应用概率统计 第二十七卷
第二期 2011年4月

Chinese Journal of Applied Probability
and Statistics Vol.27 No.2 Apr. 2011

Variable Selection and Estimation in High-Dimensional Partially Linear Models *

Yang Yiping
(*College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing, 400067*)

Xue Liugen
(*College of Applied Sciences, Beijing University of Technology, Beijing, 100124*)

Abstract

In this paper, we propose an approach for achieving simultaneously variable selection and estimation for the linear and nonparametric components in high-dimensional partially linear models. We use Dantzig selector, applied to the linear part and various derivatives of nonparametric component, to achieve sparsity in the linear part and produce nonparametric estimators. Non-asymptotic theoretical bounds on the estimator error are obtained. The finite sample properties of the proposed approach are investigated through a simulation study.

Keywords: Partially linear model, variable selection, Dantzig selector, SCAD.

AMS Subject Classification: 62G05, 62G20.

§1. Introduction

Consider the problem of simultaneous variable selection and estimation in the partially linear model

$$Y = X^T\beta + g(T) + \varepsilon,$$

where Y is a scalar response variate, X is a p-vector covariate, T is a scalar covariate and takes values in a compact interval, for simplicity, we assume this interval to be $[0,1]$, β is a $p\times 1$ column vector of unknown regression parameter, the function $g(\cdot)$ is unknown, and the model error ε is independent of (X,T) with mean 0. Traditionally, it has generally been assumed that β is finite dimension, several standard approaches, such as the Kernel method, the spline method, the local linear estimation and so on [1–3], have been proposed.

*The project supported by the National Natural Science Foundation of China (10871013, 10871217), the Natural Science Foundation of Beijing (1072004) and Research Fund of Chongqing Technology and Business University (20105609).

Received May 29, 2009. Revised August 23, 2010.

But, the situation where p is large in the sense that $p \to \infty$ as the sample size $n \to \infty$, has become increasingly common. Therefore, this encourages us to consider a high-dimensional partially linear model. We are interested in the sparse modeling problem where the true model has a sparse representation.

It seems that there is not too many studies of high-dimensional partially linear models. Xie and Huang[4] applied the SCAD penalty to achieve sparsity in the linear part and used polynomial splines to estimate the nonparametric component. We also consider the problem of simultaneous variable selection and estimation in high-dimensional partially linear models. James et al.[5] considered functional linear regression and introduced a new approach, "Functional Linear Regression That's Interpretable", to estimate a coefficient function. Our approach applies the idea of James et al.[5] to high-dimensional partially linear models. This article has two primary goals. The first goal is to select significant variables for the parametric portion when it is sparse, in the sense that many of its elements are zero. The second goal is to produce estimator of $g(t)$ that is both interpretable, flexible and accurate. The key to our procedure is to apply Dantzig selector to the linear part and various derivatives of $g(t)$. Hence, we assume that one or more derivatives of $g(t)$ are sparse, this assumption is reasonable (see [5]). The basic idea of our method is to penalize simultaneously parametric parts and various derivatives of nonparametric component. Our approach is different from that of Xie and Huang[4]. The proposed approach has some advantages over Xie and Huang[4]. First, Our method has strong empirical results on models with large values of p. Second, by choosing appropriate derivatives, we can produce a large range of highly interpretable $g(t)$ curve.

This paper is organized as follows. In Section 2, we penalize simultaneously parametric parts and various derivatives of nonparametric component and non-asymptotic theoretical bounds on the estimation error are given. In Section 3, we show how to control multiple derivatives simultaneously. Section 4 discusses the turning parameters choice and reports the simulation results. The proofs are relegated to Section 5.

§2. Estimation of the Parametric and Nonparametric Components

Assume that $\{(X_i, T_i, Y_i), 1 \le i \le n\}$ are independent and identically distributed and

$$Y_i = X_i^T \beta + g(T_i) + \varepsilon_i, \qquad i = 1, \cdots, n, \tag{2.1}$$

where ε_i is independent of (X_i, T_i) with mean 0 and variance σ^2;

$$X_i = (X_{i1}, \cdots, X_{ip})^T \in R^p.$$

Throughout the article, we always assume that T_i are strictly ordered, $0 \le T_1 < T_2 < \cdots < T_n \le 1$. Let $B(t) = \{b_1(t), b_2(t), \cdots, b_q(t)\}^T$ is a q-dimensional basis, then $g(t)$ can be approximated by

$$g(t) = B(t)^T \eta + e_q(t), \tag{2.2}$$

where $e_q(t)$ represents the deviations of the true $g(t)$. We should choose q is large so $|e_q(t)|$ can be small. We require $q \le n$ and $q \to \infty$ as the sample size $n \to \infty$. In this paper, we use a simple grid basis where $b_k(t) = 1$ if $t \in \{t : T_{k-1} \le t < T_k\}$ and 0 otherwise. Combining (2.1) and (2.2), we have

$$Y_i = Z_i \alpha + \varepsilon^*,$$

where

$$Z_i = [X_i^T | B(T_i)^T], \qquad \alpha = (\beta^T, \eta^T)^T, \qquad \varepsilon^* = \varepsilon_i + e(T_i).$$

Note that α is $p+q$ dimension, it is difficult to estimate α. One could use a variable selection procedure to estimate α, which implies η will be sparse. There is no reason to suppose that estimate η is sparse. Instead we assume that one or more derivatives of $g(t)$ are sparse. Let

$$A_{n\times q} = [D^d B(T_1), \cdots, D^d B(T_n)]^T,$$

where D^d is the dth finite difference operator i.e.

$$DB(T_j) = \frac{B(T_j) - B(T_{j-1})}{T_j - T_{j-1}}, \qquad D^2 B(T_j) = \frac{DB(T_j) - DB(T_{j-1})}{T_j - T_{j-1}}$$

etc.. Let

$$\gamma = A\eta, \tag{2.3}$$

then, γ_j provides an approximation to $g^{(d)}(T_j)$. Assuming that $g^{(d)}(T_j)$ is zero at most time points means that γ is sparse. Now we consider A is constructed using a single derivative, we will discuss the situation with multiple derivatives in Section 3. If we assume that the matrix $A^T A$ is invertible, by (2.3), we have

$$\eta = (A^T A)^{-1} A^T \gamma,$$

This together with (2.1) and (2.2) yields

$$Y = V\zeta + \varepsilon^*, \tag{2.4}$$

where

$$V = [X^T|B^T(T)(A^TA)^{-1}A^T], \qquad \zeta = (\beta^T, \gamma^T)^T.$$

Since ζ is assumed to be sparse, a class of variable selection produces for parametric models can be used to fit (2.4). We choose the Dantzig selector. The Dantzig selector[6] was designed for linear regression models with large p but a sparse set of coefficients. Many advantages of the Dantzig selector have been shown in [6]. First, the method has demonstrated strong empirical results on models with high-dimensional parameters. Second, it provides a computationally convenient by using the DASSO algorithm (see [7]).

Note that the method assumes a standardized design matrix with columns of norm one, hence, we first standardized V, the model (2.4) can be reexpressed as

$$Y = \widetilde{V}\widetilde{\zeta} + \varepsilon^*, \tag{2.5}$$

where $\widetilde{\zeta} = D_v\zeta$ and D_v is a diagonal matrix consisting of the column norms of V. Consider the model (2.5), the Dantzig selector estimate, $\widehat{\widetilde{\zeta}}_{DS}$, is defined by

$$\widehat{\widetilde{\zeta}}_{DS} = \arg\min_{\widetilde{\zeta}} \|\widetilde{\zeta}\|_1 \ \text{ subject to } \ |\widetilde{V}_j^T(Y - \widetilde{V}\widetilde{\zeta})| \le \lambda, \ \ j = 1, \cdots, (p+n), \tag{2.6}$$

where $\widetilde{V}_j$ is the jth column of $\widetilde{V}$ and $\lambda \ge 0$ is a tuning parameter.

Using the Danzig selector (2.6), we can obtain estimator of $\widetilde{\zeta}$, Let $\widehat{\widetilde{\zeta}}$ be the corresponding solution from the Dantzig selector. After the coefficients, $\widehat{\widetilde{\zeta}}$, have been obtained, we produce the estimators of β and γ, respectively

$$\widehat{\beta} = (I_{p\times p}, 0_{p\times n})D_v^{-1}\widehat{\widetilde{\zeta}}, \tag{2.7}$$

$$\widehat{\gamma} = (0_{n\times p}, I_{n\times n})D_v^{-1}\widehat{\widetilde{\zeta}}, \tag{2.8}$$

where $I_{p\times p}$ is $p \times p$ identify matrix, $0_{p\times n}$ is $p \times n$ 0 matrix. We combine (2.2), (2.3) and (2.8) to produce the estimate for $g(t)$ using

$$\widehat{g}(t) = B(t)^T(A^TA)^{-1}A^T\widehat{\gamma}.$$

Theorem 2.1 provides a non-asymptotic bound of $\widehat{\beta}$ on the L_2 error. In Theorem 2.1, the values δ, θ, $N_{n,p}$ are all known constants which will be defined in Section 5.

Theorem 2.1 Suppose that $\widetilde{\zeta}$ is an S-sparse vector with $\delta_{2S}^V + \theta_{S,2S}^V < 1$ and

$$\max \|\widetilde{V}^T \varepsilon^*\| \leq \lambda, \tag{2.9}$$

then

$$\|\widehat{\beta} - \beta\| \leq \frac{1}{\sqrt{n}} N_{n,p} \lambda \sqrt{S}. \tag{2.10}$$

In Theorem 2.2 we provide a non-asymptotic bound of $\widehat{g}(t)$ on the L_2 error.

Theorem 2.2 For a given q-dimensional basis $B_q(t)$, let $\omega_q = \sup\limits_t |e_q(t)|$. Suppose that $\widetilde{\zeta}$ is an S-sparse vector with $\delta_{2S}^V + \theta_{S,2S}^V < 1$. If (2.9) holds, then, for every $0 \leq t \leq 1$,

$$|\widehat{g}(t) - g(t)| \leq \frac{1}{\sqrt{n}} M_{n,q}(t) \lambda \sqrt{S} + \omega_q, \tag{2.11}$$

where $M_{n,q}(t)$ is defined in Section 5.

Note that both Theorem 2.1 and 2.2 need the constraint (2.9), how to choose λ such that (2.9) holds with high probability? Theorem 2.3 will solve this problem.

Theorem 2.3 Suppose that $\varepsilon_i \sim N(0, \sigma^2)$, then for any $a \geq 0$, if

$$\lambda = \sigma \sqrt{2(1+a)\log(p+n)} + \omega_q \sqrt{n},$$

then (2.9) holds with probability ar least $1 - \{(p+n)^a \sqrt{4\pi(1+a)\log(p+n)}\}^{-1}$, and hence

$$\|\widehat{\beta} - \beta\| \leq \frac{1}{\sqrt{n}} N_{n,p} \sigma \sqrt{2S(1+a)\log(p+n)} + N_{n,p} \omega_q \sqrt{S}, \tag{2.12}$$

$$|\widehat{g}(t) - g(t)| \leq \frac{1}{\sqrt{n}} M_{n,q}(t) \sigma \sqrt{2S(1+a)\log(p+n)} + \omega_q \{1 + M_{n,q}(t)\sqrt{S}\}. \tag{2.13}$$

Remark 1 Under suitable conditions, we have $N_{n,p}$ and $M_{n,q}(t)$ can converge to the constant as n, p and q grow, respectively, and ω_q declines with q. For example, when we using the piecewise constant basis, ω_q converges to 0 at a rate of $1/q$ if $g'(t)$ is bounded; when we using the piecewise polynomial basis, ω_q converges to 0 at a rate of $1/q^{d+1}$ if $g^{d+1}(t)$ is bounded.

§3. Controlling Multiple Derivatives

In Section 2, we have discussed the situation with controlling a single derivative of $g(t)$. Now, we concentrate on controlling multiple derivatives. For example, we may believe that both $g^{(0)}(t)=0$ and $g^{(2)}(t)=0$ over many regions of t. In this situation, we would let

$$A=[D^0B(T_1),\cdots,D^0B(T_n),D^2B(T_1),\cdots,D^2B(T_n)]^T.$$

Let $A_{(1)}$ be the first n rows of A and $A_{(2)}$ the remainder. We assume that A is arranged so that $A_{(1)}^TA_{(1)}$ is invertible. Since $\gamma^*=A\eta$, we have

$$\eta=(A_{(1)}^TA_{(1)})^{-1}A_{(1)}^T\gamma_{(1)}^*,\qquad \gamma_{(2)}^*=A_{(2)}(A_{(1)}^TA_{(1)})^{-1}A_{(1)}^T\gamma_{(1)}^*, \tag{3.1}$$

where $\gamma_{(1)}^*$ is the first n elements of γ^* and $\gamma_{(2)}^*$ the remaining elements. Model (2.1) is equivalent to

$$Y_i=X_i^T\beta+B(T_i)^T(A_{(1)}^TA_{(1)})^{-1}A_{(1)}^T\gamma_{(1)}^*+\varepsilon^*,\qquad i=1,\cdots,n. \tag{3.2}$$

We then use this model to estimate γ^* subject to the constraint given by (3.1). Finally, we may not wish to place equal weight on each derivative when constraining multiple derivatives, we can explore the method in (3.2).

§4. Simulation Study

We generated data from the partially linear model,

$$Y=X^T\beta+g(T)+\varepsilon,$$

where β was p-dim vector, $\beta=(0.25,1,1.5,3,\mathbf{0}_{p-4})$, $\mathbf{0}_m$ denoted an m-vector of 0s. $\varepsilon\sim N(0,0.3^2)$ and X followed a p-dim multivariate normal distribution with zero mean and covariance Σ whose (j,k) entry was $\Sigma_{j,k}=\rho^{|j-k|}$, $1\le k,\ j\le p$ with $\rho=0.75$, T is simulated from a permutation of the uniform distribution $U(0,1)$, $g(t)$ was piecewise quadratic with a "flat" region,

$$g(t)=\begin{cases}20\times(t-0.5)^2-0.5, & \text{if } 0\le t<0.342,\\ 0, & \text{if } 0.342\le t\le 0.65,\\ -20\times(t-0.5)^2+0.5, & \text{if } 0.658<t\le 1.\end{cases}$$

The estimators depend on the choice of λ and d. We choose λ and d by minimizing the cross-validation residual sum of squares

$$\mathrm{CV}(\lambda,d)=\frac{1}{n}\sum_{i=1}^n(Y_i-X_i^T\widehat{\beta}_{[-i]}-\widehat{g}_{[-i]}(T_i))^2,$$

where $\widehat{\beta}_{[-i]}$ and $\widehat{g}_{[-i]}(\cdot)$ are estimators of β and $g(\cdot)$ which are computed with all of the measurements but the ith subject deleted. In particular, d is chosen from the values $d=1,2,3,4$. We compute the cross-validation residual sum of squares for $d=1,2,3,4$ and a grid of values for λ.

We considered $p=10,50,100$ for $n=200$. For comparison, we considered our proposed method (the Dantzig selector (DS)) and the SCAD penalized regression method proposed by Xie and Huang[4]. We repeated the simulation 1000 times. The performance of estimator $\widehat{\beta}$ is assessed by using the generalized mean square error (GMSE), defined as

$$\mathrm{GMSE}=(\widehat{\beta}-\beta)^T\mathsf{E}(XX^T)(\widehat{\beta}-\beta).$$

The variable selection performance is gauged by (C,I), where "C" gives the average number of zero coefficients correctly set to zero, and "I" gives the average number of nozero coefficients incorrectly set to zero. The results are summarized in Table 1 and Figure 1. Several observations can be made:

(1) For parameter part, from Table 1, we see that when p is small, the SCAD can perform much better than Dantzig selector. However, when p is large ($p=100$), the Dantzig selector outperforms the SCAD.

(2) For nonparametric part, Figure 1 shows that the sparsity in the zeroth derivative generates the zero section while the sparsity in the third derivative ensures a smooth fit. From Figure 1 (b), we can see that the polynomial spline estimator (proposed by Xie and Huang[4]) provides a worse approximation than our method for the region where $g(t)=0$.

Table 1 variable selection and fitting results based on 1000 replications lines

p	method	C	I	GMSE
10	Truth	6	0	
	DS	5.480	0.098	0.066
	SCAD	5.920	0.015	0.011
50	Truth	46	0	
	DS	44.718	0.115	0.071
	SCAD	45.86	0.040	0.015
100	Truth	96	0	
	DS	93.958	0.209	0.076
	SCAD	87.720	0.320	0.264

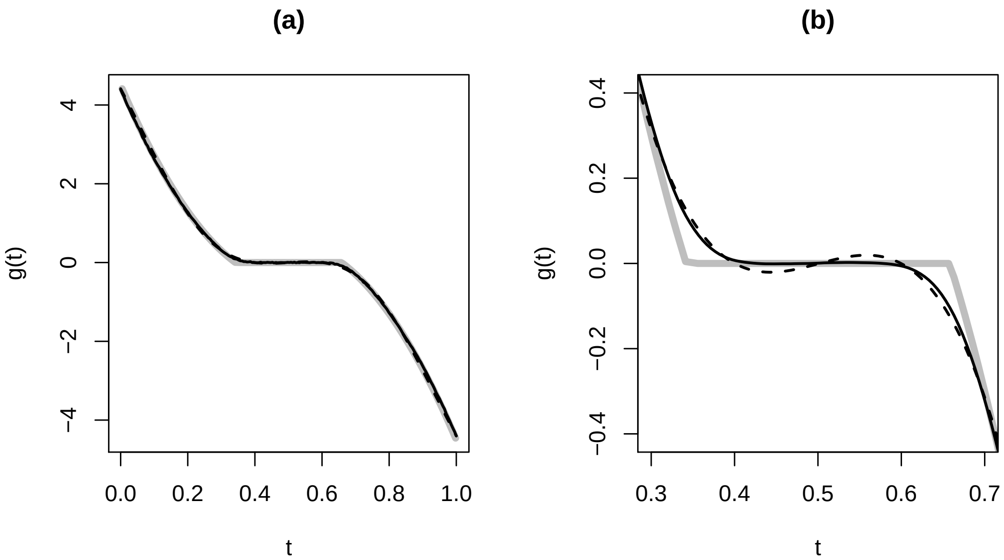

Figure 1 (a) Plots of true curve (grey) and corresponding estimators. The solid line in black is the estimator of $g(t)$ from constraining the zeroth and third derivative, the dashed line is the polynomial spline fit for $n = 200$, $p = 10$. (b) Same plot for the region $0.3 \le t \le 0.7$.

§5. Proofs of Theorems

To establish these results we define quantities δ and θ, first introduced in Candes and Tao[8], which provide measures of how far any S columns of X are from an orthogonal matrix.

Definition 5.1 Let X be an $n\times p$ matrix and let X_T, $T\subset\{1,\cdots,p\}$ be the $n\times|T|$ submatrix obtained by standardizing the columns of X and extracting those corresponding to the indices in T. Then we define δ_S^X as the smallest quantity such that

$$(1-\delta_S^X)\|c\|_2^2\le\|X_Tc\|_2^2\le(1+\delta_S^X)\|c\|_2^2$$

for all subsets T with $|T|\le S$ and all vectors c of length $|T|$.

Definition 5.2 Let T and T' be two disjoint sets with T, $T'\subset\{1,\cdots,p\}$, $|T|\le S$ and $|T'|\le S'$. Then, provided $S+S'\le p$, we define $\theta_{S,S'}^X$ as the smallest quantity such that

$$|(X_Tc)^TX_{T'}C'|\le\theta_{S,S'}^X\|c\|_2\|c'\|_2$$

for all T and T' and all corresponding vectors c and c'.

To prove the Theorem 2.1 and 2.2, we need one of the results from in [7].

Lemma 5.1 (Theorem 4 in [7]) Suppose that $\widetilde{\zeta}$ is an S-sparse vector with $\delta_{2S}^V + \theta_{S,2S}^V < 1$. Let $\widehat{\widetilde{\zeta}}$ be the corresponding solution from the Dantzig selector. If (2.10) holds, then

$$\|\widehat{\widetilde{\zeta}} - \widetilde{\zeta}\| \le \frac{4\lambda\sqrt{S}}{1-\delta_{2S}^V - \theta_{S,2S}^V}.$$

Proof (Proof of Theorem 2.1) By Lemma 5.1, we have

$$\begin{aligned} \|\widehat{\beta} - \beta\| &= \|(I_{p\times p}, 0_{p\times n})D_V^{-1}(\widehat{\widetilde{\zeta}} - \widetilde{\zeta})\| \\ &\le \|(I_{p\times p}, 0_{p\times n})D_V^{-1}\|\|\widehat{\widetilde{\zeta}} - \widetilde{\zeta}\| \\ &= \frac{1}{\sqrt{n}C_{n,p}\|\widehat{\widetilde{\zeta}} - \widetilde{\zeta}\|} \\ &\le \frac{1}{\sqrt{n}}\frac{4C_{n,p}\lambda\sqrt{S}}{1-\delta_{2S}^V - \theta_{S,2S}^V} \\ &= \frac{1}{\sqrt{n}N_{n,p}\lambda\sqrt{S}}, \end{aligned}$$

where

$$C_{n,p} = \sqrt{\max_{1\le j\le p}\frac{1}{\frac{1}{n}\sum\limits_{i=1}^{n}X_{ij}^2}}, \qquad N_{n,p} = \frac{4C_{n,p}}{1-\delta_{2S}^V - \theta_{S,2S}^V}.$$

We need some notions. Let

$$H_{q\times n} = (A^TA)^{-1}A^T, \qquad M_{n,q}(t) = \frac{4E_{n,q}(t)}{1-\delta_{2S}^V - \theta_{S,2S}^V},$$

where

$$E_{n,q}(t) = \sqrt{\sum_{j=1}^{n}\frac{(B(t)^TH_j)^2}{\frac{1}{n}\sum\limits_{i=1}^{n}(B(T_i)^TH_j)^2}},$$

H_j is the jth column of H. $\square$

Proof (Proof of Theorem 2.2) Since

$$g(t) = B(t)^T\eta + e_q(t) = B(t)^TH_{q\times n}(0_{n\times p}, I_{n\times n})D_V^{-1}\widetilde{\zeta} + e_q(t),$$

we can derive

$$\begin{aligned}|\widehat{g}(t)-g(t)| &\le |\widehat{g}(t)-B(t)^T\eta|+|e_q(t)|\\ &= |B(t)^T H_{q\times n}(0_{n\times p}, I_{n\times n})D_V^{-1}(\widehat{\widetilde{\zeta}}-\widetilde{\zeta})|+|e_q(t)|\\ &\le \|B(t)^T H_{q\times n}(0_{n\times p}, I_{n\times n})D_V^{-1}\|\|\widehat{\widetilde{\zeta}}-\widetilde{\zeta}\|+\omega_q\\ &= \frac{1}{\sqrt{n}E_{n,q}(t)\|\widehat{\widetilde{\zeta}}-\widetilde{\zeta}\|}+\omega_q\\ &\le \frac{1}{\sqrt{n}}\frac{4E_{n,q}(t)\lambda\sqrt{S}}{1-\delta_{2S}^V-\theta_{S,2S}^V}+\omega_q\\ &= \frac{1}{\sqrt{n}M_{n,q}(t)\lambda\sqrt{S}}+\omega_q. \qquad \square\end{aligned}$$

Proof (Proof of Theorem 2.3) Substituting

$$\lambda=\sigma\sqrt{2(1+a)\log(p+n)}+\omega_q\sqrt{n}$$

into (2.10) and (2.11) gives (2.12) and (2.13), respectively. Note that

$$|\widetilde{V}_j^T\varepsilon^*|=|\widetilde{V}_j^T\varepsilon+\widetilde{V}_j^Te_q(T_j)|\le|\widetilde{V}_j^T\varepsilon|+|\widetilde{V}_j^Te_q(T_j)|\le\sigma|Z_j|+\omega_q\sqrt{n},$$

where $Z_j\sim N(0,1)$. This result follows from the fact that $\widetilde{V}_j$ is norm one and $\varepsilon_i\sim N(0,\sigma^2)$, it will be the case that $\widetilde{V}_j^T\varepsilon\sim N(0,\sigma^2)$. Hence

$$\begin{aligned}\mathsf{P}\Big(\max_j|\widetilde{V}_j^T\varepsilon^*|>\lambda\Big) &= \mathsf{P}\Big(\max_j|\widetilde{V}_j^T\varepsilon^*|>\sigma\sqrt{2(1+a)\log(p+n)}+\omega_q\sqrt{n}\Big)\\ &\le \mathsf{P}\Big(\max_j|Z_j|>\sqrt{2(1+a)\log(p+n)}\Big)\\ &\le (p+n)\frac{1}{\sqrt{2\pi}}\exp\{-(1+a)\log(p+n)\}/\sqrt{2(1+a)\log(p+n)}\\ &= \{(p+n)^a\sqrt{4\pi(1+a)\log(p+n)}\}^{-1}.\end{aligned}$$

The penultimate line follows from the fact that

$$\mathsf{P}\Big(\sup_j|Z_j|>u\Big)\le\frac{p}{u}\frac{1}{\sqrt{2\pi}}\exp(-u^2/2). \qquad \square$$

References

[1] Speckman, P., Kernel smoothing in partial linear model, *J. Roy. Statist. Soc. Ser. B*, **50**(1998), 413–436.

[2] Heckman, N.E., Spline smoothing in a partly linear model, *J. Roy. Statist. Soc. Ser. B*, **48**(1996), 244–248.

[3] Fan, J. and Li, R., New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis, *J. Amer. Statist. Assoc.*, **99**(2004), 710–723.

[4] Xie, H.L. and Huang, J., SCAD-Penalized regression in high-dimensional partially linear models, *Ann. Statist.*, **37**(2009), 673–696.

[5] James, G., Wang, J. and Zhu, J., Functional linear regression that's interpretable, *Ann. Statist.*, **37**(2009), 2083–2108.

[6] Candes, E. and Tao, T., The Dantzig selector: statistical estimation when p is much larger than n, *Ann. Statist.*, **35**(2007), 2313–2351.

[7] James, G., Radchenko, P. and Lv, J., DASSO: connections between the Dantzig selector and Lasso, *J. Roy. Statist. Soc. Ser. B*, **71**(2009), 127–142.

[8] Candes, E. and Tao, T., Decoding by linear programming, *IEEE Trans. Inform. Theory*, **51**(2005), 4203–4215.

高维部分线性模型的变量选择和估计

杨宜平 薛留根

(重庆工商大学数学与统计学院, 重庆, 400067) (北京工业大学应用数理学院, 北京, 100124)

考虑高维部分线性模型, 提出了同时进行变量选择和估计兴趣参数的变量选择方法. 将Dantzig变量选择应用到线性部分及非参数部分的各阶导数, 从而获得参数和非参数部分的估计, 且参数部分的估计具有稀疏性, 证明了估计的非渐近理论界. 最后, 模拟研究了有限样本的性质.

关键词: 部分线性模型, 变量选择, Dantzig选择, SCAD.

学科分类号: O212.7.