

On Multiresponse Semiparametric Regression Model

¹Wahyu Wibowo, ²Sri Haryatmi and ¹I. Nyoman Budiantara

¹Department of Statistics, Faculty of Mathematics and Natural Sciences,
Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

²Department of Mathematics, Faculty of Mathematics and Natural Sciences,
Gadjah Mada University, Yogyakarta, Indonesia

Received 2012-08-13, Revised 2012-12-04; Accepted 2013-02-20

ABSTRACT

The modeling between predictors and response in statistics sometimes deals with more than one response or multiresponse situation. Furthermore, it can be happen that some predictors have linear relationship with the responses and the others predictor have unknown relationship. To overcome this modeling problem we proposed multiresponse semiparametric regression model. This model has more than one response and contains both parametric and nonparametric model. This study focuses on how to estimate parameter in multiresponse semiparametric regression. The weighted penalized least squares method is used to fit the model. This method produce partial spline estimator for nonparametric model and by applying some assumptions the estimator is polynomial natural spline. The performance of this estimator depends on smoothing parameter. So, we also proposed G criteria as modification of generalized cross validation in the context of multiresponse semiparametric regression to choose the optimal smoothing parameter. Using simulation data, it can be shown that this model can work well to describe relationship between some predictors and several responses.

Keywords: Weighted Penalized Least Square, Multiresponse Semiparametric Regression, Partial Spline, Generalized Cross Validation

1. INTRODUCTION

Statistical modeling especially in regression analysis sometimes deals with two or more response variables. For example, bank manager wants to know impact of credit size on company performance. Company performance can be measured based on hours of activity, labor, asset, revenue and benefit. Another example is from sport. A coach wants to measure impact of exercise duration on quality of fitness. Fitness can be measured using endurance, speed, power, strength, flexibility and balance. In addition, in the field of emission gas analysis, an engineer wants to know how impact of vehicle age on emission gas, which can be measured using concentration of CO₂, CO, NO_x and SO_x.

Several authors have developed multiresponse statistical modeling. Wegman (1981); Miller and Wegman (1987) and Fessler (1991) developed algorithm for spline smoothing. Yee and Wild (1996) used additive model for multiresponse data from exponential family. Soo and Bates (1996) developed multiresponse spline regression by using Gauss-Newton algorithm as method for estimating model's parameter. Wang *et al.* (2000) developed bivariate spline nonparametric regression using assumption that covariance matrix is unknown and they proved that the joint estimates have smaller posterior variances than those of function-by-function estimates and are therefore more efficient. You *et al.* (2007) used two stages estimation on nonparametric seemingly unrelated regression.

Corresponding Author: Wahyu Wibowo, Department of Statistics, Faculty of Mathematics and Natural Sciences, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

Semiparametric regression model is also called partially linear model. This model is fusion between parametric and nonparametric regression. It is more flexible than fully parametric or nonparametric model because some of the predictors are linear and the rest being modeled nonparametrically. Engle *et al.* (1986) used this model firstly and for several decades it has been developed extensively and its application spread in many field. Further detailed about application of semiparametric model can be seen in Hardle *et al.* (2000). Lin and Carroll (2001) used semiparametric model for data clustering by developing semiparametric partially generalized linear model. Next, Qin and Zhu (2008) and Qin *et al.* (2009) used semiparametric model for longitudinal data. You and Zhou (2009) used polynomial spline for semiparametric regression in panel data. In addition, Ruppert *et al.* (2009) reviewed semiparametric regression development and its application during 2003-2007.

The estimator in semiparametric regression can be obtained by penalized least square method. This estimator is solution of optimization problem below Equation 1:

$$\text{Min}_{f \in W_2^2[a,b]} \{R(f) + \lambda J(f)\} \tag{1}$$

where, $R(f)$ and $J(f)$ indicate goodness of fit and smoothness level respectively. The quantity λ controls the tradeoff between goodness of fit $R(f)$ and smoothness $J(f)$. The quantity λ is also called smoothing parameter.

Wahba (1990) has showed that if the goodness of fit is sum of squares residual and smoothness level is integral of second derivative square, then the estimator which minimize (1) in the context of semiparametric regression is partial spline. However, this estimator still depends on smoothing parameter λ . If λ value is small ($\lambda \rightarrow 0$), then the estimator tend to interpolate the data, fitting every data point exactly. Conversely, if λ value is large ($\lambda \rightarrow \infty$), then the estimator tend to be forced linear regression. So, the optimal smoothing parameter must be chosen to get the best estimator. Several authors had developed method for choosing the optimal smoothing parameter, such as Wahba (1990) that proposed generalized cross validation method. Wang (1998) also proposed Unbiased Risk (UBR) criteria to choose the optimal smoothing parameter.

Motivated by applied of multiresponse statistical modeling and the result of partial spline as in Wahba (1990), in this study we proposed multiresponse semiparametric regression model. Regarding to this model, the study focused on how to estimate parameter of the model. After that, we consider G criteria as method to

choose the optimal smoothing parameter. This criteria is modification of generalized cross validation in the context of multiresponse semiparametric regression. Then, parameter estimation and choosing of optimal smoothing parameter will be demonstrated using the simulation data.

2. MATERIALS AND METHODS

2.1. Statistical Model

Suppose there are n sample random and each sample is observed for r response and some predictors. Suppose y_{ji} denote response j th and observation i th, with $j = 1, 2, \dots, r$ and $i = 1, 2, \dots, n$. It is assumed that response y_{ji} have linear relationship with some predictor and unknown relationship with one predictor. Denote $\underline{x}_{ji} = (x_{ji1}, x_{ji2}, \dots, x_{jip})$ as vector dimension $1 \times p$ is predictor variable and have linear model with response y_{ji} as well as $t_{ji} \in [a, b]$ as predictor variable have unknown relationship with response.

Based on above data construction, multiresponse regression model is constructed below Equation 2:

$$y_{ji} = \underline{x}_{ji} \beta_j + f_j(t_{ji}) + \varepsilon_{ji}, \quad j = 1, \dots, r; i = 1, \dots, n \tag{2}$$

$\varepsilon_{ji}, i = 1, 2, \dots, n$ are random error independently with mean zero and variance σ_j^2 . There is correlation between error ε_{ji} and ε_{ki} . $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jp})^T \in R^p$ is $p \times 1$ parameter model as parametric component. Due to unknown relationship predictor t_{ji} and response y_{ji} is, so $f_j(t_{ji})$ is unknown function as nonparametric component and is only assumed smooth and continuous on interval $[a_j, b_j]$. Now, model (2) can be stated in matrix notation Equation 3:

$$\underline{y} = X \underline{\beta} + \underline{f} + \underline{\varepsilon} \tag{3}$$

Where:

$$\underline{y} = (\underline{y}_1^T, \underline{y}_2^T, \dots, \underline{y}_r^T)^T, \quad \underline{y}_j = (y_{j1}, y_{j2}, \dots, y_{jn})^T, \quad j = 1, 2, \dots, r$$

$$X = \text{diag}(X_1, X_2, \dots, X_r), \quad j = 1, 2, \dots, r \text{ where}$$

$$X_j = \begin{pmatrix} x_{j11} & x_{j12} & \dots & x_{j1p} \\ x_{j21} & x_{j22} & \dots & x_{j2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{jn1} & x_{jn2} & \dots & x_{jn p} \end{pmatrix}$$

$$\underline{\beta} = (\beta_1^T, \beta_2^T, \dots, \beta_r^T)^T, \quad \beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jp})^T$$

$$\underline{f} = (f_1^T, f_2^T, \dots, f_r^T)^T, \text{ where}$$

$$f_j(t_j) = (f_j(t_{j1}), f_j(t_{j2}), \dots, f_j(t_{jn}))^T$$

$$\underline{\varepsilon} = (\varepsilon_1^T, \varepsilon_2^T, \dots, \varepsilon_r^T)^T, \varepsilon_j = (\varepsilon_{j1}, \varepsilon_{j2}, \dots, \varepsilon_{jn})^T,$$

$$j = 1, 2, \dots, r$$

2.2. Estimation Method

The estimation method of the model’s parameter is the weighted penalized least square. Based on this method and model (3), it is defined a goodness of fit $R(\underline{\beta}, \underline{f})$ and smoothness $J(\underline{f})$ respectively:

$$R(\underline{\beta}, \underline{f}) = (\underline{y} - X\underline{\beta} - \underline{f})^T B^{-1}(\underline{\sigma}^2) (\underline{y} - X\underline{\beta} - \underline{f})$$

$$J(\underline{f}) = (\|P_1 f_1\|^2, \|P_1 f_2\|^2, \dots, \|P_1 f_r\|^2)^T$$

where, $f_j \in H, j = 1, 2, \dots, r$ for some Hilbert space H . H can be decomposed to be $H = H_0 \oplus H_1$ where $H_0 \perp H_1$, as in Wahba (1990). P_1 is orthogonal projection f_j onto H_1 in H space. $B^{-1}(\underline{\sigma}^2)$ is weighted for goodness of fit and it is from variance covariance matrix between response and its element is defined:

$$B^{-1}(\underline{\sigma}^2) = \begin{bmatrix} B_{11}(\sigma_1^2) & B_{12}(\sigma_{12}) & \dots & B_{1r}(\sigma_{1r}) \\ B_{21}(\sigma_{21}) & B_{22}(\sigma_2^2) & \dots & B_{2r}(\sigma_{2r}) \\ \vdots & \vdots & \dots & \vdots \\ B_{r1}(\sigma_{r1}) & B_{r2}(\sigma_{r2}) & \dots & B_{rr}(\sigma_r^2) \end{bmatrix}$$

Where:

$$\underline{\sigma}^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_r^2)$$

$$B_{jj}(\sigma_j^2) = \begin{bmatrix} \sigma_j^2 & 0 & \dots & 0 \\ 0 & \sigma_j^2 & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_j^2 \end{bmatrix}, j = 1, 2, \dots, r$$

$$B_{jk}(\sigma_{jk}) = B_{kj}(\sigma_{kj}) = \begin{bmatrix} \sigma_{jk} & 0 & \dots & 0 \\ 0 & \sigma_{jk} & \dots & 0 \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & \sigma_{jk} \end{bmatrix}$$

$$j \neq k; j, k = 1, 2, \dots, r$$

Next, using $R(\underline{\beta}, \underline{f})$ and $J(\underline{f})$, it is defined Lagrange function Equation 4:

$$R(\underline{\beta}, \underline{f}) + \underline{\lambda}J(\underline{f}) \tag{4}$$

Where:

$$\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_r)$$

Estimator of $\underline{\beta}$ and \underline{f} are minimizers of optimization problem below Equation 5:

$$(\underline{y} - X\underline{\beta} - \underline{f})^T B(\underline{\sigma}^2) (\underline{y} - X\underline{\beta} - \underline{f}) + \sum_{j=1}^r \lambda_j \|P_1 f_j\|^2 \tag{5}$$

3. RESULTS

3.1. Parameter Estimation

It would be shown that solutions of optimization (5) are estimator for $\underline{\beta}$ and \underline{f} . They are Equation 6 and 7:

$$\hat{\underline{f}} = T\underline{\hat{\gamma}} + \Sigma \hat{\underline{\alpha}} \tag{6}$$

$$\hat{\underline{\beta}} = (X^T (I - A(\underline{\lambda}))^T B X)^{-1} X^T (I - A(\underline{\lambda}))^T B \underline{y} \tag{7}$$

Where:

$$T = \text{diag}(T_1, T_2, \dots, T_r), T_j = \{\phi_{jv}(t_{ji})\},$$

$$j = 1, 2, \dots, r; i = 1, 2, \dots, n; v = 1, 2, \dots, m$$

$$\Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_r); k, l = 1, 2, \dots, n;$$

$$\Sigma_j = \{ \langle \psi_{jk}, \psi_{jl} \rangle \}, M = B \Sigma + n \Lambda$$

$$\Lambda = \text{diag}(\lambda_1 I_n, \lambda_2 I_n, \dots, \lambda_r I_n)$$

$$\hat{\underline{\gamma}} = (\hat{\gamma}_1^T, \hat{\gamma}_2^T, \dots, \hat{\gamma}_r^T)^T = (T^T M^{-1} B T)^{-1} T^T M^{-1} B (\underline{y} - X \hat{\underline{\beta}})$$

$$\hat{\underline{\alpha}} = (\hat{\alpha}_1^T, \hat{\alpha}_2^T, \dots, \hat{\alpha}_n^T)^T$$

$$= M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B) (\underline{y} - X \hat{\underline{\beta}})$$

$$A(\underline{\lambda}) = T (T^T M^{-1} B T)^{-1} T^T M^{-1} B + \Sigma M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B)$$

To show the result of (6) and (7), it was started from properties of element in Hilbert space. If $f_j \in H, j = 1, 2, \dots, r$ for some Hilbert space H , then f_j can be stated $f_j = f_{j0} + f_{j1}$ where $f_{j0} \in H_0$ and $f_{j1} \in H_1$. The f_{j1} component

is orthogonal projection of f onto H_1 in space H . Suppose that $\phi_{j1}, \phi_{j2}, \dots, \phi_{j_m}$ are basis for H_0 sub space, such that for each function $f_{j0} \in H_0$ can be stated Equation 8:

$$f_{j0} = \delta_{j1}\phi_{j1} + \delta_{j2}\phi_{j2} + \dots + \delta_{j_m}\phi_{j_m}, \quad j=1,2,\dots,r \tag{8}$$

Next, a $n \times m$ matrix T_j is defined, with:

$$T_j = \{\phi_{jk}(t_{ji})\}, \quad i=1,2,\dots,n; \quad k=1,2,\dots,m$$

Meanwhile, consider $\psi_{j1}, \psi_{j2}, \dots, \psi_{j_n}$ as basis for H_1 . So, for each $f_{j1} \in H_1$, it can be expressed Equation 9:

$$f_{j1} = \alpha_{j1}\psi_{j1} + \alpha_{j2}\psi_{j2} + \dots + \alpha_{j_n}\psi_{j_n}, \quad j=1,\dots,r \tag{9}$$

Due to $f_j = f_{j0} + f_{j1}$, then for every $f_j \in H$, it can be stated:

$$f_j(t) = \sum_{k=1}^m \gamma_{jk} \phi_{jk}(t) + \sum_{i=1}^n \alpha_{ji} \psi_{ji}(t) \tag{10}$$

In matrix notation, the Equation (10) can be stated:

$$\underline{f}_j = T_j \underline{\gamma}_j + \Sigma_j \underline{\alpha}_j$$

Where:

$$\Sigma_j = \{\langle \psi_{jk}, \psi_{ji} \rangle\}; k,l=1,2,\dots,n$$

Generally, for all j , the nonparametric component can be presented Equation 11:

$$\underline{f} = T\underline{\gamma} + \Sigma\underline{\alpha} \tag{11}$$

Where:

$$\begin{aligned} T &= \text{diag}(T_1, T_2, \dots, T_r), \quad \Sigma = \text{diag}(\Sigma_1, \Sigma_2, \dots, \Sigma_r), \\ \underline{\alpha} &= (\alpha_1^T, \alpha_2^T, \dots, \alpha_r^T)^T, \quad \underline{\alpha}_j = (\alpha_{j1}, \alpha_{j2}, \dots, \alpha_{jn})^T, \\ \underline{\gamma} &= (\gamma_1^T, \gamma_2^T, \dots, \gamma_r^T)^T, \quad \underline{\gamma}_j = (\gamma_{j1}, \gamma_{j2}, \dots, \gamma_{jm})^T, \quad j=1,2,\dots,r \end{aligned}$$

It should be noted that for quantity $J(\underline{f})$ and $j = 1,2,\dots,r$:

$$\begin{aligned} \|\underline{P}_1 \underline{f}_j\|^2 &= \langle \underline{P}_1 \underline{f}_j, \underline{P}_1 \underline{f}_j \rangle \\ &= \langle \underline{P}_1 (\underline{f}_{j0} + \underline{f}_{j1}), \underline{P}_1 (\underline{f}_{j0} + \underline{f}_{j1}) \rangle \\ &= \underline{\alpha}_j^T \Sigma_j \underline{\alpha}_j \end{aligned}$$

So, the quantity $J(\underline{f})$ can be presented:

$$J(\underline{f}) = (\underline{\alpha}_1^T \Sigma_1 \underline{\alpha}_1, \lambda_2 \underline{\alpha}_2^T \Sigma_2 \underline{\alpha}_2, \dots, \underline{\alpha}_r^T \Sigma_r \underline{\alpha}_r)^T \tag{12}$$

Next, based on Equation (5), (11) and (12), it is defined $Q(\underline{\beta}, \underline{\alpha}, \underline{\gamma})$ equal to:

$$= (\underline{y} - X\underline{\beta} - T\underline{\gamma} - \Sigma\underline{\alpha})^T B(\underline{\sigma}^2) (\underline{y} - X\underline{\beta} - T\underline{\gamma} - \Sigma\underline{\alpha}) + \underline{\alpha}^T \Lambda \Sigma \underline{\alpha} \tag{13}$$

Where:

$$\Lambda = \text{diag}(\lambda_1 I_n, \lambda_2 I_n, \dots, \lambda_r I_n)$$

Minimizing of Equation (13) can be done by differentiating over each parameter that will be estimated. Firstly, it will be derived estimator for \underline{f} . For this purpose, the $\underline{z} = \underline{y} - X\underline{\beta}$ is defined as constant. So, Equation (13) becomes:

$$\begin{aligned} Q(\underline{\alpha}, \underline{\gamma}) &= (\underline{z} - T\underline{\gamma} - \Sigma\underline{\alpha})^T B(\underline{z} - T\underline{\gamma} - \Sigma\underline{\alpha}) + n\underline{\alpha}^T \Lambda \Sigma \underline{\alpha} \\ \frac{\partial Q(\underline{\alpha}, \underline{\gamma})}{\partial \underline{\alpha}} &= -\underline{z}^T B \Sigma + \underline{\gamma}^T T^T B \Sigma - \Sigma^T B \underline{z} + \Sigma^T B T \underline{\gamma} + \\ &2 \Sigma^T B \Sigma \underline{\alpha} + 2n \Lambda \Sigma \underline{\alpha} = 0 \\ B T \underline{\gamma} + (B \Sigma + n \Lambda) \underline{\alpha} &= B \underline{z} \end{aligned}$$

It is defined that $M = B \Sigma + n \Lambda$, so Equation 14:

$$\underline{\alpha} = M^{-1} B (\underline{z} - T\underline{\gamma}) \tag{14}$$

$$B T \underline{\gamma} = B \underline{z} - M \underline{\alpha} = B (\underline{y} - X \underline{\beta}) - M \underline{\alpha}$$

$$\begin{aligned} \frac{\partial Q(\underline{\alpha}, \underline{\gamma})}{\partial \underline{\gamma}} &= -\underline{z}^T B T - T^T B \underline{z} + 2 T^T B T \underline{\gamma} + \\ T^T B \Sigma \underline{\alpha} + \underline{\alpha}^T \Sigma^T B T &= 0 \\ T^T B T \underline{\gamma} + T^T B \Sigma \underline{\alpha} &= T^T B \underline{z} \end{aligned}$$

Due to $M = B \Sigma + n \Lambda$, then:

$$B \Sigma = M - n \Lambda$$

and:

$$B \Sigma M^{-1} = (M - n \Lambda) M^{-1} = I - n \Lambda M^{-1}$$

So:

$$\begin{aligned} \underline{\gamma} &= (T^T M^{-1} B T)^{-1} T^T M^{-1} B \underline{z} \\ \underline{\alpha} &= M^{-1} B (\underline{z} - T \underline{\gamma}) \\ &= M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B) \underline{z} \\ &\text{with } \underline{z} = \underline{y} - X \underline{\beta} \end{aligned}$$

It should be noted that solution for $\underline{\gamma}$ and $\underline{\alpha}$ imply function f can be seen as smoothing result for $\underline{y} - X \underline{\beta}$. The estimator of f can be stated Equation 15:

$$\hat{f} = [T (T^T M^{-1} B T)^{-1} T^T M^{-1} B + \Sigma M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B)] (\underline{y} - X \underline{\beta}) = A(\lambda) (\underline{y} - X \underline{\beta}) \tag{15}$$

Where:

$$A(\lambda) = \frac{T (T^T M^{-1} B T)^{-1} T^T M^{-1} B + \Sigma M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B)}{\Sigma M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B)}$$

In other side, substitute Equation (11) to (14) will produce:

$$\begin{aligned} f &= \underline{y} - X \underline{\beta} - B^{-1} M \underline{\alpha} + \Sigma \underline{\alpha} = \underline{y} - X \underline{\beta} - B^{-1} n \Lambda \underline{\alpha} \\ n \Lambda \underline{\alpha} &= B (\underline{y} - X \underline{\beta} - \hat{f}) \\ &= B (I - A(\lambda)) (\underline{y} - X \underline{\beta}) \end{aligned}$$

Next, it can be shown that Equation 16:

$$\begin{aligned} T^T \underline{\alpha} &= T^T M^{-1} B (I - T (T^T M^{-1} B T)^{-1} T^T M^{-1} B) \underline{z} = 0 \\ \Sigma \underline{\alpha} &= \hat{f} - T \underline{\gamma} = A(\lambda) (\underline{y} - X \underline{\beta}) - T \underline{\gamma} \\ n \Lambda \underline{\alpha}^T \Sigma \underline{\alpha} &= n \lambda \underline{\alpha}^T (A(\lambda) (\underline{y} - X \underline{\beta}) - T \underline{\gamma}) \\ Q(\beta) &= (\underline{y} - X \underline{\beta} - \hat{f})^T (\underline{y} - X \underline{\beta} - \hat{f}) + n \lambda \underline{\alpha}^T \Sigma \underline{\alpha} \tag{16} \\ \frac{\partial Q(\beta)}{\partial \beta} &= -X^T (I - A(\lambda))^T B \underline{y} + X^T (I - A(\lambda))^T B X \underline{\beta} = 0 \\ \hat{\beta} &= (X^T (I - A(\lambda))^T B X)^{-1} X^T (I - A(\lambda))^T B \underline{y} \end{aligned}$$

From (15) and (16), the multiresponse semiparametric regression can be written as Equation 17:

$$\hat{\underline{y}} = X \hat{\beta} + \hat{f} = H(\lambda) \underline{y} \tag{17}$$

Where:

$$\begin{aligned} H(\lambda) &= (I - A(\lambda)) X (X^T (I - A(\lambda))^T B X)^{-1} \\ &\quad X^T (I - A(\lambda))^T B + A(\lambda) \end{aligned}$$

Next, consider special condition for nonparametric estimator \hat{f} when assumption for f is hold following condition:

$$f_j \in W_2^m([a_j, b_j]), \quad j = 1, 2, \dots, r$$

Where:

$$W_2^m([a_j, b_j]) = \{f_j : f_j, f_j', \dots, f_j^{(m-1)}, \text{abs. cont.,}$$

$$\int_{a_j}^{b_j} (f_j^{(m)}(t_j))^2 dt_j < \infty \}$$

$$\|P_j f_j\|^2 = \int_{a_j}^{b_j} (f_j^{(m)}(t_j))^2 dt_j$$

$$\phi_v(t) = \frac{t^{v-1}}{(v-1)!}, \quad v = 1, 2, \dots, m$$

$$\psi_j(s, t_j) = \int_{a_j}^{b_j} \frac{(s-u)_+^{m-1} (t_j-u)_+^{m-1}}{(m-1)!(m-1)!} du, \quad s, t_j \in [a_j, b_j]$$

then the function f which minimize weighted penalized least square would be:

$$\begin{aligned} &(\underline{y} - X \underline{\beta} - \hat{f})^T B (\sigma^2) (\underline{y} - X \underline{\beta} - \hat{f}) + \\ &\sum_{j=1}^r \lambda_j \int_{a_j}^{b_j} (f_j^{(m)}(t_j))^2 dt_j \end{aligned}$$

which is Natural Spline (NS) degree $2m-1$. It is denoted by NS^{2m-1} . From Wahba (1990), natural spline $f(t) = f_n^m(t)$ is a real-valued function on $[a, b]$ defined with the aid of n knot $-\infty \leq a < t_1 < \dots < t_n < b \leq \infty$ with the following properties:

- $f \in \pi^{m-1}, t \in [a, t_1], t \in [x_n, b]$
- $f \in \pi^{2m-1}, t \in [t_i, t_{i+1}], i = 1, 2, \dots, n-1$
- $f \in C^{2m-2}, t \in (-\infty, \infty)$

Where:

π^k = The polynomial of degree k or less
 C^k = The class of function with k continuous derivatives

For more detailed the reader can read that reference.

The following process shows that if f holds the above assumption, then estimator of f is natural spline degree 2m-1. For $t \in [a_j, t_{ji}]$:

$$\begin{aligned} \psi_j(t, t_j) &= \int_{a_j}^{b_j} \frac{(t-u)_+^{m-1} (t_j-u)_+^{m-1}}{(m-1)!(m-1)!} du \\ &= \int_{a_j}^t \frac{[tt_j - tu - t_ju + u^2]^{m-1}}{[(m-1)!]^2} du \end{aligned}$$

Then:

$$\psi_j(t, t_j) \in \pi^{2m-1}, t \in [a, t_{ji}]$$

For:

$$\begin{aligned} t &\in [t_{ji}, b_j], \\ \psi_j(t, t_j) &= \int_{a_j}^{b_j} \frac{(t-u)_+^{m-1} (t_j-u)_+^{m-1}}{(m-1)!(m-1)!} du \\ &= \int_{a_j}^{t_{ji}} \frac{[tt_j - tu - t_ju + u^2]^{m-1}}{[(m-1)!]^2} du \end{aligned}$$

Then:

$$\psi_j(t, t_j) \in \pi^{m-1}, s \in [t_{ji}, b_j]$$

Multiplying \hat{q} by T^T from left will lead to:

$$T^T \hat{q} = 0$$

This implies:

$$T_j \hat{q}_j = 0, \quad j = 1, 2, \dots, r$$

As a result:

$$\sum_{i=1}^n \alpha_{ji} t_{ji}^k = 0, k = 0, 1, 2, \dots, m-1$$

Next, for $t < t_{ji}$:

$$\begin{aligned} \sum_{i=1}^n \alpha_{ji} \psi_{ji}(t) &= \sum_{i=1}^n \alpha_{ji} \int_{a_j}^{b_j} \frac{(t-u)_+^{m-1} (t_{ji}-u)_+^{m-1}}{[(m-1)!]^2} du \\ &= 0 \end{aligned}$$

So:

$$f_j(t) \in \pi^{m-1}, t < t_{ji}$$

For:

$$\begin{aligned} t < t_{ji}, i = 1, 2, \dots, n, \\ \psi_{ji}(t) &= \int_a^b \frac{(t-u)_+^{m-1} (t_{ji}-u)_+^{m-1}}{[(m-1)!]^2} du \end{aligned}$$

By solving this integration:

$$\psi_{ji}(t) \in \pi^{2m-1}$$

So:

$$f_j(t) \in \pi^{2m-1}, t \in [t_{ji}, t_{j(i+1)}], \quad i = 1, 2, \dots, n-1$$

For:

$$\begin{aligned} t > t_{jn} \\ \psi_{ji}(t) &= \int_a^b \frac{(t-u)^{m-1} (t_{ji}-u)^{m-1}}{[(m-1)!]^2} du \end{aligned}$$

By solving this integration:

$$\psi_{ji}(t) \in \pi^{m-1}$$

As a result:

$$f_j(t) \in \pi^{m-1}, t \in [t_{jn}, b]$$

Based on (i), (ii) and (iii), it can be inferred that $f_j(t) \in NS^{2m-1}$.

3.2. Smoothing Parameter

The partial spline estimator still depends on smoothing parameters $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_r)^T$. The following is steps to choose the optimal smoothing parameter. Recall Equation (17) and let:

$$N = \sum_{j=1}^r n_j$$

Mean Square Error of this estimator is given by:

$$MSE(\lambda) = N^{-1}(\underline{y} - \hat{\underline{y}})^T B(\sigma^2)(\underline{y} - \hat{\underline{y}})$$

with simple algebra manipulation, the result will be:

$$N^{-1} \left\| B^{1/2}(\sigma^2) [(I - H(\lambda)) \underline{y}] \right\|^2$$

Matrix $B^{1/2}(\sigma^2)$ has properties:

$$(B^{1/2}(\sigma^2))^T B^{1/2}(\sigma^2) = B(\sigma^2)$$

Next, it is defined:

$$G(\lambda) = \frac{N^{-1} \left\| B^{1/2}(\sigma^2) [(I - H(\lambda)) \underline{y}] \right\|^2}{\left[N^{-1} \text{trace}(I - H(\lambda)) \right]^2}$$

The optimal smoothing parameter $\lambda_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)}, \dots, \lambda_{r(opt)})^T$ was obtained by solving optimization problem:

$$G_{opt}(\lambda_{(opt)}) = \text{Min}_{\lambda_j \in \square^+, j=1,2,\dots,r} \{G(\lambda)\}$$

$$= \text{Min}_{\lambda_j \in \square^+, j=1,2,\dots,r} \left\{ \frac{N^{-1} \left\| B^{1/2}(\sigma^2) [(I - H(\lambda)) \underline{y}] \right\|^2}{\left[N^{-1} \text{trace}(I - H(\lambda)) \right]^2} \right\}$$

where, $\| \underline{y} \| = \sqrt{v_1^2 + v_2^2 + \dots + v_r^2}$ is norm for some vector \underline{y} with dimension r , $\underline{y} = (v_1, v_2, \dots, v_r)^T$.

So, the optimal smoothing parameter $\lambda_{opt} = (\lambda_{1(opt)}, \lambda_{2(opt)}, \dots, \lambda_{r(opt)})^T$ was obtained by minimizing quantity $G_{opt}(\lambda_{(opt)})$ for each non negative smoothing parameter $\lambda_j \in \square^+, j = 1, 2, \dots, r$.

3.3. Numerical Example

As example for multiresponse semiparametric regression, simulation data was generated using the following equation:

$$y_{1i} = 5 + 2x_{1i} + 2\sin(2\pi t_i^2) + \varepsilon_{1i}$$

$$y_{2i} = 3 + 4x_{2i} + 2\sin(2\pi t_i^2) + \varepsilon_{2i}$$

$$y_{3i} = 1 + 6x_{3i} + 2\sin(2\pi t_i^2) + \varepsilon_{3i}$$

$$i = 1, 2, \dots, 100$$

This equation generates data for multiresponse semiparametric regression with three responses, linear parametric component and sinus pattern for

nonparametric component. Predictor variable $x_{ji}, j = 1, 2, 3$ as parametric predictor was generated from Normal (0,1). Predictor variable $t_{ji}, j = 1, 2, 3$ as nonparametric predictor was from Uniform(0,2). Random error $\underline{\varepsilon} = (\varepsilon_1^T, \varepsilon_2^T, \varepsilon_3^T)^T$ was from Multivariate Normal (0,Ω). $\varepsilon_{1i}, \varepsilon_{2i}, \varepsilon_{3i}$ are correlated with $\text{corr}(\varepsilon_{1i}, \varepsilon_{2i}) = \text{corr}(\varepsilon_{1i}, \varepsilon_{3i}) = \text{corr}(\varepsilon_{2i}, \varepsilon_{3i}) = \rho = 0.5$. Error variance σ_j^2 was generated from Uniform(0.1, 0.2), $j = 1, 2, 3$.

Figure 1 show partial scatter plot of above simulation. This figure is used to confirm what relationship between response and parametric predictor as well as response and nonparametric predictor for each response. The scatter plot shows that the relationship between parametric predictor (x) and response is linear as well as relationship between nonparametric predictor (t) that relatively follows the pattern of sinus function.

Figure 2 shows three dimension surface plot to describe relation between predictor and response in one picture. This surface plot is composed by combining predictor x and predictor t as horizontal axis as well as response y as vertical axis. We just present surface plot of response 1 due to limitation space. The surface plot of response 2 and 3 has the similar pattern with response 1.

Next, based on this data, the parametric and nonparametric estimator will be computed by implementing R software. Due to the limitation of space, the result of computing is presented simply in the following tables and figures. **Table 1** shows various smoothing parameters and their G score. We can see the changing of G score from small (line 1) to large smoothing parameter (line 11)

As mentioned before that the partial spline estimator depends on the smoothing parameter. The following figures shows impact of smoothing parameter on the shape of surface plot between predicted value of response as vertical axis as well as predictor x and t as horizontal axis. We just also present surface plot of response 1 to avoid too many figure with the same information.

Figure 3 shows that for small smoothing parameter (line 1 of **Table 1**), the spline estimated imply the predicted value tend to interpolate the actual value of response. We can compare surface plot in **Fig. 2** and surface plot in **Fig. 3** that show their shape is almost the same.

Figure 4 shows that for large smoothing parameter (line 11 of **Table 1**), the predicted value tend to be forced to fit the linear regression. It can be seen, as long both horizontal axis, the predicted value have the linear pattern. This mean both parametric and nonparametric have the linear pattern. We also conclude that for large smoothing parameter, the partial spline estimator tend to be forced linear regression fit.

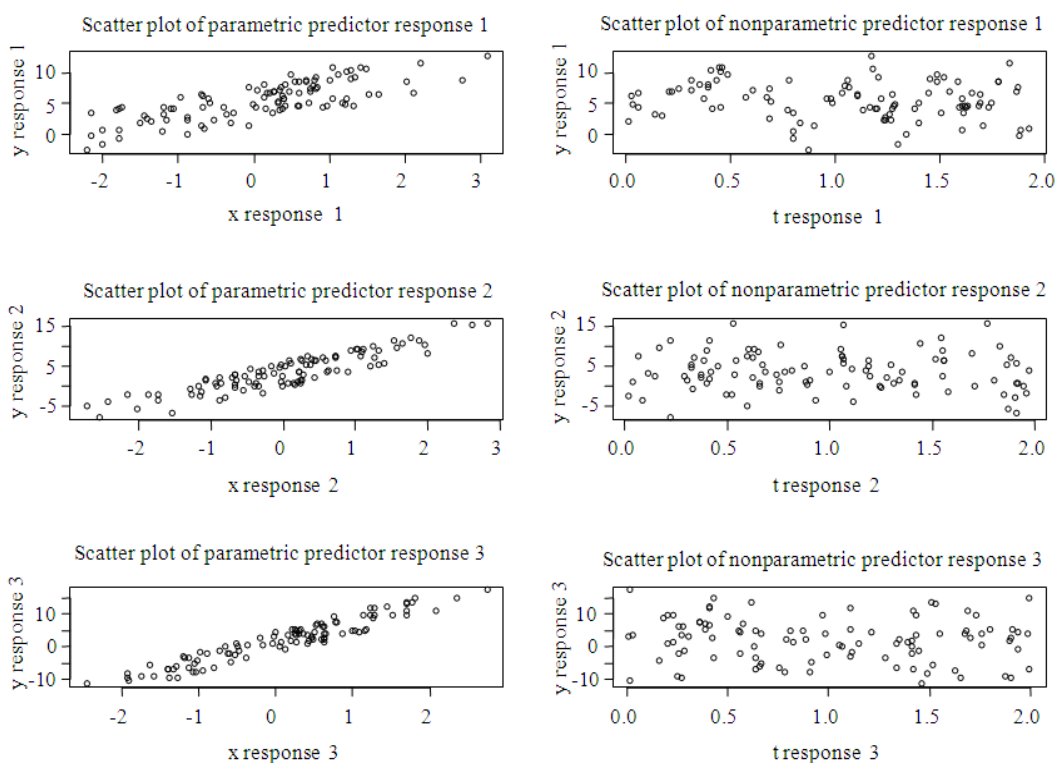


Fig. 1. Partial scatter plot of predictor and response

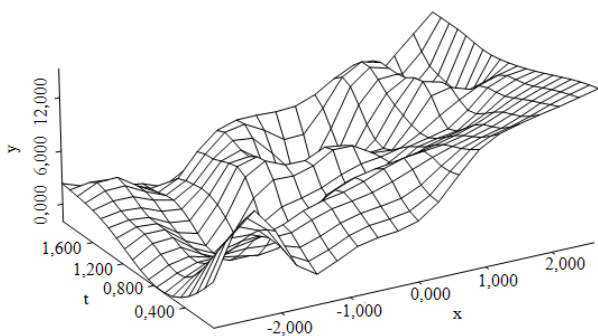


Fig. 2. Surface plot of response 1 based on actual data of predictor and response

Table 1 shows that the optimal condition is reached when the G score is minimum, equal to 0.320519 (line 6 of Table 1) and the smoothing parameters of response 1, response 2 and response 3 are 2.667644e-08, 6.223154e-09 and 2.962001e-09 respectively. Using the optimal smoothing parameter, the corresponding estimator for parametric component β is presented in Table 2.

Table 1. The smoothing parameter and G value

λ_1	λ_2	λ_3	G ($\lambda_1, \lambda_2, \lambda_3$)
2.66764e-13	6.2231e-14	2.9620e-14	0.5052072
2.6676e-12	6.2231e-13	2.9620e-13	0.5047398
2.6676e-11	6.2231e-12	2.9620e-12	0.5002114
2.6676e-10	6.2231e-11	2.9620e-11	0.5695718
2.6676e-09	6.2231e-10	2.9620e-10	0.3798685
2.6676e-08	6.2231e-09	2.9620e-09	0.3205190
2.6676e-07	6.2231e-08	2.9620e-08	0.5434078
2.6676e-06	6.2231e-07	2.9620e-07	1.5095615
2.6676e-05	6.2231e-06	2.9620e-06	2.6249815
2.6676e-04	6.2231e-05	2.9620e-05	3.3163426
2.6676e-03	6.2231e-04	2.9620e-04	3.5232653

Table 2. Parametric estimator for each response

Response	$\hat{\beta}_0$	$\hat{\beta}_1$
1	5.245424	2.011274
2	3.513463	3.750386
3	1.418387	5.881660

Figure 5 shows the surface plot of predicted value of each response using the optimal smoothing parameter.

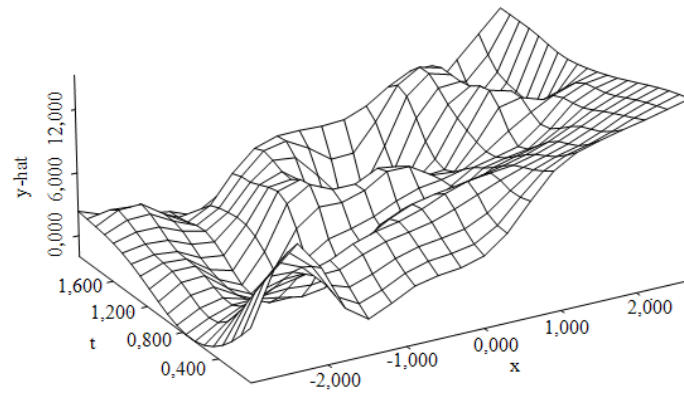


Fig. 3. Surface plot of predicted value response 1 using the small smoothing parameter

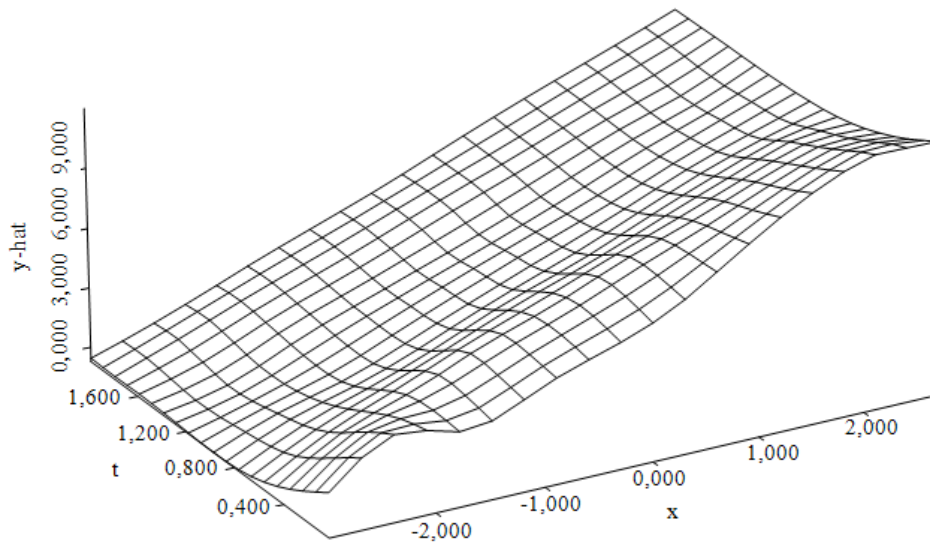
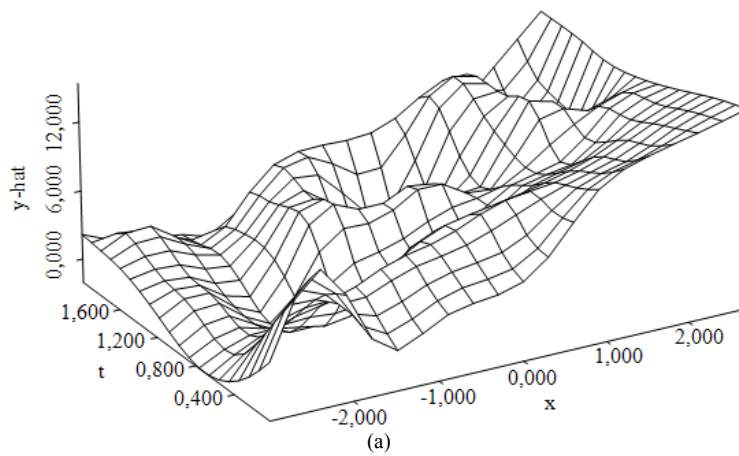


Fig. 4. Surface plot of predicted value response 1 using the large smoothing parameter



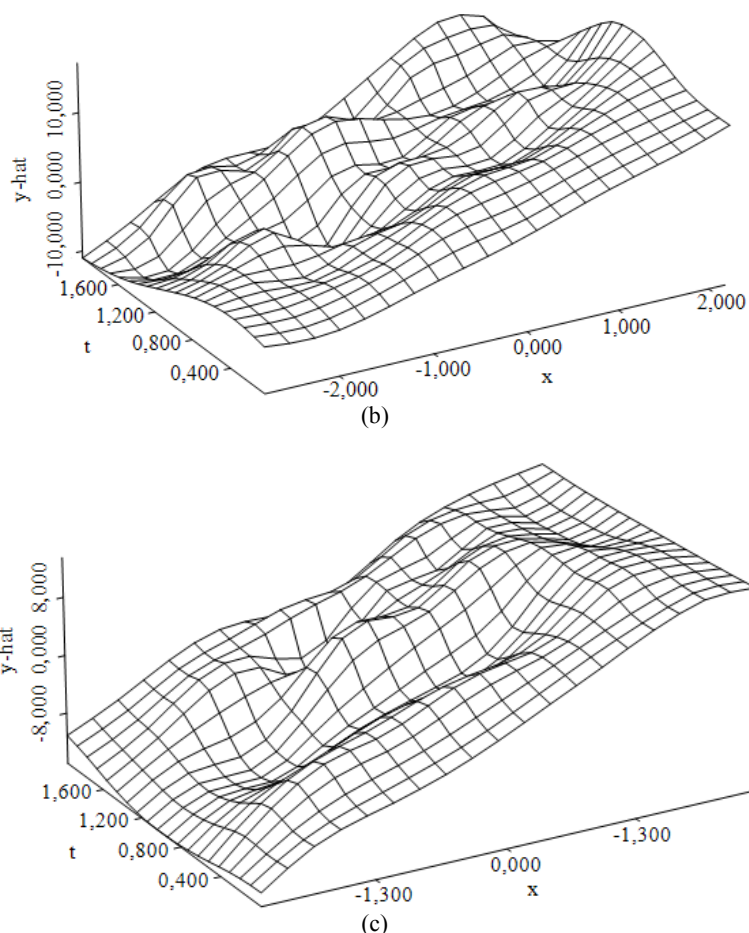


Fig. 5. Surface plot of predicted value of response 1 (a), response 2 (b) and response 3 (c) using the optimal smoothing parameter

4. DISCUSSION

There are several issues regarding the multiresponse semiparametric regression. The first one is about estimator of nonparametric component. In this study, the nonparametric estimator as in Equation (6) is general form of smoothing spline, which is produced by weighted penalized least square method. However, if we apply some assumptions for function f , the estimator will be special form of spline smoothing, such as polynomial natural spline in this study. In other words, the different assumption of f and its norm will produce different form of spline. We did not discuss further about this, but further discussion can be found in Wahba (1990). Another method also can be used to estimate parameter, such as penalized likelihood. Nevertheless, this method requires distribution assumption of residual.

The second issue is smoothing parameter especially method to choose the optimal smoothing parameter and how many smoothing parameter must be chosen. In this study, we modify generalized cross validation to choose the optimal smoothing parameter in the context of multiresponse semiparametric regression. Actually, there are other methods to select the optimal smoothing parameter, such as Unbiased Risk (UBR), Generalized Maximum Likelihood (GML) or Cross Validation (CV). Wang (1998) did comparison of these method. In our case, the model involves the multiple smoothing parameters and this situation will imply the complicated of optimization problem. Therefore, to make easier the optimization problem, we choose GCV methods that relatively easy to be modified and optimized.

5. CONCLUSION

Nonparametric and parametric estimator in multiresponse semiparametric regression model can be obtained by using weighted penalized least square. More specific, estimator for nonparametric component is partial spline function. Especially, if nonparametric component hold some assumption, then the kind of spline function is polynomial natural spline. This estimator depends on smoothing parameter and this imply that the predicted values also depend on smoothing parameter. However, the optimal smoothing parameter can be chosen by using G criteria. Numerical example shows that the model can be applied well using simulation data. The problem remaining is to find statistical properties of the estimator and to apply this model in real life problem.

6. REFERENCES

- Engle, R.F., C.W.J. Granger, J. Rice and A. Weiss, 1986. Semiparametric estimates of the relation between weather and electricity sales. *J. Am. Stat. Assoc.*, 81: 310-320. DOI: 10.1080/01621459.1986.10478274
- Fessler, J.A., 1991. Nonparametric fixed-interval smoothing with vector splines. *IEEE Trans. Signal Process.*, 39: 852-859. DOI: 10.1109/78.80907
- Hardle, W., H. Liang and J. Gao, 2000. *Partially Linear Models*. 1st Edn., Springer, Wurzburg.
- Lin, X.H. and R.J. Carroll, 2001. Semiparametric regression for clustered data. *Biometrika*, 88: 1179-1185. DOI: 10.1093/biomet/88.4.1179
- Miller, J.J. and E.J. Wegman, 1987. Vector function estimation using splines. *J. Stat. Plann. Inference*, 17: 173-180. DOI: 10.1016/0378-3758(87)90110-8
- Qin, G., Z. Zhu and W.K. Fung, 2009. Robust estimation of covariance parameters in partial linear model for longitudinal data. *J. Stat. Plann. Inference*, 139: 558-570. DOI: 10.1016/j.jspi.2008.03.042
- Qin, G.Y. and Z.Y. Zhu, 2008. Robust estimation in partial linear mixed model for longitudinal data. *Acta Math. Sci.*, 28B: 333-347.
- Ruppert, D., M.P. Wand and R.J. Carroll, 2009. *Semiparametric regression during 2003-2007*. *Elect. J. Stat.*, 3: 1193-1256. DOI: 10.1214/09-EJS525
- Soo, Y.W. and D.M. Bates, 1996. Multiresponse spline regression. *Comput. Stat. Data Anal.*, 22: 619-631. DOI: 10.1016/0167-9473(96)00014-X
- Wahba, G., 1990. *Spline Models for Observational Data*. 1st Edn., SIAM, Philadelphia, ISBN-10: 0898712440, pp: 180.
- Wang, Y., 1998. Smoothing spline models with correlated random errors. *J. Am. Stat. Assoc.*, 93: 341-348.
- Wang, Y., W. Guo and M.B. Brown, 2000. Spline smoothing for bivariate data with applications to association between hormones. *Stat. Sinica*, 10: 377-397.
- Wegman, E.J., 1981. Vector splines and the estimation of filter functions. *Technometrics*, 23: 83-89. DOI: 10.1080/00401706.1981.10486241
- Yee, T.W. and C.J. Wild, 1996. Vector generalized additive models. *J. Royal Stat. Soc. Series B*, 58: 481-493.
- You, J. and X. Zhou, 2009. Partially linear models and polynomial spline approximations for the analysis of unbalanced panel data. *J. Stat. Plann. Inference*, 139: 679-695. DOI: 10.1016/j.jspi.2007.04.037
- You, J., S. Xie and Y.D. Zhou, 2007. Two-stage estimation for seemingly unrelated nonparametric regression models. *J. Syst. Sci. Comp.*, 20: 509-520. DOI: 10.1007/s11424-007-9048-8