TECHNICAL REPORT NO. 945

February 1995

MULTI RESPONSE SPLINE REGRESSION

by

Yuh-Wen Soo  Douglas M. Bates
American Cyanamid Company  University of Wisconsin
Princeton, NJ  Madison, WI
Multiresponse Spline Regression

Yuh-Wen Soo

American Cyanamid Company, Princeton, NJ, USA

Douglas M. Bates

University of Wisconsin, Madison, WI, USA

Abstract: In this paper we propose using free-knot splines with a common knot vector to approximate multiresponse model functions and their associated error variance-covariance matrix. A "loose coupling" type of algorithm for minimizing the determinant parameter estimation criterion is constructed for multiresponse spline models. The algorithm evaluates the increment on the parameters of the common knot positions first, then evaluates the increment on spline coefficients individually. One constraint on using the determinant criterion for multiresponse estimation is that the number of design points has to be greater than the number of parameters. This makes spline models infeasible when a large number of responses are measured. We loosen the constraint on the number of parameters allowed and propose a more appropriate residual degrees of freedom for the spline models.

Keywords: Multiresponse regression; Free-knot splines; Generalized Gauss-Newton algorithm; "Loose coupling" structure.
1 Introduction

In some experimental circumstances it is possible to measure more than one response at the same experimental conditions and the same run. The response variables can have different, linear or nonlinear, functional relations with the independent variables but if the different functional relations have some common parameters they should be estimated by multireponse regression. Although it may be reasonable to assume that the noise terms in a model for the multiple responses are independent between runs, it is usually unrealistic to assume that the noise terms for different responses on the same run are independent. Box and Draper (1965) introduced a parameter estimation criterion for such multireponse data, based on the determinant of a matrix derived from the residuals. This criterion is nonlinear in the parameters and requires iterative Newton-Raphson type procedures to optimize it. Bates and Watts (1987) presented a generalized Gauss-Newton (GGN) method for this optimization. Further computational adjustments and approximate inference methods were suggested by Kang (1989). An important special class of the multireponse nonlinear model is that in which the responses are described by a linear system of differential equations. These models are used in chemical kinetics and in pharmacokinetics, where they are called compartment models. Bates and Watts (1988, chapter 4) gave a general description of estimating parameters for such models using multireponse regression techniques.

When model functions for the responses are not available, or even when model
functions are given but they are not suitable because they cannot follow some features in the data, we need approximating functions that describe the responses simultaneously. In this paper we propose using splines (de Boor, 1978) to provide a class of approximating functions which are flexible, involve few assumptions, and can describe nonlinear curves with relatively low computational complexity. Specifically, we suggest using free-knot splines to approximate the underlying model functions, and propose a modified generalized Gauss-Newton algorithm for parameter estimation using the determinant criterion. In the next section, we formulate the multiresponse spline regression model. In section 3, we give a modified generalized Gauss-Newton algorithm for optimizing the determinant for the general multiresponse model. In particular, the "loose coupling" structure of the parameters in the multiresponse spline models is utilized in the modified generalized Gauss-Newton algorithm. In section 4, we discuss one restriction on the dimensions and the determination of error degrees of freedom. Numerical examples and the results of simulations on the error degrees of freedom are given in section 5, and some discussion and conclusions are given in section 6.

2 Model

In spline regression the number of knots and their positions control the trade-off between smoothness and goodness-of-fit. If too few knots are chosen, the spline is
too smooth, while if too many knots are chosen the fitted curve begins to follow the data points too closely. Friedman and Silverman (1989) have proposed a knot selection strategy for piecewise linear fitting. By carefully defining their basis function to allow updating, Friedman and Silverman provide us with an $O(N)$ method, where $N$ is the number of design points, for selecting the number of knots and their positions in univariate data. This makes the required amount of computing time reasonable. Although Friedman and Silverman's knot selection strategy can be extended to high-order splines, it is time-consuming, since we expect there is no simple updating formula to be found. Thus, we feel that it is not worthwhile to apply a Friedman and Silverman type of knot selection procedure, especially in a multiresponse model. Usually, if the data are not very noisy, a spline with a small number of interior knots can provide a reasonable fit, so we could start by using just one or two knots in fitting multiple curves to the multiresponse data. After the number of knots is decided, the corresponding optimal knot positions can be easily obtained by using free-knot spline regression techniques.

We now formulate the multiresponse spline regression model as

$$Y_{nm} = s_m(x_n, h_m, \xi_m) + Z_{nm}, \quad n = 1, \ldots, N, m = 1, \ldots, M, \quad (1)$$

where $Y_{nm}$ is the value of the $m$-th response for the $n$-th case,

$$s_m(x_n, h_m, \xi_m) = \sum_{i=1-k}^{r_m} h_{m,i+k} B_i(x_n, \xi_m)$$
is the spline function for the \( m \)-th response, \( x_n \) is the \( n \)-th experimental setting, \( h_m \) is the spline coefficient for the \( m \)-th response, \( \xi_m \) is the knot vector for the \( m \)-th response, which defines the location of the knots and is neither constrained nor redundant (Jupp, 1978), \( r_m \) is the number of interior knots for the \( m \)-th response, \( k \) is the order of the spline, \( B_l \) is the \( l \)-th b-spline basis function (de Boor, 1978), and \( Z_{nm} \) is the disturbance term which is assumed to follow a multivariate normal distribution with the properties

\[
E[Z_{nm}] = 0
\]

and \( E[Z_{nm}Z_{ij}] = \delta_{ni}\{\Sigma\}_{mj} \), where \( \delta_{ni} \) is 1 if \( n = i \) and is 0 otherwise. That is, responses from different experimental runs are statistically independent, but the \( M \) vector of responses within each experimental run are usually statistically correlated and have a fixed, unknown variance-covariance matrix \( \Sigma \).

As in parametric multiresponse regression where the mechanistic models for the different responses may have parameters in common, we use a common knot vector in multiresponse spline regression. That is, we require that \( \xi_1 = \xi_2 = \ldots = \xi_M = \xi \) and \( r_1 = r_2 = \ldots = r_M = r \). Hence the model becomes

\[
Y_{nm} = \sum_{l=1-k}^{r} h_{m,i+k} B_l(x_n, \xi) + Z_{nm}, \quad n = 1, \ldots, N, m = 1, \ldots, M, \tag{2}
\]

and the total number of parameters is \( P = M(r + k) + r \).
3 Parameter Estimation

One approach for estimating the parameters is to find the parameter vector that minimizes the residual sum of squares from all the responses. However, this approach is appropriate only when the disturbance terms are all uncorrelated and all have equal variance. In the analysis of correlated multiresponse data, every measured response provides some information for parameter estimation. Therefore, we would expect that combining the information from all the responses will provide more precise parameter estimation. Under the assumption on the disturbance terms, a determinant parameter estimation criterion was developed by Box and Draper (1965) based upon the marginal posterior density of the unknown parameter vector $\theta$. That is, the Bayesian estimates are obtained by optimizing

$$S(\theta) = \left| Z^T Z \right|$$

(3)

with respect to $\theta$. A maximum likelihood argument can also be used (Kang, 1989) to derive this estimation criterion. As pointed out in Box and Tiao (1973), the determinant criterion is remarkably general, since it does not depend on the form of the model function. Further, the estimates are invariant under scaling transformations or even general linear transformations of the responses.

The increment, $\delta$, for optimizing the determinant $\left| Z^T Z \right|$ in parameter estimation
for the multiresponse model, is given in Bates and Watts (1987) as the solution to

\[ (H + 2gg^T) \delta = -g, \]

where \( g \) and \( H \) are the gradient and the approximate Hessian of \( \ln |Z^T Z| / 2 \). The derivation of this increment assumes that \( H + 2gg^T \) is positive definite. This is not always true, so some modifications are needed to the iterative scheme. The usual modification is to calculate the increment as

\[ \delta^i = -\left( H + 2gg^T + \alpha I \right)^{-1} g, \]

(4)

where \( \alpha \) is chosen large enough so that \( H + 2gg^T + \alpha I \) is positive definite. Such modifications are discussed in Dennis and Schnabel (1983) and Bates and Watts (1987). This may require determining the eigenvalues and eigenvectors of the approximate Hessian of \( |Z^T Z| \), which could make convergence slow because the computation of the eigenvalues and eigenvectors is usually computationally extensive. Therefore, Kang (1989) proposed a modified algorithm to reduce this difficulty. He first formed the orthogonal-triangular decomposition of the residual matrix \( Z \) as

\[ Z = QR = [Q_1, Q_2][R_1^T, 0^T]^T = Q_1 R_1, \]

where \( Q \) is an orthogonal matrix, \( Q_1 \) is the first \( M \) columns of \( Q \), \( Q_2 \) is the last \( N - M \) columns of \( Q \) and \( R_1 \) is an \( M \times M \) upper triangular matrix, and created the \( N \times M \) matrix

\[ J_p = \begin{bmatrix} J_{p1} \\ J_{p2} \end{bmatrix} = \begin{bmatrix} Q_1^T Z(p) R_1^{-1} \\ Q_2^T Z(p) R_1^{-1} \end{bmatrix}, \quad p = 1, \ldots, P, \]

(5)
where $Z_{(p)} = \partial Z/\partial \theta_{p}$. Then the $(p, q)$-th entry of $H$ becomes

$$\{H\}_{pq} = \text{tr}(J_p J_q^T) - \text{tr}(J_{p1} J_{q1}) - \text{tr}(J_{p1} J_{q1}^T)$$

$$\equiv \{A\}_{pq} - \{G_1\}_{pq} - \{G_2\}_{pq}. \quad (6)$$

Because

$$\{A\}_{pq} = \text{tr}(Z_{(p)} R_{1}^{-1} R_{1}^{-T} Z_{(q)}) = \left[\text{vec} \left(Z_{(p)} R_{1}^{-1}\right)\right]^T \left[\text{vec} \left(Z_{(q)} R_{1}^{-1}\right)\right],$$

where $\text{vec}$ is a matrix operation which stacks the columns of a matrix one under the other to form a single column, and

$$\{G_2\}_{p} = \text{tr}(Q_1^T Z_{(p)} R_{1}^{-1} R_{1}^{-T} Z_{(q)} Q_1) = \left[\text{vec} \left(Q_1^T Z_{(p)} R_{1}^{-1}\right)\right]^T \left[\text{vec} \left(Q_1^T Z_{(q)} R_{1}^{-1}\right)\right]$$

we obtain

$$\Lambda = W^T W, \quad (7)$$

$$G_2 = W_1^T W_1 \quad (8)$$

and

$$g = W^T \text{vec}(Q_1), \quad (9)$$

where

$$W = \left[\text{vec} \left(Z_{(1)} R_{1}^{-1}\right), \ldots, \text{vec} \left(Z_{(p)} R_{1}^{-1}\right)\right]$$

and

$$W_1 = \text{diag} \left(Q_1^T, \ldots, Q_1^T\right) W.$$
Hence, it is easy to see that both $\mathbf{A}$ and $\mathbf{G}_2$ are positive definite, and $\mathbf{A} - \mathbf{G}_2 = \mathbf{W}_2^T \mathbf{W}_2$, where $\mathbf{W}_2 = \text{diag} \left( \mathbf{Q}_2^T, \ldots, \mathbf{Q}_2^T \right) \mathbf{W}$, is also positive definite. Thus it is only $\mathbf{G}_1$ that causes the approximate Hessian to be non-positive definite. Kang (1989) proposed a modified iteration method, where the increment is obtained by solving

$$\mathbf{A}\delta^i = -\mathbf{g};$$

that is, $\delta^i$ is the least squares solution to $\mathbf{W}\delta = -\text{vec} (\mathbf{Q}_1)$. He suggested this modification because the magnitudes of $\mathbf{G}_1$ and $\mathbf{G}_2$ are negligible compared with $\mathbf{H}$ when $N$ is significantly larger than $M$ and $\mathbf{g}$ is also negligible near the optimum.

In practice, $N$ could be close to $M$, so eliminating $\mathbf{G}_1$ and $\mathbf{G}_2$ might cause the increment to be too short, and more iterations will be needed to reach the optimum. Therefore, we consider an increment $\delta$ determined from

$$(\mathbf{A} - \mathbf{G}_2)\delta = -\mathbf{g}. \tag{11}$$

The increment from (11) is larger than the increment from (10) since $\mathbf{G}_2$ is non-negative definite. This results in faster convergence near the optimum. However, far away from the optimum the increment from (11) can be much too large and might cause the algorithm to fail to converge. To shrink the size of the increment in the early stages, we re-introduce the factor $2\mathbf{g}\mathbf{g}^T$ in the approximate Hessian but with a factor $\alpha$ that is initially 1 and decreases to zero near the optimum. We determine the
parameter increment at the i-th iteration, \( \delta^i \), from

\[
\left( \Lambda - G_2 + \alpha_{i-1} 2gg^T \right) \delta^i = -g,
\]

where \( \alpha_0 = 1 \) and

\[
\alpha_i = \alpha_{i-1} \frac{S(\theta^{i-1}) - S(\theta^i)}{S(\theta^{i-1})}.
\]

It is easy to show that the increment for the parameters is then evaluated by

\[
\delta^i = \frac{-(\Lambda - G_2)^{-1}g}{1 + 2\alpha_{i-1}g^T(\Lambda - G_2)^{-1}g}.
\]

If the change in the objective function \( S(\theta) = |Z^TZ| \) is small then \( \alpha \) is decreased, otherwise \( \alpha \) is close to the value from the previous iteration. As \( \theta \) approaches the optimum, \( \alpha \) goes to zero, so the procedure ends with \( (\Lambda - G_2)\delta = -g \). The quantities computed at the final iteration can be used for other inferences (Kang, 1989).

**Algorithm 1 (The Modified Generalized Gauss-Newton Algorithm)**

1. Evaluate the \( N \times M \) residual matrix \( Z \) and the \( N \times M \times P \) derivative array at the current value of the parameters.

2. Decompose \( Z \) as \( QR \). Evaluate \( |Z^TZ| \) as the square of the product of the diagonal elements of \( R_1 \), the first \( M \) rows of \( R \). That is,

\[
|Z^TZ| = \prod_{m=1}^{M} \{r_{1m}\}^2.
\]
3. Post-multiply each face of the derivative array by $R^{-1}_1$ to produce

$$W = \left[ vec\left(Z_{(1)}R^{-1}_1\right), \ldots, vec\left(Z_{(P)}R^{-1}_1\right) \right]$$

and

$$W_2 = \left[ vec\left(Q_2^T Z_{(1)}R^{-1}_1\right), \ldots, vec\left(Q_2^T Z_{(P)}R^{-1}_1\right) \right],$$

where $Q_2$ is the last $N - M$ columns of $Q$.

4. Decompose $W_2$ as $ST$, a QR decomposition of $W_2$, and let $S_1$ be the first $P$ columns of $S$ and $T_1$ be the first $P$ rows of $T$.

5. Solve

$$T_1^T \left(T_1 \delta^1\right) = -W vec \left(Q_1\right)$$

for $\delta^1$ by back-substitution. Evaluate $d = \left(T_1 \delta^1\right)^T \left(T_1 \delta^1\right)$ and finally

$$\delta = \delta^1/(1 + 2\alpha d).$$

6. Evaluate the orthogonality convergence criterion (Bates and Watts, 1981)

$$\left(\frac{d}{1 + 2\alpha d}\right)/PS^2,$$

where $s^2 = \left|Z^T Z\right|/(N - P)$.

Soo (1991) compared the generalized Gauss-Newton (GGN) algorithm, Kang's algorithm and the modified algorithm by fitting compartment models to two sets of
multiresponse data using three sets of initial values for each data set. The results show that in one case the generalized Gauss-Newton algorithm fails to converge, and in other cases a non-positive definite Hessian occurred very often thus slowing down the estimation procedure as it requires the evaluation of eigenvalues at each iteration. Also the examples show that the elimination of $G_2$ is quite unreasonable since its relative magnitude, $||G_2||$, may be large compared to $||A||$. In some cases this causes Kang's algorithm to fail to converge or to take a long time to converge. This agrees with our expectation that when $N$ is close to $M$, $G_2$ has a very significant effect, and the increment from Kang's algorithm is too short. When the number of parameters is small, the modified approach and Kang's method show very similar performance. Since we usually do not know if the starting values are close to the optimal values before the optimum has been obtained, Soo (1991) concluded that the modified approach is the most reasonable choice because it requires fewer function evaluations than the other approaches.

Since the free-knot spline regression model is a multiresponse model, the modified algorithm could be used to estimate the spline coefficients and the optimal knot positions. But a “loose coupling” (Soo and Bates, 1992) type of algorithm, which leads to savings in storage and computing time, can be constructed by taking advantage of the special structure of the parameters in the spline models, even though the derivative array has a full structure.
Let $\theta = (h_1^T, \ldots, h_M^T, \xi^T)^T$ and $P = M(r + k) + r$. Form the orthogonal-triangular decomposition of $Z$ as $Z = QR = Q_1R_1$. Let $B$ be the $N \times (r + k)$ spline design matrix, then

$$W_2 = -\left[ (K^T \otimes Q_2^T B), (K^T \otimes Q_2^T \dot{B}_1), \ldots, (K^T \otimes Q_2^T \dot{B}_r) h \right]$$

where $\dot{B}_r = \partial B(x, \xi)/\partial \xi_r$, $K = R_1^{-1}$, and $h = \left( h_1^T, \ldots, h_M^T \right)^T$.

We then obtain, algebraically,

$$\delta^1_\xi = \Gamma^{-1} \left[ \begin{array}{c} \sum_{i=1}^M \sum_{j=1}^r \{K\}_{ij}h_i^T \dot{B}_i^T P q_i \\ \sum_{i=1}^M \sum_{j=1}^r \{K\}_{ij}h_i^T \dot{B}_i^T P q_i \\ \vdots \\ \sum_{i=1}^M \sum_{j=1}^r \{K\}_{ij}h_i^T \dot{B}_r^T P q_i \end{array} \right]$$

(15)

where $q_i$ is the $i$-th column of $Q_1$,

$$\{\Gamma\}_{ij} = h^T [KK^T \otimes \dot{B}_i^T PQ_2Q_2^T \dot{B}_j] h$$

and $P = I - Q_2Q_2^T B \left( B^T Q_2Q_2^T B \right)^{-1} B^T$. Also

$$\delta^1_{h_m} = \left( B^T Q_2Q_2^T B \right)^{-1} B^T \left( \sum_{i=1}^m \{R\}_{im} q_i - Q_2Q_2^T \sum_{j=1}^r \dot{B}_j h_m \delta^1_\xi \right)$$

(16)

$m = 1, 2, \ldots, M$. After obtaining $\delta^1$, we evaluate the adjustment term

$$d = \text{vec}(Q_1)^T W \left( W_2^T W_2 \right)^{-1} W^T \text{vec}(Q_1)$$

$$= \sum_{j=1}^M \left\| \sum_{i=1}^r \{K\}_{ij} Q_2^T \left( B \delta^1_{h_i} + \sum_{l=1}^r \dot{B}_l h_i \delta^1_{\xi_l} \right) \right\|^2$$

(17)
and the increment on the spline coefficients and the knot positions is evaluated by

\[ \delta = \frac{\delta^1}{1 + 2\alpha_d}. \]  

(18)

When the number of knots is properly selected, the spline function provides a good fit to the multiresponse data, and then \( \Sigma \) is estimated by the maximum likelihood estimate

\[ \hat{\Sigma}_{ML} = \frac{\hat{Z}^T\hat{Z}}{N}, \]  

(19)

or its adjustment

\[ \hat{\Sigma}_{AML} = \frac{\hat{Z}^T\hat{Z}}{\nu}, \]  

(20)

where \( \nu \) is the error degrees of freedom, which is discussed in the next section.

4 Error Degrees of Freedom

A disadvantage of using the determinant parameter estimation criterion in the general multiresponse model is that the number of design points, \( N \), has to be greater than the number of parameters, \( P \), since otherwise the determinant \( |Z^TZ| \) can be made zero by fitting any one response perfectly, or even by fitting a linear combination of the responses perfectly (Bates and Watts, 1988). By this argument \( N \) must be greater than \( M(r + k) + r \) for a multiresponse spline model, which means that a spline is not a feasible model for multiresponse data when a large number of responses are measured.
To overcome this drawback, we use the special structure of the spline where the coefficients $h_m$ only affect the $m$-th response. For example, it can be observed in the computational scheme described above that once the increment on the common knot vector is obtained, we can evaluate the increment on the spline coefficients individually for each response. Thus, although there are $M(r + k)$ spline coefficients in the spline model, they are estimated separately and require at least $r + k$ observations for each spline fit. In contrast, the $r$ common knots affect all the responses and require at least $r$ observations for this estimation. With this justification, we establish the constraint on the number of observations as $N > 2r + k$, and use $N - (2r + k)$ for the residual degrees of freedom.

Alternatively, if we express the multiresponse spline model in terms of a “plus function” representation as

$$Y_{nm} = \sum_{i=0}^{k-1} \alpha_{mi} x_n^i + \sum_{i=1}^{r} \beta_{mi} (x_n - t_i)_+^k + Z_{nm}, \quad n = 1, \ldots, N, \quad m = 1, \ldots, M,$$

where the knot vector $t = (t_1, t_2, \ldots, t_r)^T$ is fixed, then the $m$-th response is mainly affected by the parameter $(\alpha_m^T, \beta_m^T)^T$, hence we require at least $k + r$ observations for fitting fixed-knot splines (de Boor, 1978) to each response. When the common knot vector, $t$, is also unknown, each of the spline model functions for the responses consists of one individual polynomial term and a truncated power function term which depends on the parameter $(\beta_m^T, t^T)^T$. For the polynomial terms, we require at least $k$ observations to estimate the coefficients of the polynomials. It requires $r$ degrees of
freedom to estimate the optimal knot placements and, since the coefficients of the plus functions, the $\beta_m$'s, are affected by the common knot placements, it might require $Mr$ additional degrees of freedom to estimate the $\beta_m$'s. Therefore the constraint on the number of observations becomes to $N > Mr+k+r$, and this leads to $N-(Mr+k+r)$ residual degrees of freedom.

5 Examples and Simulations

In order to choose the most appropriate degrees of freedom, two simulation studies are performed. We first define the profile $t$ function (the signed square root log-likelihood ratio), $t(\theta_l)$ as

$$t(\theta_l) = \text{sign}(\theta_l - \hat{\theta}_l) \frac{\sqrt{\tilde{S}(\theta_l) - S(\hat{\theta})}}{\sqrt{S(\hat{\theta})}}$$

where

$$\tilde{S}(\theta_l) = \min_{\hat{\theta}_{-l}} S\left(\left(\theta_l, \hat{\theta}_{-l}^T\right)^T\right) = S\left(\left(\theta_l, \hat{\theta}_{-l}^T\right)^T\right)$$

is the profiled objection function and $\hat{\theta}_{-l} = (\tilde{\theta}_1, \ldots, \tilde{\theta}_{l-1}, \tilde{\theta}_{l+1}, \ldots, \tilde{\theta}_P)^T$ is the estimate of $\theta_{-l} = (\theta_1, \ldots, \theta_{l-1}, \theta_{l+1}, \ldots, \theta_P)^T$ conditional on $\theta_l$. Consider the multiresponse b-spline model with 1 free knot and $\theta = (h_1^T, \ldots, h_M^T, \xi)^T$, where $\xi$ is the parameter vector introduced by Jupp (1978) which contains the information of the knot placement. Thus the objective function $S(\theta)$ is the determinant $|Z(\theta)^TZ(\theta)|$, and the
profile t function for $\xi$ is

$$t(\xi) = \text{sign}(\xi - \hat{\xi}) \sqrt{[\hat{Z}(\xi)^T \hat{Z}(\xi)] - [\hat{Z}^T \hat{Z}]} / \sqrt{[\hat{Z}^T \hat{Z}] / \nu},$$

(24)

where $\hat{Z}(\xi)$ is the estimated residual matrix of the fixed-knot multiresponse spline model and $\hat{Z}$ is that of the free-knot multiresponse spline model. By analogy to the linear model, the profile $t$ function is approximately distributed as $t$ with $\nu$ degrees of freedom.

Example 1 (Aspartame pH3)

Aspartame is a synthetic sweetener that offers people a sweet taste in their diet without the calories of sugar. Aspartame mixed in soft drinks decomposes into some by-products. The principal product of the decomposition of aspartame (APM) is DKP (diketopiperazine). Other decomposition products are AP, $\beta$-APM, $\beta$-AP, PME and PHE. Jeffery Stamp of the University of Minnesota conducted experiments to measure the concentrations (units are mMoles $\times$ 100) of these components over time under various combinations of temperature and pH. At temperature 100°C and pH 3, the concentrations corresponding to APM and the six by-products were measured. The data are shown in Table 1.

Table 1 goes here

There were $N = 15$ observations and $M = 7$ responses. By the argument of Bates and Watts (1988), it is impossible to fit a cubic spline ($k = 4$) with one free-knot
(r = 1) model to this multireponse data using the determinant criterion since N is less than the number of parameters, 36. Using our computation scheme, we obtained  \( \hat{\xi} = -0.142 \) (knot position is 6.4244) and  \( \hat{h} \) as well as  \( \hat{\Sigma}_{ML} \). Figure 1 shows that the multiple spline fit is fairly reasonable.

**Figure 1 goes here**

We then generated a multinormal sample from  \( N(0, \hat{\Sigma}_{ML}) \) and created a simulated  \( N \times M \) data matrix  \( Y_S \). Two maximum likelihood analyses were carried out on the simulated data (with a fixed knot corresponding to  \( \xi = -0.142 \), and with a free knot), and the value of  \( t(\xi) \) was determined from 500 trials. The Kolmogorov-Smirnov test was applied to test the goodness-of-fit between the simulated  \( t(\xi) \) values and a  \( t \) distribution with  \( \nu \) degrees of freedom. The results shown in Table 2 and histograms of  \( t(\xi) \), together with the  \( t \) density curve for several values of  \( \nu \), shown in Figure 2 indicate that that  \( \nu = 3 \) is an appropriate choice.

**Table 2 goes here**

**Figure 2 goes here**

Example 2 (Aspartame pH7)

In this aspartame experiment, the concentration corresponding to APM and two by-products DKP and AP were measured at temperature 100°C and pH 7. The data are shown in Table 3.
Table 3 goes here

The number of observations, $N$, is 24 and the number of responses, $M$, is 3. We first fit a cubic ($k = 4$) spline model with $r = 1$ free-knot to this multiresponse data and obtained $\hat{\xi} = 1.03197$ and $\hat{h}$ as well as $\hat{\Sigma}_{ML}$.

The simulation study described in the last example was then repeated 1000 times. Histograms of $t(\xi)$, together with the $t$ density curve for several values of $\nu$ (shown in Figure 3), and Kolmogorov-Smirnov test (Table 2) indicate that $\nu = 16$ is an appropriate choice. The degrees of freedom used in the usual multiresponse regression, $N - P = 8$, seems not to be a suitable choice in our multiresponse spline setting.

Figure 3 goes here

The simulation studies indicate that the appropriate value of the residual degrees of freedom is

$$N - [(M + 1)r + k],$$

hence, from the viewpoint of statistical inference, the constraint on the number of observations becomes

$$N > (M + 1)r + k.$$ 

However, further theoretical investigations are needed to support these simulation studies.
6 Summary

6.1 Discussion

The determinant surface in the parameter space often has numerous local minima. To increase the chances of success in the parameter estimation, good starting values are crucial. A simple but useful principle can be used for determining starting values: By fitting a free-knot spline to the primary response, we can obtain the initial estimates of the knot positions and the spline coefficients for this response. Then, fitting fixed-knot splines, with the estimated knot locations fixed, to each one of the other responses, we can obtain the initial estimates of the remaining spline coefficients. For a uniresponse fixed-knot spline regression, the model function is a linear combination of b-spline basis functions, so the spline coefficients can be easily estimated using linear least squares.

Free-knot spline estimates often show inappropriate behavior at the boundaries of the estimation region. A possible solution to this problem is to restrict the flexibility of the splines near the boundaries by constraining all but the first derivative of the splines to be equal to zero at the boundaries. The resulting splines, known as natural splines, reduce the dimension of the parameter space from \( M(r+k)+r \) to \( M(r+2)+r \) in multiresponse spline regression.
6.2 Conclusions

Although model functions are usually provided for the analysis of multiresponse data, they can be unsuitable if they cannot follow the same features in the data. Also, the selection of a correct model is not a trivial job in the multiresponse regression problems. If we are interested in the predictive power of the fit rather than in the estimated parameters, a nonparametric type of regression technique can provide a very informative fit. We proposed using free-knot splines with a common knot vector to approximate the underlying multiresponse model functions and their associated error variance-covariance matrix. Since the multiresponse spline model is also a multiresponse model, the estimation methods proposed here and in other places can be applied to estimate the spline coefficients and the optimal knot positions. However, a more computationally efficient "loose coupling" type of algorithm was constructed by taking the advantage of the special structure of parameters in spline models which evaluates the increment on the Jupp parameter $\xi$ first then evaluates the increment on the spline coefficients $h_m$ individually.

The choice of residual degrees of freedom is an important issue in nonparametric regression. For the usual multiresponse models, the residual degrees of freedom is $N - P$, where $N$ is the number of observations and $P$ is the number of parameters, hence we require $N > P$. In our free-knot spline setting, the residual degrees of freedom would be $N - [M(r + k) + r]$, where $M$ is the number of responses, $r$ is
the number of interior knots and \( k \) is the order of splines. This leads to a constraint \( N > [M(r + k) + r] \) and makes our spline models infeasible when a large number of responses are measured. After carefully examining the spline models we find the individual spline coefficients \( h_m \) affect only the \( m \)-th response and do not affect the other responses, so we carried out simulation studies and established the constraint on the number of observations as \( N > Mr + k + r \). This leads to \( N - (Mr + k + r) \) residual degrees of freedom for the multiresponse free-knot spline models.
References


Figure 1: The multiresponse free-knot spline fit for the Aspartame (pH3) data. 6.4244 is the optimal knot position.
Table 1: *Data Set for Aspartame (pH3) Example*.

<table>
<thead>
<tr>
<th>Time</th>
<th>APM</th>
<th>DKP</th>
<th>AP</th>
<th>β-APM</th>
<th>β-AP</th>
<th>PME</th>
<th>PHE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>203.47</td>
<td>1.69</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.5</td>
<td>193.49</td>
<td>3.78</td>
<td>1.33</td>
<td>0.00</td>
<td>0.00</td>
<td>4.40</td>
<td>0.00</td>
</tr>
<tr>
<td>1.0</td>
<td>177.27</td>
<td>8.16</td>
<td>1.81</td>
<td>0.00</td>
<td>0.00</td>
<td>9.42</td>
<td>0.00</td>
</tr>
<tr>
<td>1.5</td>
<td>158.49</td>
<td>11.99</td>
<td>2.15</td>
<td>0.00</td>
<td>0.00</td>
<td>13.19</td>
<td>0.00</td>
</tr>
<tr>
<td>2.0</td>
<td>154.47</td>
<td>15.73</td>
<td>2.45</td>
<td>6.06</td>
<td>0.00</td>
<td>17.62</td>
<td>0.00</td>
</tr>
<tr>
<td>3.0</td>
<td>135.24</td>
<td>28.50</td>
<td>2.77</td>
<td>7.47</td>
<td>1.79</td>
<td>24.68</td>
<td>0.00</td>
</tr>
<tr>
<td>4.0</td>
<td>125.39</td>
<td>36.62</td>
<td>5.19</td>
<td>10.65</td>
<td>6.75</td>
<td>30.39</td>
<td>6.16</td>
</tr>
<tr>
<td>5.0</td>
<td>112.70</td>
<td>41.31</td>
<td>2.95</td>
<td>12.16</td>
<td>3.09</td>
<td>36.07</td>
<td>5.58</td>
</tr>
<tr>
<td>6.0</td>
<td>98.57</td>
<td>46.52</td>
<td>2.88</td>
<td>13.55</td>
<td>3.93</td>
<td>39.16</td>
<td>8.23</td>
</tr>
<tr>
<td>7.0</td>
<td>82.73</td>
<td>49.12</td>
<td>1.91</td>
<td>12.01</td>
<td>4.51</td>
<td>40.91</td>
<td>10.54</td>
</tr>
<tr>
<td>8.0</td>
<td>76.59</td>
<td>57.05</td>
<td>1.83</td>
<td>15.99</td>
<td>8.02</td>
<td>45.22</td>
<td>15.58</td>
</tr>
<tr>
<td>9.0</td>
<td>62.99</td>
<td>59.19</td>
<td>2.52</td>
<td>16.22</td>
<td>9.60</td>
<td>45.54</td>
<td>18.64</td>
</tr>
<tr>
<td>10.0</td>
<td>60.42</td>
<td>64.82</td>
<td>1.49</td>
<td>14.80</td>
<td>8.93</td>
<td>48.87</td>
<td>21.18</td>
</tr>
<tr>
<td>11.0</td>
<td>49.90</td>
<td>65.87</td>
<td>1.38</td>
<td>17.98</td>
<td>9.06</td>
<td>48.71</td>
<td>23.58</td>
</tr>
<tr>
<td>12.0</td>
<td>43.37</td>
<td>69.19</td>
<td>1.20</td>
<td>18.53</td>
<td>9.93</td>
<td>49.67</td>
<td>27.24</td>
</tr>
</tbody>
</table>
Table 2: Kolmogorov-Smirnov goodness-of-fit test for the simulated $t(\xi)$ values in the simulation studies.

<table>
<thead>
<tr>
<th></th>
<th>Kolmogorov-Smirnov Goodness-of-Fit Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\nu$</td>
</tr>
<tr>
<td>pH3 test statistic</td>
<td>.088</td>
</tr>
<tr>
<td>$p$-value</td>
<td>$&lt;.02$</td>
</tr>
<tr>
<td></td>
<td>$\nu$</td>
</tr>
<tr>
<td>pH7 test statistic</td>
<td>.091</td>
</tr>
<tr>
<td>$p$-value</td>
<td>$&lt;.02$</td>
</tr>
</tbody>
</table>
Table 3: Data set for the aspartame (pH7) example.

<table>
<thead>
<tr>
<th>Time</th>
<th>APM</th>
<th>DKP</th>
<th>AP</th>
<th>Time</th>
<th>APM</th>
<th>DKP</th>
<th>AP</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.000</td>
<td>208.00</td>
<td>10.70</td>
<td>0.29</td>
<td>0.100</td>
<td>94.46</td>
<td>99.37</td>
<td>8.91</td>
</tr>
<tr>
<td>0.000</td>
<td>205.23</td>
<td>10.74</td>
<td>0.28</td>
<td>0.100</td>
<td>95.87</td>
<td>99.60</td>
<td>8.97</td>
</tr>
<tr>
<td>0.017</td>
<td>172.29</td>
<td>36.78</td>
<td>2.71</td>
<td>0.133</td>
<td>64.71</td>
<td>123.30</td>
<td>11.87</td>
</tr>
<tr>
<td>0.017</td>
<td>172.14</td>
<td>37.41</td>
<td>2.75</td>
<td>0.133</td>
<td>65.11</td>
<td>123.83</td>
<td>11.92</td>
</tr>
<tr>
<td>0.033</td>
<td>149.71</td>
<td>54.36</td>
<td>4.58</td>
<td>0.167</td>
<td>45.30</td>
<td>142.44</td>
<td>13.62</td>
</tr>
<tr>
<td>0.033</td>
<td>146.33</td>
<td>53.62</td>
<td>4.51</td>
<td>0.167</td>
<td>45.21</td>
<td>142.36</td>
<td>13.60</td>
</tr>
<tr>
<td>0.050</td>
<td>127.78</td>
<td>74.81</td>
<td>6.67</td>
<td>0.200</td>
<td>38.08</td>
<td>150.87</td>
<td>14.58</td>
</tr>
<tr>
<td>0.050</td>
<td>129.78</td>
<td>74.44</td>
<td>6.65</td>
<td>0.200</td>
<td>38.11</td>
<td>150.54</td>
<td>14.54</td>
</tr>
<tr>
<td>0.067</td>
<td>105.91</td>
<td>87.10</td>
<td>7.96</td>
<td>0.250</td>
<td>26.24</td>
<td>163.85</td>
<td>15.58</td>
</tr>
<tr>
<td>0.067</td>
<td>105.19</td>
<td>86.59</td>
<td>7.91</td>
<td>0.250</td>
<td>25.93</td>
<td>164.07</td>
<td>15.47</td>
</tr>
<tr>
<td>0.083</td>
<td>93.36</td>
<td>99.96</td>
<td>9.27</td>
<td>0.417</td>
<td>5.01</td>
<td>178.51</td>
<td>17.92</td>
</tr>
<tr>
<td>0.083</td>
<td>93.46</td>
<td>99.44</td>
<td>9.18</td>
<td>0.417</td>
<td>4.97</td>
<td>177.14</td>
<td>17.76</td>
</tr>
</tbody>
</table>
Figure 2: Histograms of simulated profile $t$ values for the pH3 data. The dotted lines represent the density function of a $t$ distribution with various degrees of freedom.
Figure 3: Histograms of simulated profile t values for the pH7 data. The dotted lines represent the density function of a t distribution with various degrees of freedom.