INFLUENTIAL NONNEGLIGENCE PARAMETERS UNDER THE SEARCH LINEAR MODEL (U) CALIFORNIA UNIV RIVERSIDE DEPT OF STATISTICS S GHOSH 25 APR 86 TR-141 AFOSR-TR-86-0733

UNCLASSIFIED AFOSR-86-0048
INFLUENTIAL NONNEGLIGIBLE PARAMETERS UNDER THE SEARCH LINEAR MODEL

BY Subir Ghosh

University of California, Riverside

IN THIS PAPER SOME RESULTS USEFUL IN DETECTING THE INFLUENTIAL NONNEGLIGIBLE PARAMETERS UNDER THE SEARCH LINEAR MODEL ARE PRESENTED. AN ESTIMATOR OF THE NUMBER OF NONNEGLIGIBLE PARAMETERS WHICH ARE SIGNIFICANT AND INFLUENTIAL IS ALSO GIVEN.

AMS 1970 Subject Classifications: Primary and Secondary: 62J05, 62K05, 62K15, 62K99

Keywords and Phrases: Factorial Experiments, Influential Parameters, Linear Models, Residuals, Search, Treatment.

*The work of the author is sponsored by the Air Force Office of Scientific Research under Grant AFOSR-86-0048.
1. Introduction

Consider a general factorial experiment with the design consisting of $t$ treatments and corresponding to the $u$th treatment there are $n_u (\geq 1)$ observations and $\sum_{u=1}^{t} n_u = N$. Let $y_{uv}$ be the observation corresponding to the $u$th replication of the $v$th treatment and $\bar{y}_u$ be the mean of all observations corresponding to the $u$th treatment. The model for this experiment is

$$E(y) = X_1 \beta_1 + X_2 \beta_2,$$

$$V(y) = \sigma^2 I,$$

$$\text{Rank } X_1 = \nu_1,$$

where $\beta_1(v_1 \times 1)$ is a vector of specified lower order interactions and $\beta_2(v_2 \times 1)$ is a vector of some or all of the higher order interactions, $X_1(N \times v_1)$ and $X_2(N \times v_2)$ are known matrices. It is known that $K$ (very small compared to $v_2$) elements of $\beta_2$ are nonzero and the other are zero; however the value of $K$ and the nonzero elements of $\beta_2$ are unknown. The problem is to search the nonzero elements of $\beta_2$ and draw inferences on them in addition to the elements of $\beta_1$. Such a model is called the search linear model and was introduced in Srivastava (1975). Suppose $K_1$ is an initial guess on $K$. Note the three possibilities $K_1 > K$, $K_1 = K$ and $K_1 < k$. We consider $n_{1}\begin{pmatrix} v_2 \\ K_1 \end{pmatrix}$ models

$$E(y) = X_1 \beta_1 + X_2^{(1)} \beta_2^{(1)}, i=1,\ldots, v_2,$$

$$V(y) = \sigma^2 I,$$

$$\text{Rank}[X_1, X_2^{(1)}] = v_1 + K_1,$$
where $X_2^{(i)}(N \times K_i)$ is a submatrix of $X_2$ and $\beta_2^{(i)}(K_i \times 1)$ is a subvector of $\beta_2$. It can be seen from Srivastava (1975) that we in fact need

$$\text{Rank} \left[ X_1, X_2^{(i)}, X_2^{(i')} \right] = (v_i + 2K_i), \text{ for all } i \neq i' \). This implies that $N \geq (v_i + 2K_i)$. In case $K_1 = K$, one of $v_2$ models is the correct model. If $K_1 > K$, then $v_2-K_1$ models out of $v_2$ models include the true model as a submodel in the expectation forms of the models. The methods discussed in this paper will not only identify $K$ nonzero parameters but also find how many of them have significant effects and, finally, rank the significant noneglligible parameters in the order of their influence on the fitted values. In case $K_1 < K$, the methods will identify from $K_1$ parameters the parameters which are significant and influential. We also propose an estimator of $K$ in the Section 3.

In some industrial experiments, it is often easy to find replications ($n_u > 1$) in observations corresponding to a particular (the $u$th) treatments, see Taguchi and Wu (1985). There are also situations in industrial experiments where it is impossible to get replication in observations for a treatment, see Daniel (1976) and Box and Meyer (1985). The methods discussed in this paper consider both situations. In all Taguchi design methods, the higher order interactions (2-factor and higher order in most plans) are assumed to be zero. A few of those higher order interactions may be nonnegligible, significant and influential. The use of the search linear models may be a potential tool in improving upon the Taguchi design methods.
2. **Influential Nonnegligible Parameters.**

Let \( Z^{(1)}_1 \sim \) be such that \( \text{Rank} \ Z^{(1)}_1 = (N-v_1-K_1) \), \( Z^{(1)}_1(1) = I \) and \( Z^{(1)}_1[X_1, X_2^{(1)}] = 0 \). Let \( Z^{(1)}(K_1 \times N) \) be such that

\[
\text{Rank} \begin{bmatrix} \mathbf{Z}_1^{(1)} \\ \mathbf{Z}^{(1)} \end{bmatrix} = (N-v_1), \ Z^{(1)}(1) = I, \ Z^{(1)}(1) = 0 \text{ and } \mathbf{Z}^{(1)}K_1 = 0. 
\]

It can be seen that under the \( i \)th model in (2), the minimum variance unbiased estimator (MVUE) of \( \beta^{(1)}_2 \) is

\[
\hat{\beta}_2^{(1)} = (\mathbf{Z}_2^{(1)}X_2^{(1)})^{-1} \mathbf{Z}_2^{(1)} \mathbf{y}. \tag{3}
\]

In fact we can write \( Z^{(1)}(1) = P_1X_2^{(1)}D^{(1)} \), where \( D^{(1)} \) is a nonsingular (and triangular) matrix so that \( Z^{(1)}_1D^{(1)} = I \) and \( P_1 = I - X_1(X_1X_1)^{-1}X_1^{(1)} \). From the \( i \)th model in (2), the MVUE for \( \beta_1 \) is

\[
\hat{\beta}_1^{(1)} = (X_1X_1)^{-1}X_1^{(1)} \mathbf{y} - (X_1X_1)^{-1}X_1^{(1)}X_2^{(1)} \hat{\beta}_2^{(1)}. \tag{4}
\]

The fitted value of \( \mathbf{y} \) from the \( i \)th model in (2) is

\[
\hat{\mathbf{y}}^{(1)} = X_1\hat{\beta}_1^{(1)} + X_2^{(1)}\hat{\beta}_2^{(1)}. \tag{5}
\]

The residuals from the \( i \)th model in (2) are

\[
\mathbf{R}^{(1)} = \mathbf{y} - \hat{\mathbf{y}}^{(1)} = P_1(\mathbf{y} - X_2^{(1)}\hat{\beta}_2^{(1)})
\]

\[
= P_1[I - X_2^{(1)}(X_2^{(1)})^{-1}X_2^{(1)}]P_1\mathbf{y}. \tag{6}
\]

The sum of squares due to error under the \( i \)th \( (i = 1, \ldots, (V^2)_{K_1}) \) model in (2) is

\[
\text{SSE}^{(1)} = \mathbf{R}^{(1)}X^{(1)} = \mathbf{y}Z^{(1)}X^{(1)}X^{(1)}\mathbf{y}. \tag{7}
\]
The residuals under the model (1), when \( \beta_2 = 0 \), are

\[
R^{(o)} = y - \hat{y} = P_1 y.
\]  \( \text{(8)} \)

It can be seen that

\[
P_1 = Z_1^{(1)'} Z_1^{(1)} + Z^{(1)'} Z^{(1)}.
\]  \( \text{(9)} \)

Therefore, for \( i = 1, \ldots, \left( \frac{v_2}{K_1} \right) \),

\[
\text{SSE}^{(o)} = R^{(o)'} R^{(o)} = \text{SSE}^{(1)} + y' Z^{(1)'} Z^{(1)} y.
\]  \( \text{(10)} \)

For \( i = 1, \ldots, \left( \frac{v_2}{K_1} \right) \), we define

\[
P(i) = \frac{y' Z^{(i)'} Z^{(i)} y / K_1}{\text{SSE}^{(i)}/(N - v_i - K_1)}.
\]  \( \text{(11)} \)

Let \( \bar{y}^{(1)}_u \) be the fitted value of the observation corresponding to the \( u \)-th \( (u = 1, \ldots, w) \) treatment under the \( i \)-th model in (2). We write the sum of squares due to lack of fit as

\[
\text{SSLOF}^{(1)} = \sum_{u=1}^{w} n_u (\bar{y}^{(1)}_u - \bar{y}_u)^2,
\]  \( \text{(12)} \)

and the sum of squares due to pure error as

\[
\text{SSPE} = \sum_{u=1}^{w} \sum_{v=1}^{n_u} (y_{uv} - \bar{y}_u)^2.
\]  \( \text{(13)} \)

For \( i = 1, \ldots, \left( \frac{v_2}{K_1} \right) \), we define

\[
P_{\text{LOF}}(i) = \frac{\text{SSLOF}^{(1)}/(w - v_i - K_1)}{\text{SSPE}/(N - w)}.
\]  \( \text{(14)} \)
Theorem 1. For $i \in \{1, ..., [v_2] \}$, the following statements are equivalent.

(a) $\text{SSE}(i)$ is a minimum,
(b) $F(i)$ is a maximum,
(c) $\text{SSLOF}(i)$ is a minimum,
(d) $F_{LO}$ is a minimum,
(e) The Euclidean distance between $\tilde{\mathbf{y}}(i)$ and $\tilde{\mathbf{y}}(o)$ is a maximum,
(f) The square of the (sample) simple correlation coefficient between
the elements of $\mathbf{R}(i)$ and $\mathbf{R}(o)$ is a minimum.

Proof. We have from (10) and (11) that

$$
\left( \frac{K_1}{(N-v_1-K_1)} F(i) \right)^{+1} = \frac{\text{SSE}(o)}{\text{SSE}(i)}.
$$

Noting that the numerator on the RHS of the above expression does not depend on $i$, we get the equivalence of (a) and (b). Again,

$$
\text{SSE}(i) = \text{SSPE} + \text{SSLOF}(i),
$$

and SSPE does not depend on $i$. Therefore (a) and (c) are equivalent.

From (14), the equivalence of (c) and (d) is clear. From (3), (6),
(8) and (9), it follow that

$$
\mathbf{y}' \mathbf{z}^{(i)} \mathbf{z}^{(i)} = \mathbf{z}^{(i)} \mathbf{y}' \mathbf{z}^{(i)} = \frac{\beta_2^{(i)}}{\beta_2} \mathbf{x}_2' \mathbf{z}^{(i)} \mathbf{z}^{(i)} \mathbf{x}_2
$$

$$
= \frac{\beta_2^{(i)}}{\beta_2} \mathbf{x}_2' \left[ \mathbf{P}_1 \mathbf{z}_1^{(1)} \mathbf{z}_1^{(1)} \right] \mathbf{x}_2 \beta_2^{(i)}
$$

$$
= \frac{\beta_2^{(i)}}{\beta_2} \mathbf{x}_2' \mathbf{P}_1 \mathbf{x}_2 \beta_2^{(i)}
$$

$$
= \left( \mathbf{R}^{(i)} - \mathbf{R}(o) \right) \left( \mathbf{R}^{(i)} - \mathbf{R}(o) \right)
$$

$$
= \left( \mathbf{R}^{(i)} - \mathbf{R}(o) \right) \left( \mathbf{R}^{(i)} - \mathbf{R}(o) \right)
$$

$$
= \left( \mathbf{S}^{(i)} + \mathbf{S}(o) \right) \left( \mathbf{S}^{(i)} + \mathbf{S}(o) \right).
$$

(15)
The equivalence of (a) and (e) is now easy to see from (10) and (15).

It follows from (10) and (15) that \( R^{(1)} R^{(1)} = R^{(1)} R^{(o)} \). We thus have

\[
\frac{\text{SSE}^{(1)}}{R^{(o)' R^{(o)}}} = \frac{R^{(1)' R^{(1)}}}{R^{(o)' R^{(o)}}} = \frac{(R^{(1)' R^{(o)})^2}{(R^{(1)' R^{(1)}) (R^{(o)' R^{(o)})})}
\]

= the square of the (sample) Simple correlation Coefficient between \( R^{(1)} \) and \( R^{(o)} \).

The equivalence of (a) and (f) is now clear from (16). This completes the proof of the theorem.

**Proposition 1.** Under the \( i \)th model in (2),

\[
Z^{(1)} R^{(1)} = 0.
\]

**Proof.** It follows from (3) and (5) that

\[
Z^{(1)} Y = Z^{(1)} X_2 P_2 = Z^{(1)} Y.
\]

This completes the proof.

We have

\[
V(R^{(1)}) = \sigma^2 P_1 \left[ I - X_2^{(1)} (X_2^{(1)})' P_1 X_2^{(1)} \right]^{-1} X_2^{(1)} P_1.
\]

The residual in \( R^{(1)} \) are correlated and the question may be asked about the appropriateness in combining the elements of \( R^{(1)} \) in \( \text{SSE}^{(1)} \).

If we take the transformed residuals as \( Z^{(1)} R^{(1)} \), we then have

\[
E(Z^{(1)} R^{(1)}) = 0 \quad \text{and} \quad V(Z^{(1)} R^{(1)}) = \sigma^2 I.
\]

The sum of squares of these transformed residuals is

\[
\sum Z^{(1)} R^{(1)} Z^{(1)} R^{(1)}
\].
Proposition 2. For \( i = 1, \ldots, (\nu_2^2) \),

\[
\text{SSE}^{(1)} = R^{(1)} Z_1^{(1)} Z_1^{(1)} R^{(1)}. \tag{20}
\]

Proof. We write the RHS using (9) as

\[
R^{(1)} Z_1^{(1)} Z_1^{(1)} R^{(1)} = R^{(1)} P_1 R^{(1)} - R^{(1)} Z^{(1)} Z^{(1)} R^{(1)}. \tag{21}
\]

It can be checked that \( P_1 R^{(1)} = R^{(1)} \). By using the Proposition 1, the rest of the proof is clear. This completes the proof.

Proposition 2 thus supports the use of \( \text{SSE}^{(1)} \). Theorem 1 gives various interpretations of a search procedure, discussed in Srivastava (1975), of selecting \( \beta_2^{(1)} \) as the influential set of \( K_1 \) nonnegligible parameters.

We now denote

\[
\begin{align*}
\beta_2^{(1)} &= [\beta_2^{(1)}, \ldots, \beta_2^{(1)}, \ldots, \beta_2^{(1)}], \\
x_2^{(1)} &= [x_2^{(1)}, \ldots, x_2^{(1)}, \ldots, x_2^{(1)}], \\
x_2^{(1j)} &= \text{the matrix obtained from } x_2^{(1)} \text{ by} \\
&\quad \text{deleting the } j \text{th column of } x_2^{(1)}. \\
x_1^{(1j)} &= [x_1, x_2^{(1j)}], \\
p_1^{(1j)} &= I - x_1^{(1j)} (x_1^{(1j)})' (x_1^{(1j)})^{-1} x_1^{(1j)}', \\
z_{1j}^{(1)} &= \frac{p_{1j}^{(1j)}}{\sqrt{x_2^{(1j)} p_{1j}^{(1j)}}}, \\
z_0^{(1)} &= [z_0^{(1)}, \ldots, z_0^{(1)}, \ldots, z_0^{(1)}].
\end{align*}
\]
It can be seen that
\[ \text{rank} \begin{bmatrix} Z_1^{(1)} \\ \cdot \\ \cdot \\ Z_{\nu K}^{(1)} \end{bmatrix} = (N - \nu_1 - K_1 + 1), \]

\[ Z_{1j}^{(1)} Z_{1j}^{(1)'} = 1, \quad Z_{1j}^{(1)'} Z_1^{(1)} = 0', \quad (23) \]

\[ Z_1^{(1)'} X_{12} = 0', \quad \text{rank} Z_0^{(1)} = K_1. \]

There exists a nonsingular (triangular) matrix \( D_0^{(1)} \) such that
\[ Z^{(1)} = D_0^{(1)} Z_0^{(1)}. \quad (24) \]

From (3) and (24), we have
\[ \hat{\beta}_2^{(1)} = (Z_0^{(1)'} X_2^{(1)})^{-1} Z_0^{(1)} Y. \quad (25) \]

Now
\[ Z_0^{(1)'} X_2^{(1)} = \text{diag} \left[ Z_{11}^{(1)'} X_{21}, \ldots, Z_{1j}^{(1)'} X_{2j}, \ldots, Z_{1K}^{(1)'} X_{2K} \right] \]

is a diagonal matrix. Thus
\[ \beta_2^{(1)} = \frac{Z_{1j}^{(1)'} Y}{Z_{1j}^{(1)'} X_{2j}}. \quad (27) \]

Let \( R_{ij}, i = 1, \ldots, K_1, j = 1, \ldots, K_1 \), be the residuals obtained from ith model in (2) assuming \( \beta_2^{(1)} = 0 \). Then the sum of squares due to error is
\[ \text{SSE}^{(1j)} = R_{ij}^{(1j)'} R_{ij}^{(1j)} = (Z_{1j}^{(1)'} Y)^2 + \text{SSE}^{(1)}. \quad (28) \]

We now define, for \( i = 1, \ldots, K_1 \) and \( j = 1, \ldots, K_1 \),
Proposition 3. For a fixed \( k \) in \( \{1, \ldots, (\nu_{2}/k_{1})\} \) and an \( m \) in \( \{1, \ldots, k_{1}\} \), the following statements are equivalent.

(a) \( \text{SSE}(i,m) \) is a minimum,

(b) \( \tau(i,m) \) is a maximum.

Proof. The proof can be easily seen from (28) and (29).

In the set \( \beta_{2}^{(k)} \) of influential nonnegligible parameters, \( \beta_{2m}^{(k)} \) is the most influential nonnegligible parameters. The influential nonnegligible parameters may or may not have significant effects on observations.
3. Influential Significant Nonnegligible Parameters

We now assume the normality in (2) and therefore for
\[ i = 1, \ldots, (v_2), \sum \underbrace{N(x_1 \beta_1^i + x_2 \beta_2^i)}_{\text{independent}}, \sigma^2 I. \]

Under the null hypothesis \( H_0: \beta_2 = 0, F(1) \) has the central \( F \) distribution with \((K, N - v_1 - K)\) d.f. and under the null hypothesis \( H_0: \beta_2 = 0, t(ij) \) has the central \( t \) distribution with \((N - v_1 - K)\) d.f. We now present a further development of a procedure suggested in Srivastava (1975).

Case I. If \( \max F(i) \leq F_{a;K,N-v_1-K} \), we then conclude that there is no significant nonnegligible parameter. \( F_{a;K,N-v_1-K} \) is the upper \( a \) percent point of the central \( F \) distribution with \((K,N-v_1-K)\) d.f.

Case II. Suppose for \( i = 1, \ldots, v_1 \), we have \( F(i) > F_{a;K,N-v_1-K} \). We denote for \( j = 1, \ldots, v_2 \),
\[ \gamma_j = \text{the number of } i \text{ in } [1, \ldots, v_1] \text{ for which } |t_{ij}| > \frac{a}{2}, N-v_1-K. \]

Note that \( 0 \leq \gamma_j \leq s \). We now arrange \( \gamma_j \)'s in decreasing order of magnitude and write \( \gamma_1 \geq \gamma_2 \geq \ldots \geq \gamma_{v_2} \). If there are at least \( K \) nonzero \( \gamma_j \)'s, we select the influential significant parameters as \( \beta(1), \ldots, \beta(K) \); otherwise we pick the influential \( \beta(j) \)'s corresponding to nonzero \( \gamma_j \)'s (Note that the number of influential parameters is then less than \( K \)). The parameter \( \beta(1) \) is the most influential significant nonnegligible parameter. An estimator of the unknown \( K \) is
\[ \hat{K} = \text{the number of nonzero } \gamma_j \text{'s}, j = 1, \ldots, v_2. \]
4. Miscellaneous Results

4.a. Let us denote the unknown nonzero elements of $\beta_2$ in (1) by $\beta_{2c}$ ($K\times1$) and the zero elements of $\beta_2$ by $\beta_{2d}((v-K)\times1)$, the corresponding columns in $X_2$ matrix are $X_{2c}$ and $X_{2d}$. The unknown true expectation form of (1) is thus

$$E(y) = X_{1-1} + X_{2c}\beta_{2c}.$$

(30)

The expectation form of the $i$th model in (2) can be written as

$$E(y) = X_{1-i} + X_{2c}(i)\beta_{2c} + X_{2d}\beta_{2d},$$

(31)

where $X_{2c}(i) (Nx_{i1})$ is a submatrix of $X_{2c}$, $X_{2d}(i) (Nx(1-K_{i1}))$ is a submatrix of $X_{2d}$, $\beta_{2c}(i)$ is a subvector of $\beta_{2c}$ and $\beta_{2d}$ is a subvector of $\beta_{2d}$. Let $\beta_{2c}(i)$ is the vector of elements in $\beta_{2c}$ which are not in $\beta_{2c}$ and $X_{2c}(i)$ is the matrix whose columns are in $X_{2c}$ but not in $X_{2c}$. The following result, a counterpart of the result in (10) for the population, can be verified very easily.

**Proposition 4.** Under (30),

$$E(SSE(i)) = E(SSLOF(i)) + \sigma^2(N-w)$$

$$= \sigma^2(N-v_1-K_1) + \beta_{2c}^t X_{2c} Z_{1}(i)' Z_{1} X_{2c} \beta_{2c}$$

$$= \sigma^2(N-v_1-K_1) + \beta_{2c}^t X_{2c} Z_{1}(i)' Z_{1} X_{2c} \beta_{2c}$$

(32)

$$= E(SSE(0)) - [\sigma^2 K_1 + \beta_{2c}^t X_{2c} Z_{1}(i)' Z_{1} X_{2c} \beta_{2c}] .$$
4.b. The model obtained from (2)

\[ E(Z^{(1)}_Y) = Z^{(1)} X^{(1)} \beta^{(1)}_2, \]

\[ V(Z^{(1)}_Y) = \sigma^2 I, \]

is called the pure search model (Srivastava (1976)). In fact, Srivastava (1976) considered a special form of \( Z^{(1)} \).

4.c. The influential nonnegligible parameter may depend on noise, i.e., a parameter may be influential under one noise but may not be influential under another noise.

4.d. The replicated observations will surely improve the chances of detecting the correct influential nonnegligible parameters.

4.e. In presence of outliers in observations, one may combine residuals with unequal weights, or in other words, may use transformed residuals (see, Cook and Weisberg (1982)).

4.e.1. An example of transformed residual is the vector 

\[ M^{(1)} R^{(1)} \] where \( M^{(1)} \) is a diagonal matrix whose \( u \)th diagonal element is \( \frac{1}{\sqrt{m_u^{(1)}}} \) with \( m_u^{(1)} \) being the \( u \)th diagonal element of \( \sigma^{-2} V(R^{(1)}) \).

4.e.2. Suppose the underlying design is robust against the unavailability of any single observation [see, Ghosh (1980)] in the sense that the estimation of \( \beta_1 \) and \( \beta_2^{(1)} \) is possible under (2) when any single observation is unavailable. We find the predicted value of the \( u \)th observation from the remaining \( (N-1) \) observations (i.e.,
by deleting the $u$th observation). The difference between the $u$th observation and its predicted value is called the $u$th predicted residual (using the idea of cross validation). It can be verified algebraically that the vector of predicted residuals is $[M^{(1)}]^2 \hat{R}^{(1)}$. The predicted residual sum of squares (PRESS) from the $i$th model under (2) is

$$PRESS^{(1)} = \hat{R}^{(1)'} [M^{(1)}]^4 \hat{R}^{(1)}$$

In presence of outliers, one may take $PRESS^{(1)}$ as an alternative to $SSE^{(1)}$. 
BIBLIOGRAPHY


In this paper some results useful in detecting the Influential Nonnegligible parameters under the search linear model are presented. An estimator of the number of nonnegligible parameters which are significant and influential is also given.
END
D Tic
10-86