Model Selection in Additive Partial Linear Models Using Local False discovery Rate

Changhwan Lee[a], Seonghyun Jeong[b], Taeyoung Park[a,*]

[a]*Department of Applied Statistics, Yonsei University, Seoul, Korea*
[b]*Department of Statistics, North Carolina State University, Raleigh, North Carolina, USA*

Abstract

1. Introduction

As an intermediate model between multiple linear regression models and additive nonparametric regression models, an additive partial linear model(APLMs) consists of two models to represent parametric components and nonparametric components. Therefore, APLMs has advantages of multiple linear regression models such as easy interpretation and those of additive nonparametric regression models such as flexibility needed when fitting the model.

Especially, when estimating appropriate forms for complicated functions, the nonparametric regression method has been appealed to many researchers because it helps the researchers to pursuit more flexibility and accuracy. Also, it has enabled them to overcome the barriers caused by the parametric assumptions, which means the nonparametric regression method relaxes the strong model assumption included in the parametric regression methods.

One of the famous methods having been studied by many researchers is kernel smoothing and local polynomials. These methods estimate true functions by imposing weights locally. Also, regression splines and smoothing splines are used frequently in nonparametric regression. When using the regression spline method, how to select knots for accurate curve fitting is an important issue. There have been suggested various methods for it. One group of methods is conventional knot selection such as stepwise selection or branch and bound. Others are stochastic knot selection such as free-knot selection method, which means the number and location of knots are determined by data. It shows much better accurate results than those of conventional methods. However, it also has drawbacks that it is time-consuming for convergence (Miyata and Shen, 2003).

When fitting curve using spline and the number and locations of knots are predetermined to improve the shortcomings of free-knot selection method, it is clear that selecting appropriate knots to fit curve means deciding which basis function's coefficient corresponding to each knot is significant. Therefore, knot selecting problem can be replaced to variable selection and multiple testing in that situation. Traditionally, there have been many methods to select meaningful variables using various penalties. Ridge regression suggested by Hoerl and Kennard (1970) and Lasso regression suggested by Tibshirani (1996) select variables imposing penalty on each coefficient's size. Especially, Lasso regression is perceived as a useful tool for variable selection. It used absolute

*Corresponding author
Email address: tpark@yonsei.ac.kr (Taeyoung Park)

value as a penalty when choosing the coefficients in a model. That model was further extended by other researchers to develop some methods such as Group Lasso (Yuan and Lin, 2006), Fused Lasso (Tibshirani et al., 2005) and Elastic Net (Zou and Hastie, 2005) to solve more complicated problems.

By assigning appropriate prior density, Lasso can work in a Bayesian method as Park and Casella (2008) suggested. They constructed the Bayesian model for the parameters in the Lasso model and suggested that the square of tuning parameter follows gamma distribution. Based on the Bayesian Lasso, Zhao and Sarkar (2015) suggested a model constructing reliable confidence interval in the sparsity situation by changing the prior density for the coefficients. In addition, as a criterion to decide which variables are significant, Zhao and Sarkar (2015) used the local false discovery rate representing how significant a variable is. This concept was originated from false discovery rate suggested by Benjamini and Hochberg (1995). False discovery rate is associated with the Bonferroni correction in multiple testing in the way of controlling the overall error rate. Efron et al. (2001) focused on whether each coefficient is significant and extended false discovery rate to locally one, local false discovery rate. So, we proposed the model selecting coefficients based on the prior suggested by Zhao and Sarkar suggested in additive nonparametric regression in the Bayesian method.

The rest of this article is organized as follows.

2. Model Description

2.1. Zero Inflated Mixture Prior

Zhao and Sarkar (2015) developed a model based on the Bayesian Lasso model suggested by Park and Casella (2008). The model has a particular prior called zero inflated mixture prior(ZIMP) to deal with the problem of sparsity that modern data have.

$$\mathbf{Y}|\boldsymbol{\beta} \sim f(\mathbf{y}|\boldsymbol{\beta})$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$ and $\boldsymbol{\beta}$ has the prior as follows.

$$\beta_i \sim \begin{cases} 0 & \text{with probability } \pi_0 \\ \psi(\beta_i) & \text{with probability } 1-\pi_0 \end{cases}$$

where $\psi(\beta_i)$ is a known distribution when $\beta_i \neq 0$. The prior was applied to $\boldsymbol{\beta}$ commonly in Bayesian Lasso model to consider the situation that many variables are not significant because of lack of the number of data. We consider that situation is related to the fitting smooth function because all of βs are not used and a few βs shows significance. Therefore, to appproximate some unknown smooth functions, we propose a model with zero inflated mixture prior and different bases as following.

$$\mathbf{Y} - \bar{y}\mathbf{1}_n = \mathbf{b}^*(\mathbf{X})\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{\epsilon} \overset{i.i.d}{\sim} N(0, \sigma^2) \tag{1}$$

where $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_p)^\top$ is regression parameters, $\mathbf{Y}$ is n $\times$ 1 vector, $\mathbf{X}$ is the $n \times p$ matrix, σ^2 is unknown variance of ϵ and $\mathbf{b}^*(x)$ is standardized $\mathbf{b}(x)$. $\mathbf{b}(x)$ is defined as follows.

$$y_i = \beta_0 + \sum_{j=1}^{J} f_j(X_j) + \sum_{k=1}^{K} \beta_k Z_k + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} \mathrm{N}(0, \sigma^2)$$

$$\mathbf{b}(x) = \{x, |x-\tau_1|^2 log(|x-\tau_1|^2), \cdots, |x-\tau_\kappa|^2 log(|x-\tau_\kappa|^2)\}, \tag{2}$$

where τ_κs are fixed knots.

In this situation, we can consider selcting knots as a way to fit the given function with the polynomial function. However, our model approximates the given functions by determining which β of the polynomial function is 0 instead of selecting the location of knots when considering that all β are not used to fit the smooth functions.

2.2. Local False Discovery Rate

Benjamini and Hochberg (1995) introduced false discovery rate(fdr) to control the familywise error rate caused by repetitive single inference procedures. It was defined a proportion as following.

$$\frac{\text{The number of null hypotheses that is true although they are declared significant}}{\text{The number of hypotheses declared significant}}$$

In many cases, we are interested in the each probability that β_i is true under H_0 instead of controlling overall error rate. Therefore, Efron et al. (2001) introduced local fdr which implies a probability that hypothesis is true given the data. In other words, a local fdr is defined as a posterior probability that the corresponding null hypothesis can not be rejected. Suppose that a hypothesis testing as follows.

$$X \sim \begin{cases} f_0(X) & \text{with probability } p_0 \\ f_1(X) & \text{with probability } 1 - p_0 (= p_1) \end{cases}$$

where x has a density $f_0(x)$ if the H_0 is true otherwise x does $f_1(x)$. So we can calculate the local false rate as following.

$$P(H_0 \text{ is true}|x) = \frac{p_0 f_0(x)}{p_0 f_0(x) + p_1 f_1(x)}$$

Combined with the ZIMP model (2.1), we can consider fitting the given smooth functions as multiple hypothesis testing of $\boldsymbol{\beta}$. The null hypothesis is that $\beta_i = 0$ and the alternative hypothesis is that β_i follows normal distribution with mean 0 and unknown variance. We use the local fdr as a criterion to determine whether each β is 0 or not to approximate the given smooth functions. To be specific, the significance of each parameter is determined by multiple testing procedure controlling Bayes fdr(Sarkar et al., 2008) at significant level 0.05. To select appropriate parameters, we conducted variable selection based on the local false discovery rate(refer). First, conduct MCMC simulation. Then, we can notice which variables are used to build the model and transform the sample to 1 if the coefficients are not drawn from 0. As a result, we can compute the following p_j for each jth parameter.

$$p_j \equiv \frac{1}{K} \sum_{k=1}^{K} \mathrm{I}(\gamma_j \in X^{(k)})$$

γ_j means jth covariate and $X^{(k)}$ represents the set of covariates used to the build the model in the kth iteration. Therefore, $1 - p_j$ is the estimate of the local false discovery rate. Next, set the global FDR bound α in (0,1). Sort the local fdr calculated frome the previous step in ascending order and choose m local fdrs such that

$$\underset{1\leq m\leq p}{\arg\max}\left\{m\Big|\frac{1}{m}\sum_{i=1}^{m} fdr_{(i)}(\mathbf{X},\mathbf{Y}) \leq \alpha\right\} \tag{3}$$

where $fdr_{(i)}(\mathbf{X},\mathbf{Y})$ represents ith smallest local fdr. Let $p_{(m)} = \phi_\alpha$. Next, Claim set $\mathcal{X}_{\phi_\alpha} = \{j : p_j > \phi_\alpha\}$ is significant. Finally, the m parameters corresponding the chosen local fdrs are selected. In other words, we select the m βs not being likely to 0 when considering the data.

3. Bayesian Inference

3.1. Prior Specification

Park and Casella (2008) established the Bayesian model to determine which β is significant. The prior for β changed to mixture distribution used the ZIMP prior suggested by Zhao and Sarkar (2015). Therefore, considering the similarity of our model to estimate appropriate β, we choose priors for some parameters as follows based on the the model with ZIMP. The prior of μ is a flat prior. In case of σ^2, although an improper prior is used for σ^2, the posterior density for σ^2 is proper.

$$\begin{aligned}\pi(\mu) &= 1\\ \pi(\sigma^2) &= \text{Inv} - \text{Gamma}(a,b), a = 0, b = 1\end{aligned}$$

The prior probability that β is equal to 0 follows Beta distribution. k and η are fixed constants.

$$\pi_0 \sim Beta\left(k\eta, k(1-\eta)\right)$$

β_i follows a prior mixture distribution as follows.

$$\beta_i \sim \pi_0 I(\beta_i = 0) + (1-\pi_0)N(0,\sigma^2\tau_i^2)$$

τ_i follows the

$$\tau_1^2, \cdots, \tau_p^2 \sim \prod_{j=1}^{p} \frac{\lambda^2}{2} e^{-\lambda^2 \frac{\tau_j^2}{2}} d\tau_j^2$$

The square of Lasso parameter follows Gamma distribution. γ and δ are fixed constants.

$$\lambda^2 \sim Gamma\left(\gamma, \frac{1}{\delta}\right)$$

3.2. Posterior Computation

We draw samples from the below target posterior density. The columns of $\mathbf{X}$ are standardized and $\tilde{\mathbf{Y}}$ is defined as $\mathbf{Y} - \bar{y}\mathbf{1}_n$. We usually marginalize out μ because it is out of our interest.

$$
\begin{aligned}
&P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\Sigma}, \boldsymbol{\pi}, \boldsymbol{\lambda}^2, \sigma^2 | \tilde{\mathbf{Y}}) \\
&= \int P(\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \boldsymbol{\Sigma}, \boldsymbol{\pi}, \boldsymbol{\lambda}^2, \sigma^2, \mu | \mathbf{Y}) d\mu \\
&\propto p(\tilde{\mathbf{Y}} | \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}^2, \sigma^2) \prod_{g=1}^{G} \prod_{j=1}^{m_g} p(\alpha_{gj} | \boldsymbol{\tau}^2, \sigma^2, \pi_g) p(\boldsymbol{\beta} | \boldsymbol{\Sigma}, \sigma^2) \\
&\prod_{g=1}^{G} \prod_{j=1}^{m_g} p(\tau_{gj}^2 | \lambda_g^2) \prod_{g=1}^{G} p(\pi_g | \alpha_{\pi_g}, \beta_{\pi_g}) \prod_{g=1}^{G} p(\lambda_g^2 | \gamma, \delta) p(\sigma^2 | a, b)
\end{aligned}
$$

Step 1. Draw θ_{gj} from the following a mixture distribution,

$$
\theta_{gj} | \tilde{\mathbf{Y}}, \boldsymbol{\theta}_{-gj}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}^2, \boldsymbol{\pi}, \sigma^2 \sim \begin{cases} 0 & \text{with probability } l_{gj} \\ \mathrm{N}\left(\dfrac{(\tilde{\mathbf{Y}} - \mathbf{X}_{-gj}\boldsymbol{\theta}_{-gj})^\top \mathbf{X}_{gj}}{\mathbf{X}^\top{}_{gj}\mathbf{X}_{gj} + \dfrac{1}{\tau_{gj}^2}}, \dfrac{\sigma^2}{\mathbf{X}^\top{}_{gj}\mathbf{X}_{gj} + \dfrac{1}{\tau_{gj}^2}} \right) & \text{with probability } 1 - l_{gj} \end{cases}
$$

where

$$
l_{gj} = \frac{\pi_g}{\pi_g + (1 - \pi_g)(1 + \tau_{gj}^2 \mathbf{X}_{gj}^\top \mathbf{X}_{gj})^{-\frac{1}{2}} \exp\left(\dfrac{((\tilde{\mathbf{Y}} - \mathbf{X}_{-gj}\boldsymbol{\theta}_{-gj})^\top \mathbf{X}_{gj})^2}{2\sigma^2 \left(\mathbf{X}_{gj}^\top \mathbf{X}_{gj} + \dfrac{1}{\tau_{gj}^2} \right)} \right)}
$$

Step 2. Draw π_0 from Draw π_g from $p(\pi_g | \tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}^2, \boldsymbol{\pi}_{-g}, \sigma^2)$ which is an beta distribution,

$$
\pi_g | \tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}^2, \boldsymbol{\pi}_{-g}, \sigma^2 \sim \mathrm{Beta}\left(k\eta + \sum_{j=1}^{m_g} Z_{gj}, k(1 - \eta) + m_g - \sum_{j=1}^{m_g} Z_{gj} \right)
$$

where

$$
Z_{gj} \equiv \mathrm{I}(\theta_{gj} = 0)
$$

Step 3. Draw σ^2 from $p(\sigma^2|\tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}^2, \boldsymbol{\pi})$ which is an inverse gamma distribution,

$$\sigma^2|\tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}^2, \boldsymbol{\pi} \sim \text{IG}\left(\frac{N}{2}, \frac{\|\tilde{\mathbf{Y}} - \mathbf{X}\boldsymbol{\theta}\|_2^2 + \boldsymbol{\theta}^\top \mathbf{D}_{\boldsymbol{\tau}}{}^{-1}\boldsymbol{\theta}}{2}\right)$$

where

$$N \equiv n + q - 1 + \sum_{g=1}^{G} g m_g - \sum_{g=1}^{G}\sum_{j=1}^{m_g} Z_{gj}$$
$$\mathbf{D}_{\boldsymbol{\tau}} \equiv diag(\tau_{g1}^2, \tau_{g2}^2, \cdots, \tau_{Gm_G}^2)$$

Step 4. Draw τ_{gj}^2 from $(\tau_{gj}^2)^{-1}|\tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}_{-gj}, \boldsymbol{\lambda}^2, \boldsymbol{\pi}, \sigma^2 \sim$

$$\begin{cases} \text{Inv-Gamma}\left(1, \dfrac{\lambda_g^2}{2}\right), & \text{if } \theta_{gj} = 0 \\ \text{Inv-Gaussian}\left(\sqrt{\dfrac{\lambda_g^2\sigma^2}{\theta_{gj}^2}}, \lambda_g^2\right), & \text{if } \theta_{gj} \neq 0 \end{cases}$$

Step 5. Draw λ_g^2 from $p(\lambda_g^2|\tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}_{-g}^2, \boldsymbol{\pi}, \sigma^2)$ which is a gamma distribution,

$$\lambda_g^2|\tilde{\mathbf{Y}}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\tau}, \boldsymbol{\lambda}_{-g}^2, \boldsymbol{\pi}, \sigma^2 \sim \text{Gamma}\left(m_g + \gamma, \left(\frac{\sum_{j=1}^{m_g}\tau_{gj}^2}{2} + \delta\right)^{-1}\right)$$

4. Simulation Study

We conducted a simulation study to assess the performance of the proposed model using various functions from simple linear functions to complicated nonlinear functions. We set $n = 600$ and $p = 31$. 30 knots are located with the same interval in [0,1]. The locations of knots are fixed. σ for ϵ was set to 1. The above panels show the probabilities that each β is chosen and the below panels show the true functions(dashed lines), point-wise medians of the functions(solid lines) and point-wise 95% intervals.(gray regions). Each function in nonparametric components represents non significant, linearly significant, non-linearly significant relationship between explanatory variables and dependent variable. Also, parametric components consist of non significant and linearly significant relationship.

$$y_i = \beta_0 + \sum_{j=1}^{3} f_j(X_j) + \sum_{k=1}^{2} \beta_k Z_k + \epsilon_i,\ \epsilon_i \overset{iid}{\sim} \text{N}(0, \sigma^2)$$

$$f_1(x) = 1$$
$$f_2(x) = 2x + 1$$
$$f_3(x) = \sin(8(x - 0.5)) + 2\exp(-16^2(x - 0.5)^2)$$
$$(\beta_1, \beta_2)^\top = (0, -3)^\top$$

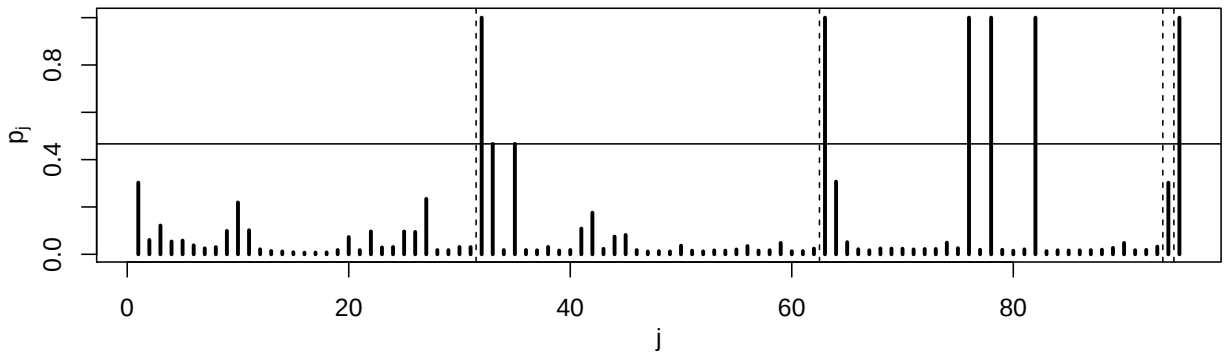

Figure 1: plot of estimate of all local fdrs

j		$p*$	True	Decision
1–31	f_1	0	not significant	not significant
31–62	f_2	1	linearly significant	linearly significant
63–93	f_3	4	non-linearly significant	non-linearly significant
94	β_1	0	not significant	not significant
95	β_2	1	linearly significant	linearly significant

Table 1: $p*$:The number of selected covariates

Figure 1 represents the plot with whole local false discovery rates. Vertical dotted lines are used to identify local false discovery rate group corresponding each function defined above. Therefore, there are 31 local false discovery rates in each group and the last 2 local false discovery rates correspond to parametric components. The horizontal lines means phi.alpha determined based on the false discovery rate. As a result, covariates corresponding to f_1 and β_1 are not selected, which means there is no signifcant coefficients in the groups. Although some samples frome the MCMC simulations are not 0 in the groups, we don't use the sample and determine all coefficients in the groups as 0. f_2 has only one significant coefficient and the coefficient corresponds to the first basis term representing linear relationship between x and y. Therefore we don't consider other coefficients corresponding to other basis term when building the model. β_2 is chosen by the decision rule based on the false discovery rate. Finally, f_3 which has non linearly relationship shows that there are 4 significant covariates. Therefore, we can build f_3 with only 4 covariates. As a result, among the 3 nonparametric components and 2 parametric components 3 components are choses as significant. In addition, we started the whole model with 95 covariates but have built the model with 6 covariates. Figure 2 shows the posterior distributions of parametric components and error variance. The dashed line represents the true value of each paramter. In the Figure 2, the marginal posterior distribution are well covered the true vertical line, which illustrates the validity of the model that we proposed.

The first row of Figure 3 shows the The second row of Figure 3 shows fitted nonparametric regression plot before choosing significant coefficients using local false discovery rate. Before applying fdr based variable selection, we can identify the overfitting problem. Finally, the third row of Figure 3 shows fitted nonparametric regression plot after choosing significant coefficients using local false discovery rate which means our proposed model. By using local false discovery rate, we

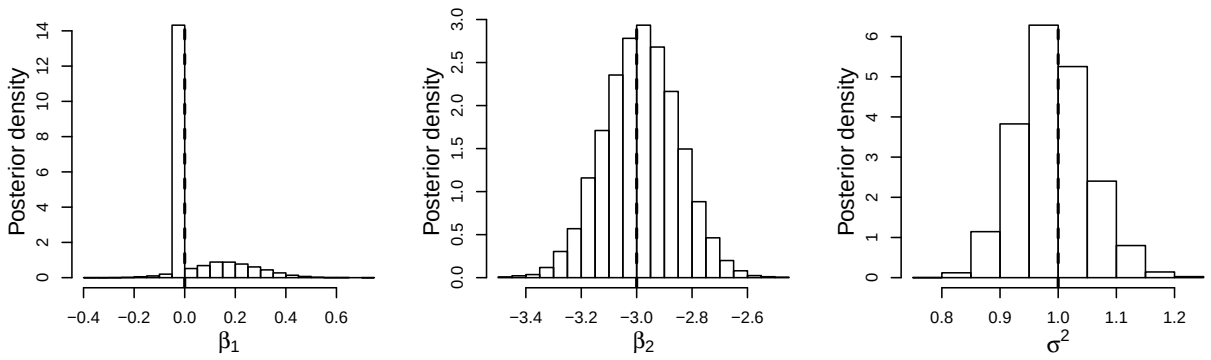

Figure 2: Marginal peosterior distributions for parametric components and error variance

can not only build the additive partial linear model parsimoniously, but also improve the ovefitting problem.

5. Real data analysis

There have been many kinds of research that discuss the relationship between cancers and beta-carotene. Such epidemiological studies have shown the influential antioxidant properties which beta-carotene has can help prevent cancer. In addition, beta-carotene has the ability to purge the body of free radicals which can lead to cancer. Therefore, many researchers such as clinicians and nutritionists have tried to investigate the relationship between concentrations of beta-carotene and other factors like age, alcohol consumption, dietary intake, smoking status that can help them to make a clinical decision and to customize the therapy. For example, Nierenberg et al found that there was a positive relationship between beta-carotene and not only sex but also indietary carotene, whereas smoking status and body mass index(BMI) has a negative relationship. However, Faure et al. found that beta-carotene was related to age,gender,smoking status, In sum, there have been various results which make them less convinicing.

The dependent variable is concentrations of beta-carotene from 315 subjects and it was transformed to log(beta-carotene+1). numerical error when applying logarithm casued by 0 value in the beta-carotene variable by adding 1 to the variable. The factors to investigate the relationship are as follows: AGE, BMI, CALORIES, FAT, FIBER, ALCOHOL, CHOL, BETADIET, RETDIET, SEX, SMOKE.

$$\begin{aligned}\log(\text{beta-carotene}+1) = \beta_0 &+ f_1(\text{AGE}) + f_2(\text{BMI}) + f_3(\text{CALORIES}) + f_4(\text{FAT}) \\ &+ f_5(\text{FIBER}) + f_6(\text{ALCOHOL}) + f_7(\text{CHOL}) + f_8(\text{BETADIET}) \\ &+ f_9(\text{RETDIET}) + \beta_1\text{SEX} + \beta_2\text{SMOKE} + \epsilon\end{aligned}$$

The number of knots is 9. Therefore, the number of basis term for each nonparametric components is 10. Except for the categorical variables such as SEX and SMOKE, we considers all variables into nonparametric components expecting the proposed model can detect the linear relationship although the variabels are fitted under the nonparametric components.

Xiang Liu et al. predetermined the variales and the relationship between the variables. Ther result shows that AGE and RETDIET has non-linear significant relationship.

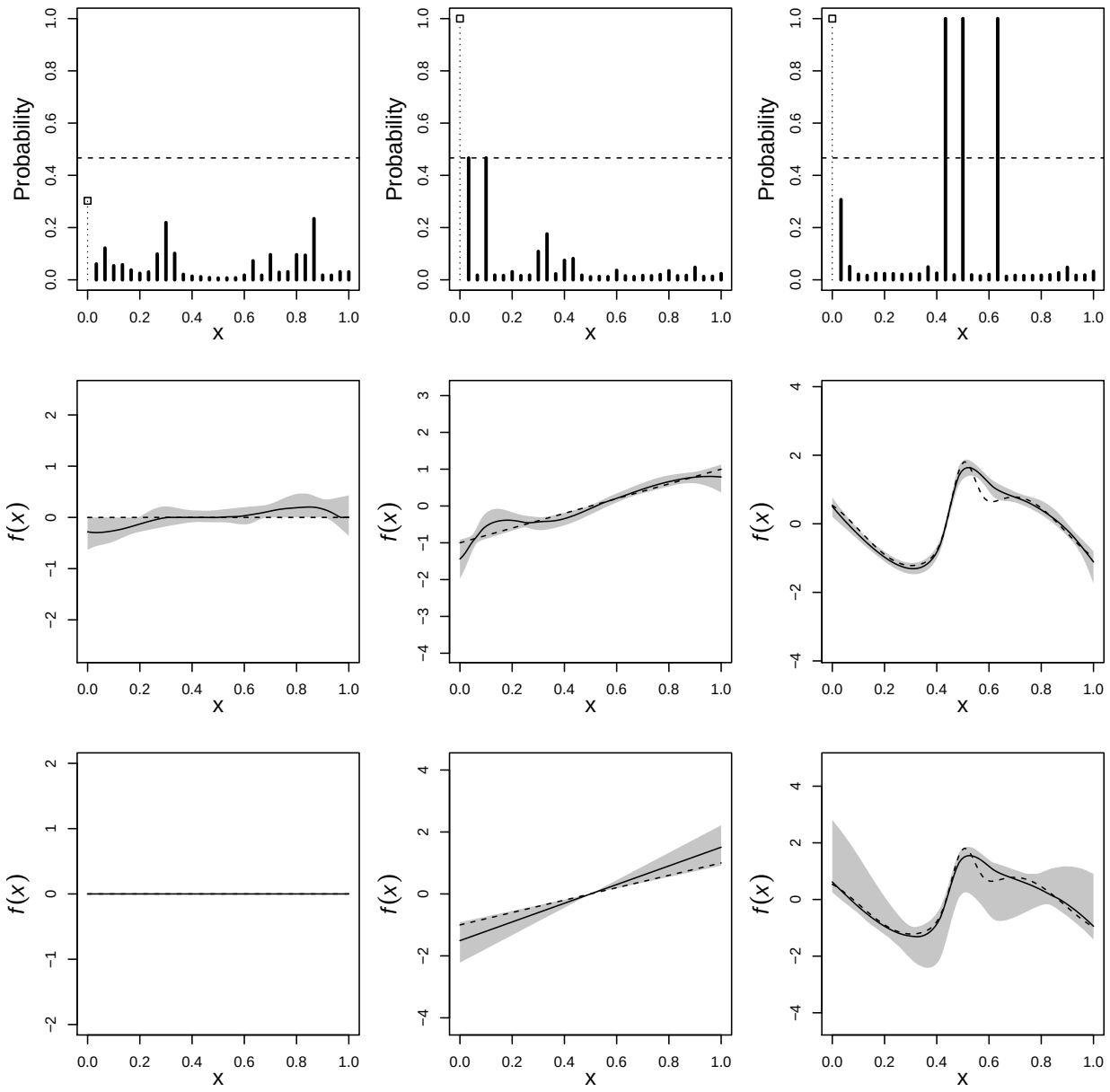

Figure 3: simulation result plot

Acknowledgements

T. Park's research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A1B03033536).

References

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the royal statistical society. Series B (Methodological), 289–300.

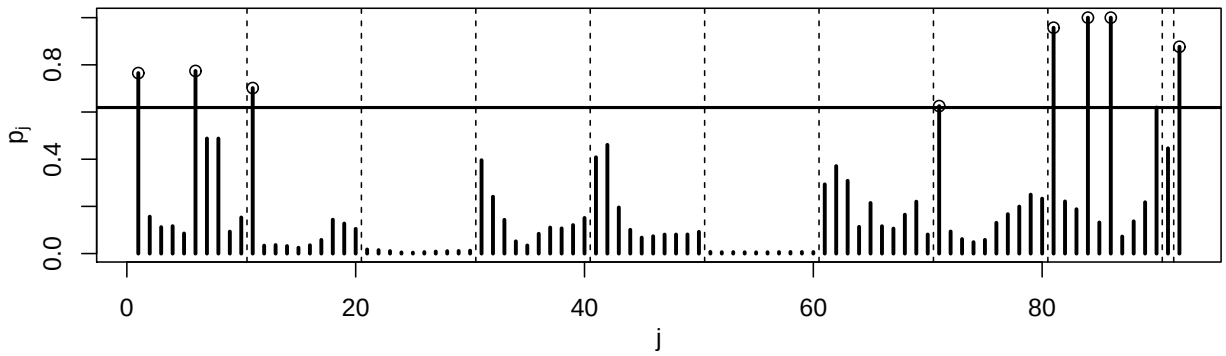

Figure 4: real data result plot

j	Covariates	$p*$	Decision	Previous study		
				SCAD	LASSO	BIC
1–10	AGE	2	non-linearly	non-linearly	non-linearly	non-linearly
11–20	BMI	1	linearly	linearly	lineary	lineary
21–30	CALORIES	0	not	not	not	not
31–40	FAT	0	not	not	not	not
41–50	FIBER	0	not	not	linearly	not
51–60	ALCOHOL	0	not	not	not	not
61–70	CHOL	0	not	non-linearly	non-linearly	non-linearly
71–80	BETADIET	1	linearly	linearly	linearly	not
81–90	RETDIET	3	non-linearly	-	-	-
91	SEX	0	not	linearly	linearly	not
92	SMOKE	1	linearly	linearly	linearly	linearly

Table 2: $p*$:The number of selected covariates

Parameter	mean	median	95% lower	95% upper
Intercept	4.953	4.953	4.874	5.032
σ^2	0.505	0.505	0.416	0.628
Covariate				
BMI	-0.033	-0.033	-0.052	0
BETADIET	0.001	0	-0.005	0.014
SMOKE	-0.156	-0.167	-0.286	0

Table 3: summary for parameters and covariates determined as linearly significant

Efron, B., Tibshirani, R., Storey, J. D., Tusher, V., 2001. Empirical bayes analysis of a microarray experiment. Journal of the American statistical association 96 (456), 1151–1160.

Hoerl, A. E., Kennard, R. W., 1970. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics 12 (1), 55–67.

Miyata, S., Shen, X., 2003. Adaptive free-knot splines. Journal of Computational and Graphical Statistics 12 (1), 197–213.

Park, T., Casella, G., 2008. The bayesian lasso. Journal of the American Statistical Association 103 (482), 681–686.
Sarkar, S. K., Zhou, T., Ghosh, D., 2008. A general decision theoretic formulation of procedures controlling fdr and fnr from a bayesian perspective. Statistica Sinica, 925–945.
Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society. Series B (Methodological), 267–288.
Tibshirani, R., Saunders, M., Rosset, S., Zhu, J., Knight, K., 2005. Sparsity and smoothness via the fused lasso. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (1), 91–108.
Yuan, M., Lin, Y., 2006. Model selection and estimation in regression with grouped variables. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68 (1), 49–67.
Zhao, Z., Sarkar, S. K., 2015. A bayesian approach to constructing multiple confidence intervals of selected parameters with sparse signals. Statistica Sinica, 725–741.
Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 67 (2), 301–320.