应用概率统计　第二十五卷
第五期　2009年10月

Chinese Journal of Applied Probability
and Statistics Vol.25 No.5 Oct. 2009

Inference in Varying-Coefficient Mixed Models by Using Smoothing Spline *

Lu Yiqiang

(*Institute of Electronic Technology, the PLA Information Engineering University, Zhengzhou, 450004*)

Xu Wangli

(*School of Statistics, Renmin University of China, Beijing, 100872*)

Abstract

Varying-coefficients mixed model (VCMM) is proposed for longitudinal data and the other correlated data. This model allows flexible functional dependence of the response variable on the covariates by using varying-coefficients linear part to present the covariates effects, while accounting the within-subject correlation by using random effect. In this article, the coefficient functions are estimated by using smoothing spline and restricted maximum likelihood is used to estimate the smoothing parameters and the variance components simultaneously. The performance of the proposed method is evaluated though some simulation studies, which show that both the coefficient functions and variance components could be estimated well for the VCMMs with all kinds of covariance structures.

Keywords: Varying-coefficients mixed model, smoothing spline estimation, restricted maximum likelihood, linear mixed effect model.

AMS Subject Classification: Primary 62G05; secondary 62G08.

§1. Introduction

The varying coefficient model (VCM) is a useful extension of classical linear models. There are extensive studies on it after the seminal work of Hastie and Tibshirani (1993). The appeal of this model is that via allowing coefficients to vary with some covariate, the modeling bias can significantly be reduced and "curse of dimensionality" can be avoided. Another advantage of this model is its interpretability. It arise naturally when one is interested in exploring how regression coefficients change over different groups such as age. It is particularly appealing in longitudinal studies where it allows one to examine the extent to which covariates affect response over time. See Hoover et al. (1998), Fan and Zhang (2000), Chiang et al. (2001) and Huang, Wu and Zhou (2002), Zhang and

*The research of Lu is supported by the National Natural Science Foundation of China (10501053). The research of Xu is supported by the National Natural Science Foundation of China (10701079).

Received July 10, 2007. Revised November 15, 2007.

Lu (2004) for the details on the applications of varying-coefficient models to longitudinal data. Additional work is cited in these references.

A challenge in the analysis of longitudinal data is that the data are correlated, as multiple observations are measured for each individual. This correlation should be taken into account in the analysis to yield the valid inference. However, in the least square estimation of VCM, whether smoothing spline estimation or local method, this correlation is often neglected and only considered in the study of the asymptotic property of the estimators. In the application of VCM, another critical issue is how to select good estimators of smoothing parameters or bandwidth parameters. Although cross-validation is a reasonable approach to select the smoothing parameters for clustered data, it is often computational expensive and subsequent inference on the correlation parameters is difficult. It is hence of substantial interest to develop a systematic procedure to make inference on all model parameters.

A popular parametric way within the likelihood frame is to use linear mixed models (LMM) to analyze the longitudinal data (Laird and Ware, 1982). When simple parametric forms are insufficient, however, nonparametric approaches allowing arbitrary functional forms must be considered. There are some papers in the recent literature extending LMM to the analysis of modeling replicated functional data. For example, Lin and Zhang (1999) proposed generalized additive mixed models, where additive covariate effects were used to model the covariate effects in GLMMs. Wu and Liang (2004) proposed a random varying-coefficient model in which the time-varying coefficients are assumed to be subject-specific, and can be considered as realizations of stochastic processes. Guo (2002) introduced a functional mixed model allowing functional fixed and random-effect functions of arbitrary form, with the modeling done by using smoothing splines. Morris and Carroll (2006) used a Bayesian wavelet-based approach to fit the functional mixed model in order to suite for modeling irregular functional data. In this paper, we consider the varying-coefficients mixed model (VCMM), which is an extension of LMM in the spirit of Hastie and Tibshirani (1993). This new class of model uses the varying-coefficient linear part to model covariate effects while accounting for overdispersion and correlation by adding random effects to VCM. The smoothing spline is used to fit the coefficient functions. We treat the smoothing parameters as the extra variance components and estimate them jointly with the other variance components by the using restricted maximum likelihood (REML) (Harville, 1977). It is shown that all model parameters can be estimated from a modified linear mixed model.

This paper is organized as follows. Section 2 states model. Section 3 contains the estimation procedures. The algorithms are summarized in Section 4. Section 5 reports the results from some simulation studies designed to evaluate the performance of the proposed estimation. Section 6 presents some discussion.

§2. The Varying-Coefficient Mixed Effect Models

Let the data be consist of m subjects with ith subject having n_i observations over time. Suppose that Y_{ij} $(i=1,\cdots,m,\ j=1,\cdots,n_i)$ are the response for the ith subject at time point t_{ij} and can be modeled by the varying-coefficients mixed model

$$Y_{ij}=X_{ij}^{\tau}\beta(t_{ij})+Z_{ij}^{\tau}b_i+\epsilon_{ij}, \tag{2.1}$$

where $\beta(t)=(\beta_1(t),\cdots,\beta_p(t))^{\tau}$ is a $p\times 1$ vector of unknown coefficient functions associated with covariates $X_{ij}=(X_{ij1},\cdots,X_{ijp})^{\tau}$; b_i are independent $q\times 1$ vectors of random effects associated with covariates $Z_{ij}=(Z_{ij1},\cdots,Z_{ijq})^{\tau}$; and the ϵ_{ij} are independent measure errors. Without a loss of generality, let $t\in[0,1]$. Furthermore, suppose that b_i is distributed as Normal$(0,D(\theta))$ where $D(\theta)$ is a positive matrix depending on parameter vector θ, and ϵ_{ij} is distributed as Normal$(0,\sigma^2)$ and independent of b_i. $(\theta^{\tau},\sigma^2)^{\tau}$ is called to be the vector of variance components.

A key feature of VCMM (2.1) is that varying-coefficients part is used to model covariate effects and random effects are used to model correlation between observations. If $\beta_j(\cdot)$ are constant functions of t, the VCMM (2.1) reduces to the linear mixed model (LMM). If $X_1\equiv 1$ and $\beta_j(t)$ except of $\beta_1(t)$ are constant function of t, the VCMM (2.1) reduces to the semiparametric mixed model (Zeger and Diggle, 1994). In the functional mixed model (Morris and Carroll, 2006), random effect b_i is extended to the realization of statistic process.

Denote $Y_i=(Y_{i1},\cdots,Y_{in_i})^{\tau}$ and Z_i,ϵ_i similarly $(i=1,\cdots,m)$. Let $t^0=(t_1^0,\cdots,t_r^0)^{\tau}$ be a vector of ordered distinct values of the time points t_{ij} $(i=1,\cdots,m,\ j=1,\cdots,n_i)$ and N_i be the incidence matrix for the ith subject connecting $t_i=(t_{i1},\cdots,t_{in_i})^{\tau}$ and t^0 such that (j,l)th element of N_i is 1 if $t_{ij}=t_l^0$ and 0 otherwise $(j=1,\cdots,n_i,\ l=1,\cdots,r)$. The model (2.1) can be written as

$$Y_i=D_{i1}N_i\beta_1+D_{i2}N_i\beta_2+\cdots+D_{ip}N_i\beta_p+Z_ib_i+\epsilon_i, \tag{2.2}$$

where $D_{ik}=\mathrm{diag}(X_{i1k},\cdots,X_{in_ik})$ and $\beta_k=(\beta_k(t_1^0),\cdots,\beta_k(t_r^0))^{\tau}$ for $k=1,\cdots,p$.

Let $G_i=(D_{i1}N_i,D_{i2}N_i,\cdots,D_{ip}N_i)$, $\beta=(\beta_1^{\tau},\cdots,\beta_p^{\tau})^{\tau}$ and $Z=\mathrm{diag}(Z_1,\cdots,Z_m)$. Further denoting $Y=(Y_1^{\tau},\cdots,Y_m^{\tau})^{\tau}$ and G,b,ϵ similarly, (2.2) reduces to

$$Y=G\beta+Zb+\epsilon, \tag{2.3}$$

where b is distributed as Normal$(0,\mathcal{D}(\theta))$ with $\mathcal{D}(\theta)=\mathrm{diag}(\mathcal{D}(\theta),\cdots,\mathcal{D}(\theta))$ and ϵ is distributed as Normal$(0,\sigma^2I_n)$ with I_n denoting an identity matrix of dimension $n=\sum\limits_{i=1}^m n_i$.

§3. Estimation for the Mean Components

Statistical inference in VCMM (2.1) involves inference on the coefficient functions $\beta_j(t)$, which often requires the estimation of the smoothing parameter, say λ_j, and inference on the variances $(\theta^\tau, \sigma^2)^\tau$. In this section, We shall first discuss how to construct natural cubic smoothing estimators of $\beta_j(t)$ when $\lambda = (\lambda_1, \cdots, \lambda_p)^\tau$ and $(\theta^\tau, \sigma^2)^\tau$ are known; and the inference on λ and $(\theta^\tau, \sigma^2)^\tau$ will be studied in the next section.

3.1 Estimation of Coefficient Functions

For given variance component, the loglikelihood function of $\beta_j(\cdot)$ is, apart from a constant,

$$l(\beta; Y) = -\frac{1}{2}\log|R| - \frac{1}{2}(Y - G\beta)^\tau R^{-1}(Y - G\beta), \tag{3.1}$$

where $R = Z\mathcal{D}(\theta)Z^\tau + I\sigma^2$. Since $\beta_j(t)$ are infinite dimensional smoothing functions, we estimate $\beta_j(t)$ by maximizing the penalized loglikelihood function

$$l_p(\beta_1(t), \cdots, \beta_p(t); Y) = l(\beta; Y) - \frac{1}{2}\sum_{j=1}^{p}\lambda_j\int[\beta_j''(t)]^2\mathrm{d}t, \tag{3.2}$$

where $l(\beta; Y)$ is the loglikelihood function defined in equation (3.1), the λ_j, $j = 1, \cdots, p$ are smoothing parameters that control the smoothness of $\widehat{\beta}_j(t)$ and goodness-of-fit of the model to the data. The maximizer $\widehat{\beta}_j(t)$ of $l_p(\beta; Y)$ are natural cubic splines, which can be expressed as

$$\beta_j(t) = \sum_{v=1}^{2}\delta_{jv}\phi_v + \sum_{l=1}^{r}a_{jl}R(t_l^0, t),$$

where ϕ_v is a $(v-1)$th polynomial (e.g. $\phi_{jv} = t^{v-1}/(v-1)!$) and $R(x, y)$ is defined as

$$R(x, y) = K_2(x)K_2(y) - K_4(x - y).$$

For

$$K_2(x) = \frac{1}{2}\Big(k_1^2(x) - \frac{1}{12}\Big), \qquad k_4(x) = \frac{1}{24}\Big(k_1^4(x) - \frac{k_1^2(x)}{2} + \frac{7}{240}\Big)$$

and $k_1(x) = x - 0.5$ on $x \in [0, 1]$. See Gu (2002, P. 37).

Denote $\delta_j = (\delta_{j1}, \delta_{j2})^\tau$ and $a_j = (a_{j1}, \cdots, a_{jr})^\tau$. Then β_j and penalties in (3.2) can be expressed as

$$\beta_j = T\delta_j + \Sigma a_j \qquad \text{and} \qquad \frac{1}{2}\lambda_j\int[\beta_j''(t)]^2\mathrm{d}t = \frac{\lambda_j}{2}a_j^\tau\Sigma a_j, \tag{3.3}$$

where T is $r \times 2$ with (l, s)th element equal to $\phi_s(t_l^0)$ and Σ is a positive definite matrix with (l, s)th element equal to $R(t_l^0, t_s^0)$. Let $G_1^{(i)} = (D_{i1}N_iT, \cdots, D_{ip}N_iT)$, $G_2^{(i)} =$

$(D_{i1}N_i\Sigma,\cdots,D_{ip}N_i\Sigma)$, $G_1=(G_1^{(1)^\tau},\cdots,G_1^{(m)^\tau})^\tau$ and G_2 similarly. The model (2.3) reduces to

$$Y=G_1\delta+G_2a+Zb+\epsilon,$$

where $\delta=(\delta_1^\tau,\cdots,\delta_p^\tau)^\tau$ and $a=(a_1^\tau,\cdots,a_p^\tau)^\tau$. Therefore, the penalized loglikelihood (3.2) becomes

$$\begin{aligned} l_p(\delta,a;Y) &= -\frac{1}{2}\log|R|-\frac{1}{2}(Y-G_1\delta-G_2a)^\tau R^{-1}(Y-G_1\delta-G_2a)-\frac{1}{2}\sum_{j=1}^{p}\lambda_j a_j^\tau\Sigma a_j \\ &= -\frac{1}{2}\log|R|-\frac{1}{2}(Y-G_1\delta-G_2a)^\tau R^{-1}(Y-G_1\delta-G_2a)-\frac{1}{2}a^\tau\Omega^{-1}(\lambda)a, \quad (3.4) \end{aligned}$$

where $\lambda=(\lambda_1,\cdots,\lambda_p)'$ and $\Omega^{-1}(\lambda)=\mathrm{diag}(\lambda_1\Sigma,\cdots,\lambda_p\Sigma)$.

The maximizing penalized loglikelihood estimates (MPLEs) $(\widehat{\delta},\widehat{a})$ can be obtained by solving the linear system

$$\begin{pmatrix} G_1^\tau R^{-1}G_1 & G_1^\tau R^{-1}G_2 \\ G_2^\tau R^{-1}G_1 & \Omega^{-1}+G_2^\tau R^{-1}G_2 \end{pmatrix}\begin{pmatrix}\delta\\ a\end{pmatrix}=\begin{pmatrix}G_1R^{-1}Y\\ G_2^\tau R^{-1}Y\end{pmatrix}. \quad (3.5)$$

Estimating of the subject-specific random effect b_i can proceed by calculating their conditional expectations given the data Y_i, while estimating δ and a by their MPLEs. This gives

$$\widehat{b}_i=D(\theta)Z_i^\tau(Z_iD(\theta)Z_i^\tau+I_q\sigma^2)^{-1}(Y_i-G_1^{(i)}\widehat{\delta}-G_2^{(i)}\widehat{a}).$$

This suggests that the following theorem holds.

Theorem 3.1 If the coefficient functions in (2.1) are estimated by smoothing spline. Then the maximizing penalized loglikelihood estimates (MPLEs) are identical to the best linear unbiased predictors (BLUPs) of the following linear mixed model

$$Y=G_1\delta+G_2a+Zb+\epsilon, \quad (3.6)$$

where δ are the regression coefficient and a and b are mutually independent random effect with $a\sim N(0,\Omega(\lambda))$ and $b\sim N(0,\mathcal{D}(\theta))$.

Theorem 3.1 can be easily proved by the comparison between the MPLEs (3.5) and BLUPs defined in Harville (1977).

3.2 Bayesian Formulation and Inference

For the classical nonparametric model $y_i=f(t_i)+\epsilon_i$, where ϵ_i are independent and distributed as $N(0,\sigma^2)$, the regression f can be estimated by smoothing spline and f have the form

$$f=T\delta+\Sigma a,$$

where T and Σ are defined in (3.3). Suppose that δ have a flat prior and a have a normal prior $N(0,(\lambda\Sigma)^{-1})$, one can easily show that the posterior model and mean of $(\delta^\tau,a^\tau)^\tau$ under the Bayesian model are identical to the MPLEs $(\widehat{\delta}^\tau,\widehat{a}^\tau)^\tau$. Wahba (1985) suggested estimating the covariance of $\widehat{f}$ using the posterior covariance of $\widehat{f}$ under Bayesian model. She showed that the resulting Bayesian confidence intervals of f calculated using the Bayesian standard errors have good coverage probability when the true $f(x)$ is a fixed smooth function.

Similar to the smoothing spline estimate of nonparametric regression, the covariance of $\widehat{\beta}_j$ can be estimated by their posterior covariance under Bayesian models. Note that β_j can be expressed as (3.3). Assuming that δ_j have a flat prior and a_j are independent and have a normal prior $N(0,(\lambda_j\Sigma)^{-1})$, some calculation shows that the MPLEs are identical to posterior mean of δ and a. Let $\alpha=(\delta^\tau,a^\tau)^\tau$ and $\widehat{\alpha}=(\widehat{\delta}^\tau,\widehat{a}^\tau)^\tau$. The Bayesian covariance matrix of MPLEs $\widehat{\alpha}$ is

$$\mathsf{Cov}_B(\widehat{\alpha})=H^{-1},$$

where H is the coefficient matrix on the left-hand side of equation (3.5). It follows that the Bayesian covariance of $\widehat{\beta}_j$ is

$$\mathsf{Cov}(\widehat{\beta}_j)=(T,\Sigma)\mathsf{Cov}_B(\widehat{\alpha}_j)(T,\Sigma)^\tau, \tag{3.7}$$

where $\widehat{\alpha}_j=(\widehat{\delta}_j^\tau,\widehat{a}_j^\tau)^\tau$ and $\mathsf{Cov}_B(\widehat{\alpha}_j)$ can easily obtained from the corresponding blocks of H^{-1}.

§4. Inference on the Smoothing Parameters and Variance Components

We assume in Section 3 the smoothing parameters λ_j and the variance component vector are known when we make inference on the coefficient functions β_j in VCMM (2.1). However, they are often need to be estimated from the data. Under the linear mixed models, the restricted maximum likelihood (REML), which takes into account the loss in degree of freedom resulting from estimating fixed effects, is often used to estimate the variance components (Harville, 1977). Due to the smoothing spline estimator of nonparametric function has often a representation of linear mixed model, REML has also been used for selecting the smoothing parameter in the nonparametric regression. For example, Ansley, Kohn and Wang (1993) and Lin and Zhang (1999).

In this section, using the connection between the smoothing spline estimation of varying-coefficient mixed models and the linear mixed model established in Section 3, we propose to estimate the soothing parameter λ and variance component $(\theta^\tau, \sigma^2)^\tau$ simultaneously using REML by treating λ as an extra variance component vector in the linear mixed model (3.6). For the convenience of calculation, let $\tau_j = 1/\lambda_j$ and $\tau = (\tau_1, \cdots, \tau_p)^\tau$. By Harville (1977), the REML of $\gamma = (\tau', \theta', \sigma^2)'$ in the LMM (3.6) is

$$l_R(\tau, \theta, \sigma^2; Y) = -\frac{1}{2}\log(|V|) - \frac{1}{2}\log(|G_1^\tau V^{-1} G_1|) - \frac{1}{2}(Y - G_1\widehat{\delta})^\tau V^{-1}(Y - G_1\widehat{\delta}), \quad (4.1)$$

where $\widehat{\delta}$ is the estimator in Section 3 and

$$V = R + G_2\Omega(\lambda)G_2^\tau = I_n\sigma^2 + Z\mathcal{D}(\theta)Z^\tau + G_2\Omega(\lambda)G_2^\tau.$$

Differentiating $l_R(\tau, \theta, \sigma^2; Y)$ with respect to γ, the REML estimating equation is

$$s(\gamma) = -\frac{1}{2}\mathrm{tr}\Big(P\frac{\partial V}{\partial \gamma_k}\Big) + \frac{1}{2}(Y - G_1\widehat{\delta})^\tau V^{-1}\frac{\partial V}{\partial \gamma_k}V^{-1}(Y - G_1\widehat{\delta}) = 0, \quad (4.2)$$

where

$$P = R^{-1} - R^{-1}(G_1, G_2)H^{-1}(G_1, G_2)^\tau R^{-1}.$$

Using the identity $V^{-1}(Y - G_1\widehat{\delta}) = R^{-1}(Y - G_1\widehat{\delta} - G_2\hat{a})$ (Harville (1977), Eq. 5.2), we have

$$-\frac{1}{2}\mathrm{tr}\Big(PG_2\frac{\partial \Omega}{\partial \tau_k}G_2^\tau\Big) + \frac{1}{2}(Y - G_1\widehat{\delta} - G_2\widehat{a})^\tau R^{-1}G_2\frac{\partial \Omega}{\partial \tau_k}G_2^\tau R^{-1}(Y - G_1\widehat{\delta} - G_2\widehat{a}) = 0,$$
$$-\frac{1}{2}\mathrm{tr}\Big(PZ\frac{\partial \mathcal{D}}{\partial \theta_k}Z^\tau\Big) + (Y - G_1\widehat{\delta} - G_2\widehat{a})^\tau R^{-1}Z\frac{\partial \mathcal{D}}{\partial \theta_k}Z^\tau R^{-1}(Y - G_1\widehat{\delta} - G_2\widehat{a}) = 0,$$

and

$$-\frac{1}{2}\mathrm{tr}(P) + (Y - G_1\widehat{\delta} - G_2\widehat{a})^\tau R^{-2}(Y - G_1\widehat{\delta} - G_2\widehat{a}) = 0,$$

where $\widehat{\delta}$ and $\widehat{a}$ are the estimators in Section 3. The Fisher information of $\widehat{\gamma}$ is

$$I(\gamma) = \Big(\frac{1}{2}\mathrm{tr}\Big(P\frac{\partial V}{\partial \gamma_i}P\frac{\partial V}{\partial \gamma_j}\Big)\Big) = \begin{pmatrix} I_{\tau\tau} & I_{\tau\theta} & I_{\tau\sigma^2} \\ I_{\theta\tau} & I_{\theta\theta} & I_{\theta\sigma^2} \\ I_{\sigma^2\tau} & I_{\sigma^2\theta} & I_{\sigma^2\sigma^2} \end{pmatrix}. \quad (4.3)$$

Denoting by A^{jk} the (j,k)th element of the matrix A, we have

$$I_{\tau\tau}^{jk} = 0.5\mathrm{tr}\Big(G_2^\tau PG_2\frac{\partial \Omega}{\partial \tau_j}G_2^\tau PG_2\frac{\partial \Omega}{\partial \tau_k}\Big), \qquad I_{\tau\theta}^{jk} = 0.5\mathrm{tr}\Big(Z^\tau PG_2\frac{\partial \Omega}{\partial \tau_j}G_2^\tau PZ\frac{\partial \mathcal{D}(\theta)}{\partial \theta_k}\Big),$$
$$I_{\theta\theta}^{jk} = 0.5\mathrm{tr}\Big(Z^\tau PZ\frac{\partial \mathcal{D}(\theta)}{\partial \theta_j}Z^\tau PZ\frac{\partial \mathcal{D}(\theta)}{\partial \theta_k}\Big), \qquad I_{\tau\sigma^2}^{j1} = 0.5\mathrm{tr}\Big(PG_2\frac{\partial \Omega}{\partial \tau_j}G_2^\tau P\Big),$$
$$I_{\theta\sigma^2}^{j1} = 0.5\mathrm{tr}\Big(PZ\frac{\partial \mathcal{D}(\theta)}{\partial \tau_j}Z^\tau P\Big) \quad \text{and} \quad I_{\sigma^2\sigma^2} = 0.5\mathrm{tr}(P).$$

The covariance of $\widehat{\gamma}$ can be estimated using the inverse of Fisher information matrix. To estimate γ, we can now proceed as follows. Beginning with some starting values $\gamma^{(0)}$, iterate

$$\gamma^{(k+1)} = \gamma^{(k)} + I(\gamma^{(k)})^{-1}s(\gamma^{(k)})$$

until convergence.

§5. Summary of Inference in VCMM

Since the estimation of coefficient functions depends on the variance parameters and smoothing parameters and vice versa, both may be estimated through cycling between the two estimation step, using the current values of $(\widehat{\delta}^{\tau}, \widehat{a}^{\tau})^{\tau}$ and $(\widehat{\tau}', \widehat{\theta}', \sigma^2)'$ in the respective formulae. In this section, we will summarize the presented results in a algorithm that describes the simultaneous estimation of both coefficient functions and variance parameters.

(1) Choose starting value $(\delta^{(0)}, a^{(0)})$, $(\tau^{(0)}, \theta^{(0)}, \sigma^{2(0)})$, a termination criterion ϵ and define the iteration index $k=0$. In the following simulation studies, we select the starting value $\sigma^{2(0)} = 1$, $\tau_j^{(0)} = 1$. Furthermore, we select $\theta^{(0)}$ such that $\mathsf{Var}\,(b_i) = 1$, $\mathrm{corr}(b_i, b_j) = 0.5$. $(\delta^{(0)}, a^{(0)})$ is chosen to be the estimator of linear model $Y = G_1\delta + G_2 a + \epsilon$.

(2) Determine $(\widehat{\delta}^{(k+1)}, \widehat{a}^{(k+1)})$ as the solution of the linear system (3.5). Note that all involved matrices are evaluated at the current estimates.

(3) Compute the score vector $s(\gamma^{(k)})$ and the expected Fisher-information matrix $I(\gamma^{(k)})$. Update $\gamma^{(k+1)}$ via

$$\gamma^{(k+1)} = \gamma^{(k)} + I^{-1}(\gamma^{(k)})s(\gamma^{(k)}).$$

(4) Compute the distance measures

$$d_1 = d((\widehat{\delta}^{(k)}, \widehat{a}^{(k)}), (\widehat{\delta}^{(k+1)}, \widehat{a}^{(k+1)})) = \frac{\|(\widehat{\delta}^{(k+1)}, \widehat{a}^{(k+1)}) - (\widehat{\delta}^{(k)}, \widehat{a}^{(k)})\|}{\|(\widehat{\delta}^{(k)}, \widehat{a}^{(k)})\|}$$

and

$$d_2 = d(\gamma^{(k)}, \gamma^{(k+1)}) = \frac{\|\gamma^{(k)} - \gamma^{(k+1)}\|}{\|\gamma^{(k)}\|}.$$

If $d_1 > \epsilon$ or $d_2 > \epsilon$, replace k by $k+1$ and go back to (2). Otherwise stop the estimation procession.

§6. Simulation Study

Some simulation studies were carried out to evaluate of the performance of MPLEs of coefficient functions and the REML estimates of the smoothing parameters and variance components under all kinds of the different covariance construction. Each data set was composed of 100 clusters of size $n_i = 5$. Denote the coefficient functions

$$\beta_1(t) = 15 + 20\sin\Big(\frac{\pi t}{60}\Big), \quad \beta_2(t) = 4 - \Big(\frac{t-20}{10}\Big)^2, \quad \beta_3(t) = 2 - 3\cos\Big(\frac{(t-25)\pi}{15}\Big), \tag{6.1}$$

where $t \in [0, 30]$.

Example 6.1 Let

$$Y_{ij} = X_{ij1}\beta_1(t_{ij}) + X_{ij2}\beta_3(t_{ij}) + Z_{ij}b_i + \epsilon_{ij}, \tag{6.2}$$

where $\beta_1(\cdot)$ and $\beta_3(\cdot)$ were defined in (6.1); $X_{ij1} \equiv 1$, $X_{ij2} \sim \text{Normal}(1, 0.25)$ and $Z_{ij}^\tau = (Z_{ij1}, Z_{ij2}) = (X_{ij1}, X_{ij2})$. The covariate t varied within each cluster with 100 equally spaced knots in $[0, 30]$, that is, $t_{ij} = 30\lfloor(i+4)/5\rfloor/100 + 6(j-1)$ for $i = 1, \cdots, 100$, $j = 1, \cdots, 5$, where $\lfloor\cdot\rfloor$ denotes a truncation operator. The random effect b_i were supposed to be independent and distributed $\text{Normal}(0, D(\theta))$, where $D(\theta) = \text{diag}(\theta_1, \theta_2)$ with $\theta_1 = 1$ and $\theta_2 = 0.5$. The random error ϵ_{ij} were supposed to be independent and distributed as $\text{Normal}(0, \sigma^2)$ with $\sigma^2 = 1$.

Example 6.2 Example 6.2 is the same as Example 6.1 except of the covariance matrix of the random effect $D(\theta) = \text{diag}(\theta_1, \theta_1)$ with $\theta_1 = 0.5$.

Example 6.3 Let

$$Y_{ij} = X_{ij1}\beta_2(t_{ij}) + X_{ij2}\beta_3(t_{ij}) + Z_{ij}^\tau b_i + \epsilon_{ij}, \tag{6.3}$$

where $\beta_2(\cdot)$ and $\beta_3(\cdot)$ were defined in equation (6.1); X_{ij}, Z_{ij} and t_{ij} were generated by the same method as the one used in Example 6.1. The random effect was supposed to be distributed as

$$b_i \sim N(0, D(\theta)), \qquad D(\theta) = \begin{pmatrix} \theta_1 & \theta_2 \\ \theta_2 & \theta_3 \end{pmatrix},$$

where $\theta_1 = 0.8$, $\theta_3 = 0.3$, $\theta_2 = -0.5\sqrt{\theta_1\theta_3} = -0.2450$, that is, the correlation of b_{i1} and b_{i2} is $\text{corr}(b_{i1}, b_{i2}) = -0.5$. The error ϵ_{ij} were suppose to be $N(0, \sigma^2)$ with $\sigma^2 = 0.25$.

It is easy to see that the main difference among Example 6.1–6.3 is in the structure of the covariance matrix $D(\theta)$ of random effect b_i. Note that $D(\theta)$ is a diagonal positive-definite matrix in Example 6.1, a multiple of the identity positive-definite matrix in Example 6.2 and a general positive-definite matrix in Example 6.3. Three thousand data sets were generated for each example and the estimation procedure developed in Section 3 was applied to each dataset. For comparison, varying-coefficient models were also fitted by assuming the random effects were absent.

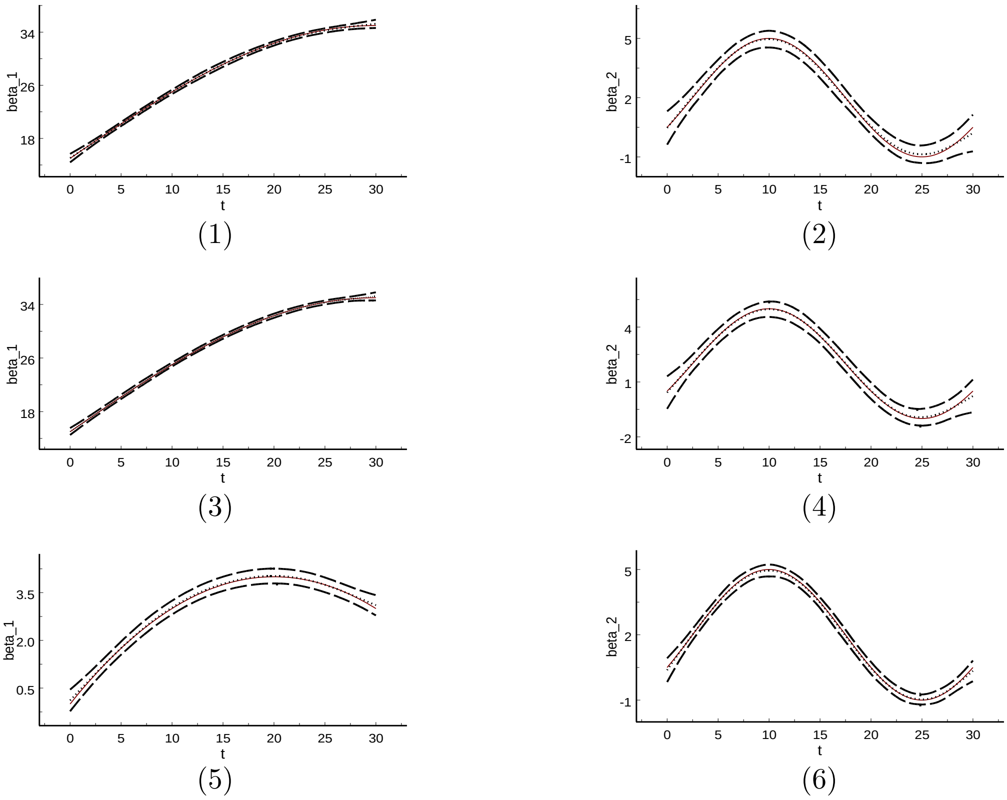


Figure 1 True and estimated coefficient functions of Example 6.1–6.3 based on 3000 replications: solid curve — true coefficient functions; shot-dashed curve — the average of the estimated coefficient functions; long-dashed curve — empirical 95% confidence intervals. Figure (1–2) is the estimate for Example 6.1, Figure (3–4) for Example 6.2, and Figure (5–6) for Example 6.3.

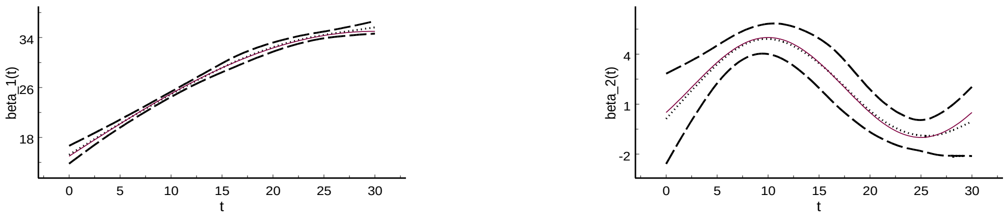
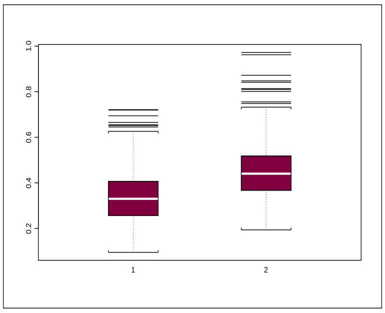


Figure 2 A typical result of Example 6.1: solid curve — true coefficient functions; shot-dashed curve — the the estimates of the coefficient functions; long-dashed curve — 95% confidence intervals.

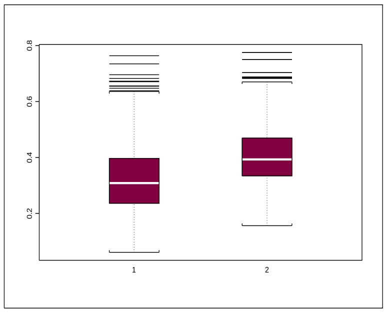
Table 1 The means (and standard errors) of the estimates of variance components over 3000 replications of Example 6.1–6.3

		θ_1	θ_2	θ_3	σ^2
Eg. 1	True value	1	0.5		1
	Estimation	1.011(0.210)	0.502(0.165)		0.999(0.073)
Eg. 2	True value	0.5			1
	Estimation	0.502(0.085)			0.996(0.078)
Eg. 3	True value	0.8	-0.245	0.3	0.25
	Estimation	0.791(0.168)	-0.242(0.102)	0.298(0.084)	0.249(0.020)

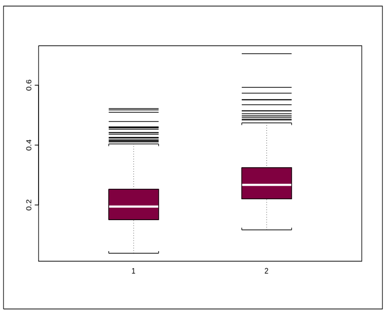
Figure 1 presents the average of the estimated coefficient functions and empirical 95% confidence intervals from 3000 simulation runs. Figure 2 gives a typical result of Example 6.1 drawn from 3000 simulations at random. The other two examples were also estimated well in each simulation and the typical results were omitted. Table 1 gives the estimated variance components and their empirical standard errors. From Figure 1–2 and Table 1, it suggests that both the coefficient functions and variance components of VCMM with all kinds of covariance structures were estimated well by the procedure developed in Section 3.



Example 6.1



Example 6.2



Example 6.3

Figure 3 The boxplots of ADE of Example 6.1–6.3: “1” — the ADE for estimates of VCMM; “2” — the ADE for estimates of VCM in which the random effect were neglected.

The performance of a fit for the coefficient functions can be measured by the absolute deviation error(ADE) from the true coefficient curves. Let $u_1, \cdots, u_T$ be T grid points. The ADE is defined as

$$\mathrm{ADE} = \sum_{l=1}^{p} \mathrm{ADE}_l = \sum_{l=1}^{p} \sum_{j=1}^{T} |\widehat{\beta}_l(u_j) - \beta_l(u_j)|, \tag{6.4}$$

where $\mathrm{ADE}_l = \sum\limits_{j=1}^{T} |\widehat{\beta}_l(u_j) - \beta_l(u_j)|$. Figure 3 shows the comparatione between the ADEs for the estimate of VCMM and that of VCM. The first column of Figure 3 presents the boxplots of ADE for the estimate of VCMM over 3000 replications, where Example 6.1–6.3 were estimated by the use of the procedures developed in Section 3. The second column presents the boxplots of ADE for the estimate of VCM from the assumption that the random effect were absent. Figure 3 suggests that the procedure developed in Section 3 would be more efficient than the estimates of VCM when the random effect exists.

§7. Discussion

In this article, we proposed a VCMM for correlated data. Smoothing spline was used to estimate the coefficient functions and the restricted maximum likelihood (REML) was used to estimate the smoothing parameters and the variance components simultaneously. A key feature of this approach is that it allows us to make systematic inference on all model parameters of VCMMs, including coefficient functions, smoothing parameters and variance components.

The simulation studies in Section 6 show that the proposed method performs well in estimating the coefficient functions and variance components of VCMM with all kinds of covariance structures, such as, a diagonal matrix, a multiple of the identity matrix or a general positive-definite matrix. According the characteristics of smoothing spline, the proposed procedure can be easily extended to the other penalized spline estimation, such as, penalized B-spline (Eilers and Marx, 1996).

References

[1] Ansley, C.F., Kohn, R. and Wong, C., Nonparametric spline regression with prior information, *Biometrika*, **80**(1993), 75–88.

[2] Chiang, C.-T., Rice, J.A. and Wu, C.O., Smoothing spline estimation for varying coefficient models with repeatedly measure dependent variables, *J. Amer. Statist. Assoc.*, **96**(2001), 605–619.

[3] Eilers, P.H.C. and Marx, B.D., Flexible smoothing with B-spline and penalties, *Statistical Science*, **11**(1996), 89–121.

[4] Fan, J. and Zhang, J.T., Two-step estimation of functional linear models with applications to longitudinal data, *J. Royal Statist. Soc. B*, **62**(2000), 303–322.

[5] Gu, C., *Smoothing Spline ANOVA Models*, New York: Springer-Verlag, 2002.

[6] Harville, D.A., Maximum likelihood approaches to variance component estimation and to related problems, *J. Amer. Statist. Assoc.*, **72**(1977), 320–338.

[7] Hastie, T. and Tibshirani, R.J., Varying-coefficient model, *J. R. Statist. Soc. Ser. B*, **55**(1993), 757–796.

[8] Hoover, D.R., Rice, J.A., Wu, C.O. and Yang, L.P., Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data, *Biometrika*, **85**(1998), 809–822.

[9] Huang, J.Z., Wu, C.O. and Zhou, L., Varying-coefficient models and basis functions approximations for the analysis of repeated measurements, *Biomatrika*, **89**(2002), 111–128.

[10] Lin, X. and Zhang, D., Inference in generalized additive mixed models using smoothing splines, *J. R. Statist. Soc. Ser. B*, **61**(1999), 381–400.

[11] Laird, N.M. and Ware, J.H., Random-effects models for longitudinal data, *Biometrics*, **38(4)**(1982), 963–974.

[12] Morris, J.S. and Carroll, R.J., Wavelet-based functional mixed models, *J. R. Statist. Soc. B*, **68**(2006), 179–199.

[13] Wahba, G., A comparison of GCV and GML for choosing the smoothing parameters in the generalized spline smoothing problem, *Ann. Statist.*, **13**(1985), 1378–1402.

[14] Wu, H. and Liang, H., Backfitting random varying-coefficient models with time-dependent smoothing covariates, *Scand. J. Statist.*, **31**(2004), 3–19.

[15] Zhang, R. and Lu, Y., Varying-Coefficient Model, Beijing: China Science Press, 2004.

[16] Zeger, S.L. and Diggle, P.J., Semiparametric models for longitudinal data with application to CD4 cell numbers in HIV seroconverters, *Biometrics*, **50**(1994), 689–699.

变系数混合模型的光滑样条推断

卢一强 许王莉

(解放军信息工程大学电子技术学院, 郑州, 450004) (中国人民大学统计系, 北京, 100872)

为了拟合纵向数据和其他相关数据, 本文提出了变系数混合效应模型(VCMM). 该模型运用变系数线性部分来表示协变量对响应变量的影响, 而用随机效应来描述纵向数据组内的相关性, 因此, 该模型允许协变量和响应变量之间存在十分灵活的泛函关系. 文中运用光滑样条来估计均值部分的系数函数, 而用限制最大似然的方法同时估计出光滑参数和方差成分, 我们还得到了所提估计的计算方法. 大量的模拟研究表明对于具有各种协方差结构的变系数混合效应模型, 运用本文所提出的方法都能够十分有效地估计出模型中的系数函数和方差成分.

关键词: 变系数混合效应模型, 光滑样条估计, 限制最大似然, 线性混合效应模型.

学科分类号: O212.7.