer,Fuyang Teachers College,Anhui 236032,China

Abstract: In this paper,based on the theory of parameter estimation,we give a selection method and ,in a sense of a good character of the parameter estimation,we think that it is very reasonable. Moreover,we offer a calculation method of selection statistic and an applied example.

Key words: parameter estimation;linear regression model;selection criterion;mean square error

CLC number: O 212.1

1 Introduction

According to the professional knowledge and experience of the pratical problems,we preliminarily estimate that altogether there are p (containing constants) possible variables concerned with functions;and all the variables with functions are suited to the linear regression model. Given the practical observation data , we have the model:

$$Y = X\beta + e, Ee = 0, \mathrm{cov}(e) = \sigma^2 I \tag{1.1}$$

Where,Y is the observation vector of $n\times 1$,X is the designed matrix of $n\times p$,β is the parameter vector of $p\times 1$,e is the random vector of $n\times 1$.

Let's write X into the divided form:$X=(X_q,X_t)$ and correspondingly,$\beta'=(\beta_q',\beta_t')$,so,(1.1) can be rewritten as :

$$Y = X_q\beta_q + X_t\beta_t + e \tag{1.2}$$

Suppose $R(X_q)=q,R(X_t)=t,q+t=p$.

So we face a problem about the selection of independent variables of (1.2):

1) Suppose the real model is $Y=X\beta+e$,if we think the model is $Y=X_q\beta_q+e$, we will lose some independent variables by mistake.

2) Suppose the real model is $Y=X_q\beta_q+e$, if we think the model is $Y=X\beta+e$, we will introduce some unnecessary variables into the model.

So far,there are many solutions to solve the problem ,such as the C_p criterion,stepwise regression criterion[1],AIC criterion[2],BIC criterion[3]etc. These criterions are all sorts of model criterions based on the estimation's residual sum of squares,and each of them has its own rationality and convenient calculation ways. While in this paper,based on the theory of parameter estimation,we'll establish a linear regression model and in a sense of a good character of the parameter estimation,we think that the selection method of the independent variables is reasonable. So the established regression model and the parameter estimation are supportive for each other,which makes the model become more representative.

Received date:1999-12-28

Foundation item: Supported by the Natural Science Foundation of Anhui Education Committee

Biography:Sun Dao-de(1957-),male,associate professor,research interest:theory and applications of mathematical statistics.

For convenience, we call (1.1) as the complete model and name

$$Y = X_q\beta_q + e \tag{1.3}$$

the selective model.

Under the complete model, the least square estimation of β is recorded as $\hat{\beta}=(X'X)^{-1}X'Y=\begin{pmatrix}\hat{\beta}_q\\ \hat{\beta}_t\end{pmatrix}$;

and under the selective model, the least square estimation of β_q is recorded as

$$\hat{\beta}_q=(X_q'X_q)^{-1}X_q'Y \tag{1.4}$$

From the theorem 2.1 of Ref. [1], we know if the complete model (1.1) is right, then

$$E\,\tilde{\beta}_q=\beta_q+A\,\beta_t$$

Where
$$A=(X_q'X_q)^{-1}X_q'X_t \tag{1.5}$$
$$(X_q'X_q)^{-1}=\begin{pmatrix}(X_q'X_q)^{-1}+ADA' & -AD\\ -DA & D\end{pmatrix}$$
$$D^{-1}=X_t'(I-X_q(X_q'X_q)^{-1}X_q')X_t$$

then
$$\hat{\beta}_q=((X_q'X_q)^{-1}+ADA')X_q'Y-ADX_t'Y \tag{1.6}$$
$$\tilde{\beta}_q=(-DA'X_q'+DX_t')Y$$

Using the big or small which is the mean square error of β_q with it's two estimators $\tilde{\beta}_q$ and $\hat{\beta}_q$ we can establish a rule to select the independent variables. In part 2 we will introduce the rule.

2 The Selection Criterion of the Independent Variable

To construct and select statistics according to the idea above, we firstly study the mean square error of β_q with its two estimators.

Theorem
$$E(\|\tilde{\beta}_q-\beta_q\|^2)=\sigma^2\mathrm{tr}((X_q'X_q)^{-1})+\beta_t'A'A\,\beta_t \tag{2.1}$$
$$E(\|\hat{\beta}_q-\beta_q\|^2)=\sigma^2\mathrm{tr}((X_q'X_q)^{-1})+\sigma^2\mathrm{tr}(ADA') \tag{2.2}$$

Proof The demonstration of (2.1). From (1.4), (1.2) and theorem 2.1 in Ref. [1],

$$\begin{aligned}
E(\|\tilde{\beta}_q-\beta_q\|^2) &= E(\|(X_q'X_q)^{-1}X_q'Y-\beta_q\|^2)\\
&= E(\|(X_q'X_q)^{-1}X_q'(X_q\beta_q+X_t\beta_t+e)-\beta_q\|^2)\\
&= E(\|(X_q'X_q)^{-1}X_q'X_t\beta_t+(X_q'X_q)^{-1}X_q'e\|^2)\\
&= \beta_t'X_t'X_q(X_q'X_q)^{-1}(X_q'X_q)^{-1}X_q'X_t\beta_t+E(e'X_q(X_q'X_q)^{-1}(X_q'X_q)^{-1}X_q'e)\\
&= \beta_t'A'A\,\beta_t+\sigma^2\mathrm{tr}((X_q'X_q)^{-1}).
\end{aligned}$$

The demonstration of (2.2). From $\mathrm{cov}\hat{\beta}=E(\hat{\beta}-\beta)(\hat{\beta}-\beta)'=\sigma^2(X'X)^{-1}$ and (1.6), we have

$$\mathrm{cov}(\hat{\beta}_q)=\sigma^2((X_q'X_q)^{-1}+ADA'),$$

so,
$$\begin{aligned}
E(\|\hat{\beta}_q-\beta_q\|^2) &= E((\hat{\beta}_q-\beta_q)'(\hat{\beta}_q-\beta_q))=E(\mathrm{tr}((\hat{\beta}_q-\beta_q)'(\hat{\beta}_q-\beta_q)))\\
&= E(\mathrm{tr}((\hat{\beta}_q-\beta_q)(\hat{\beta}_q-\beta_q)'))=\mathrm{tr}(E((\hat{\beta}_q-\beta_q)(\hat{\beta}_q-\beta_q)'))
\end{aligned}$$

$$= \sigma^2\mathrm{tr}(\mathrm{cov}\ \hat{\underset{\sim}{\beta}}_q) = \sigma^2\mathrm{tr}((X_q'X_q)^{-1}) + \sigma^2\mathrm{tr}(ADA').$$

From (2.1),(2.2),we know ,if $\sigma^2\mathrm{tr}(ADA') \geqslant \underset{\sim}{\beta}_t' A' A \underset{\sim}{\beta}_t$,then

$$E(\| \tilde{\underset{\sim}{\beta}}_q - \underset{\sim}{\beta}_q \|^2) \leqslant E(\| \hat{\underset{\sim}{\beta}}_q - \underset{\sim}{\beta}_q \|^2)$$

To the proper parameter $\underset{\sim}{\beta}_q$, using the least square estimation $\tilde{\beta}_q$ of the selective model is smaller than the least square estimation β_q of the complete model and the mean square error of the proper parameter $\underset{\sim}{\beta}_q$. Now there is :

$$\triangle = \sigma^2\mathrm{tr}(ADA') - \underset{\sim}{\beta}_t' A' A \underset{\sim}{\beta}_t = \sigma^2\mathrm{tr}(ADA') - \| A \underset{\sim}{\beta}_t \|^2 \tag{2.3}$$

If the independent variables' entering into the selective model makes $\triangle$ become bigger,we think that the factor affects the model notably,otherwise the factor can be rejected from the selective model. Thus,we can use '$\triangle$ is the bigger the better' as a criterion to select the regression model.

3 $\triangle$'s Estimation Problem

From(2,3),we know that there are parameters σ^2 and $\| A \underset{\sim}{\beta}_t \|^2$ in $\triangle$,now let's discuss their estimation problem.

1) From theorem 2.5 in Ref. [1],we can get σ^2's unbiased estimation,which is:

$$\hat{\sigma}^2 = \| Y - X \hat{\underset{\sim}{\beta}} \|^2/(n-p) \tag{3.1}$$

2) The estimation of $\| A\underset{\sim}{\beta}_t \|^2$.

From $\mathrm{cov}\ \hat{\beta} = \sigma^2(X'X)^{-1}$, and according to (1.6) we know $\mathrm{cov}\ \hat{\beta}_t = \sigma^2 D$ and thus

$$E(\| A \hat{\underset{\sim}{\beta}}_t \|^2) = E(\| A \hat{\underset{\sim}{\beta}}_t - A \underset{\sim}{\beta}_t + A \underset{\sim}{\beta}_t \|^2 = E(\| A(\hat{\underset{\sim}{\beta}}_t - \underset{\sim}{\beta}_t \|^2) + \| A \underset{\sim}{\beta}_t \|^2$$

but, $$E(\| A(\hat{\underset{\sim}{\beta}}_t - \underset{\sim}{\beta}_t \|^2) = E(\mathrm{tr}(A'A(\hat{\underset{\sim}{\beta}}_t - \underset{\sim}{\beta}_t)(\hat{\underset{\sim}{\beta}}_t - \underset{\sim}{\beta}_t)'))$$

$$= \mathrm{tr}(A'AE(\hat{\underset{\sim}{\beta}}_t - \underset{\sim}{\beta}_t)(\hat{\underset{\sim}{\beta}}_t - \underset{\sim}{\beta}_t)') = \mathrm{tr}(A'A\sigma^2 D) = \sigma^2\mathrm{tr}(ADA')$$

So, $$E(\| A \hat{\underset{\sim}{\beta}}_t \|^2) = \| A \underset{\sim}{\beta}_t \|^2 + \sigma^2\mathrm{tr}(ADA') \tag{3.2}$$

From (3.2) we can see that using $\| A \hat{\underset{\sim}{\beta}}_t \|^2$ to estimate $\| A \underset{\sim}{\beta}_t \|^2$ is somewhat bigger. To compress the statistics $\| A \hat{\underset{\sim}{\beta}}_t \|^2$,we can use the theorem of compression estimation in Ref. [4].

let $$g(c) = E(c \| A \hat{\underset{\sim}{\beta}}_t \|^2 - \| A \underset{\sim}{\beta}_t \|^2)^2 = E(c^2 \| A \hat{\underset{\sim}{\beta}}_t \|^4 - 2c \| A \hat{\underset{\sim}{\beta}}_t \|^2 \| A \underset{\sim}{\beta}_t \|^2 + \| A \underset{\sim}{\beta}_t \|^4)$$

$$= c^2 E \| A \hat{\underset{\sim}{\beta}}_t \|^4 - 2cE \| A \hat{\underset{\sim}{\beta}}_t \|^2 \| A \underset{\sim}{\beta}_t \|^2 + E \| A \underset{\sim}{\beta}_t \|^4$$

then $g'(c) = 2cE \| A \hat{\underset{\sim}{\beta}}_t \|^4 - 2E \| A \hat{\underset{\sim}{\beta}}_t \|^2 \cdot \| A \underset{\sim}{\beta}_t \|^2$. According to (3.2),

$$g'(1) = 2E \| A \hat{\underset{\sim}{\beta}}_t \|^4 - 2E \| A \hat{\underset{\sim}{\beta}}_t \|^2 \cdot \| A \underset{\sim}{\beta}_t \|^2$$

$$> E \| A \hat{\underset{\sim}{\beta}}_t \|^4 - 2 \| A \underset{\sim}{\beta}_t \|^2 E \| A \hat{\underset{\sim}{\beta}}_t \|^2 + \| A \underset{\sim}{\beta}_t \|^4$$

$$= E(\| A \hat{\underset{\sim}{\beta}}_t \|^4 - 2 \| A \hat{\underset{\sim}{\beta}}_t \|^2 \cdot \| A \underset{\sim}{\beta}_t \|^2 + \| A \underset{\sim}{\beta}_t \|^4)$$

$$= E(\| A \hat{\underset{\sim}{\beta}}_t \|^2 - \| A \underset{\sim}{\beta}_t \|^2)^2 > 0$$

It shows that when $c<1$ and fully approaching 1,we will obtain:

$$E(c\|A\hat{\underset{\sim}{\beta}}_t\|^2-\|A\underset{\sim}{\beta}_t\|^2)^2<E(\|A\hat{\underset{\sim}{\beta}}_t\|^2-\|A\underset{\sim}{\beta}_t\|^2)^2,$$

and that is to say compressing $\|A\hat{\underset{\sim}{\beta}}_t\|^2$ properly is helpful to low mean square error.

Since $g'(c)=0$ and paying attention to the formular (3.2), we have:

$$c=\frac{\|A\underset{\sim}{\beta}_t\|^2E\|A\hat{\underset{\sim}{\beta}}_t\|^2}{E\|A\hat{\underset{\sim}{\beta}}_t\|^4}=\frac{\|A\underset{\sim}{\beta}_t\|^4+\sigma^2\|A\tilde{\beta}_t\|^2\mathrm{tr}(ADA')}{E(\|A\hat{\underset{\sim}{\beta}}_t\|^4)} \tag{3.3}$$

while, $$E\|A\hat{\underset{\sim}{\beta}}_t\|^4=E\|A\hat{\underset{\sim}{\beta}}_t-A\underset{\sim}{\beta}_t+A\underset{\sim}{\beta}_t\|^4=E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)+A\underset{\sim}{\beta}_t\|^4$$
$$=E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^4+4\|A\hat{\underset{\sim}{\beta}}_t\|\cdot E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^3$$
$$+6\|A\hat{\underset{\sim}{\beta}}_t\|^2\cdot E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^2+4\|A\underset{\sim}{\beta}_t\|^3\cdot E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|$$

According to Ref. [5],

$$E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^3=0 \text{ and } E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|=0$$

then

$$E\|A\hat{\underset{\sim}{\beta}}_t\|^4=E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^4+6\|A\underset{\sim}{\beta}_t\|^2E(\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t\|^2)+\|A\underset{\sim}{\beta}_t\|^4$$
$$=E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^4+6\sigma^2\|A\underset{\sim}{\beta}_t\|^2\mathrm{tr}(ADA')+\|A\underset{\sim}{\beta}_t\|^4$$

Now, we investigate $E(A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^4)$, for $E(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)=0$,

$$\mathrm{cov}(D^{-\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)=D^{\frac{1}{2}}\mathrm{cov}(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)(D^{-\frac{1}{2}})'=\sigma^2D^{-\frac{1}{2}}DD^{-\frac{1}{2}}=\sigma^2I_n,$$

Therefore:

$$E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)^4=E\|AD^{\frac{1}{2}}D^{-\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)\|^4=E((D^{-\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t))'D^{\frac{1}{2}}A'AD^{\frac{1}{2}}(D^{-\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)))^2$$

let $M=D^{\frac{1}{2}}A'AD^{\frac{1}{2}}$, $\underset{\sim}{z}=D^{-\frac{1}{2}}(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)$, then $E\|A(\hat{\underset{\sim}{\beta}}_t-\underset{\sim}{\beta}_t)^4\|=E(\underset{\sim}{z}'M\underset{\sim}{z})^2$, pay attention to $\underset{\sim}{z}\sim N(\underset{\sim}{0},\sigma^2I)$,

$$\underset{\sim}{z}'M\underset{\sim}{z}=\sum_{i,j=1}m_{ij}z_iz_j,\ (\underset{\sim}{z}'M\underset{\sim}{z})^2=\sum_{i,j=1}\sum_{k,l=1}m_{ij}m_{kl}z_iz_jz_kz_l$$

$$E(z_iz_jz_kz_l)=\begin{cases}3\sigma^4, & \text{when all following labels is equal;}\\ \sigma^4, & \text{when we divide following labels into two files, it is equal in each}\\ & \text{file, but unequal between the two;}\\ 0, & \text{the others.}\end{cases}$$

therefore

$$E(\underset{\sim}{z}'M\underset{\sim}{z})^2=3\sigma^4\sum_{i=1}m_{ij}^2+\sigma^4\sum_{i\neq j}(m_{ii}m_{jj}+m_{ij}^2+m_{ij}m_{ji})$$
$$=\sigma^4[2\sum_{i,j=1}m_{ij}^2+(\sum_{i=1}m_{ii})^2]=2\sigma^4\mathrm{tr}(M^2)+\sigma^4(\mathrm{tr}(M))^2$$
$$=2\sigma^4\mathrm{tr}(ADA')^2+\sigma^4(\mathrm{tr}(ADA'))^2$$

thus $$c=\frac{\|A\underset{\sim}{\beta}_t\|^4+\sigma^2\|A\underset{\sim}{\beta}_t\|^2\mathrm{tr}(ADA')}{\|A\underset{\sim}{\beta}_t\|^4+6\sigma^2\|A\underset{\sim}{\beta}_t\|^2\mathrm{tr}(ADA')+2\sigma^4\mathrm{tr}(ADA')+\sigma^4(\mathrm{tr}(ADA'))^2}$$

For $g(c)$ is quadratic function and $g(c)=\min$, let

$$c^{*} = \frac{\| A \underset{\sim}{\beta_t} \|^2 + \sigma^2 \mathrm{tr}(ADA')}{\| A \underset{\sim}{\beta_t} \|^2 + 6\sigma^2 \mathrm{tr}(ADA')} \tag{3.4}$$

then $c < c^{*} < 1, g(c) < g(c^{*}) < g(1)$.

Associating (3.2) with (3.3), we can use the following formular to estimate c^{*},

$$c^{**} = \frac{\| A \underset{\sim}{\hat{\beta}} \|^2}{\| A \underset{\sim}{\hat{\beta}_t} \|^2 + 5\sigma^2 \mathrm{tr}(ADA')} \tag{3.5}$$

We can see using it to estimate c^{*} is suitable. Combining with (2.3), we can define a selection criterion statistic which is

$$\Delta^{*} = \hat{\sigma}^2 \mathrm{tr}(ADA') - c^{**} \| A \underset{\sim}{\hat{\beta}_t} \|^2 \tag{3.6}$$

as Δ's estimation. In this paper, we take "the bigger Δ^{*} is, the better it'll become" as the selection criterion of the linear regression model.

4 Δ^{*}'s Calculation and Application

According to (3.5) and (3.6), if we want to calculate Δ^{*}, we have to solve the calculation problem of tr(ADA') and $\| A \underset{\sim}{\hat{\beta}_t} \|^2$ firstly. On the basis of the scanning algorithm in Ref. [1], we suppose $A = (a_{ij})_{n\times n}$, if $a_{ii} \neq 0$, we define a new square matrix $B = (b_{ij})_{n\times n}$, in it:

$$b_{ii} = \frac{1}{a_{ii}}, b_{ij} = \frac{a_{ij}}{a_{ii}},\ j \neq i, b_{ji} = -\frac{a_{ji}}{a_{ii}},\ j \neq i, b_{kl} = a_{kl} - \frac{a_{il}a_{ki}}{a_{ii}}, k \neq i, l \neq j.$$

The transformation from A to B is called S operation with the pivot of a_{ii} and is recorded as $B = S_i A$. According to S' arithmetic properties (theorem 7.1, 7.2 in Ref. [1]), we assume that X_q consists of the NO. $1 \leqslant i_1, i_2, \cdots\cdots, i_q \leqslant p$ row in X;

$$C = S_{i1} S_{i2} \cdots S_{iq} \begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix}$$

λ_i and $\tau_i (i = 1, 2, \cdots, p)$ is NO. i diagonal element of $(X'X)^{-1}$ and C; V is a matrix which consists of element C's NO. $i_1, i_2, \cdots, i_q$ column, the front p row arranged in C's order.

$\underset{\sim}{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_p)$ is a p-vector, in it:

$$\alpha_j = \begin{cases} 0, & j = i_l; \\ \hat{\beta}_j, & j \neq i_l \end{cases} \quad l = 1, 2, \cdots, q,\ j = 1, 2, \cdots, p$$

So, we can introduce:

$$\mathrm{tr}(ADA') = \sum_{l=1}^{q} (\lambda_{il} - \tau_{il}) \tag{4.1}$$

$$A \underset{\sim}{\hat{\beta}_t} = V \underset{\sim}{\alpha} \tag{4.2}$$

According to the scanning algorithm and (4.1), (4.2), we can quickly figure out Δ^{*} and realize the selection of linear regression model.

Let's adopt a classical example: Hald cement problem[1], which is mostly used in documents of the regression analysis to illustrate the application of the variable selection.

When some cement becomes solid it releases heat (calorie) and contains the following four types of chemical composition:

x_1 equate the content (%) of the $3CaO \cdot Al_2O_3$, x_2 equate the content (%) of the $3CaO \cdot SiO_2$, x_3 equate the content (%) of the $4CaO \cdot Al_2O_3 \cdot Fe_2O_3$, x_4 equate the content (%) of the $2CaO \cdot SiO_2$.

The problem is to investigate the relation between the released heat per gram (noted as Y) and these four types of composition. Table 1 shows the experimental datum.

Table 1 The experimental datum

x_1	7	1	11	11	7	11	3	1	2	21	1	11	10
x_2	26	29	56	31	52	55	71	31	54	47	40	66	68
x_3	6	15	8	8	6	9	17	22	18	4	23	9	8
x_4	60	52	20	47	33	22	6	44	22	26	34	12	12
Y	78.5	74.3	104.3	47.8	95.9	109.2	102.7	72.5	93.1	115.9	83.8	113.3	109.4

$$X' = \begin{pmatrix} 7 & 1 & 11 & 11 & 7 & 11 & 3 & 1 & 2 & 21 & 1 & 11 & 10 \\ 26 & 29 & 56 & 31 & 52 & 55 & 71 & 31 & 54 & 47 & 40 & 66 & 68 \\ 6 & 15 & 8 & 8 & 6 & 9 & 17 & 22 & 18 & 4 & 23 & 9 & 8 \\ 60 & 52 & 20 & 47 & 33 & 22 & 6 & 44 & 22 & 26 & 34 & 12 & 12 \end{pmatrix}$$

$$Y' = (78.5, 74.3, 104.3, 47.8, 95.9, 109.2, 102.7, 72.5, 93.1, 115.9, 83.8, 113.3, 109.4)$$

After the scanning algorithm of $\begin{pmatrix} X'X & X'Y \\ Y'X & Y'Y \end{pmatrix}$, we can introduce LS estimation of Hald cement problem.

From Table 2, we know that Δ^* will come to the greatest valuation at (x_1, x_2), and it's also bigger at x_2. It proves that $3CaO \cdot SiO_2$ is the primary element of cement's releasing heat. The amount of released heat and the content of $3CaO \cdot Al_2O_3$ and $3CaO \cdot SiO_2$ are most close to each other. Under the Δ^* criterion, the best linear regression model follows:

$Y = 52.577 + 1.468x_1 + 0.662x_2$, it tallies with example 3.2 in Ref. [1].

Table 2 LS estimation of Hald cement problem and valuation of Δ^*

Independent variable in the model	β_0	β_1	β_2	β_3	β_4	Δ^*
x_1	81.4794	1.8687				4903.9443
x_2	57.4237		0.7891			4905.2429
x_3	110.2026			−1.2558		4713.6000
x_4	117.5680				−0.7382	4572.7207
x_1x_2	52.5774	1.4682	0.6623			4905.2551
x_1x_3	72.3491	2.3124		0.4945		4896.4444
x_1x_4	103.0974	1.4399			−0.6140	4802.6622
x_2x_3	72.0746		0.7313	−1.0080		4902.8431
x_2x_4	94.1601		0.3109		−0.4569	4648.5257
x_3x_4	131.2824			−1.2000	−0.7246	4138.1392
$x_1x_2x_3$	48.1937	1.6958	0.6570	0.5000		4892.7450
$x_1x_2x_4$	71.6484	1.4518	0.4162		−0.2365	4686.6689
$x_2x_3x_4$	203.6420		−0.9235	−1.4480	−1.5570	−4666.7750
$x_1x_3x_4$	111.6844	1.0517		−0.4100	−0.6428	4669.8517
$x_1x_2x_3x_4$	62.4051	1.5510	0.5102	0.1020	−0.1441	0

References:

[1] Chen Xi-ru, Wang Song-gui. *Modern Regression Analysis*. Hefei: Anhui Education Press, 1987(Ch).
[2] Akaike H. A New Look at the Statistical Model Indentification. *IEEE Trans Automatic control*, 1974, **19**: 714-723.
[3] Schwarz G. Estimating the Dimension of a Model. *Ann statist*, 1987, **6**: 416-44.
[4] Wang Song-gui. *Theory and Applications of Linear Model*. Hefei: Anhui Education Press, 1987(Ch).
[5] Zhang Yao-ting, Fang Kai-tai. *Introduction to Multivariate Statistical Analysis*. Beijing: Science Press, 1997, 71-72(Ch).