The Combination of Spline and Kernel Estimator for Nonparametric Regression and its Properties

I Nyoman Budiantara, Vita Ratnasari, Madu Ratna and Ismaini Zain

Statistics Department, Sepuluh Nopember Institute of Technology
Kampus ITS, Sukolilo, Surabaya, Indonesia

Abstract

Consider additive nonparametric regression model with two predictor variables components. In the first predictor component, the regression curve is approached using Spline regression, and in the second predictor component, the regression curve is approached using Kernel regression. Random error of regression model is assumed to have independent normal distribution with zero mean and the same variance. This article provides an estimator of Spline regression curve, estimator of Kernel regression curve, and an estimator of a combination of Spline and Kernel regressions. The produced estimators are biased estimators, but all estimators are classified as linear estimators in observation. Estimator of a combination of Spline and Kernel regression depended on knot points and bandwidth parameter. The best estimator of a combination of Spline and Kernel regression is found by minimizing Generalized Cross Validation (GCV) function.

Keywords: Nonparametric Regression, Spline, Kernel, Mixed Estimator, GCV

1. Introduction

In recent decades, nonparametric regression has received a lot of attention from researchers. Nonparametric regression is a regression model approached used if the pattern of relation between predictor variable and response isn’t known, or if there is no complete past information on the shape of data pattern [1], [2], [3]. The nonparametric regression models which receive a lot of attention from re-
searchers are Kernel [4],[5],[6], Spline smoothing [2],[7],[8],[9],[10], Fourier Series [11],[12],[13] and Local Polynomial [14],[15]. Nonparametric regression approach has high flexibility because data is expected to look for its shape of regression curve estimation without being influenced by researchers subjective factors [1], [3], [6]. In several cases, response variable may have linear relation with one of the predictor variables, but the relation patterns with other predictor variables aren’t known. In this situation, Wahba [7] suggests using semi-parametric regression approach. Several researchers such as Eubank [1], Diana et. al. [10] and Wibowo, et. al. [16] have developed partial spline smoothing estimator to estimate semiparametric regression curve. Meanwhile, Hardle [4] and Cheng et. al.[6] use Kernel approach to estimate semiparametric regression curve.

Among nonparametric and semiparametric models above, spline truncated is one of the models with very specific and good statistical interpretation and visual interpretation [3], [9]. Estimator spline truncated has high flexibility [1]. Spline truncated also has very good ability to handle data with changeable behaviors on certain sub-intervals [1],[3],[8],[9],[10]. Kernel estimator in nonparametric and semiparametric regression lately also receive a lot of attention from researchers. Kernel estimator has a good ability to model data with no particular pattern [4]. Kernel estimator has relatively faster convergence speed than other nonparametric regression curve estimators such as Local Polynomial, Fourier Series, k-NN, or Spline [4].

When the nonparametric and semiparametric regression models developed by the researchers above are explored, essentially there are two very heavy and basic assumptions, i.e. first, the pattern of each predictor in the multivariable nonparametric regression model of the predictor is considered to have the same pattern. The second assumption is researchers insist on using only one shape of model estimator for every predictor variable. The two assumptions used in nonparametric regression models are only theoretical. In applications on cases, data pattern is often different from each predictor variable. Moreover, by using only one estimator is estimating a multivariable nonparametric regression curve, the produced estimators won’t fit the data pattern. As a result, the regression model estimation produced isn’t correct and tends to produce large errors.

Based on the research results above and to obtain regression curve model estimations which fit data patterns, this article used two estimator models which are a combination of Spline and Kernel regression to estimate multivariable nonparametric regression curve of the predictor, as well as properties related to the combination estimator.

2. The Shape of Estimator of Combination of Spline and Kernel Regression

Consider paired data \((x_i, t_i, y_i)\) and the relations between predictor variables \(x_i, t_i\) and response variable \(y_i\) are assumed to follow nonparametric regression model:
\[ y_i = \mu(x_i, t_i) + \epsilon_i, \quad i = 1, 2, \ldots, n. \]  

(1)

The shape of regression curve \( \mu(x_i, t_i) \) is assumed to be unknown and smooth, meaning continuous and differentiable. Random error \( \epsilon_i \) has normal distribution with zero mean and \( E(\epsilon_i^2) = \sigma^2 \). Then regression curve \( \mu(x_i, t_i) \) is assumed to be additive, meaning \( \mu(x_i, t_i) \) can be written as:

\[ \mu(x_i, t_i) = f(x_i) + g(t_i), \]

(2)

with \( f(x_i) \) and \( g(t_i) \) being smooth functions. The main problem in the estimator of combination of nonparametric regression curves is obtaining the shape of the estimation of regression curve \( \mu(x_i, t_i) \) which is:

\[ \hat{\mu}_{\alpha, \lambda}(x_i, t_i) = \hat{f}_{\alpha, \lambda}(x_i, t_i) + \hat{g}_{\alpha}(t_i). \]

(3)

Parameter \( \alpha \) is bandwidth parameter and \( \lambda \) is knot point. To obtain an estimator of combination of Spline and Kernel regression, first, regression curve \( f(x_i) \) is approached using \( m \) degree Spline truncated function and knot point \( \lambda = (\lambda_1, \lambda_2, \ldots, \lambda_r)' \). Then, regression curve \( g(t_i) \) is approached using Kernel’s function. For example, given a basis for Spline space:

\[ \{ 1, x, x^2, \ldots, x^n, (x - \lambda_1)I(x \geq \lambda_1), \ldots, (x - \lambda_r)I(x \geq \lambda_r) \} \]

with \( I \) being indicator function. Regression curve \( f(x_i) \) can be written as:

\[ f(x_i) = \theta_0 + \theta_1 x_i + \ldots + \theta_m x_i^m + \phi_i (x_i - \lambda_1)^m I(x_i \geq \lambda_1) + \ldots + \phi_i (x_i - \lambda_r)^m I(x_i \geq \lambda_r) \]

(4)

with \( \theta_0, \theta_1, \ldots, \theta_m, \phi_1, \phi_2, \ldots, \phi_i \) being unknown parameters. Then, estimation of regression curve \( g(t) \) can be presented as:

\[ \hat{g}_{\alpha}(t) = n^{-1} \sum_{i=1}^{n} \left( \frac{K_{\alpha}(t-t_i)}{n^{-1} \sum_{j=1}^{n} K_{\alpha}(t-t_j)} \right) y_i = n^{-1} \sum_{i=1}^{n} W_{\alpha}(t)y_i \]

(5)

\[ W_{\alpha}(t) = \frac{K_{\alpha}(t-t_i)}{n^{-1} \sum_{j=1}^{n} K_{\alpha}(t-t_j)}, \quad K_{\alpha}(t-t_i) = \frac{1}{\alpha} K \left( \frac{t-t_i}{\alpha} \right) \]

with \( K \) being Kernel function. Kernel function \( K \) can be Kernel Gaussian, Kernel Uniform, Kernel Epanechnikov, Kernel Triweight, or other Kernels [4]. Estimator of regression curve (5) depend on two things, i.e. bandwidth parameter and the Kernel function. Using estimator (5), the estimator of combination of Spline and Kernel Regression in equation (3) will be searched using maksimum likelihood method (MLE). For this purpose, the following theorems and lemmas are presented.
Lemma 1
If regression curve \( f(x_i) \) is given by equation (4), then:
\[
\tilde{f} = X_1 \tilde{\theta} + X_2(\tilde{\lambda}) \tilde{\phi},
\]

vectors \( \tilde{f}, \tilde{\theta} \) and \( \tilde{\phi} \) respectively are given by:
\[
\tilde{f} = (f(x_1), f(x_2), \ldots, f(x_n))^T, \quad \tilde{\theta} = (\theta_0, \theta_1, \ldots, \theta_m)^T, \quad \tilde{\phi} = (\phi_1, \phi_2, \ldots, \phi_r)^T,
\]
\( \tilde{\lambda} = (\lambda_1, \lambda_2, \ldots, \lambda_r)^T \) and matrices \( X_1 \) dan \( X_2(\tilde{\lambda}) \) respectively are given by:
\[
X_1 = \begin{pmatrix}
1 & x_1 & \cdots & x_1^m \\
1 & x_2 & \cdots & x_2^m \\
\vdots & \vdots & \ddots & \vdots \\
1 & x_n & \cdots & x_n^m
\end{pmatrix}, \quad \text{and}
\]
\[
X_2(\tilde{\lambda}) = \begin{pmatrix}
(x_1 - \lambda_1)^m I(x_1 \geq \lambda_1) & \cdots & (x_1 - \lambda_r)^m I(x_1 \geq \lambda_r) \\
\vdots & \ddots & \vdots \\
(x_n - \lambda_1)^m I(x_n \geq \lambda_1) & \cdots & (x_n - \lambda_r)^m I(x_n \geq \lambda_r)
\end{pmatrix}
\]

Proof:
Equation (4) gives:
\[
f(x_i) = \theta_0 + \theta_1 x_i + \ldots + \theta_m x_i^m + \phi_1 (x_i - \lambda_1)^m I(x_i \geq \lambda_1) + \ldots + \phi_r (x_i - \lambda_r)^m I(x_i \geq \lambda_r),
\]
For every \( i = 1, 2, \ldots, n \). Regression curve \( f(x_i) \) can be written into two components, which are polynomial component and truncated component, giving:
\[
f(x_i) = \sum_{j=0}^m \theta_j x_i^j + \sum_{j=1}^r \phi_j (x_i - \lambda_j)^m I(x_i \geq \lambda_j).
\]
Because it applies to each \( i = 1, 2, \ldots, n \), equation (8) can be written as matrix:
\[
\tilde{f} = X_1 \tilde{\theta} + X_2(\tilde{\lambda}) \tilde{\phi}
\]
with matrices \( X_1 \) and \( X_2(\tilde{\lambda}) \) respectively given by equations (6) and (7).

Lemma 2
If given regression model (1) and the estimator for Kernel is given by equation (5), total squared error is given by:
\[
\|\tilde{\varepsilon}\|^2 = \left\| (I - D(\alpha)) \tilde{y} - X(\tilde{\lambda}) \tilde{\phi} \right\|^2
\]
with \( \|\tilde{\varepsilon}\| \) being the length of vector \( \tilde{\varepsilon}, \tilde{y} = (y_1, y_2, \ldots, y_n)^T \),
\[
X(\tilde{\lambda}) = \begin{pmatrix} X_1 & X_2(\tilde{\lambda}) \end{pmatrix}, \quad \tilde{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^T, \quad \tilde{\beta} = \begin{pmatrix} \tilde{\theta} \\ \tilde{\phi} \end{pmatrix}, \quad \text{and}
\]
\[ D(\alpha) = \left\{ n^{-1}W_{a_i}(t_j) \right\}, \quad j,i = 1,2,\ldots,n. \]

**Proof:**

Equation (5) gives:

\[
\hat{g}_a(t) = n^{-1} \sum_{i=1}^{n} W_{a_i}(t)y_i
\]

Because it applies to each \( t = t_1, t_2, \ldots, t_n \), then:

\[
\begin{pmatrix}
\hat{g}_a(t_1) \\
\hat{g}_a(t_2) \\
\vdots \\
\hat{g}_a(t_n)
\end{pmatrix} =
\begin{pmatrix}
n^{-1}W_{a_1}(t_1) & n^{-1}W_{a_2}(t_1) & \cdots & n^{-1}W_{a_n}(t_1) \\
n^{-1}W_{a_1}(t_2) & n^{-1}W_{a_2}(t_2) & \cdots & n^{-1}W_{a_n}(t_2) \\
\vdots & \vdots & \ddots & \vdots \\
n^{-1}W_{a_1}(t_n) & n^{-1}W_{a_2}(t_n) & \cdots & n^{-1}W_{a_n}(t_n)
\end{pmatrix}
\begin{pmatrix}
y_1 \\
y_2 \\
\vdots \\
y_n
\end{pmatrix}
\]

\[= D(\alpha)\tilde{y} \tag{9} \]

Regression model (1), equation (9) and Lemma 1, give:

\[
\tilde{y} = \tilde{f} + \tilde{g}_a(t) + \tilde{\varepsilon}
\]

\[= X_1\tilde{\theta} + X_2(\tilde{\lambda})\tilde{\phi} + D(\alpha)\tilde{y} + \tilde{\varepsilon}
\]

\[= \left( X_1 \quad X_2(\tilde{\lambda}) \right) \begin{pmatrix} \hat{\theta} \\ \hat{\phi} \end{pmatrix} + D(\alpha)\tilde{y} + \tilde{\varepsilon}
\]

\[= X(\tilde{\lambda})\tilde{\beta} + D(\alpha)\tilde{y} + \tilde{\varepsilon}. \tag{10} \]

Equation (10) gave total squared error:

\[\| \tilde{\varepsilon} \|^2 = \| \tilde{y} - X(\tilde{\lambda})\tilde{\beta} - D(\alpha)\tilde{y} \|^2 = \| [I - D(\alpha)]\tilde{y} - X(\tilde{\lambda})\tilde{\beta} \|^2. \]

**Theorem 1**

If total squared error of nonparametric regression model given by Lemma 2, error model with normal multivariate distribution with zero mean and \( E(\tilde{\varepsilon}^2) = \sigma^2 I \) and \( L\left( \tilde{\beta}, \sigma^2 \mid \alpha, \tilde{\lambda} \right) \) are likelihood function, MLE estimator for parameter vector \( \tilde{\beta} \) is found from optimization:

\[\max_{\tilde{\beta}, \sigma^2} \left\{ L\left( \tilde{\beta}, \sigma^2 \mid \alpha, \tilde{\lambda} \right) \right\} = \min_{\tilde{\beta}, \sigma^2} \left\{ \| [I - D(\alpha)]\tilde{y} - X(\tilde{\lambda})\tilde{\beta} \|^2 \right\}. \]

**Proof:**

Consider nonparametric regression model as in equation (1). Random error \( \tilde{\varepsilon} \) with multivariate normal distribution with \( E(\tilde{\varepsilon}) = 0 \) and \( E(\tilde{\varepsilon}\tilde{\varepsilon}') = \sigma^2 I \), then Likelihood function \( L\left( \tilde{\beta}, \sigma^2 \mid \alpha, \tilde{\lambda} \right) \) is given by:
Lemma 2 gave likelihood function:
\[
L(\hat{\beta}, \sigma^2 | \alpha, \lambda) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \left\| (I-D(\alpha)) \tilde{y} - X(\lambda) \hat{\beta} \right\|^2 \right).
\]

Based on MLE method, the estimator for parameter \( \hat{\beta} \) is found from optimization:
\[
\text{Max}_{\hat{\beta} \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ L(\hat{\beta}, \sigma^2 | \alpha, \lambda) \right\} = \text{Max}_{\hat{\beta} \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \left\| (I-D(\alpha)) \tilde{y} - X(\lambda) \hat{\beta} \right\|^2 \right) \right\}
\]

Then log likelihood function is given by:
\[
\log L(\hat{\beta}, \sigma^2 | \alpha, \lambda) \ni -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\| (I-D(\alpha)) \tilde{y} - X(\lambda) \hat{\beta} \right\|^2.
\]

Resulting:
\[
\text{Max}_{\hat{\beta} \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ L(\hat{\beta}, \sigma^2 | \alpha, \lambda) \right\} = \text{Max}_{\hat{\beta} \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ -\frac{n}{2} \log(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left\| (I-D(\alpha)) \tilde{y} - X(\lambda) \hat{\beta} \right\|^2 \right\}
\]

Equation (11) reaches maximum value if log likelihood component:
\[
\left\| (I-D(\alpha)) \tilde{y} - X(\lambda) \hat{\beta} \right\|^2
\]
has minimum value, resulting:
\[
\text{Max}_{\hat{\beta} \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ L(\hat{\beta}, \sigma^2 | \alpha, \lambda) \right\} = \text{Min}_{\hat{\beta} \in \mathbb{R}^m \times \mathbb{R}^n} \left\{ \left\| (I-D(\alpha)) \tilde{y} - X(\lambda) \hat{\beta} \right\|^2 \right\}.
\]

Then, based on Lemma 1, Lemma 2 and Theorem 1, the estimator of combination of Spline and Kernel Regression can be found. The detail is presented in Theorem 2.

**Theorem 2**

If given regression model (1), total squared error given by Lemma 2, error model has mutivariate normal distribution with zero mean and \( E(\tilde{e}\tilde{e}') = \sigma^2 I \), and MLE estimator for parameter \( \hat{\beta} \) is found from optimization as in Theorem 1, then MLE estimator for combined regression curve \( \hat{\mu}(x,t) \) is given by:
\[
\hat{\mu}_{a,\alpha}(x,t) = \hat{f}_{a,\lambda}(x,t) + \hat{g}_{a}(t)
\]
where $\hat{f}_{\alpha,\lambda}(x,t) = X(\tilde{\lambda})\hat{\beta}(\tilde{\lambda},\alpha)$, $\hat{g}_{\alpha}(t) = D(\alpha)\tilde{y}$ and

$$\hat{\beta}(\tilde{\lambda},\alpha) = [(X(\tilde{\lambda}))'X(\tilde{\lambda}))^{-1}(X(\tilde{\lambda}))'(I-D(\alpha))\tilde{y}.$$  

with matrices $X(\tilde{\lambda})$ and $D(\alpha)$ given by Lemma 2.

**Proof:**

To get MLE estimator for parameter vector $\tilde{\beta}$, Theorem 2 produces optimization:

$$\min_{\beta \in \mathbb{R}^{m \cdot r}} \left\{ \left\| [I-D(\alpha)]\tilde{y} - X(\tilde{\lambda})\tilde{\beta} \right\|^2 \right\} = \min_{\beta \in \mathbb{R}^{m \cdot r}} \left\{ Q(\tilde{\beta},\sigma^2 | \alpha,\tilde{\lambda}) \right\}$$

with:

$$Q(\tilde{\beta},\sigma^2 | \alpha,\tilde{\lambda}) = \left\| (I-D(\alpha))\tilde{y} - X(\tilde{\lambda})\tilde{\beta} \right\|^2 = \left( (I-D(\alpha))\tilde{y} - X(\tilde{\lambda})\tilde{\beta} \right)^T \left( (I-D(\alpha))\tilde{y} - X(\tilde{\lambda})\tilde{\beta} \right) = \left\| I-D(\alpha) \right\|^2 \tilde{y}^2 - 2\tilde{\beta}^T(X(\tilde{\lambda}))'(I-D(\alpha))\tilde{y} + \tilde{\beta}^T(X(\tilde{\lambda}))'X(\tilde{\lambda})\tilde{\beta}$$

Then with partial derivative:

$$\frac{\partial}{\partial \tilde{\beta}} \left( Q(\tilde{\beta},\sigma^2 | \alpha,\tilde{\lambda}) \right) = 0$$

gives normal equation:

$$(X(\tilde{\lambda}))'(X(\tilde{\lambda}))\tilde{\beta} = (X(\tilde{\lambda}))'(I-D(\alpha))\tilde{y}$$  \hspace{1cm} (12)

MLE estimator for $\tilde{\beta}$ is given by:

$$\hat{\beta}(\tilde{\lambda},\alpha) = [(X(\tilde{\lambda}))'(X(\tilde{\lambda}))^{-1}(X(\tilde{\lambda}))'(I-D(\alpha))\tilde{y} = C(\tilde{\lambda},\alpha)\tilde{y}.$$  \hspace{1cm} (13)

with $C(\tilde{\lambda},\alpha) = [(X(\tilde{\lambda}))'(X(\tilde{\lambda}))^{-1}(X(\tilde{\lambda}))'(I-D(\alpha))$.

Considering equation (13) and invariance characteristic of MLE method, the estimator for spline regression curve $\hat{f}(x,t) = X(\tilde{\lambda})\tilde{\beta}$ is given by:

$$\hat{f}_{\alpha,\lambda}(x,t) = X(\tilde{\lambda})\hat{\beta}(\tilde{\lambda},\alpha).$$

Then equation (9) gives the estimator of Kernel:

$$\hat{g}_{\alpha}(t) = D(\alpha)\tilde{y}.$$  

Lemma 1 gives regression curve of the combination of Spline and Kernel Regression:

$$\hat{\mu}(x,t) = \hat{f}(x) + \hat{g}(t).$$
As a result, the estimator of the combination of Spline and Kernel Regression \( \hat{\mu}(x,t) \) are given by:

\[
\hat{\mu}_{\alpha,\lambda}(x,t) = \hat{f}_{\alpha,\lambda}(x,t) + \hat{g}_{\alpha}(t).
\]

**Lemma 3**

If estimators \( \hat{\beta}(\lambda,\alpha), \hat{g}_{\alpha}(t), \hat{f}_{\alpha,\lambda}(x,t) \) and \( \hat{\mu}_{\alpha,\lambda}(x,t) \) are given into Theorem 2, then:

\[
\hat{f}_{\alpha,\lambda}(x,t) = A(\lambda,\alpha)\tilde{y}, \quad \text{and} \quad \hat{\mu}_{\alpha,\lambda}(x,t) = B(\lambda,\alpha)\tilde{y}
\]

with

\[
A(\lambda,\alpha) = X(\lambda)[(X(\lambda))'X(\lambda)]^{-1}(X(\lambda))'(I-D(\alpha)), \quad \text{and}
\]

\[
B(\lambda,\alpha) = A(\lambda,\alpha) + D(\alpha).
\]

**Proof:**

Theorem 2 gives:

\[
\hat{f}_{\alpha,\lambda}(x,t) = X(\lambda)\hat{\beta}(\lambda,\alpha)
\]

Equation (13) and Equation (14) give:

\[
\hat{f}_{\alpha,\lambda}(x,t) = X(\lambda)[(X(\lambda))'X(\lambda)]^{-1}(X(\lambda))'(I-D(\alpha))\tilde{y} = A(\lambda,\alpha)\tilde{y}
\]

with matrix:

\[
A(\lambda,\alpha) = X(\lambda)[(X(\lambda))'X(\lambda)]^{-1}(X(\lambda))'(I-D(\alpha))
\]

Then Theorem 2 and Equation (9) and Equation (15) give:

\[
\hat{\mu}_{\alpha,\lambda}(x,t) = \hat{f}_{\alpha,\lambda}(x,t) + \hat{g}_{\alpha}(t) = A(\lambda,\alpha)\tilde{y} + D(\alpha)\tilde{y} = \left(A(\lambda,\alpha) + D(\alpha)\right)\tilde{y} = B(\lambda,\alpha)\tilde{y}
\]

with \( B(\lambda,\alpha) = A(\lambda,\alpha) + D(\alpha) \).

### 3. The Properties of Estimator of Combination of Spline and Kernel Regression

The following investigates the properties of estimator \( \hat{\beta}(\lambda,\alpha) \), estimator of Kernel regression \( \hat{g}_{\alpha}(t) \), estimator of Spline regression \( \hat{f}_{\alpha,\lambda}(x,t) \) and estimator of combination of Spline and Kernel regressions \( \hat{\mu}_{\alpha,\lambda}(x,t) \). Estimators \( \hat{\beta}(\lambda,\alpha), \hat{g}_{\alpha}(t), \hat{f}_{\alpha,\lambda}(x,t) \) and \( \hat{\mu}_{\alpha,\lambda}(x,t) \) as other nonparametric regression estimators, are biased. However, these estimators are still in linear estimator class in observation. Complete properties of estimators are presented in Theorem 3.
Theorem 3

If \( \hat{\beta}(\lambda, \alpha) \), \( \hat{f}_{\alpha, \lambda}(x, t) \), \( \hat{g}_\alpha(t) \) and \( \hat{\mu}_{\alpha, \lambda}(x, t) \) are estimators given in Theorem 2 and Lemma 3, then:

(a). \( \hat{\beta}(\lambda, \alpha) \), \( \hat{f}_{\alpha, \lambda}(x, t) \), \( \hat{g}_\alpha(t) \) and \( \hat{\mu}_{\alpha, \lambda}(x, t) \) respectively are biased estimators for \( \beta_{\lambda} \), \( f_{\lambda}(x, t) \), \( g_\alpha(t) \), and \( \mu_{\lambda}(x, t) \).

(b). \( \hat{\beta}(\lambda, \alpha) \), \( \hat{f}_{\alpha, \lambda}(x, t) \), \( \hat{g}_\alpha(t) \) and \( \hat{\mu}_{\alpha, \lambda}(x, t) \) are linear estimator class in observation \( \tilde{y} \).

Proof:

(a). Based on Theorem 2 and Lemma 3:

\[
E\left[ \hat{\beta}(\tilde{\lambda}, \alpha) \right] = \left[ (X(\tilde{\lambda})'X(\tilde{\lambda}))^{-1}(X(\tilde{\lambda}))'(I-D(\alpha))E(\tilde{y}) \right] \\
= \left[ (X(\tilde{\lambda})'X(\tilde{\lambda}))^{-1}(X(\tilde{\lambda}))'(I-D(\alpha)) \left[ \bar{f}(x) + \bar{g}(t) \right] \right] \\
= \hat{\beta} - [(X(\tilde{\lambda}))'X(\tilde{\lambda}))^{-1}(X(\tilde{\lambda}))'(I-D(\alpha)) \bar{g}(t) \neq \hat{\beta} \\

E\left[ \hat{f}_{\alpha, \lambda}(x, t) \right] = E\left[ A(\tilde{\lambda}, \alpha)\tilde{y} \right] = A(\tilde{\lambda}, \alpha) \left[ \bar{f}(x) + \bar{g}(t) \right] \\
= A(\tilde{\lambda}, \alpha) \bar{f}(x) + A(\tilde{\lambda}, \alpha) \bar{g}(t) \neq \bar{f}(x) \\

E\left[ \hat{g}_\alpha(t) \right] = E\left[ D(\alpha)\tilde{y} \right] = D(\alpha) \left[ \bar{f}(x) + D(\alpha) \bar{g}(t) \right] \neq \bar{g}(t) \\

E\left[ \hat{\mu}_{\alpha, \lambda}(x, t) \right] = E\left[ \hat{f}_{\alpha, \lambda}(x, t) + \hat{g}_\alpha(t) \right] = E\left[ \hat{f}_{\alpha, \lambda}(x, t) \right] + E\left[ \hat{g}_\alpha(t) \right] \\
= A(\tilde{\lambda}, \alpha) + D(\alpha) \left[ \bar{f}(x) + D(\alpha) \bar{g}(t) \right] \\
\neq \bar{f}(x) + \bar{g}(t) = \bar{\mu}(x, t) \\

(b). Theorem 2 and Lemma 3 give:

\( \hat{\beta}(\tilde{\lambda}, \alpha) = C(\tilde{\lambda}, \alpha)\tilde{y} \), \( \hat{f}_{\alpha, \lambda}(x, t) = A(\tilde{\lambda}, \alpha)\tilde{y} \), \\
\( \hat{g}_\alpha(t) = D(\alpha)\tilde{y} \), dan \( \hat{\mu}_{\alpha, \lambda}(x, t) = B(\tilde{\lambda}, \alpha)\tilde{y} \).

It appears that estimators \( \hat{\beta}(\tilde{\lambda}, \alpha) \), \( \hat{f}_{\alpha, \lambda}(x, t) \), \( \hat{g}_\alpha(t) \) and \( \hat{\mu}_{\alpha, \lambda}(x, t) \) are linear estimator class in observation \( \tilde{y} \).

The estimator of the combination of Spline and Kernel regressions \( \hat{\mu}_{\alpha, \lambda}(x, t) \) highly depends on many (a). locations of knot points \( \tilde{\lambda} = (\lambda_1, \lambda_2, ..., \lambda_r)' \), and (b). bandwidth parameter \( \alpha \). To get the best estimator of the combination of Spline and Kernel regressions, knot points and bandwidth parameter should be selected optimally. For this purpose various methods can be used. One of the famous methods is Generalized Cross Validation (GCV). GCV function is given by [7]:

\[ GCV(\alpha) = ... \]
\[
G(\lambda, \alpha) = \frac{n^{-1} \left\| \tilde{y} - \hat{\mu}_{\alpha, \lambda}(x, t) \right\|^2}{\left( n^{-1} \text{trace} \left[ I - A(\lambda, \alpha) - D(\alpha) \right] \right)^2}.
\]

Optimal knot point \( \lambda_{\text{opt}} = (\lambda_{1(\text{opt})}, \lambda_{2(\text{opt})}, \ldots, \lambda_{r(\text{opt})})' \) and optimal bandwidth parameter \( \alpha_{\text{opt}} \) is found from optimization:

\[
G(\lambda_{\text{opt}}, \alpha_{\text{opt}}) = \min_{\lambda, \alpha} \left\{ G(\lambda, \alpha) \right\}.
\]

4. Conclusion

If given additive nonparametric regression model:
\( \tilde{y} = \mu(x, t) + \varepsilon = \tilde{f}(x) + \tilde{g}(t) + \varepsilon \)

a). Estimator of the combination of Spline and Kernel regressions is given by:
\[
\hat{\mu}_{\alpha, \lambda}(x, t) = \hat{f}_{\alpha, \lambda}(x, t) + \hat{g}_{\alpha, \lambda}(t), \quad \text{where :}
\]
\[
\hat{f}_{\alpha, \lambda}(x, t) = A(\lambda, \alpha)\tilde{y}, \quad \hat{g}_{\alpha, \lambda}(t) = D(\alpha)\tilde{y} \quad \text{and} \quad \hat{\mu}_{\alpha, \lambda}(x, t) = B(\lambda, \alpha)\tilde{y}
\]

b). Estimators \( \hat{f}_{\alpha, \lambda}(x, t), \hat{g}_{\alpha, \lambda}(t) \) and \( \hat{\mu}_{\alpha, \lambda}(x, t) \) are biased estimators, but they are linear estimator class in observation.

c). Estimator of the combination of Spline and Kernel regressions \( \hat{\mu}_{\alpha, \lambda}(x, t) \) highly depend on knot locations, many knot points and bandwidth parameter. The best estimator of combination is related to knot points and optimal bandwidth parameter. Knot points and optimal bandwidth parameter around found from the smallest GCV value.

Acknowledgements. This article is a part of the research result of Unggulan Perguruan Tinggi funded by DP2M-DIKTI in 2015. The author will like to thank the Director General of DIKTI in Jakarta for the fund provided.

References


Combination of spline and kernel estimator


http://dx.doi.org/10.3844/jmssp.2012.342.347


Received: August 24, 2015; Published: October 2, 2015