Scientific
Research

Variable Selection for Partially Linear Varying Coefficient Transformation Models with Censored Data[*]

Jiang Du[1][#], Zhongzhan Zhang[1], Ying Lu[1,2]

[1]College of Applied Sciences, Beijing University of Technology, Beijing, China
[2]College of Applied Sciences, Communication University of China, Beijing, China
Email: [#]dujiang84@163.com

ABSTRACT

In this paper, we study the problem of variable selection for varying coefficient transformation models with censored data. We fit the varying coefficient transformation models by maximizing the marginal likelihood subject to a shrinkage-type penalty, which encourages sparse solutions and hence facilitates the process of variable selection. We further provide an efficient computation algorithm to implement the proposed methods. A simulation study is conducted to evaluate the performance of the proposed methods and a real dataset is analyzed as an illustration.

Keywords: Variable Selection; Maximum Likelihood Estimation; Spline Smoothing

1. Introduction

To model the effects of X on the failure time T, the following linear transformation models are considered

$$H_0(T) = -X^T \beta_0 + \varepsilon , \qquad (1)$$

where H_0 is an unknown, strictly increasing function with $H_0(0) = 0$, β_0 is a $p \times 1$ vector of unknown regression parameters, and the error term ε has a known continuous distribution that is independent of X. In the presence of censoring, we observe the event time $\tilde{T} = \min(T, C)$ and the censoring indicator $\delta = I(T \leq C)$, where C is the censoring time and $I(\cdot)$ is the indicator function. It is usually assumed that the censoring variable C is independent of T given X. Linear transformation model (1), encompassing Cox's model and the proportional odds model as special cases, provide a popular tool in analyzing time-to-failure data, and have recently attracted considerable attention due to their high flexibility [1-4].

In practice, in order to accurately predict the survival time, many explanatory variables are generally collected and need to be assessed during the initial analysis. Deciding which covariates to have significant effect on the survival time and which variables are non-informative is practically interesting, but is always a tricky task for survival analysis. Variable selection is therefore of fundamental interest for survival analysis. Variable selection procedures for survival data have been studied extensively in the literature; see [5-7] for cox's model; see [8]

for proportional odds model.

Partially linear varying coefficient regression models have played an increasingly important role in statistical analysis for a good balance between flexibility and parsimony. Recently, varying coefficient regression models have been studied by many authors. [9] proposed a local least-squares method with a kernel weight function. [10] proposed an estimation procedure based on the local polynomial fitting method. [11] proposed a profile least-squares technique for estimating the parametric components and applied the generalized likelihood ratio techniques to the testing problems for the nonparametric components. [12] proposed a class of efficient estimation method to estimate a partially linear varying coefficient model. Under quantile loss function, based on B-spline basis function approximations, [13] proposed the estimator and test for partially linear varying coefficient model. As for variable selection, [14] adopted the SCAD variable selection procedure to select important variables in the parametric components, and adopted the generalized likelihood ratio tests to select important variables in the nonparametric components of semi-parametric model. However, very little work has been done for partially linear varying coefficient transformation models with censored data, which are defined as following

$$\log H_0(T) = -X^T \beta_0 - Z^T \theta_0(W) + \varepsilon , \quad (2)$$

where H_0 is an unknown, strictly increasing function with $H_0(0) = 0$, θ_0 comprises q unknown smooth functions, $Z = (Z_1, \cdots, Z_q)$ is a q dimensional covariate, β_0 is the regression parameters, $X = (X_1, \cdots, X_p)$ is a

p dimensional covariate, W ranges over a non-degenerate compact interval, without loss of generality, that is assumed to be the unit interval $[0,1]$ and ε is the error with known continuous density, f_ε, and is independent of the covariates. In the absence of Z and W, model (2) reduces to a linear transformation model (1). When Z is a constant, model (2) reduces to a partial linear transformation model. [15] studied the partially linear transformation models using martingale representation based estimating equation. [16] proposed variable selection for the partially linear transformation models. To the best of our knowledge, there is no paper considering the variable selection for varying coefficient transformation models with censored data. To solve this problem, in this paper, we propose the penalized likelihood function method for variable selection in varying coefficient transformation models with censored data. The penalized likelihood function with adaptive Lasso is proposed. We only consider here the application of adaptive Lasso, because the convexity of L_1 penalty enables us to demonstrate the method easily. Nevertheless, the same idea is readily applicable to other shrinkage methods, such as SCAD proposed by [17], or one-step sparse estimator of [18]. The major advantages of the proposed procedures over the existing ones are their computational expediency and the most existing procedures are our special cases for survival data. Simulation experiments have provided evidence of the superiority of the proposed procedures.

The paper is organized as follows. In Section 2, we present the penalized maximum likelihood method with the adaptive Lasso (ALasso) penalty. A penalized maximum likelihood estimator is proposed for optimizing the penalized log marginal likelihood functions. In Section 3, we present an algorithm to implement the penalized maximum likelihood estimator. In Section 4, we conduct a simulation to detect the behavior of the proposed method. In Section 5, we illustrate the proposed procedure via analyzing lung cancer data.

2. Penalized Maximum Likelihood Estimation

Let $\left(x_i, z_i, w_i, \delta_i, T_i, C_i, \tilde{T}_i, \varepsilon_i \right)$ be independent and identically distributed copies of $\left(X, Z, W, \delta, T, C, \tilde{T}, \varepsilon \right)$. The observations are $\left\{ \left(x_i, z_i, w_i, \delta_i, \tilde{T}_i \right) : i = 1, \cdots, n \right\}$. Then, from model (2), the log-likelihood function of (β, θ, H) can be written as

$$
\begin{aligned}
L(\beta, \theta, H) \\
= \sum_{i=1}^{n} \Big(\delta_i \Big[\log \lambda_\varepsilon \left\{ \log \left(\tilde{T}_i \right) + x_i^T \beta + z_i^T \theta(w_i) \right\} \\
- \log \left(H \left(\tilde{T}_i \right) \right) + \log \left(\dot{H} \left(\tilde{T}_i \right) \right) \Big] \\
- \Lambda_\varepsilon \left\{ \log \left(\tilde{T}_i \right) + x_i^T \beta + z_i^T \theta(w_i) \right\} \Big)
\end{aligned}
$$

where λ_ε and Λ_ε are the baseline and cumulative hazard functions of ε, respectively, and H is an increasing function. Throughout the paper, $\dot{f}$ and $\ddot{f}$ respectively denote the first and second derivatives of f provided that they exist. Following [19] and [20], we approximate $\theta_l(t), l = 1, \cdots, q$ by the B-spline method. Let $\pi(t) = \left(B_1(t), \cdots, B_{k_n+1+\iota}(t) \right)$ be a set of B-spline basis functions of order $\iota + 1$ with k_n quasi-uniform internal knots. Then, $\theta_l(t)$ can be approximated by

$$
\theta_l(t) \approx \pi(t)^T \alpha_l, \tag{3}
$$

where $\alpha_l = \left(\alpha_{l,1}, \cdots, \alpha_{l,k_n+1+\iota} \right)^T$ is the spline coefficient vector, substituting (3) into the model (2), we can get log-likelihood function:

$$
\begin{aligned}
L(\beta, \alpha, H) = \sum_{i=1}^{n} \Big(\delta_i \Big[\log \lambda_\varepsilon \left\{ \log \left(\tilde{T}_i \right) + x_i^T \beta + \Pi(w_i) \alpha \right\} \\
- \log \left(H \left(\tilde{T}_i \right) \right) + \log \left(\dot{H} \left(\tilde{T}_i \right) \right) \Big] \\
- \Lambda_\varepsilon \left\{ \log \left(\tilde{T}_i \right) + x_i^T \beta + \Pi(w_i) \alpha \right\} \Big)
\end{aligned}
$$

where $\Pi(t) = \left(z_1 \pi(t), \cdots, z_q \pi(t) \right), \alpha = \left(\alpha_1, \cdots, \alpha_q \right)^T$.

In order to obtain the sparsity solution, this motivates us to consider a penalized log-likelihood function. The penalized log-likelihood function with adaptive Lasso is defined as

$$
Q(\beta, \alpha, H) = L(\beta, \alpha, H) - n \sum_{j=1}^{p} \lambda_{jn} \tilde{\tau}_j \left| \beta_j \right| \tag{4}
$$

where λ_{jn} is tuning parameters. Denote the maximum likelihood estimate (MLE) as $\left(\tilde{\beta}, \tilde{\alpha}, \tilde{H} \right)$. Since $\tilde{\beta}$ are shown to be consistent [21]. Therefore, we choose $\tilde{\tau}_j = \left| \tilde{\beta} \right|_j^{-1}$.

3. Computation Algorithm

We provide an iterative computation algorithm to solve (4). In the spirit of nonparametric maximum likelihood, we restrict $H(\cdot)$ to be a non-decreasing step function with $H(0) = 0$ and with jumps only at $t_1 < \cdots < t_{m_n}$, where t_j is the jth smallest observed failure time and $m_n = \sum_{i=1}^{n} \delta_i$ is the total number of observed failures. Consider

$$
\begin{aligned}
Q(\beta, \alpha, H) = \sum_{i=1}^{n} \Big(\delta_i \Big[\log \lambda_\varepsilon \left\{ \log \left(\tilde{T}_i \right) + x_i^T \beta + \Pi(w_i) \alpha \right\} \\
- \log \left(H \left(\tilde{T}_i \right) \right) + \log \left(\Delta H \left(\tilde{T}_i \right) \right) \Big] \\
- \Lambda_\varepsilon \left\{ \log \left(\tilde{T}_i \right) + x_i^T \beta + \Pi(w_i) \alpha \right\} \Big) \\
- n \sum_{j=1}^{p} \lambda_{jn} \tilde{\tau}_j \left| \beta_j \right|
\end{aligned}
$$

where $\Delta H(t) = H(t) - H(t-)$ and α is $q \times (k_n + 1 + \iota)$. A direct numerical maximization of the above $Q(\beta, \alpha, H)$ may be problematic since the number of parameters, $p + q \times (k_n + 1 + \iota) + m_n$, may be very large. We recommend the following iteration procedure, which has been stable and effective in our numerical experience.

Step 1. Choose an intial value $(\beta^{(0)}, \alpha^{(0)}, H^{(0)})$.

Step 2. Set the estimator to be $(\beta^{(k)}, \alpha^{(k)}, H^{(k)})$ at the kth iteration. By differentiating $Q(\beta, \alpha, H)$ with repect to $\Delta H(t_j), j = 1, \cdots, m_n$, we can obtain

$$\Delta H^{(k+1)}(t_j) = \sum_{i=1}^{n} \delta_i I(\tilde{T}_i = t_j) \Big/ \psi_n(t_j; \beta^{(k)}, \alpha^{(k)}, H^{(k)})$$

where

$$\psi_n(t; \beta, \alpha, H) = \sum_{i=1}^{n} I(\tilde{T}_i \geq t) \Big[\lambda_\varepsilon \{\log H(\tilde{T}_i) + x_i^T \beta + \Pi(w_i)\alpha\} + \delta_i - \delta_i \dot{\lambda}_\varepsilon \{\log H(\tilde{T}_i) + \alpha\}$$
$$+ \delta_i - \delta_i \dot{\lambda}_\varepsilon \{\log H(\tilde{T}_i) + x_i^T \beta + \Pi(w_i)\alpha\} \Big/ \{\lambda_\varepsilon \log H(\tilde{T}_i) + x_i^T \beta + \Pi(w_i)\alpha\} \Big] \Big/ H(\tilde{T}_i)$$

Then, $H^{(k+1)}(t) = \sum_{j=1}^{m_n} \Delta H^{(k+1)}(t_j) I(t \geq t_j)$.

Step 3. By differentiating $Q(\beta, \alpha, H)$ with respect to (β, α), we obtain the score function, $\dot{Q}(\beta, \alpha, H)$ and obtain the root $(\beta^{(k+1)}, \alpha^{(k+1)})$ to the equation $\dot{Q}(\beta, \alpha, H^{(k+1)}) = 0$ by using the Newton-Raphson iterative algorithm.

Step 4. Repeat Steps 2 and 3 until convergence.

4. Simulation Studies

In this section, we examine the finite sample performance of the proposed procedures. The model is

$$\log H_0(T) = -X^T \beta_0 + Z\theta_0(W) + \varepsilon, \quad (6)$$

where $H_0(t) = t/2, \theta_0(t) = 8t(1 - t^2), \beta_0 = (3, 0, 1.5, 0, 0)^T$ and the components of X is standard normal distribution. The correlation between X_i and X_j is $\rho^{|i-j|}$ with $\rho = 0$ and $\rho = 0.5$. Z is uniformly distributed on $[0,2]$. The distribution of W is independent of the Z and $W \sim Un[0,1]$. The baseline hazard function of error ε, is $\exp(t)/\{1 + \upsilon \exp(t)\}$, where υ is constant; see [2] and [21]. The model (6) reduces to the proportional hazards model if $\upsilon = 0$ and to the proportional odds model if $\upsilon = 1$. We consider $\upsilon = 0, 0.5$ and 1. The censoring time, C, follows the exponential distribution with mean 1. The largest follow-up time is set to 2.

In our simulation, we use the cubic B-spline with two interior knots to approximate the functional coefficient. The performance of $\hat{\theta}(\cdot)$ will be assessed by using the square root of average errors (RASE)

$$\text{RASE} = \left\{ \frac{1}{N} \sum_{s=1}^{N} \left[\hat{\theta}(t_s) - \theta(t_s)\right]^2 \right\}^{1/2},$$

where $t_s, s = 1, \cdots, N$ are the grid points at which the function $\hat{\theta}(t)$ is evaluated. In this simulation, $N = 200$ is used. To improve the computational efficiency, we choose λ_{jn} by the method proposed by [22], which is $\lambda_{jn} = \frac{\log(n)}{n|\tilde{\beta}_{nj}|}$, where $\tilde{\beta}_{nj}$ is maximum likelihood estimation without any penalty. To evaluate variable selection performance of adaptive Lasso and Oracle method, we report the results over 100 simulations in **Table 1**. Both columns "C" and "IC" are measures of model complexity. Column "C" shows the average number of zero coefficients correctly estimated to be zero, and column "IC" presents the average number of nonzero coefficients incorrectly estimated to be zero. As in [14], the performance of estimator $\hat{\beta}$ will be assessed by using the generalized mean square error (GMSE), defined as

$$\text{GMSE} = (\hat{\beta} - \beta_0)^T E(XX^T)(\hat{\beta} - \beta_0).$$

The column "GMSE" and "RASE" report both the mean of 100 GMSEs and RASEs, respectively.

We now summarize the main findings from **Table 1**.

1) The estimation errors of the ALasso method approach rapidly those of the oracle estimators as the sample size increases.

2) GMSE and RASE are decreasing with increase of sample size.

Figure 1 demonstrates the performance of the curve estimation of $\hat{\theta}(w)$ for I with $n = 100$. From the **Figure 1** we can see that the estimated curves are very close to the true curve $\theta_0(w)$. We may conclude that the proposed estimation of the function $\hat{\theta}(w)$ performs reasonably well. For other cases, the figures perform similarly with **Figure 1**, to save space, we omit them.

5. An Application

Now we illustrate the proposed procedures by an analysis of the lung cancer data from the Veteran's Admini-

Table 1. Simulation results for different *v*, *ρ* and different sample size.

Method	ρ	v	n = 100				n = 300			
			IC	C	GMSE	RASE	IC	C	GMSE	RASE
ALasso	0	0	0	2.5	0.191	0.441	0	2.58	0.039	0.217
Oracle	0	0	0	3	0.146	0.413	0	3	0.034	0.216
ALasso	0	0.5	0.01	2.26	0.257	0.481	0	2.25	0.081	0.237
Oracle	0	0.5	0	3	0.185	0.476	0	3	0.050	0.241
ALasso	0	1	0	2.20	0.273	0.614	0	2.35	0.127	0.291
Oracle	0	1	0	3	0.162	0.619	0	3	0.057	0.291
ALasso	0.5	0	0	2.45	0.191	0.442	0	2.62	0.062	0.220
Oracle	0.5	0	0	3	0.140	0.434	0	3	0.057	0.223
ALasso	0.5	0.5	0	2.42	0.204	0.563	0	2.46	0.077	0.238
Oracle	0.5	0.5	0	3	0.150	0.548	0	3	0.044	0.240
ALasso	0.5	1	0.01	2.48	0.344	0.550	0	2.36	0.096	0.280
Oracle	0.5	1	0	3	0.216	0.558	0	3	0.047	0.284
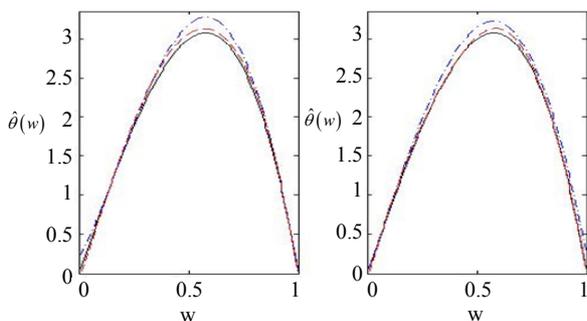
Figure 1. Estimated and curves (solid line) for *v* = 0 (dash-dot line), and *v* = 1 (dash line). The left panel is the estimator of $\theta_0(w)$ for *ρ* = 0. The right panel is the estimator of $\theta_0(w)$ for *ρ* = 0.5.

stration lung cancer trail. In this trial, 137 males with advanced inoperable lung cancer were randomized to either a standard treatment or chemotherapy. Details of the study design and method have been given by [23]. The data set has been analyzed by many authors, for example, [5] fitted the proportional hazards model and [8] considered the proportional odds model. In both methods, all covariates were assumed linear, which may not be true, particularly for the age effect. It is well known that age is a complex confounding factor, and its effect usually shows a nonlinear trend [15].

In our analysis, we take X_1 to be the treatment: (1 = standard, 2 = test), X_2 to be cell type: (1 = squamous, 2

= small cell, 3 = adeno, 4 = large), X_3 to be Karnofsky score, X_4 to be months from diagnosis, X_5 to be prior therapy (0 = no, 1 = yes) and *W* to be age. So we consider the following model:

$$\log H_0(T) = -X_1\beta_1 - X_2\beta_3 - X_3\beta_3 - X_4\beta_4 \\ -X_5\beta_5 - \theta_1(W) + \varepsilon \tag{7}$$

For estimation, we first rescale age between 0 and 1, and use cubic B-spline with two interior knots to approximate the nonparametric function as in simulations. We further apply the penalized likelihood function approach to select significant variable. For tuning parameters, we choose λ_{jn} by the method proposed by [22] as conducted in simulations. For comparison, we show the regression results for the proportional hazards model with $\upsilon = 0$ and the proportional odds model with $\upsilon = 1$, respectively.

Table 2 gives estimators for the regression coefficient and corresponding standard errors. From **Table 2**, only the variable X_3 is selected, in other words, only Karnofsky score has significant effect on survival time. Our findings is the same with [5] expect the age. **Figure 2** gives the estimator of the nonlinear components. From **Figure 2**, it is easy to see that age has nonlinear effect on the survival function. Therefore, the partially linear transformation models can be more powerful in discovering significant covariates than those assuming simply linear covariate

Table 2. Estimated coefficients for model (6).

	$v = 0$		$v = 0.5$	
Variable	MLE	ALasso	MLE	ALasso
X_1	0.179(0.185)	0(0)	0.106(0.244)	0(0)
X_2	0.103(0.078)	0(0)	0.058(0.113)	0(0)
X_3	−0.031(0.005)	−0.026(0.005)	−0.041(0.006)	−0.035(0.006)
X_4	0.003(0.009)	0(0)	0.002(0.012)	0(0)
X_5	−0.070(0.222)	0(0)	−0.118(0.293)	0(0)
	$v = 1$		$v = 2$	
Variable	MLE	ALasso	MLE	ALasso
X_1	0.086(0.287)	0(0)	0.110(0.354)	0(0)
X_2	−0.006(0.143)	0(0)	−0.127(0.189)	0(0)
X_3	−0.046(0.007)	−0.040(0.007)	0.053(0.009)	−0.045(0.009)
X_4	0.000(0.015)	0(0)	−0.0025(0.018)	0(0)
X_5	−0.108(0.350)	0(0)	−0.063(0.437)	0(0)
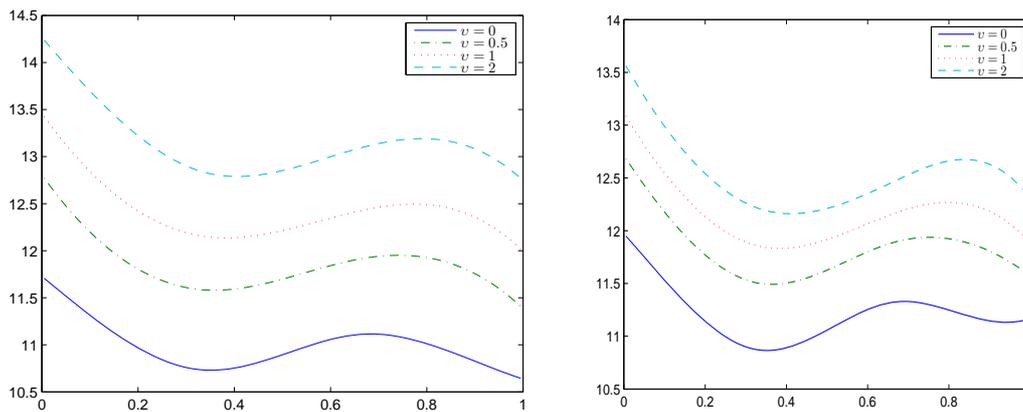
Figure 2. Estimated of $\hat{\theta}(w)$ for v = 0, 0.5, 1 and 2.

effects, which is consistent with the conclusions of [16].

REFERENCES

[1] S. C. Cheng, L. J. Wei and Z. Ying, "Analysis of Transformation Models with Censored Data," *Biometrika*, Vol. 82, No. 4, 1995, pp. 835-845. doi:10.1093/biomet/82.4.835

[2] K. Chen, Z. Z. Jin and Z. L. Ying, "Semiparametric Analysis of Transformation Model with Censored Data," *Biometrika*, Vol. 89, No. 3, 2002, pp. 659-668. doi:10.1093/biomet/89.3.659

[3] D. Zeng and D. Y. Lin, "Efficient Estimation in the Accelerated Failure Time Model," *Journal of the American Statistical Association*, Vol. 102, No. 480, 2007, pp. 1387-1396. doi:10.1198/016214507000001085

[4] D. Zeng and D. Y. Lin, "Efficient Estimation of Semiparametric Transformation Models For Counting Processes," *Biometrika*, Vol. 93, No. 3, 2006, pp. 627-640. doi:10.1093/biomet/93.3.627

[5] R. Tibshirani, "The Lasso Method for Variable Selection in the Cox Model," *Statistics in Medicine*, Vol. 16, No. 4, 1997, pp. 385-395. doi:10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3

[6] J. Fan and R. Li, "Variable Selection for Cox's Proportional Hazards Model and Frailty Model," *Annals of Statistics*, Vol. 30, No. 1, 2002, pp. 74-99.

[7] H. H. Zhang and W. Lu, "Adaptive-Lasso for Cox's Proportional Hazards Model," *Biometrika*, Vol. 94, No. 3, 2007, pp. 691-703. doi:10.1093/biomet/asm037

[8] W. Lu and H. H. Zhang, "Variable Selection for Proportional Odds Model," *Statistics in Medicine*, Vol. 26, No. 20, 2007, pp. 3771-3781. doi:10.1002/sim.2833

[9] Q. Li, C. J. Huang, D. Li and T. T. Fu, "Semiparametric Smooth Coefficient Models," *Journal of Business & Economic Statistics*, Vol. 20, No. 3, 2002, pp. 412-422. doi:10.1198/073500102288618531

[10] W. Zhang, S. Lee and X. Song, "Local Polynomial Fitting Semivarying Coefficient Model," *Journal of Multivariate*

Analysis, Vol. 82, No. 1, 2002, pp. 166-188. doi:10.1006/jmva.2001.2012

[11] J. Fan and T. Huang, "Profile Likelihood Inferences on semiparametric Varying-Coefficient Partially Linear Models," *Bernoulli*, Vol. 11, No. 6, 2005, pp. 1031-1057. doi:10.3150/bj/1137421639

[12] I. Ahmad, S. Leelahanon and Q. Li, "Efficient Estimation of a Semiparametric Partially Linear Varying Coefficient Model," *Annals of Statistics*, Vol. 33, No. 1, 2005. pp. 258-283. doi:10.1214/009053604000000931

[13] H. Wang, Z. Zhu and J. Zhou, "Quantile Regression in Partially Linear Varying Coefficient Models," *Annals of Statistics*, Vol. 37, No. 6B, 2009, pp. 3841-3866. doi:10.1214/09-AOS695

[14] R. Li and H. Liang, "Variable Selection in Semiparametric Regression Modeling," *Annals of Statistics*, Vol. 36, No. 1, 2008, pp. 261-286. doi:10.1214/009053607000000604

[15] W. Lu and H. H. Zhang, "On Estimation of Partially Linear Transformation Models," *Journal of the American Statistical Association*, Vol. 105, No. 490, 2010, pp. 683-691. doi:10.1198/jasa.2010.tm09302

[16] H. H. Zhang, W. Lu and H. Wang, "On Sparse Estimation for Semiparametric Linear Transformation Models," *Journal of Multivariate Analysis*, Vol. 101, No. 7, 2010, pp. 1594-1606. doi:10.1016/j.jmva.2010.01.015

[17] J. Fan and R. Li, "Variable Selection via Nonconcave Penalized Likelihood and Its Oracle Properties," *Journal of American Statistical Association*, Vol. 96, No. 456, 2001, pp. 1348-1360. doi:10.1198/016214501753382273

[18] H. Zou and R. Li, "One-Step Sparse Estimates in Nonconcave Penalized Likelihood Models," *Annals of Statistics*, Vol. 36, No. 4, 2008, pp. 1509-1533. doi:10.1214/009053607000000802

[19] X. He, W. K. Fung and Z. Y. Zhu, "Robust Estimation in Generalized Partial Linear Models for Clustered Data," *Journal of the American Statistical Association*, Vol. 100, No. 472, 2005, pp. 1176-1184.

[20] X. He, Z. Y. Zhu and W. K. Fung, "Estimation in a Semiparametric Model for Longitudinal Data with Unspecified Dependence Structure," *Biometrika*, Vol. 89, No. 3, 2002, pp. 579-590. doi:10.1093/biomet/89.3.579

[21] K. Chen and X. Tong, "Varying Coefficient Transformation Models with Censored Data," *Biometrika*, Vol. 97, No. 4, 2010, pp. 969-976. doi:10.1093/biomet/asq032

[22] H. Wang, G. Li and G. Jiang, "Robust Regression Shrinkage and Consistent Variable Selection via the LAD-LASSO," *Journal of Business & Economics Statistics*, Vol. 25, No. 3, 2007, pp. 347-355. doi:10.1198/073500106000000251

[23] J. D. Kalbfleish and R. L. Prentice, "The Statistical Analysis of Failure Time Data," 2nd Edition. Wiley, Hoboken, 2002. doi:10.1002/9781118032985