**Spline Estimator in Multi-Response Nonparametric Regression Model**

Budi Lestari 1), I Nyoman Budiantara 2), Sony Sunaryo 2), Muhammad Mashuri 2)

1)Department of Mathematics, Faculty of Mathematics and Natural Sciences, University of Jember
2)Department of Statistics, Faculty of Mathematics and Natural Sciences, Sepuluh Nopember Institute of Technology.

**ABSTRACT**

In many applications two or more dependent variables are observed at several values of the independent variables, such as at time points. The statistical problems are to estimate functions that model their dependences on the independent variables, and to investigate relationships between these functions. Nonparametric regression model, especially smoothing splines provides powerful tools to model the functions which draw association of these variables. Penalized weighted least-squares is used to jointly estimate nonparametric functions from contemporaneously correlated data. In this paper we formulate the multi-response nonparametric regression model and give a theoretical method for both obtaining distribution of the response and estimating the nonparametric function in the model. We also estimate the smoothing parameters, the weighting parameters and the correlation parameter simultaneously by applying three methods: generalized maximum likelihood (GML), generalized cross validation (GCV) and leaving-out-one-pair cross validation (CV).

**Keywords**: Multi-response Nonparametric Regression Model, Penalized Weighted Least-Squares, Generalized Maximum Likelihood, Generalized Cross Validation, leaving-out-one-pair cross validation

**INTRODUCTION**

There are many writers who have studied spline estimators for estimating regression curve of nonparametric regression models. Kimeldorf & Wahba (1971), Craven & Wahba (1979) and Wahba (1990) proposed original spline estimator to estimates regression curve of smooth data. Cox (1983) and Cox & O'Sullivan (1996) used M-type spline to overcome outliers in nonparametric regression. Wahba (1983) proposed polynomial spline to obtain confidence interval based on posterior covariance function.


In the real cases, we are frequently faced to the problem in which two or more dependent variables are observed at several values of the independent variables, such as at time points. Multi-response nonparametric regression model provide powerful tools to model the functions which draw association of these variables. Many authors have considered nonparametric models for multiresponse data. Wegman (1981), Miller & Wegman (1987) and Flessler (1991) proposed algorithms for spline smoothing. Wahba (1992) developed the theory of general smoothing splines using reproducing kernel Hilbert spaces. Gooijer et al. (1991) and Fernez & Opsomer (2005) proposed methods of estimating nonparametric regression models with serially and spatially correlated errors, respectively. Wang et al. (2000) proposed spline smoothing for estimating nonparametric functions from bivariate data. Lestari (2007) studied spline smoothing for estimating three responses nonparametric regression models with the same variances of errors for the same response. Lestari (2008a) developed spline estimator in biresponse nonparametric regression model with unequal variances of errors and Lestari (2008b) developed penalized weighted least-squares estimator for bivariate nonparametric regression model with correlated errors. All, except Wang et al. (2000) and Lestari (2007, 2008a, & 2008b),
assumed that the covariance matrix is known, which is usually not the case in practice. When the covariance matrix is unknown, it has to be estimated from the data and this can affect the estimates of the smoothing parameters (Wang 1998).

In this paper, we study mathematical statistics methods for obtaining distribution of responses, and estimating the nonparametric functions and the parameters in the multi-response nonparametric regression model. Here, we assume that the covariance parameters are unknown, and errors of the same responses have the same variances. Based on the multi-response nonparametric regression model given, we estimate multi-response nonparametric regression function by using penalized weighted least-squares. Next, we describe three methods: generalized maximum likelihood (GML), generalized cross validation (GCV) and leaving-out-one-pair cross validation (CV) to estimate the smoothing parameters, the weighting parameters and the correlation parameter simultaneously.

RESULTS AND DISCUSSION

Multi-response nonparametric regression models

Assume that data \( \{y_{ik}, t_{ik}\} \) follows multi-response nonparametric regression model:

\[
y_{ik} = f_k(t_{ik}) + \epsilon_{ik}
\]

(1)

where \( k = 1, 2, ..., p \); \( i = 1, 2, ..., n_i \). It means that the \( i^{th} \) response of the \( k^{th} \) variable \( y_{ik} \) is generated by the \( k^{th} \) function \( f_k \) evaluated at the design point \( t_{ik} \) plus a random error \( \epsilon_{ik} \).

Assume \( \epsilon_{ik} \sim N(0, \sigma^2_k) \) for fixed \( k = 1, 2, ..., p \); \( \text{and} \) \( \text{Corr}(\epsilon_{ik}, \epsilon_{il}) = \rho_k \) for \( k \neq l \) and zero otherwise. It is a special case of our other paper, i.e., \( \text{Corr}(\epsilon_{ik}, \epsilon_{il}) = \rho_k \), which has been submitted for an international journal.

Here, for simplicity of notation, we assume that the domain of the functions are \([0,1]\) and \( f_k \) is element of Sobolev space \( W^p \), i.e., \( f_k \in W^p = \{f: f, f' \text{ absolutely continuous}, \int_0^1 (f''(t))^2 \, dt < \infty \} \). Our methods can be easily extended to the general smoothing spline models where the \( p \) domains are arbitrary (thus could be different) and the observations are linear functionals instead of evaluations (Wahba 1990, 1992).

Distribution of the responses

Suppose that we denote \( t_k = (t_{k1}, ..., t_{k_n})^T \); \( y_k = (y_{1k}, ..., y_{nk})^T \); \( \epsilon_k = (\epsilon_{1k}, ..., \epsilon_{nk})^T \); \( f_k = (f_k(t_k1), f_k(t_k2), ..., f_k(t_{k_n}))^T \); \( f = (f^T, ..., f^T) \) and \( y = (y^T, ..., y^T) \), where the superscript \( T \) refers to transpose. For for \( k = 1, 2, ..., p \), let \( r_j = \frac{\sigma_j}{m}, \quad m = \prod_{j=k}^p \sigma_j \);

\[
\theta = \prod_{k=1}^p \sigma_k;
\]

\[
\gamma_j = \frac{\rho}{\sigma_k}(i, j = 1, 2, ..., p; k \neq i \neq j);
\]

and

\( J_{qs} \) be a \( n_q \times n_s \) matrix with \((i, j)^{th} \) element equal to 1 if the \( i^{th} \) element of \( y_q \) and the \( j^{th} \) element of \( y_s \) is a pair, and zero otherwise. Note that \( J = I \), the identity matrix, when all observations come in pairs. By taking \( E(y) \) and \( Var(y) \), we obtain the distribution of responses, i.e., \( y \sim N(f, \theta W^{-1}) \), where

\[
W^{-1} = \begin{bmatrix}
    r_1 I_{n_1} & \gamma_{12} J_{12} & ... & \gamma_{1p} J_{1p} \\
    \gamma_{12} J_{12} & r_2 I_{n_2} & ... & \gamma_{2p} J_{2p} \\
    \vdots & \vdots & & \vdots
    \\
    \gamma_{1p} J_{1p} & \gamma_{2p} J_{2p} & ... & r_p I_{n_p}
\end{bmatrix}
\]

(2)
Spline estimator of multi-response nonparametric model

The nonparametric functions \( f_k \) are estimated by carrying out the following penalized weighted least-squares:

\[
\begin{align*}
\text{Min} & \quad \{(y - f)^TW(y - f)\} \\
& \quad + \lambda \int (f_k'(t))^2 dt + \lambda \int (f_k''(t))^2 dt + \ldots \\
& \quad + \lambda \int (f_k''''(t))^2 dt \\
\end{align*}
\]  
(3)

The parameters \( \lambda_k \) (\( k = 1, 2, \ldots, p \)) control the trade-off between goodness-of-fit and the smoothness of the estimates and are referred to as smoothing parameters.

We extend method as in Wang (1998) (i.e., in case of single-response nonparametric regression model) to multi-response nonparametric regression model. Let \( \phi (t) = \Gamma^{-1} \dot{\phi} (v - 1)! \), \( v = 1, 2, \ldots, p \); \( R'(s, t) = k_p(s)k_p(t) - k_p(s-t) \), where \( k_p(.) \) is the \( p^{th} \) Bernoulli polynomial. Let \( T_k = \{ \phi (t_k) \}^{n, n} \); \( T = \text{diag}(T_1, \ldots, T_p) \); \( \Sigma = \text{diag}(\Sigma_1, \ldots, \Sigma_p) \). By extending method as in both Wang (1998) and Wahba (1990) to multi-response case, we can show that for fixed \( \lambda_k, \gamma_k \) for \( k = 1, 2, \ldots, p \) and \( \rho \), the solution to (3) is

\[
\hat{f}_k(t) = \sum_{i=1}^{n} d_i \phi (t) + \sum_{i=1}^{n} c_{ij} R'(t, t_k) \\
\]  
(4)

where \( k = 1, 2, \ldots, p \); and

\[
\xi = (c_{11}, \ldots, c_{1n}, c_{21}, \ldots, c_{2n}, \ldots, c_{p1}, \ldots, c_{pn})^T; \\
d = (d_1, \ldots, d_n, d_2, \ldots, d_n, \ldots, d_p, \ldots, d_n)^T \\
\]

are solutions to

\[
\begin{pmatrix}
T'W & T' \Sigma + \text{diag}(\lambda_1, \ldots, \lambda_p) \\
\Sigma W & T \end{pmatrix} \\
\]  
(5)

Note that \( \hat{T}_k = \{ \hat{f}_k(t_{i1}), \ldots, \hat{f}_k(t_{i1}), \hat{f}_k(t_{i2}), \ldots, \hat{f}_k(t_{in}), \ldots, \hat{f}_k(t_{pn}) \} \) is always unique when \( T \) is of full column rank, which are assumed to be true in this paper. It can be verified that a solution to

\[
\begin{pmatrix}
\Sigma + W^{-1} \text{diag}(\lambda_1, \ldots, \lambda_p) \\
\end{pmatrix} \xi + T d = y \\
T' \xi = 0 \\
\]  
(6)

is also a solution to (5). Thus we need to solve simultaneous equation (6) for \( \xi \) and \( d \). In fact, \( W^{-1} \text{diag}(\lambda_1, \ldots, \lambda_p) \) is asymmetric if \( \lambda_1 \neq \lambda_i \neq \ldots \neq \lambda_p \) and \( \rho 
eq 0 \). To calculate the coefficients \( \xi \) and \( d \), we use the following transformations:

\[
\widetilde{\Sigma} = \text{diag}(\lambda_1, \ldots, \lambda_p) \\
\widetilde{\beta} = \text{diag}(\lambda_1, \ldots, \lambda_p) \xi. \\
\]  
(7)

Let \( T_k = (Q_{i1}, Q_{i2}) \left( \begin{array}{c} R_i \\ 0 \end{array} \right), k = 1, 2, \ldots, p \) be the QR decompositions. Let \( Q_1 = \text{diag}(Q_{i1}, Q_{i2}, \ldots, Q_{ip}); \)

\( Q_2 = \text{diag}(Q_{i1}, Q_{i2}, \ldots, Q_{ip}); \)

\( R = \text{diag}(R_1, R_2, \ldots, R_p); \) and \( B = \widetilde{\Sigma} + W^{-1} \). It can be shown that the solutions to (7) are

\[
\tilde{\xi} = Q_1 (Q_1^T B Q_1)^{-1} Q_1^T y, \\
\]  
(8)

\[
R d = Q_1 (y - B \tilde{\xi}) \\
\]  
(9)

Note that \( \tilde{T}_k = A y \) where

\[
A = I - W^{-1} Q_1 (Q_1^T B Q_1)^{-1} Q_1^T \\
\]  
(9)

is the “hat matrix”. Here, \( A \) is not symmetric, which is different from the usual independent case.

Estimations of parameters

We have assumed that the parameters \( \lambda_k, \gamma_k \) (for \( i, j, k = 1, 2, \ldots, p; k \neq i \neq j \)) and \( \rho \) are fixed. In practice it is very important to estimate these parameters from the data. Since observations are correlated, popular methods such as the usual generalized maximum likelihood (GML) method and the generalized cross validation (GCV) method may underestimate the smoothing parameters.
In this section we propose the following three methods to estimate the smoothing parameters $\lambda$, the weighting parameters $r$, and $\gamma$; and the correlation parameter $\rho$ simultaneously, i.e. an extension of the GML method based on a Bayesian model; an extension of the GCV method; and leaving-out-one-pair cross validation.

Wang (1998) proposed the GML and GCV methods for correlated observations with one smoothing parameter. Wang et al. (2000) proposed the GML and GCV methods for correlated observations with two smoothing parameters. In multi-response (with $p$ responses) nonparametric regression model, there are $p$ smoothing parameters which need to be estimated simultaneously together with the covariance parameters. Following an extension of derivation, we extend the GML and GCV in both Wang (1998) and Wang et al. (2000) as follows.

The GML estimates of $\lambda$, $\gamma$, $r$, and $\rho$ are minimizers of the following GML function:

$$
M(\lambda, \gamma, r, \rho) = -\frac{y^T W(I - A)y}{\left[ \text{det}'(W(I - A)) \right]^{1/4}}
= -\frac{\tilde{z}^T (Q^T B Q)^{-1} \tilde{z}}{\left[ \text{det}(Q^T B Q) \right]^{1/4}}
$$

where $n = n_1 + n_2 + ... + n_p$; $\text{det}'$ is the product of the nonzero eigen values and $\tilde{z} = Q^T y$. The minimizers of $M(\lambda, \gamma, r, \rho)$ are called GML estimates.

The GCV estimates of $\lambda$, $\gamma$, $r$, and $\rho$ are minimizers of the following GCV function:

$$
V(\lambda, \gamma, r, \rho) = \frac{\| W(I - A)y \|}{\left[ \text{Tr}(W(I - A)) \right]^{1/4}}
= \frac{\tilde{z}^T (Q^T B Q)^{-1} \tilde{z}}{\left[ \text{Tr}(Q^T B Q) \right]^{1/4}}\quad (11)
$$

In the following we propose a cross validation method based on leaving-out-one-pair procedure. Suppose there are a total of $N$ ($N \geq \max \{n_1, n_2, ..., n_p\}$) distinct time points and thus $N$ pairs of observations. Any one observation in a pair may be missing. These pairs are numbered from 1 to $N$. We use the following notation: superscripts $(i)$ to denote the collection of elements corresponding to the $i^{th}$ pair; superscripts $[i]$ to denote the collection of elements after deleting the $i^{th}$ pair; superscripts $(i)$ to denote solution of $f_i$ without the $i^{th}$ pair. When one observation in a pair is missing, superscripts indicate a single observation instead of a pair. The solutions to:

$$
\begin{align*}
\text{Min}_1^{N} \{(y_i^{(1)} - f_i^{(1)})^T W_1^{(1)} (y_i^{(1)} - f_i^{(1)}) + \\
\lambda_1 \int_0^1 (f_i^{(1)}(t))^2 dt + \lambda_2 \int_0^1 (f_i^{(1)}(t))^2 dt + ... + \\
\lambda_N \int_0^1 (f_i^{(1)}(t))^2 dt \} \quad (12)
\end{align*}
$$

are $\hat{f}_i^{(1)}$, $\hat{f}_i^{(2)}$, ..., $\hat{f}_i^{(N)}$. Assume that there are $p$ elements in the $i^{th}$ pair (it is simple if there is only one). Denote $i_1$, $i_2$, ..., $i_p$ as the row numbers of this pair in $\nu_1$, $\nu_2$, ..., $\nu_p$, respectively. Define:

$\hat{y}_i = \left\{ \begin{array}{l}
y_{i,j}, j \neq i_i \\
\hat{f}_i^{(1)}(t_{i_1}), j = i_1, k = 1,2, ..., p
\end{array} \right.$

Suppose that we denote $\nu^*_i = (\nu_1^*, ..., \nu_p^*)^T$, $\nu^*_i = (\nu_1^T, ..., \nu_p^T)^T$, and $\hat{f}_i^{(1)}(t_i)$ = $(\hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_p}), ...)$, $\hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_p}), ...$, $\nu^*_i$ = $(\nu_1^*, ..., \nu_p^*)^T$, $\nu^*_i = (\nu_1^T, ..., \nu_p^T)^T$, and $\hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_p}), ...$, $\nu^*_i$ = $(\nu_1^*, ..., \nu_p^*)^T$, $\nu^*_i = (\nu_1^T, ..., \nu_p^T)^T$, and $\hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_p}), ...$, $\nu^*_i$ = $(\nu_1^*, ..., \nu_p^*)^T$, $\nu^*_i = (\nu_1^T, ..., \nu_p^T)^T$, and $\hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_p}), ...$. Then we have the following leaving-out-one-pair lemma.

Lemma. For fixed $\lambda$, $\gamma$, $r$, $\rho$, and $i$, we have $\hat{f}_i^{(1)}(t_i) = A \nu_i^*$

**Proof:**

Let $f(t) = (f_1(t_i), ..., f_p(t_i), f_1(t_{i_1}), ..., f_2(t_{i_1}), ..., f_p(t_{i_1}), f_1(t_{i_2}), ..., f_p(t_{i_2}), ..., f_2(t_{i_2}), ..., f_p(t_{i_2}), ..., f_2(t_{i_3}), ..., f_p(t_{i_3}), ..., f_2(t_{i_3}), ..., f_p(t_{i_3}), ..., f_2(t_{i_4}), ..., f_p(t_{i_4}), ..., f_2(t_{i_4}), ..., f_p(t_{i_4}), ..., f_2(t_{i_5}), ..., f_p(t_{i_5}), ..., f_2(t_{i_5}), ..., f_p(t_{i_5}), ...)$ and $\hat{f}_i^{(1)}(t_i) = (\hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_1}), ..., \hat{f}_i^{(1)}(t_{i_2}), ..., \hat{f}_i^{(1)}(t_{i_2}), ..., \hat{f}_i^{(1)}(t_{i_3}), ..., \hat{f}_i^{(1)}(t_{i_3}), ..., \hat{f}_i^{(1)}(t_{i_4}), ..., \hat{f}_i^{(1)}(t_{i_4}), ..., \hat{f}_i^{(1)}(t_{i_5}), ..., \hat{f}_i^{(1)}(t_{i_5}), ...$. Similarly define $f(t_i)$ and $\hat{f}_i^{(1)}(t_i)$ as $f(t)$ and $\hat{f}_i^{(1)}(t_i)$ respectively without the elements corresponding to the $i^{th}$ pair. For any function $f_1, f_2, ..., f_p$ in $W_i$, we have :
The last section is called cross validation. Therefore

\[ f + \sum_{i=1}^{n} \left( f^*(i) - f^*(i) \right)^T W \left( f^*(i) - f^*(i) \right) \]

\[ \geq \left( y^* - f^*(i) \right)^T W \left( y^* - f^*(i) \right) \]

The second inequality holds because \( f_1, \ldots, f_p \) are solutions to (12). The last equality holds because of the definition of \( y^* \).

The inequality at (13) indicates that \( f_1, \ldots, f_p \) are solutions to (3) with \( y \) replaced by \( y^* \). Therefore \( f^*(i) = A y^* \).

As a consequence of this lemma, we do not need to solve separate minimization problems (12) for each deleting-one-pair set. All we need to do is to solve the following equations

\[
\begin{pmatrix}
m_{11} & m_{12} & \ldots & m_{1p} \\
m_{21} & m_{22} & \ldots & m_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
m_{p1} & m_{p2} & \ldots & m_{pp}
\end{pmatrix}
\begin{pmatrix}
s_1 \\ s_2 \\ \vdots \\ s_p
\end{pmatrix}
= \begin{pmatrix}
u_1 \\ u_2 \\ \vdots \\ u_p
\end{pmatrix}
\]

(14)

for \( f_k^*(i) - y_{ik} \), where \( m_{11} = 1 - a(i, i) \);

\[ m_{1n} = -a(i, n+i) ; \ldots ; m_{pn} = -a(n+i, i) ; \]

\[ m_{111} = -a(n_i, i) ; \]

\[ m_{121} = 1 - a(n_i, i) ; \]

\[ \ldots ; m_{p1} = -a(n_i, i) ; \]

\[ m_{p2} = -a(n_i, i) ; \]

\[ \ldots ; \]

\[ m_{m_p} = -a(n_i, i) ; \]

\[ s_1 = f_k^*(i) - y_{ik} ; \]

\[ s_2 = f_k^*(i) - y_{ik} ; \]

\[ \ldots ; s_p = f_k^*(i) - y_{ik} ; \]

\[ u_1 = f_k^*(i) - y_{ik} ; \]

\[ \ldots ; u_p = f_k^*(i) - y_{ik} ; \]

\[ a(i, j) \text{ are elements of the matrix A. If there is only one} \]

\[ \text{observation in the } i^{th} \text{ pair, for example } y_{ik} \text{, we} \]

\[ \text{then have the following equation} \]

\[ (1-a(i, i))(f_k^*(i) - y_{ik}) = f_k^*(i) - y_{ik} \]

(15)

Note that (15) is exactly the same as the "leaving-out-one" lemma in the independent case.

Let \( f^\sim = (f^\sim_1 \ldots f^\sim_p) \) and \( i^\sim = (i^\sim_1 \ldots i^\sim_p) \), where \( i^\sim_k \) denotes the index of the pair for observation \( y_{ik} \). Define

\[ C \left( \mathbf{\lambda}^\sim, \mathbf{y}^\sim, \mathbf{r}^\sim, \rho \right) = \frac{1}{n} \left\| F \left( y - f^\sim \right) \right\|^2 \]

(16)

Here, \( C \) estimates the weighted mean-square errors (WMSE) (Wang 1998). The minimizers of \( C \left( \mathbf{\lambda}^\sim, \mathbf{y}^\sim, \mathbf{r}^\sim, \rho \right) \) are called cross validation estimates of the parameters.

CONCLUSION

The distribution of vector responses \( y \) is Multivariate Normal with mean \( f \) and variance \( \sigma W \). General smoothing spline models provide flexibility for estimating nonparametric functions and are widely used in many areas. With multiple correlated responses it is better to estimate these functions jointly using the penalized weighted least-squares.
DAFTAR PUSTAKA


