STUDIA UNIV. "BABEŞ–BOLYAI", MATHEMATICA, Volume **XLVIII**, Number 2, June 2003

FITTING OF SOME LINEARISABLE REGRESSION MODELS

NICOLETA BREAZ AND DANIEL BREAZ

Abstract. In this paper, we obtain fitting conditions for some linearisables regressional models. These conditions are refering to the matrix of sample data.

1. **Introduction**

The fitting condition for those models, which, by substitution, can be reduced to a linear model, is refering to the matrix of new sample data/variables that results by substitution. In this paper, we consider models such as the polynomial, spline and piecewise linear model and we give for these, fitting conditions in the matrix of initial sample data/variables.

Let be the multiple linear model

$$Y = \alpha_1 X_1 + ... + \alpha_p X_p + \varepsilon \tag{1}$$

and a sample data

$$\mathbf{y}^T = (y_1, y_{2,}..., y_n) \in \Re^n,$$

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_1 & , \mathbf{x}_2, ..., & \mathbf{x}_p \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & ... & x_{1p} \\ x_{21} & x_{22} & ... & x_{2p} \\ ... & ... & ... & ... \\ x_{n1} & x_{n2} & ... & x_{np} \end{pmatrix} \in \mathbf{M}_{n,p}, n > p.$$

Denoting $\boldsymbol{\alpha}^T = \begin{pmatrix} \alpha_1, & \alpha_2, & ..., & \alpha_p \end{pmatrix} \in \Re^p, \boldsymbol{\varepsilon}^T = (\varepsilon_1, \varepsilon_2, ..., \varepsilon_n) \in \Re^n$ from (1) we obtain the matriceal form $\mathbf{y} = \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$.

The principle of least squares leads to the fitting model:

$$\mathbf{y} = \mathbf{x}\mathbf{a} + \mathbf{e},$$

with

$$\mathbf{a}^T = (a_1, a_2, ..., a_p) \in \Re^p, \mathbf{e}^T = (e_1, e_2, ..., e_n) \in \Re^n \text{and} \sum_{i=1}^{n} e_i^2 = \min.$$

In case of a linear model which contains a constant term we have

$$\mathbf{y} = \mathbf{x}\boldsymbol{\alpha} + \boldsymbol{\varepsilon} = \mathbf{x}_0\boldsymbol{\alpha}_0 + \mathbf{u}\boldsymbol{\alpha}_p + \boldsymbol{\varepsilon} \qquad (2)$$

with $\mathbf{x}_0 = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{p-1}), \boldsymbol{\alpha}_0^T = (\alpha_1, \alpha_2, ..., \alpha_{p-1}), \mathbf{u}^T = (1, 1, ..., 1) \in \Re^n$, $\mathbf{x} = (\mathbf{x_0}, \mathbf{u}), \boldsymbol{\alpha} = (\boldsymbol{\alpha}_0^T, \alpha_p)$.

The following result is well known in the literature:

Theorem 1.

Received by the editors: 21.10.2002.

i)For model (1) if $\mathbf{x}$ has full column rank (the $\mathbf{x}_j$ are linearly independent) the least squares estimators a_i for $\alpha_i, i=\overline{1,p}$ are uniquely defined by

$$\mathbf{a}=\left(\mathbf{x}^T\mathbf{x}\right)^{-1}\mathbf{x}^T\mathbf{y}, \mathbf{a}^T=(a_1,a_2,...,a_p)\in\Re^p.$$

ii) For model (2) if $\mathbf{x}$ has full column rank (the $\mathbf{x}_j$ are linearly independent) the least squares estimators a_i for $\alpha_i, i=\overline{1,p}$ are uniquely defined by

$$\mathbf{a}_0=(a_1,a_2,...,a_{p-1})^T=\left(\widehat{\mathbf{x}}_0^T\widehat{\mathbf{x}}_0\right)^{-1}\widehat{\mathbf{x}}_0^T\widehat{\mathbf{y}},\quad a_p=\overline{y}-\sum_{k=1}^{p-1}a_k\overline{x}_k$$

where

$$\overline{y}=\frac{1}{n}\sum_{i=1}^{n}y_i, \overline{x}_k=\frac{1}{n}\sum_{i=1}^{n}x_{ik}, \widehat{\mathbf{x}}_0=P\mathbf{x}_0, \widehat{\mathbf{y}}=P\mathbf{y}, P=I-\frac{1}{n}\mathbf{u}\mathbf{u}^T.$$

Remark 2. The Theorem 1 ii) holds for any of the conditions

$$\operatorname{rank}(\mathbf{x})=p, \text{or } \operatorname{rank}(\mathbf{x}_0)=p-1$$

because P is a linear transformation and we have

$$\operatorname{rank}(\mathbf{x})=p\Rightarrow\operatorname{rank}(\mathbf{x}_0)=p-1\Rightarrow\operatorname{rank}(P\mathbf{x}_0)=\operatorname{rank}(\widehat{\mathbf{x}}_0)=p-1.$$

2. Main results

We consider the polynomial model

$$Y=\alpha_0+\alpha_1X+...+\alpha_rX^r+\varepsilon \qquad (3)$$

with a sample data $(x_i,y_i), i=\overline{1,n}$.

By replacing $X^j=Z_j,\ j=\overline{1,r}$ the model becomes

$$Y=\alpha_0+\alpha_1Z_1+...+\alpha_rZ_r+\varepsilon.$$

According to Theorem 1, if rank$(z)=r+1$, then the fitting solution for (3) is given by

$$\mathbf{a}=\left(\widehat{\mathbf{z}}_0^T\widehat{\mathbf{z}}_0\right)^{-1}\widehat{\mathbf{z}}_0^T\widehat{\mathbf{y}},\quad a_0=\overline{y}-\sum_{k=1}^{r}a_k\overline{z}_k$$

where

$$\mathbf{z}_0=\begin{pmatrix} z_{11} & z_{12} & ... & z_{1r}\\ z_{21} & z_{22} & ... & z_{2r}\\ ... & ... & ... & ...\\ z_{n1} & z_{n2} & ... & z_{nr}\end{pmatrix},\quad \mathbf{z}=\begin{pmatrix} 1 & z_{11} & z_{12} & ... & z_{1r}\\ 1 & z_{21} & z_{22} & ... & z_{2r}\\ ... & ... & ... & ... & ...\\ 1 & z_{n1} & z_{n2} & ... & z_{nr}\end{pmatrix}.$$

In order to give for model (3) a theorem similar to Theorem 1 we search for a relation between the sample data matrix

$$\mathbf{x}=\begin{pmatrix} x_1\\ x_2\\ ...\\ x_n\end{pmatrix}$$

and "the substitution matrix"

$$\mathbf{z} = \begin{pmatrix} 1 & x_1 & x_1^2 & ... & x_1^r \\ 1 & x_2 & x_2^2 & ... & x_2^r \\ ... & ... & ... & ... & ... \\ 1 & x_n & x_n^2 & ... & x_n^r \end{pmatrix}.$$

Theorem 3. If there are at least $r+1$ distinct values of the variable X in the sample data matrix, then the least squares fitting solutions of (3) can be written uniquely as

$$\mathbf{a} = \left(\widehat{\mathbf{z}}_0^T \widehat{\mathbf{z}}_0\right)^{-1} \widehat{\mathbf{z}}_0^T \widehat{\mathbf{y}}, \mathbf{a}^T = (a_1, a_2, ..., a_r) \in \Re^r, a_0 = \overline{y} - \sum_{k=1}^{r} a_k \overline{z}_k,$$

where

$$\widehat{\mathbf{z}}_0 = P\mathbf{z}_0, \widehat{\mathbf{y}} = P\mathbf{y}, P = I - \frac{1}{n}\mathbf{u}\mathbf{u}^T$$

and z_0 is the Vandermonde type matrix with n lines, each containing the first r integer powers of the n sample values, without the column which contains the vector $\mathbf{u}^T = (1, 1, ..., 1) \in \Re^n$.

Proof. We assume that the $r+1$ distinct values of X, are the first $r+1$ values, without limiting the generality. Obviously it is necessary that $r+1 \leq n$. If rank$(\mathbf{z}) = r+1$, where $\mathbf{z}$ is the Vandermonde matrix attached to the n values data for X then the theorem holds. Thus it is enough to prove that rank$(\mathbf{z}) = r+1$.

We consider in $\mathbf{z}$ the $r+1$ order minor formed with the rows which contain the $r+1$ distinct values:

$$d = \begin{vmatrix} 1 & x_1 & x_1^2 & ... & x_1^r \\ 1 & x_2 & x_2^2 & ... & x_2^r \\ ... & ... & ... & ... & ... \\ 1 & x_{r+1} & x_{r+1}^2 & ... & x_{r+1}^r \end{vmatrix}.$$

Since the $r+1$ values of the Vandermonde discriminant d are distinct, it follows that $d \neq 0$,and rank$(\mathbf{z}) = r+1$.$\blacksquare$

We next consider the model

$$Y = f(X_1, X_2, ..., X_p) + \varepsilon \qquad (4)$$

where

$$f(X_1, X_2, ..., X_p) = \begin{cases} a_1X_1 + a_2X_2 + ... + a_pX_p \ , (X_1, X_2, ..., X_p) \in I \\ b_1X_1 + b_2X_2 + ... + b_pX_p \ , (X_1, X_2, ..., X_p) \in J \end{cases}$$

with I and J, two subsets of $\Re^p$ such as $I \cup J = \Re^p$ and $I \cap J = \emptyset$.

We use the notations:

-$\mathbf{x}_I$ the matrix containing those rows from the sample data matrix which belong to I ,as vectors in $\Re^p$

-$\mathbf{x}_J$ the matrix containing those rows from the sample data matrix which belong to J, as vectors in $\Re^p$

-$\mathbf{y}_I$ the vector containing those components y_i for which $\mathbf{x}_i = (X_1, X_2, ..., X_p)_i \in I$

-$\mathbf{y}_J$ the vector containing those components y_j for which $\mathbf{x}_j = (X_1, X_2, ..., X_p)_j \in J$.

Theorem 4. If rank$(\mathbf{x}_I) = p$, rank$(\mathbf{x}_J) = p$ and $2p \leq n$ then the least squares fitting solution of model (4) is uniquely given by
$\mathbf{a} = \left(\mathbf{x}_I^T \mathbf{x}_I\right)^{-1} \mathbf{x}_I^T \mathbf{y}_I$ and $\mathbf{b} = \left(\mathbf{x}_J^T \mathbf{x}_J\right)^{-1} \mathbf{x}_J^T \mathbf{y}_J$, with $\mathbf{a}^T, \mathbf{b}^T \in \Re^p$.
Proof. Using the least squares criteria we have

$$S = \sum_{i=1}^{n} \left[y_i - f\left(x_{i1}, x_{i2}, ..., x_{ip}\right)\right]^2 = \min,$$

$$\frac{\partial S}{\partial a_j} = 0, \frac{\partial S}{\partial b_j} = 0, j = \overline{1, p},$$

$$\frac{\partial S}{\partial a_j} = \frac{\partial}{\partial a_j} \left(\sum_{i=1}^{n} \left[y_i - f\left(x_{i1}, x_{i2}, ..., x_{ip}\right)\right]^2 \right).$$

Denoting $A = \{i \mid (x_{i1}, x_{i2}, ..., x_{ip}) \in I\}$ and $B = \{i \mid (x_{i1}, x_{i2}, ..., x_{ip}) \in J\}$ we obtain

$$\frac{\partial S}{\partial a_j} = \frac{\partial}{\partial a_j} \left(\sum_{i \in A} \left[y_i - f\left(x_{i1}, x_{i2}, ..., x_{ip}\right)\right]^2 \right) + \frac{\partial}{\partial a_j} \left(\sum_{i \in B} \left[y_i - f\left(x_{i1}, x_{i2}, ..., x_{ip}\right)\right]^2 \right) =$$

$$= 2 \sum_{i \in A} \left[y_i - f\left(x_{i1}, x_{i2}, ..., x_{ip}\right)\right] \cdot \left(- \frac{\partial f}{\partial a_j} \right) + 2 \sum_{i \in B} \left[y_i - f\left(x_{i1}, x_{i2}, ..., x_{ip}\right)\right] \cdot \left(- \frac{\partial f}{\partial a_j} \right)$$

From

$$\frac{\partial S}{\partial a_j} = 0$$

we obtain
$\sum\limits_{i \in A} \left[y_i - \left(a_1 x_{i1} + ... + a_p x_{ip}\right)\right] \cdot x_{ij} + \sum\limits_{i \in B} \left[y_i - \left(b_1 x_{i1} + ... + b_p x_{ip}\right)\right] \cdot 0 = 0.$
Finally, we have

$$\sum_{i \in A} \left[y_i - \left(a_1 x_{i1} + ... + a_p x_{ip}\right)\right] \cdot x_{ij} = 0.$$

Similarly, from

$$\frac{\partial S}{\partial b_j} = 0$$

we obtain

$$\sum_{i \in B} \left[y_i - \left(b_1 x_{i1} + ... + b_p x_{ip}\right)\right] \cdot x_{ij} = 0.$$

Then the following holds $\mathbf{x}_I^T \mathbf{x}_I \mathbf{a} = \mathbf{x}_I^T \mathbf{y}_I$, $\mathbf{x}_J^T \mathbf{x}_J \mathbf{b} = \mathbf{x}_J^T \mathbf{y}_J$. From hypothesis we have rank$(\mathbf{x}_I)$ =rank$(\mathbf{x}_J) = p$, so follows that

$$\mathbf{a} = \left(\mathbf{x}_I^T \mathbf{x}_I\right)^{-1} \mathbf{x}_I^T \mathbf{y}_I \text{ and } \mathbf{b} = \left(\mathbf{x}_J^T \mathbf{x}_J\right)^{-1} \mathbf{x}_J^T \mathbf{y}_J.$$

We observe that $\mathbf{x}_I \in \mathbf{M}_{n_1,p}$, $\mathbf{x}_J \in \mathbf{M}_{n_2,p}$, $n_1 + n_2 = n$ where
n_1 =card$\{(x_{i1}, x_{i2}, ..., x_{ip}) \mid (x_{i1}, x_{i2}, ..., x_{ip}) \in I\}$
n_2 =card$\{(x_{i1}, x_{i2}, ..., x_{ip}) \mid (x_{i1}, x_{i2}, ..., x_{ip}) \in J\}$
where "card" denotes the number of elements of a given set.

Moreover, it is necessary that $p \leq n_1$, $p \leq n_2$ so the condition $2p \leq n$ is required.■

Theorem 5. If in model (4) the function f is given on k subdomains $I_1, I_2, ..., I_k$, then the principle of the least squares leads to

$$\mathbf{a}^S = \left(\mathbf{x}_{I_S}^T \mathbf{x}_{I_S}\right)^{-1} \mathbf{x}_{I_S}^T \mathbf{y}_{I_S},$$

where

$\left(\mathbf{a}^S\right)^T = \left(a_1^S, a_2^S, ..., a_p^S\right) \in \Re^p$, $\forall s \in \overline{1,k}$ and $\mathbf{x}_{I_S}, \mathbf{y}_{I_S}$ are defined as above.

Finally we consider

$$Y = f(X) + \varepsilon \qquad (5)$$

with f a spline function of order $r, r \geq 1$ and with m nodes, $m \geq 1$.

The spline function f has the form

$$f(X) = \alpha_0 + \alpha_1 X + ... + \alpha_r X^r + \sum_{k=1}^{m} \beta_k (X - v_k)_+^r$$

where $v_1 < v_2 < ... < v_m$ are its nodes.

We denote that after substituting $X^j = Z_j, j = \overline{1,r}$, $(X - v_k)_+^r = t_k, k = \overline{1,m}$, the model becomes a linear model with $m + r$ variables and a constant term.

Remark 6.

i) We assume that the nodes $v_k, k = \overline{1,m}$ are given. These can be taken such that in any open interval generated there is at least one value from the n values given for X. In this case $m + 1 \leq n$.

ii) Also we can define a spline function whose nodes are among the sample data of X. If $m < n$ we consider m values of X increasingly ordered as nodes of f such that in any interval $(-\infty, v_1), (v_1, v_2), ..., (v_m, \infty)$ at least one values of X exists. In this case $2m + 1 \leq n$.

In the next theorem we use the notations:

$$V^r(q_1, q_2, ..., q_s) = \begin{pmatrix} q_1 & q_1^2 & ... & q_1^r \\ q_2 & q_2^2 & ... & q_2^r \\ ... & ... & ... & ... \\ q_s & q_s^2 & ... & q_s^r \end{pmatrix},$$

$$V_1^r(q_1, q_2, ..., q_s) = \begin{pmatrix} 1 & q_1 & q_1^2 & ... & q_1^r \\ 1 & q_2 & q_2^2 & ... & q_2^r \\ ... & ... & ... & ... & ... \\ 1 & q_s & q_s^2 & ... & q_s^r \end{pmatrix}$$

$$V'(q_1, q_2, ..., q_s) = \begin{pmatrix} (q_1 - v_1)_+^r & (q_1 - v_2)_+^r & ... & (q_1 - v_m)_+^r \\ (q_2 - v_1)_+^r & (q_2 - v_2)_+^r & ... & (q_2 - v_m)_+^r \\ ... & ... & ... & ... \\ (q_s - v_1)_+^r & (q_s - v_2)_+^r & ... & (q_s - v_m)_+^r \end{pmatrix}.$$

Theorem 7. If $m + r + 1 \leq n$ and among the n values of X there is at least one value situated in each of the $m+1$ open intervals delimited by nodes and there are another r distinct values situated in $(-\infty, v_1)$ then the model is uniquelly fitted by $a_j = c_j, j = \overline{1,r}$, $b_k = c_{r+k}, k = \overline{1,m}$, $\mathbf{c}^T = (c_1, ..., c_{m+r}) \in \Re^{m+r}$, $\mathbf{c} = \left(\widehat{\mathbf{z}}_0^T \widehat{\mathbf{z}}_0\right)^{-1} \widehat{\mathbf{z}}_0^T \widehat{\mathbf{y}}$ where

$$\widehat{\mathbf{z}}_0 = P\mathbf{z}_0, \widehat{\mathbf{y}} = P\mathbf{y}, P = I - \frac{1}{n}\mathbf{u}\mathbf{u}^T, \mathbf{u} = (1, 1, ..., 1) \in \Re^n,$$

$$\mathbf{z}_0 = \left(V^r(x_1, x_2, ..., x_n) : V'(x_1, x_2, ...x_s)\right).$$

Proof. We note that model (5) is a linear model with $m+r$ variables and a constant term. In order for the theorem to remain valid one of the conditions rank($\mathbf{z}$) $= m+r+1$ or rank($\mathbf{z}_0$) $= m+r$ is required. Taking into account that the rank of a matrix is not affected by swaping some rows we consider the values of variable to be ordered as $x_1 \leq x_2 \leq ... \leq x_n$. From hypothesis, there are $r+1$ distinct values in $(-\infty, v_1)$ and in the other intervals there is at least one value. Without loss of generality we take the first $m+r+1$ values such that

$$x_1, x_2, ..., x_{r+1} \in (-\infty, v_1), x_{r+2} \in (v_1, v_2), x_{r+3} \in (v_2, v_3), ..., x_{r+n+1} \in (v_m, \infty) \tag{6}$$

We denote with d the minor formed with the first $m+r+1$ rows of $\mathbf{z}$, so we have $d =$det$M$ with

$$M = \left(V_1^r(x_1, x_2, ..., x_{m+r+1}) : V'(x_1, x_2, ..., x_{m+r+1})\right).$$

Since:

$$(x_i - v_k)_+^r = \begin{cases} (x_i - v_k)^r, x_i \geq v_k \\ 0, x_i < v_k \end{cases}$$

we obtain $d =$detM' where

$$M' = \begin{pmatrix} V_1^r(x_1, x_2, ..., x_{r+1}) & O_{r+1,m} \\ V_1^r(x_{r+2}, x_{r+3}, ..., x_{m+r+1}) & V \end{pmatrix}$$

and

$$V = \begin{pmatrix} (x_{r+2} - v_1)^r & ... & 0 & 0 \\ ... & ... & ... & ... \\ (x_{m+r} - v_1)^r & ... & (x_{m+r} - v_{m-1})^r & 0 \\ (x_{m+r+1} - v_1)^r & ... & (x_{m+r+1} - v_{m-1})^r & (x_{m+r+1} - v_m)^r \end{pmatrix}$$

Further we obtain

$$d = \left[(x_{m+r+1} - v_m)^r (x_{m+r} - v_{m-1})^r \cdot ... \cdot (x_{r+2} - v_1)^r\right] \cdot$$
$$\cdot \left[(x_{r+1} - x_r)(x_{r+1} - x_{r-1}) \cdot ... \cdot (x_2 - x_1)\right].$$

Since the first $r+1$ values are distinct it follows from (6) that $d \neq 0$ and rank($\mathbf{z}$) $= m+r+1$.

Remark 8. If those m nodes are among the values of X then $2m+r+1 \leq n$.

References

[1] Stapleton J.H., *Linear Statistical Models*, Willey-Interscience, Series in Probability and Statistics, New York,1995.

[2] Saporta G., *Probabilités, analyse des données et statistique*, Editions Technip, Paris, 1990.

[3] Micula Gh., *Funcţii spline şi aplicaţii*, Editura Tehnică, Bucureşti, 1979.

Department of Mathematics, 1 Decembrie 1918 University of Alba Iulia, Romania