

Nonlinear Regression Modeling for Longitudinal Data and its Applications

Mathematical Course, Toshihiro Misumi

Abstract — Longitudinal data analysis has made a significant progress over the last three decades in various fields of natural and social sciences, including bioinformatics, medicine, pharmaceuticals and systems engineering. Although linear mixed models provide a useful tool for analyzing such data sets, more flexible modeling procedures are required to extract information from data with complex structures. We propose several nonlinear regression modeling strategies for longitudinal data based on varying coefficient models and functional mixed models. We utilize maximum penalized likelihood methods for model estimation, and derive model selection criteria for the evaluation of estimated models. The proposed functional modeling procedure is applied for analyzing the longitudinal gene expression data.

I. Introduction

In longitudinal study, the data are characterized by repeated observations of a response variable over time for each subject and they have possibly different time points among subjects. Under linear regression models for such a longitudinal data, the linear mixed models (Laird and Ware, 1982) are quite practicable and have achieved a number of successful outcomes in medical and social sciences. When, however, the data have a complicated structure or substantial longitudinal heterogeneity between subjects, more appropriate models are required. In order to overcome such issues, we develop two varying coefficient modeling and a functional mixed modeling procedures through the nonlinear regression approach. The essential idea behind the varying coefficient model (VCM; Hastie and Tibshirani, 1993) is that the coefficients of the regression model are represented by time-dependent functions. It enables us to effectively describe the relationship between the predictors and responses which are repeatedly measured. On the other hand, the functional mixed model (FMM; Rice and Wu, 2001) was proposed to estimate population mean functions and subject specific functional random effects for the longitudinal data with large heterogeneity between subjects.

We also introduce a novel VCM, called a mixed effects historical varying coefficient model (ME-HVCM), for evaluating historical dose-response relationship in flexible-dose clinical trials. Our modeling procedures are based on basis expansion techniques, estimating by maximum penalized likelihood methods. Model selection criteria are derived for evaluating the estimated models from the viewpoints of information-theoretic and Bayesian approach. We present an application of our functional mixed modeling procedure to the analysis of a longitudinal gene expression data.

II. Varying Coefficient Models

Suppose we have p sets of predictors \mathbf{X}_k ($k = 1, \dots, p$) and a response \mathbf{Y} varying with time t , and denote i -th observations

at time points $j = 1, \dots, J_i$ as x_{ijk} , and y_{ij} , respectively. Then the VCM has the form

$$y_{ij} = \beta_0(t_{ij}) + x_{ij1}\beta_1(t_{ij}) + \dots + x_{ijp}\beta_p(t_{ij}) + \varepsilon_{ij}, \quad (1)$$

where $\beta_0(\cdot), \beta_1(\cdot), \dots, \beta_p(\cdot)$ are functions of coefficient parameters and ε_{ij} are random noises whose vector $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ_i})'$ is normally distributed with mean vector $\mathbf{0}$ and a variance-covariance matrix $\sigma^2 \mathbf{I}_{J_i}$ with unknown scalar σ^2 . We assume that coefficient functions $\beta_0(\cdot), \beta_1(\cdot), \dots, \beta_p(\cdot)$ are expressed by basis expansions as follows:

$$\beta_k(t_{ij}) = \sum_{m=1}^{M_k} \gamma_{km} \phi_m^{(k)}(t_{ij}) = \boldsymbol{\gamma}'_k \boldsymbol{\phi}^{(k)}(t_{ij}), \quad (2)$$

where $\boldsymbol{\gamma}_k = (\gamma_{k1}, \dots, \gamma_{kM_k})'$ are parameters to be estimated and $\boldsymbol{\phi}^{(k)}(t_{ij}) = (\phi_1^{(k)}(t_{ij}), \dots, \phi_{M_k}^{(k)}(t_{ij}))'$ are basis functions. As for the basis functions, we can apply B -spline basis functions, Gaussian radial basis functions and so on. Using this assumption and denoting $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ_i})'$, $\mathbf{D}_{ik} = \text{diag}(x_{i1k}, \dots, x_{iJ_ik})$ ($k = 1, \dots, p$), $\mathbf{D}_{i0} = \mathbf{I}_{J_i}$ and $\boldsymbol{\Phi}_{ik} = (\phi^{(k)}(t_{i1}), \dots, \phi^{(k)}(t_{iJ_i}))'$, the VCM in (1) can be rewritten as

$$\mathbf{y}_i = \sum_{k=0}^p \mathbf{D}_{ik} \boldsymbol{\Phi}_{ik} \boldsymbol{\gamma}_k + \boldsymbol{\varepsilon}_i, \quad \boldsymbol{\varepsilon}_i \sim N_{J_i}(\mathbf{0}, \sigma^2 \mathbf{I}_{J_i}). \quad (3)$$

Then we have a probability density function

$$f(\mathbf{y}_i | \boldsymbol{\theta}) = \frac{1}{(2\pi\sigma^2)^{J_i/2}} \exp \left\{ -\frac{1}{2\sigma^2} \left(\mathbf{y}_i - \sum_{k=0}^p \mathbf{D}_{ik} \boldsymbol{\Phi}_{ik} \boldsymbol{\gamma}_k \right)' \left(\mathbf{y}_i - \sum_{k=0}^p \mathbf{D}_{ik} \boldsymbol{\Phi}_{ik} \boldsymbol{\gamma}_k \right) \right\}, \quad (4)$$

denoting a vector of unknown parameters by $\boldsymbol{\theta} = \{\boldsymbol{\gamma}_0, \dots, \boldsymbol{\gamma}_p, \sigma^2\}$.

II.A Varying Coefficient Modeling via Regularized Basis Expansions

Unknown parameters involved in the VCM in (3) are estimated by the maximum penalized likelihood method, that is, maximizing the penalized log-likelihood function given by

$$\ell_\lambda(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \frac{n}{2} \sum_{k=1}^p \lambda_k \boldsymbol{\gamma}'_k \boldsymbol{\Omega}_k \boldsymbol{\gamma}_k, \quad (5)$$

where $\ell(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(\mathbf{y}_i | \boldsymbol{\theta})$ and $\boldsymbol{\Omega}_k$ is a positive semi-definite matrix. Since the VCM estimated by the maximum penalized likelihood method depends on tuning parameters, M_k and λ_k s, it is essential to choose appropriate values of them. Konishi and Kitagawa (1996) derived a generalized information criterion (GIC) for evaluating models estimated by various procedures including the maximum penalized likelihood method. Matsui *et al.* (2013) derived the GIC for evaluating the VCM estimated by the maximum penalized likelihood given by

$$\text{GIC} = -2 \log f(\mathbf{Y} | \hat{\boldsymbol{\theta}}) + 2 \text{tr} \left\{ \mathbf{R}^{-1}(\hat{\boldsymbol{\theta}}) \mathbf{Q}(\hat{\boldsymbol{\theta}}) \right\}, \quad (6)$$

where $\mathbf{R}(\hat{\boldsymbol{\theta}})$ and $\mathbf{Q}(\hat{\boldsymbol{\theta}})$ are, respectively,

$$\mathbf{R}(\hat{\boldsymbol{\theta}}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \{\ell_{\lambda}^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (7)$$

$$\mathbf{Q}(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial \{\ell_{\lambda}^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}} \frac{\partial \{\ell^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}'} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \quad (8)$$

Here, $\ell_{\lambda}^{(i)}(\boldsymbol{\theta}) = \ell^{(i)}(\boldsymbol{\theta}) - (1/2) \sum_{k=1}^p \lambda_k \boldsymbol{\gamma}'_k \boldsymbol{\Omega}_k \boldsymbol{\gamma}_k$ with the log-likelihood function of the i -th subject $\ell^{(i)}(\boldsymbol{\theta})$.

Konishi *et al.* (2004) proposed a generalized Bayesian information criterion (GBIC), for evaluating models estimated by maximum penalized likelihood method. Using this result, the GBIC for evaluating the VCM in (4) is proposed by Matsui *et al.* (2013) and given as follows:

$$\begin{aligned} \text{GBIC} = & -2 \log f(Y|\hat{\boldsymbol{\theta}}) + n \sum_{k=1}^p \lambda_k \hat{\boldsymbol{\gamma}}'_k \boldsymbol{\Omega}_k \hat{\boldsymbol{\gamma}}_k + \log |\mathbf{R}(\hat{\boldsymbol{\theta}})| \\ & + \left(\sum_{k=1}^p r_k + 1 \right) \log \left(\frac{n}{2\pi} \right) - \sum_{k=1}^p (M_k - r_k) \log \lambda_k - \sum_{k=1}^p \log |\boldsymbol{\Omega}_k|, \end{aligned} \quad (9)$$

where $r_k = M_k - \text{rank}(\boldsymbol{\Omega}_k)$.

II.B Sparse Varying Coefficient Modeling via Adaptive Elastic Net

We consider maximizing the following penalized log-likelihood function to estimate unknown parameters involved in the VCM (3) by sparse regularization,

$$\ell_{\lambda}(\boldsymbol{\theta}) = \sum_{i=1}^n \log f(\mathbf{y}_i|\boldsymbol{\theta}) - n\lambda \sum_{k=1}^p P_{\alpha}(\|\boldsymbol{\gamma}_k\|), \quad (10)$$

where $P_{\alpha}(\cdot)$ is a penalty function, $\|\cdot\|$ is an L_2 (Euclid) norm and $\lambda > 0$ is a regularization parameter which controls the degree of the penalty. We impose a penalty composed by a sum of penalty functions of L_2 norms of the coefficient vectors $\boldsymbol{\gamma}_k$ instead of its components separately. Then we can shrink all elements of the vector $\boldsymbol{\gamma}_k$ towards exactly zeros when the corresponding predictor seems to be less relevant to the response. For the penalty function P_{α} we use an combination of the adaptive elastic net penalty (Zou and Zhang, 2009) and the group lasso (Yuan and Lin, 2006) given by

$$P_{\alpha}(\|\boldsymbol{\gamma}_k\|) = \frac{1}{2}(1-\alpha)\|\boldsymbol{\gamma}_k\|^2 + \alpha \hat{w}_k \|\boldsymbol{\gamma}_k\|, \quad (11)$$

where $\alpha \in [0, 1]$ tunes the type of the penalty between the ridge ($\alpha = 0$) and the lasso ($\alpha = 1$). Furthermore, $\hat{w}_k > 0$ is an adaptive weight which is given in the form of $\hat{w}_k = (\sqrt{M_k} \|\boldsymbol{\gamma}_k\|)^{-\rho}$ ($\|\boldsymbol{\gamma}_k\| \neq 0$), $= \infty$ ($\|\boldsymbol{\gamma}_k\| = 0$) with a positive constant ρ , where we use $\rho = 1$. We apply the BIC for selecting the regularization parameter λ , tuning parameter α and the number of basis functions M_k . Model selection criterion BIC has the form of

$$\begin{aligned} \text{BIC} = & -2 \sum_{i=1}^n \log f(\mathbf{y}_i|\hat{\boldsymbol{\theta}}) + edf \log n \\ = & -\log(2\pi\hat{\sigma}^2) \sum_{i=1}^n J_i - \sum_{i=1}^n J_i + edf \log n, \end{aligned} \quad (12)$$

where edf is an effective degrees of freedom for the VCM. The effective degrees of freedom for the VCM estimated by the adaptive elastic net regularization is proposed by Matsui and Misumi (2015) and given as follows:

$$edf = \sum_{i=1}^n \sum_{k=1}^p \text{tr} \left\{ \left(1 - \frac{n\hat{\sigma}^2 \alpha \lambda \hat{w}_k}{\|\boldsymbol{\zeta}_k\|} \right)_+ \frac{1}{1 + n(1-\alpha)\lambda} D_{ik} \Phi_{ik} U_k^{-1} (U_k^{-1})^T \Phi_{ik}^T D_{ik} \right\}. \quad (13)$$

III. Functional Mixed Model

Suppose we have n independent repeated measurements $\{y_{ij}, t_{ij}; i = 1, \dots, n, j = 1, \dots, J_i\}$, where y_{ij} denotes a response variable at t_{ij} which intends each subject i and time-point j , that is, we consider the unbalanced data situation. For representing the relationship between these measurements, concurrently with considering the longitudinal heterogeneity between subjects, the FMM via Gaussian Process Regression is defined as

$$y_{ij} = f(t_{ij}) + \gamma_i(t_{ij}) + \varepsilon_{ij}, \quad (14)$$

$$\gamma_i(t) \sim GP(0, r), \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ_i})' \sim N_{J_i}(\mathbf{0}, \boldsymbol{\Omega}_i),$$

where $f(t)$ is a fixed effect or a population mean function, $\gamma_i(t)$ is a random effect function for subject i ($i = 1, \dots, n$), $\boldsymbol{\varepsilon}_i$ is measurement errors, and $\boldsymbol{\Omega}_i$ is a variance-covariance matrix assuming $\boldsymbol{\Omega}_i = \sigma_{\varepsilon}^2 \mathbf{I}_{J_i}$ with a J_i -dimensional identical matrix \mathbf{I}_{J_i} . The GP stands for a Gaussian process with a mean function $m(t) = 0$, and a covariance function $r(s, t)$ which represents the variability between subjects for the times $s, t \in [0, \infty)$. It is assumed that $f(t)$ and $\gamma_1(t), \dots, \gamma_n(t)$ are expressed as $f(t) = \sum_{k=1}^K v_k \varphi_k(t) = \boldsymbol{\varphi}(t)' \boldsymbol{\alpha}$ and $\gamma_i(t) = \sum_{l=1}^L w_l^{(i)} \psi_l(t) = \boldsymbol{\psi}(t)' \mathbf{b}_i$ ($i = 1, \dots, n$), where $\boldsymbol{\varphi}(t) = (\varphi_1(t), \dots, \varphi_K(t))'$ and $\boldsymbol{\psi}(t) = (\psi_1(t), \dots, \psi_L(t))'$ are the basis functions, and $\boldsymbol{\alpha} = (v_1, \dots, v_K)'$ and $\mathbf{b}_i = (w_1^{(i)}, \dots, w_L^{(i)})'$ are their coefficients. For the covariance function $r(s, t)$, the basis expansions for $\gamma_i(t)$ give the Gaussian process

$$r(s, t) = \text{Cov}(\gamma_i(s), \gamma_i(t)) = \boldsymbol{\psi}(s)' \text{Cov}(\mathbf{b}_i) \boldsymbol{\psi}(t), \quad (15)$$

where $\text{Cov}(\mathbf{b}_i)$ is an $L \times L$ variance-covariance matrix, and let $\text{Cov}(\mathbf{b}_i) = \boldsymbol{\Delta}$ for all i . Then, the FMM in (14) can be expressed as the mixed model representation:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad (16)$$

$$\mathbf{b}_i \sim N_L(\mathbf{0}, \boldsymbol{\Delta}), \quad \boldsymbol{\varepsilon}_i \sim N_{J_i}(\mathbf{0}, \sigma_{\varepsilon}^2 \mathbf{I}_{J_i}),$$

where $\mathbf{y}_i = (y(t_{i1}), \dots, y(t_{iJ_i}))'$, $\mathbf{X}_i = (\boldsymbol{\varphi}(t_{i1}), \dots, \boldsymbol{\varphi}(t_{iJ_i}))'$ and $\mathbf{Z}_i = (\boldsymbol{\psi}(t_{i1}), \dots, \boldsymbol{\psi}(t_{iJ_i}))'$. From this derivation, we can estimate the unknown parameters included in the FMM within the framework of standard linear mixed models.

Unknown parameters, such as the coefficient vectors $\boldsymbol{\alpha}$, the variance parameter σ_{ε}^2 and the variance-covariance matrix $\boldsymbol{\Delta}$ are estimated and the random effect vectors $\mathbf{b}_1, \dots, \mathbf{b}_n$ are predicted by the maximum penalized likelihood method. The penalized marginal log-likelihood function in (16) is given by

$$\ell_{m\lambda}(\boldsymbol{\theta}) = \ell_m(\boldsymbol{\theta}) - \frac{1}{2} n \lambda_{\alpha} \boldsymbol{\alpha}' \mathbf{G}_{\alpha} \boldsymbol{\alpha}, \quad (17)$$

where $\ell_m(\boldsymbol{\theta})$ is the marginal log-likelihood function, $\boldsymbol{\theta}$ is the parameter vector with $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \mathbf{b}'_1, \dots, \mathbf{b}'_n, (\text{vech}\tilde{\boldsymbol{\Delta}})')$ with an operator vech that transforms the upper triangular elements of matrix in to a vector, $\tilde{\boldsymbol{\Delta}} = \text{diag}(\boldsymbol{\Delta}, \dots, \boldsymbol{\Delta})$, the second term represents the penalty for the roughness of the population mean function, $\lambda_\alpha (> 0)$ is the smoothing parameter which controls the degree of the penalty and \mathbf{G}_α is a $K \times K$ positive semi-definite matrix. Using the result of Konishi and Kitagawa (1996), Misumi (2014) introduced a marginal GIC (mGIC) for evaluating the FMM estimated by the maximum penalized marginal likelihood is given by

$$\text{mGIC} = -2\ell_m(\hat{\boldsymbol{\theta}}) + 2\text{tr} \left\{ \mathbf{R}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{Q}(\hat{\boldsymbol{\theta}}) \right\}, \quad (18)$$

where $\mathbf{R}(\hat{\boldsymbol{\theta}})$ and $\mathbf{Q}(\hat{\boldsymbol{\theta}})$ are, respectively, given by

$$\mathbf{R}(\hat{\boldsymbol{\theta}}) = -\frac{1}{n} \sum_{i=1}^n \frac{\partial^2 \{\ell_{m\lambda}^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}, \quad (19)$$

$$\mathbf{Q}(\hat{\boldsymbol{\theta}}) = \frac{1}{n} \sum_{i=1}^n \frac{\partial \{\ell_{m\lambda}^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}} \frac{\partial \{\ell_m^{(i)}(\boldsymbol{\theta})\}}{\partial \boldsymbol{\theta}'} \bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \quad (20)$$

Here, $\ell_{m\lambda}^{(i)}(\boldsymbol{\theta})$ and $\ell_m^{(i)}(\boldsymbol{\theta})$ represents the penalized marginal log-likelihood function and marginal log-likelihood function of the i th subject, respectively. Using the result of Konishi *et al.* (2004), Misumi (2014) introduced a marginal GBIC (mGBIC) for evaluating the FMM estimated by the maximum penalized marginal likelihood is given by

$$\begin{aligned} \text{mGBIC} = & -2\ell_m(\hat{\boldsymbol{\theta}}) + n\lambda_\alpha \hat{\boldsymbol{\alpha}}' \mathbf{G}_\alpha \hat{\boldsymbol{\alpha}} + (r_\alpha + 1) \log \left(\frac{n}{2\pi} \right) \\ & - (K - r_\alpha) \log \lambda_\alpha - \log |\mathbf{G}_\alpha| + \log |\mathbf{R}(\hat{\boldsymbol{\theta}})|, \end{aligned} \quad (21)$$

where $r_\alpha = K - \text{rank}(\mathbf{G}_\alpha)$. The matrix $\mathbf{R}(\hat{\boldsymbol{\theta}})$ is the same as that of the mGIC.

IV. Mixed Effects Historical Varying Coefficient Model

Suppose we have n independent observations $\{(y_{ij}, d_{ij}, t_{ij}); i = 1, \dots, n, j = 1, \dots, J_i\}$, where y_{ij} , d_{ij} and t_{ij} are a response and dose-level at design time-point for each subject i and time-point j , respectively. We assume a design has equidistant time-points, and an unequal number of observations per subject. For representing the historical dose-level information and longitudinal heterogeneity between subjects, the ME-HVCM (Misumi and Konishi, 2014) is defined as

$$\begin{aligned} y_{ij} = & \beta_0(t_{ij}) + \sum_{p=1}^P \beta_p(t_{ij}) d_{ij-(p-1)} + \gamma_i(t_{ij}) + \varepsilon_{ij}, \quad (22) \\ \gamma_i(t) \sim & GP(0, r), \quad \boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, \dots, \varepsilon_{iJ_i})' \sim N_{J_i}(\mathbf{0}, \boldsymbol{\Omega}_i), \end{aligned}$$

where $\beta_0(t)$ is an intercept function and $\beta_1(t), \dots, \beta_P(t)$ are historical varying coefficient functions which represent the effectiveness of the dose-level in p visit before. $\boldsymbol{\Omega}_i$ is a $J_i \times J_i$ variance-covariance matrix, and $\gamma_i(t)$ is a random effect function for subject i ($i = 1, \dots, n$). Here, $\beta_p(t)$ and $\gamma_i(t)$ are assumed to be expressed as basis expansions, $\beta_p(t) = \sum_{k=1}^{K_p} v_{pk} \varphi_{pk}(t) = \boldsymbol{\varphi}_p(t)' \boldsymbol{\alpha}_p$ ($p = 0, \dots, P$), $\gamma_i(t) = \sum_{l=1}^L w_{\gamma_l}^{(i)} \psi_l(t) = \boldsymbol{\psi}(t)' \mathbf{b}_i$ ($i = 1, \dots, n$), where $\boldsymbol{\varphi}_p(t) = (\varphi_{p1}(t), \dots, \varphi_{pK_p}(t))'$ and $\boldsymbol{\psi}(t) = (\psi_1(t), \dots, \psi_L(t))'$ are basis functions, and $\boldsymbol{\alpha}_p = (v_{p1}, \dots, v_{pK_p})'$ and $\mathbf{b}_i = (w_{\gamma_1}^{(i)}, \dots, w_{\gamma_L}^{(i)})'$ are coefficients of basis functions. We assume that the covariance function $r(s, t)$ is the same form as in (15). Then (22) can

be expressed as

$$\begin{aligned} \mathbf{y}_i = & \mathbf{X}_i \boldsymbol{\alpha} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\varepsilon}_i, \\ \mathbf{b}_i \sim & N_L(\mathbf{0}, \boldsymbol{\Delta}), \quad \boldsymbol{\varepsilon}_i \sim N_{J_i}(\mathbf{0}, \boldsymbol{\Omega}_i), \end{aligned} \quad (23)$$

where $\mathbf{y}_i = (y_{i1}, \dots, y_{iJ_i})'$, $\mathbf{X}_i = (\mathbf{x}'_{i1}, \dots, \mathbf{x}'_{iJ_i})'$, $\mathbf{Z}_i = (\mathbf{z}'_{i1}, \dots, \mathbf{z}'_{iJ_i})'$, $\mathbf{x}_{ij} = (\varphi_0(t_{ij}), d_{ij} \varphi_1(t_{ij}), \dots, d_{ij-(p-1)} \varphi_p(t_{ij}))'$, $\mathbf{z}_{ij} = \boldsymbol{\psi}(t_{ij})$ and $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_0, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_P)'$.

We estimate the ME-HVCM by the maximum penalized likelihood method. It follows from (23) that the penalized log-likelihood function is

$$\ell_\lambda(\boldsymbol{\theta}) = \ell(\boldsymbol{\theta}) - \frac{1}{2} \sum_{p=0}^P \lambda_{\alpha p} \boldsymbol{\alpha}'_p \mathbf{G}_\alpha \boldsymbol{\alpha}_p - \frac{1}{2} \sum_{i=1}^n \lambda_b \mathbf{b}'_i \mathbf{G}_b \mathbf{b}_i, \quad (24)$$

where $\ell(\boldsymbol{\theta})$ is the log-likelihood function, $\boldsymbol{\theta}$ is the vector for $\boldsymbol{\theta} = (\boldsymbol{\alpha}', \mathbf{b}'_1, \dots, \mathbf{b}'_n, (\text{vech}\tilde{\boldsymbol{\Delta}})')$, $\tilde{\boldsymbol{\Delta}} = \text{diag}(\boldsymbol{\Delta}, \dots, \boldsymbol{\Delta})$, $\lambda_{\alpha 0}, \dots, \lambda_{\alpha P}, \lambda_b (> 0)$ are smoothing parameters which control the degree of the penalty, and \mathbf{G}_α and \mathbf{G}_b are positive semi-definite matrices. The estimated ME-HVCM depends on predefined smoothing parameters $\lambda_{\alpha 0}, \dots, \lambda_{\alpha P}, \lambda_b$, the number of basis functions K_0, \dots, K_P and L , and the total number of historical varying coefficients P . In order to choose the values of these tuning parameters appropriately, we present the AIC and BIC:

$$\text{AIC} = -2\ell(\boldsymbol{\theta}) + 2\widehat{edf}, \quad \text{BIC} = -2\ell(\boldsymbol{\theta}) + \widehat{edf} \log n. \quad (25)$$

Here, \widehat{edf} is the effective degrees of freedom for ME-HVCM given by $\widehat{edf} = \sum_{i=1}^n \text{tr} \left\{ \mathbf{H}_i + \mathbf{Z}_i \hat{\boldsymbol{\Delta}}_b \mathbf{Z}'_i \hat{\mathbf{V}}_i^{-1} (\mathbf{I}_{J_i} - \mathbf{H}_i) \right\}$, where $\mathbf{H}_i = \mathbf{X}_i \left(\sum_{i=1}^n \mathbf{X}'_i \hat{\mathbf{V}}_i^{-1} \mathbf{X}_i + \tilde{\mathbf{G}}_\alpha \right)^{-1} \mathbf{X}'_i \hat{\mathbf{V}}_i^{-1}$ and $\hat{\mathbf{V}}_i = \mathbf{Z}_i \hat{\boldsymbol{\Delta}}_b \mathbf{Z}'_i + \hat{\sigma}_\varepsilon^2 \mathbf{I}_{J_i}$.

V. Application Functional Mixed Model to Longitudinal Gene Expression Data

We apply our proposed modeling procedures for the FMM to longitudinal gene expression data (Spellman *et al.* 1998). They identified 800 genes as cell cycle related genes from all 6,178 genes in the yeast genome measured by cDNA microarrays, and also grouped these genes into five classes, G1, G2, M, M/G1 and S, by considering peaks in the expression patterns. We focused on the repeatedly measurement “ α factor-based synchronization data” at 7 min intervals for 119 mins with a maximum total of 18 time-points in our analysis. Furthermore, we selected 791 genes which have two and more time-points, and in consequence, the number of genes in each class were $N^{G1} = 297$, $N^{G2} = 119$, $N^M = 193$, $N^{M/G1} = 113$ and $N^S = 69$.

Table 1 shows the comparison of mean squared errors for repeated observations among our proposed and conventional model selection criteria, such as the marginal AIC, BIC, and generalized cross validation. This result indicates that our proposed mGIC and mGBIC provide better fitting performance than conventional marginal model selection criteria. The raw data, predicted curves with functional random effects for any measurements, and estimated population mean function with 95% confidence interval for the class G1 are shown in Figure 1, where the tuning parameters selected by the mGIC are used. These results show that the predicted curves are well fitted for any measurements, and suggest that unknown population mean functions are also well estimated.

Table 1: Comparisons of mean squared errors (standard deviation) for repeated observations among model selection criteria.

Class	mGIC	mGBIC	mAIC	mBIC	mGCV
G1 [$\times 10^{-2}(\times 10^{-2})$]	1.81(1.54)	1.81(1.55)	2.69(3.12)	2.69(3.12)	2.69(3.12)
G2 [$\times 10^{-2}(\times 10^{-2})$]	1.76(1.44)	1.76(1.57)	2.56(2.66)	2.56(2.66)	2.52(2.63)
M [$\times 10^{-2}(\times 10^{-2})$]	1.93(1.48)	1.93(1.48)	2.63(2.21)	2.63(2.21)	2.63(2.21)
M/G1 [$\times 10^{-2}(\times 10^{-2})$]	2.04(1.44)	2.11(1.46)	3.16(2.71)	3.19(2.70)	3.16(2.71)
S [$\times 10^{-2}(\times 10^{-2})$]	4.21(3.99)	4.45(3.93)	5.12(4.47)	5.24(4.52)	4.66(4.01)

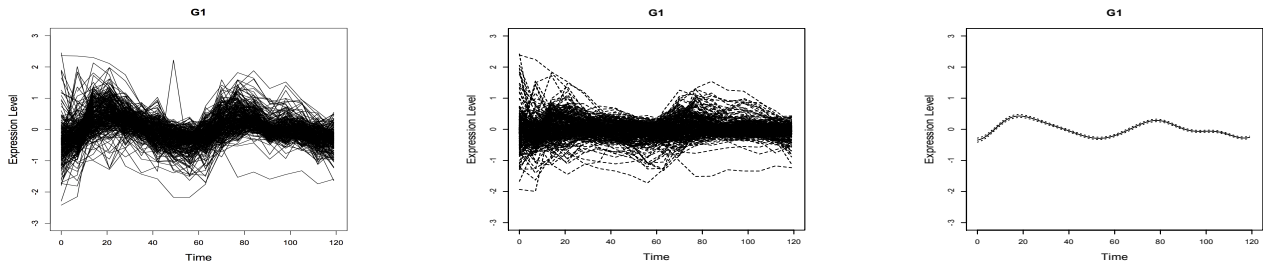


Figure 1: [Left] Raw data, [Center] Predicted curves with functional random effects, [Right] Estimated population mean functions (solid lines) with 95% confidence interval (dashed lines).

VI. Summary and Discussion

We have introduced several nonlinear regression modeling strategies for longitudinal data, and found through Monte Carlo experiments and real data analyses that our modeling procedures are useful for analyzing data with complex structure. In the future, the extension of the proposed procedures is required to the non-normal longitudinal data. The generalized linear mixed model and the generalized Gaussian process regression are helpful for the modeling of non-normal longitudinal data. In addition further work remains to be done to construct classification rules for longitudinal data.

References

- Hastie, T. and Tibshirani, R. (1993) Varying-Coefficient Models. *Journal of the Royal Statistical Society, Series B* **55**, 757–796.
- Konishi, S. and Kitagawa, G. (1996) Generalized information criteria in model selection. *Biometrika* **83**, 875–890.
- Konishi, S., Ando, T. and Imoto, S. (2004) Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika* **91**, 27–43.
- Laird, N. M. and Ware, J. H. (1982) Random-effects models for longitudinal data. *Biometrics* **38**, 963–974.
- Matsui, H., Misumi, T. and Kawano, S. (2013) Model selection criteria for the varying-coefficient modelling via regularized basis expansions. *Journal of Statistical Computation and Simulation* **84**, 2156–2165.
- Matsui, H. and Misumi, T. (2015) Variable selection for varying coefficient models with the sparse regularization. *Computational Statistics* **30**, 43–55.
- Misumi, T. (2014) Model selection for functional mixed model via Gaussian process regression. *Bulletin of Informatics and Cybernetics* **46**, 23–35.
- Misumi, T. and Konishi, S. (2014) Mixed effects historical varying coefficient model for evaluating dose-response in flexible dose trials. (Submitted and under second revision to *Journal of the Royal Statistical Society, Series C*).
- Rice, J. A. and Wu, C. O. (2001) Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253–259.
- Spellman, P. T., Sherlock, G., Zhang, M. Q., Iyer, V. R., Anders, K., Eisen, M. B., Brown, P. O., Bostein, D. and Futcher, B. (1998) Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molecular Biology of the Cell* **9**, 3273–3297.
- Yuan, M. and Lin, Y. (2006) Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society, Series B* **68**, 49–67.
- Zou, H. and Zhang, H. (2009) On the adaptive elastic-net with a diverging number of parameters. *Annals of Statistics* **37**, 1733–1751.