Efficient estimation and model selection for single-index varying-coefficient models

Peng Lai[a], Qingzhao Zhang[b], Heng Lian[c] and Qihua Wang[b]

[a]*School of Mathematics and Statistics, Nanjing University of Information Science & Technology, Nanjing 210044, China, laipeng@amss.ac.cn*
[b]*Academy of Mathematics and Systems Science, Chinese Academy of Science, Beijing 100190, China*
[c]*Division of Mathematical Sciences, SPMS, Nanyang Technological University, Singapore, 637371*

Abstract

The single-index varying-coefficient models include many types of popular semiparametric models, i.e. single-index models, partially linear models, varying-coefficient models, and so on. In this paper, we first establish the semiparametric efficiency bound for the single-index varying-coefficient model, and develop an estimation method based on the efficient estimating equations. Although our main focus is more on homoscedastic models for simplicity, the calculated efficiency bound and efficient estimating equations are for the more general heteroscedastic models. It shows that the estimator of the finite dimensional parameter is $\sqrt{n}$ consistent and asymptotically normal and attains the semiparametric efficiency bound. Moreover, for the homoscedastic model, a two-stage variable selection procedure is proposed to select the important nonparametric components and parametric components. We also find that the proposed procedures can divide the predictors into varying-coefficient predictors and constant-coefficient predictors automatically. Some simulation studies are conducted to evaluate and illustrate the proposed methods.

Keywords: Efficiency; estimating equation; group LASSO; variable selection.

1 Introduction

The varying-coefficient model has gained a lot of attentions during the past decade which is an important generalization of the linear regression model. However, for the varying-coefficient model

$$Y = \alpha(U)^\top Z + \varepsilon, \quad Z \in R^q, \tag{1}$$

if the dimension of U is high, the "curse of dimensionality" problem will suffer. This difficulty motivates us to study more flexible and more general models, the single-index varying-coefficient model,

$$Y = g^\top(X^\top \beta) Z + \varepsilon, \tag{2}$$

where $X \in R^p$ and $Z \in R^q$ are vectors of covariates, Y is the response variable, β is a $p \times 1$ vector of unknown parameters with its true value β_0, $g(\cdot)$ is a $q \times 1$ vector of unknown function and ε is a random error with $E(\varepsilon|X, Z) = 0$ and $Var(\varepsilon|X, Z) = \sigma^2(X, Z)$. Xue and Wang (2012) studied this model and developed an adjusted empirical likelihood ratio for constructing the confidence regions of parameters of interest.

Model (2) includes many important statistical models as special cases. If $g(\cdot)$ is a $q \times 1$ constant vector, (2) turns to the linear regression model. If $p = 1$ and $\beta = 1$, (2) reduces to the varying-coefficient model. Furthermore, if some elements of $g(X)$ are constants, it is the varying-coefficient partially linear model. If $q = 1$ and $Z = 1$, (2) reduces to the single-index model. If $p = q$ and $Z = X$, (2) becomes the adaptive varying-coefficient linear model. Therefore, it is meaningful to study this semiparametric model.

For any given β, model (2) reduces to (1), and many methods have been developed to estimate the regression coefficients. However, more than to obtain a consistent estimator, we would like to search for the efficient estimator of the parameter. Semiparametric efficiency problem has been studied extensively in the literature. Tsiatis (2006) gave a general introduction to semiparametric efficiency bound and efficient estimators with a focus on missing data. Ma et al. (2006) considered the heteroscedastic partially linear models, and utilized a weighted estimating equation method to define an efficient semiparametric estimator for the parameter vector in the linear part. Liang et al. (2010) studied the efficient estimators for the partially linear single-index models. The discussions for semiparametric efficient estimators also can be found in Firpo (2007); Chen et al. (2008), etc.

On the other hand, high-dimensionality is an important characteristic of many modern data sets. Then an important problem is to select the significant variables in the studied model. Most recently, variable selection using penalty functions for nonparametric or semiparametric models have been developed. For example, Li and Liang (2008) studied variable selection for varying-coefficient partially linear models, where the parametric components were identified via the SCAD but the nonparametric components were selected via a generalized likelihood ratio test. Wang and Xia (2009) studied variable selection for varying-coefficient models, the nonparametric components were estimated and selected by the shrinkage method following the group LASSO idea (Yuan and Lin (2006)) and kernel smoothing techniques. Huang et al. (2010) studied variable selection for additive models, using the adaptive group Lasso method with B-splines. Furthermore, in practice, some components of $g(\cdot)$ in the single-index varying-coefficient model (2) are non-zero constants and, some are zero constants and some other components are varying coefficients. This motivates us to develop a data driven method to identify the true model.

2 Results

2.1 Efficient estimator for β

For the sake of identifiability, we assume $\|\beta\| = 1$ and the first component of β is positive. We use the delete-one-component method proposed by Yu and Ruppert (2002). Let $\beta = (\beta_1, \ldots, \beta_p)^\top$ and $\beta^{(1)} = (\beta_2, \ldots, \beta_p)^\top$, then, we may write $\beta(\beta^{(1)}) = ((1-\|\beta^{(1)}\|^2)^{1/2}, \beta_2, \ldots, \beta_p)^\top$. Thus, the Jacobian matrix is $J_{\beta^{(1)}} = \frac{\partial \beta}{\partial \beta^{(1)}} =$

$(b_1,\ldots,b_p)^\top$, where $b_s (1 < s \le p)$ is a $(p-1)$-dimensional unit vector with sth component 1, and $b_1 = -(1-\|\beta^{(1)}\|^2)^{-1/2}\beta^{(1)}$. We can compute the efficient score function $S_\beta^{eff}(Y,X,Z)$ and the semiparametric efficient bound of model (2).

Theorem 1 *Assume that the conditional probability density function of ε given (X,Z), $p_1(\varepsilon|X,Z)$, is differentiable with respect to ε and that $0 < E(\varepsilon^2|X,Z) < \infty$ almost everywhere, the semiparametric efficient score of model (2) is given by*

$$\begin{aligned} S_{\beta_0}^{eff}(Y,X,Z) &= \varepsilon w(X,Z)\Big\{g'^\top(X^\top\beta_0)ZJ_{\beta_0^{(1)}}^\top X - E\big[w(X,Z)g'^\top(X^\top\beta_0)Z \\ &\quad \times J_{\beta_0^{(1)}}^\top XZ^\top|X^\top\beta_0\big]\big\{E[w(X,Z)ZZ^\top|X^\top\beta_0]\big\}^{-1}Z\Big\}, \qquad (3)\end{aligned}$$

where $\varepsilon = Y - Z^\top g(X^\top\beta_0)$ and $w(X,Z) = \{E(\varepsilon^2|X,Z)\}^{-1}$. Thus, the semiparametric efficiency bound of model (2) is $\big\{E(S_{\beta_0}^{eff}S_{\beta_0}^{eff\,\top})\big\}^{-1}$.

Solving the efficient estimating equations based on the efficient score vector, we can get the efficient estimators.

Theorem 2 *Under some regularity conditions, we have*

$$\sqrt{n}(\hat\beta-\beta_0) \xrightarrow{\mathcal{D}} N(0, J_{\beta_0^{(1)}}\Sigma_1^{-1}J_{\beta_0^{(1)}}^\top),$$

where $\Sigma_1 = \big\{E(S_{\beta_0}^{eff}S_{\beta_0}^{eff\,\top})\big\}$.

2.2 Variable selection procedures

Assume the number of significant variables in Z is $q_0, q_0 \le q$ and the number of varying coefficient components are d_0, $d_0 \le q_0$. A further task of variable selection reduces to identifying the nonzero varying coefficients and the nonzero constant coefficients. The subscript set of nonzero components can be defined as $\mathbf{A}_g^* = \{1,\ldots,q_0\}$, and the subscript set of nonzero varying coefficients can be defined as $\mathbf{A}_{g'}^* = \{1,\ldots,d_0\}$. Yuan and Lin (2006) used the group lasso penalty to identify zero coefficients, Wang and Xia (2009) also used group lasso idea to select the important varying coefficients for varying coefficient models. Thus, we propose the following penalized estimator

$$\hat G_{M_\lambda} = (\hat G_\lambda^\top, \hat G_\lambda'^\top)^\top = \underset{G_{ML}\in R^{2q\times n}}{argmin}\ Q_\lambda(G_{ML}), \qquad (4)$$

where

$$Q_\lambda(G_{ML}) = Q(G_{ML}) + \sum_{j=1}^q \lambda_{1j}\|c_j\| + \sum_{j=1}^q \lambda_{2j}\|d_j\|, \qquad (5)$$

$$Q(G_{ML}) = \sum_{j=1}^n\sum_{i=1}^n [Y_i - a^\top(t_j)Z_i - (hb(t_j))^\top Z_i(\frac{X_i^\top\beta - t_j}{h})]^2 K_h(X_i^\top\beta - t_j)$$

$a_j = a(t_j) = (a_1(t_j),\ldots,a_q(t_j))^\top,\ hb_j = hb(t_j) = (hb_1(t_j),\ldots,hb_q(t_j))^\top,\ j=1,\ldots,n,$
$c_k = (a_k(t_1),a_k(t_2),\ldots,a_k(t_n))^\top,\ d_k = (b_k(t_1),b_k(t_2),\ldots,b_k(t_n))^\top,\ k=1,\ldots,q,$

λ_{1j} and $\lambda_{2j}, j = 1, \ldots, q$, are different tuning parameters. Let $\phi_n = \max\{\lambda_{1j}, 1 \leq j \leq q_0\}$, $\phi'_n = \max\{\lambda_{2j}, 1 \leq j \leq q_0\}$, $\varphi_n = \min\{\lambda_{1j}, q_0 + 1 \leq j \leq q\}$, $\varphi'_n = \min\{\lambda_{2j}, q_0+1 \leq j \leq q\}$, and $\psi_n = \min\{\lambda_{2j}, d_0+1 \leq j \leq q_0\}$. Let $\mathbf{A}^*_{G_M} = \{\mathbf{A}^*_g, q_0 + \mathbf{A}^*_{g'}\}$. $\eta_{0\mathbf{A}^*_{G_M}} = (g_1(t), \ldots, g_{q_0}(t), hg'_1(t), \ldots, hg'_{d_0}(t))^\top$. Define $Z^* = (Z_1, \ldots, Z_{q_0})^\top$ and $Z^{**} = (Z_1, \ldots, Z_{d_0})^\top$.

Theorem 3 *Under some regularity conditions, as $h = O_p(n^{-1/5})$, $n^{-11/10}\phi_n \to 0$, $n^{-11/10}\phi'_n \to 0$, $n^{-11/10}\varphi_n \to \infty$, $n^{-11/10}\varphi'_n \to \infty$, and $n^{-9/10}\psi_n \to \infty$, we have*

(i). $P\big(\sup_{t\in T} \|\hat{g}_{\lambda j}(t)\| = 0\big) \to 1$ for any $q_0 < j \leq q$,

(ii). $P\big(\sup_{t\in T} \|\hat{g}'_{\lambda j}(t)\| = 0\big) \to 1$ for any $d_0 < j \leq q$.

(iii). $\sqrt{nh}\Big(\hat{\eta}_{\lambda\mathbf{A}^*_{G_M}}(t) - \eta_{0\mathbf{A}^*_{G_M}}(t) - \frac{h^2}{2}[\Sigma^*_2(t)]^{-1}\begin{pmatrix} E(Z^*Z^{*\top}|t)g''_{\mathbf{A}^*_g}(t)\mu_2 f(t) \\ 0 \end{pmatrix}\Big) \xrightarrow{\mathcal{D}} N(0, \sigma^2\Sigma_2^{*-1}(t)V_1^*\Sigma_2^{*-1}(t))$.

Thus, after the first stage, we define $\hat{g}^*_\lambda(t) = (\hat{g}_{\lambda 1}(t), \ldots, \hat{g}_{\lambda q_0}(t))^\top$ and $\hat{g}'^*_\lambda(t) = (\hat{g}'_{\lambda 1}(t), \ldots, \hat{g}'_{\lambda d_0}(t), 0^\top_{q_0-d_0})^\top$ be the penalized estimators. The penalized estimating equation on $\beta^{(1)}$ can be constructed as

$$U_{\lambda_n}(\beta^{(1)}) = \sum_{i=1}^n \tilde{S}^{eff}_\beta(Y_i, X_i, Z^*_i) - n\boldsymbol{q}_{\lambda_n}(|\beta^{(1)}|)\boldsymbol{sgn}(\beta^{(1)}). \tag{6}$$

Denote $U_{\lambda_n}(\beta^{(1)}) = (U_{\lambda_n 2}(\beta^{(1)}), \cdots, U_{\lambda_n p}(\beta^{(1)}))^\top$. We introduce a zero-crossing penalized estimating equation defined in Johnson et al. (2008). Let $\hat{\beta}^{(1)}_{\lambda_n}$ be a zero-crossing to penalized estimating equation if , for $j = 2, \cdots, p$,

$$\overline{\lim_{\epsilon\to 0+}} n^{-1}U_{\lambda_n j}(\hat{\beta}^{(1)}_{\lambda_n} + \epsilon e_j)U_{\lambda_n j}(\hat{\beta}^{(1)}_{\lambda_n} - \epsilon e_j) \leq 0,$$

where e_j is the jth canonical unit vector. Define the true nonzero components index $\mathbf{A}^*_\beta = \mathbf{A}^*_{\beta^{(1)}} \cup \{1\}$, where $\mathbf{A}^*_{\beta^{(1)}} = \{2, \ldots, p_0\}$.

Theorem 4 *Under some regularity conditions, if $nh^4 \to \infty, nh^6 \to 0$, $\lambda_n \to 0$ and $\sqrt{n}\lambda_n \to \infty$, then the following results hold:*

(i). There exists a zero-crossing $\hat{\beta}^{(1)}_{\lambda_n}$ to $U_{\lambda_n}(\beta^{(1)})$ that satisfies $\hat{\beta}^{(1)}_{\lambda_n} = \beta^{(1)}_0 + O_p(n^{-1/2})$. There exists a zero-crossing estimator $\hat{\beta}^{(1)}_{\lambda_n} = (\hat{\beta}^{(1)\top}_{\lambda_n\mathbf{A}^*_{\beta^{(1)}}}, \boldsymbol{0}^\top)^\top$ of $U_{\lambda_n}(\beta^{(1)})$ satisfies $U_{\lambda_n\mathbf{A}^*_{\hat{\beta}^{(1)}}}(\hat{\beta}^{(1)}_{\lambda_n}) = 0$.

(ii). For any root-n consistent zero-crossing estimator of $U_{\lambda_n}(\beta^{(1)})$, denoted by $\hat{\beta}^{(1)}_{\lambda_n} = (\hat{\beta}_{\lambda_n 2}, \ldots, \hat{\beta}_{\lambda_n p})^\top$, as $n \to \infty$, with probability tending to 1, $\hat{\beta}_{\lambda_n k} = 0$, $k = p_0 + 1, \ldots, p$. Moreover, let $\beta^*_0 = (\beta_{01}, \beta_{02}, \ldots, \beta_{0p_0})^\top$ and $X^* = (X_1, \ldots, X_{p_0})^\top$, then

$$\sqrt{n}(\hat{\beta}^*_{\lambda_n} - \beta^*_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2 J_{\beta^{(1)*}_0}\Sigma_3^{*-1}J^\top_{\beta^{(1)*}_0}).$$

Table 1: Variable selection for Z

	Number of selected						Proportion of function g		
n	NS	NST	NV	NVT	NC	NCT	U-fit	C-fit	O-fit
oracle	4	4	2	2	2	2	0	1	0
100	4.300	4.000	2.000	2.000	2.300	2.000	0.000	0.700	0.300
	(0.4830)	(0.0000)	(0.0000)	(0.0000)	(0.4830)	(0.0000)			
200	4.000	4.000	2.000	2.000	2.000	2.000	0.000	1.000	0.000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)	(0.0000)			

Table 2: Variable selection for X

	Number of selected		Proportion about β		
n	NS	NST	U-fit	C-fit	O-fit
oracle	3	3	0	1	0
100	3.200(0.4216)	3.000(0.000)	0.000	0.800	0.200
200	3.080(0.3253)	3.000(0.000)	0.000	0.930	0.070

3 Numerical example

Let $\{Y_i, X_i, Z_i; i = 1, \cdots, n\}$ be the i.i.d sample. $X_i = (X_{i1}, \cdots, X_{ip})^\top$ were generated from uniform distribution on $[0, 1]^p$ with independent components. $Z_{i1} = 1$ and $(Z_{i2}, \cdots, Z_{iq})^\top$ follows a multivariate normal distribution with $cov(Z_{ij_1}, Z_{ij_2}) = 0.5^{|j_1-j_2|}$ for $2 \le j_1, j_2 \le q$. Let $p = 8$ and $q = 8$. Consider the model

$$Y_i = \sum_{s=1}^{q} g_s(X_i^\top \beta) Z_{i(s+1)} + 0.5e_i, \tag{7}$$

where $\beta_0 = (3, 1.5, 0, 0, 2, 0, 0, 0)^\top/\sqrt{15.25}$, $g_1(u) = 2\sin(2\pi u), g_2(u) = \exp(2u - 1), g_3(u) = 4, g_4(u) = 1.5$ and $g_k(u) \equiv 0$ for $k = 5, \cdots, 8$, $e_i \sim N(0, 1)$. Let $n = 100$ and $n = 200$. A total of 500 simulation replications were conducted for each example setup.

To demonstrate the performance of the proposed procedure, we define the following criterions: NS shows the average number of variable selected; NST presents the average number of selected that are truly nonzero; NV means the average number of varying-coefficient components selected; NVT denotes the average number of varying-coefficient components selected that are truly nonzero and varying; NC is the average number of nonzero constant components selected while NCT is the average number of nonzero constant components selected which are truly nonzero constant. To assess the performance of two-stage variable selection procedure, following Wang and Xia (2009) we use the relative estimation error (REE)

$$REE_\beta = \frac{\sum_{j=1}^{p} |\hat{\beta}_j - \beta_j|}{\sum_{j=1}^{p} |\bar{\beta}_j - \beta_j|}, \quad REE_g = \frac{\sum_{i=1}^{n}\sum_{j=1}^{q} |\hat{g}_j(X_i^\top \hat{\beta}) - g_j(X_i^\top \beta_0)|}{\sum_{i=1}^{n}\sum_{j=1}^{q} |\bar{g}_j(X_i^\top \bar{\beta}) - g_j(X_i^\top \beta_0)|},$$

where $\bar{\beta}_j, \bar{g}_j(\cdot)$ are either the unpenalized estimators or the oracle estimators.

Table 3: Summary of Two Stages Procedure

n	Proportion of models			REE_β		REE_g	
	U-fit	C-fit	O-fit	unpenalized estimate	oracle estimate	unpenalized estimate	oracle estimate
100	0.000	0.600	0.400	0.3668(0.1372)	1.5578(0.5441)	0.3816(0.0354)	1.0713(0.0846)
200	0.000	0.930	0.070	0.3421(0.1181)	1.1346(0.3066)	0.4357(0.0412)	1.0028(0.0064)

References

Chen, X., Hong, H., and Tarozzi, A. (2008), "Semiparametric efficiency in GMM models with auxiliary data," *The Annals of Statistics*, 36, 808–843.

Firpo, S. (2007), "Efficient semiparametric estimation of quantile treatment effects," *Econometrica*, 75, 259–276.

Huang, J., Horowitz, J., and Wei, F. (2010), "Variable selection in nonparametric additive models," *Annals of statistics*, 38, 2282–2313.

Johnson, B., Lin, D., and Zeng, D. (2008), "Penalized estimating functions and variable selection in semiparametric regression models," *Journal of the American Statistical Association*, 103, 672–680.

Li, R. and Liang, H. (2008), "Variable selection in semiparametric regression modeling," *Annals of statistics*, 36, 261–286.

Liang, H., Liu, X., Li, R., and Tsai, C. (2010), "Estimation and testing for partially linear single-index models," *Annals of statistics*, 38, 3811–3836.

Ma, Y., Chiou, J., and Wang, N. (2006), "Efficient semiparametric estimator for heteroscedastic partially linear models," *Biometrika*, 93, 75–84.

Tsiatis, A. (2006), *Semiparametric theory and missing data*, Springer Verlag.

Wang, H. and Xia, Y. (2009), "Shrinkage estimation of the varying coefficient model," *Journal of the American Statistical Association*, 104, 747–757.

Xue, L. and Wang, Q. (2012), "Empirical likelihood for single-index varying-coefficient models," *Bernoulli*, 18, 836–856.

Yu, Y. and Ruppert, D. (2002), "Penalized spline estimation for partially linear single-index models," *Journal of the American Statistical Association*, 97, 1042–1054.

Yuan, M. and Lin, Y. (2006), "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68, 49–67.