ESTIMATION OF VARYING COEFFICIENTS FOR A GROWTH CURVE MODEL

Kenichi Satoh
Research Institute for Radiation Biology and Medicine
Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima 734-8553.
Hirokazu Yanagihara
Graduate School of Science, Hiroshima University
1-3-1 Kagamiyama, Higashi-Hiroshima 739-8626, JAPAN.

SYNOPTIC ABSTRACT

In this paper, a new approach of modelling growth curves is developed which uses time-varying coefficients. Since the mean structure of the growth curve model has many unknown parameters depending on both covariates and time trend designs, it can be difficult to understand and interpret. Using varying coefficient functions, the effects of covariates can be evaluated and visualized more easily. The asymptotic functional confidence intervals are derived theoretically and a procedure is proposed to test whether the effects of covariates are significant.

Key words and Phrases: Confidence interval; Growth curve model; Hypothesis testing; Longitudinal data; Repeated measurements; Varying coefficient.

1. INTRODUCTION.

Let $y_i(t)$ $(i = 1, \ldots, n)$ be an observation of the ith subject at time t, and $\varepsilon_i(t)$ $(i = 1, \ldots, n)$ be a error term with $E[\varepsilon_i(t)] = 0$. We suppose that observations for different subjects are independent, and $\boldsymbol{x}(t)$ is a q-dimensional known design vector, where n is the sample size. We write $\boldsymbol{y}_i = (y_i(t_{i1}), \ldots, y_i(t_{ip_i}))'$, $\boldsymbol{\varepsilon}_i = (\varepsilon_i(t_{i1}), \ldots, \varepsilon_i(t_{ip_i}))'$ and $\boldsymbol{X}_i = (\boldsymbol{x}(t_{i1}), \ldots, \boldsymbol{x}(t_{ip_i}))'$. Suppose that $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are independent and let the covariance matrix of $\boldsymbol{\varepsilon}_i$ be denoted by $\mathrm{Cov}[\boldsymbol{\varepsilon_i}] = \boldsymbol{\Sigma}_i$. Since our main interest is in the mean structure of observations, $\boldsymbol{\Sigma}_i$ is treated as a nuisance parameter and the structure is not specified here. Then, the traditional growth curve model (GCM) is typically expressed in two stages. The first stage consists of the "within-individual" regression model,

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_i, \quad (i = 1, \ldots, n), \tag{1}$$

where $\boldsymbol{\alpha}_i$ is a q-dimensional unknown parameter vector. It is well known that the GCM is a useful model for describing how growth varies over time. For example, if we fit a polynomial curve of degree $q-1$ to the trend over time, the design vector is taken to be $\boldsymbol{x}(t) = (1, t, \ldots, t^{q-1})'$. If a more complicated trend is required, nonparametric curves can be applied, taking basis functions to be B-splines, or utilizing a Gaussian kernel, e.g., Satoh, Yanagihara and Ohtaki (2003), Ruppert, Wand and Carroll (2003). Note that when $p_1 = \cdots = p_n = p$ and $t_{1j} = \cdots = t_{nj} = t_j$ for $j = 1, \ldots, p$, we have a so-called balanced design; otherwise we have an unbalanced design.

The second stage consists of the "between-individual" multivariate linear model, i.e., the variation between individual subjects have been attempted in terms of known covariates a_{ij} $(i = 1, \ldots, n; j = 1, \ldots, k)$. This means that $\boldsymbol{\alpha}_i$ is regressed by $\boldsymbol{a}_i = (a_{i1}, \ldots, a_{ik})'$, i.e., $\boldsymbol{\alpha}_i = \boldsymbol{\Theta}' \boldsymbol{a}_i$, where $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k)'$ is a $k \times q$ unknown parameter matrix (see Rao (1965), Potthoff and Roy (1964), Laird and Ware (1982), Vonesh and Carter (1987)). Then, the mean of observation can be written as $\boldsymbol{X}_i \boldsymbol{\Theta}' \boldsymbol{a}_i$, which describes between within and between individual variation.

However, it can be difficult to understand the effects of each covariate, because $\boldsymbol{\Theta}$ has many unknown parameters, the interpretations of which depend on both design vectors $\boldsymbol{x}(t)$ and $\boldsymbol{a}_i$. Usually our main interest is in regression coefficients of not $\boldsymbol{X}_i$ in (1), but covariates $\boldsymbol{a}_i$. Therefore, we consider a new approach of modelling the growth curve model in order to determine the effect of each covariate. The GCM in (1) can be rewritten as a model for ith subject at time t,

$$y_i(t) = \boldsymbol{a}_i'\boldsymbol{\Theta}\boldsymbol{x}(t) + \varepsilon_i(t), \quad (t = t_{i1}, \ldots, t_{ip_i}; i = 1, \ldots, n). \tag{2}$$

In the ordinary GCM, $\boldsymbol{\Theta}$ and $\boldsymbol{a}_i$ are described together as $\boldsymbol{\alpha}_i$. On the other hand, we discuss here the possibility of pairing $\boldsymbol{\Theta}$ and $\boldsymbol{x}(t)$ as the regression coefficient of $\boldsymbol{a}_i$.

A varying coefficients model (see, West, Harison and Migon (1985), Hastie and Tibshirani (1993)) is known to be an important model for analyzing longitudinal data. By defining $\boldsymbol{\beta}(t) = (\beta_1(t), \ldots, \beta_k(t))' = \boldsymbol{\Theta}\boldsymbol{x}(t)$, we can regard the growth curve model in (2) as a varying coefficients model. The varying coefficient functions are semiparametric curves, the shapes of which are controlled by $\boldsymbol{x}(t)$. The mean structure of the observations can be rewritten as

$$E[y_i(t)] = \boldsymbol{a}_i'\boldsymbol{\Theta}\boldsymbol{x}(t) = \boldsymbol{a}_i'\boldsymbol{\beta}(t) = \sum_{j=1}^{k} a_{ij}\beta_j(t). \tag{3}$$

By considering $\beta_j(t)$, we can easily explain the effects of covariates. Thus, meaningful interval estimation and hypothesis testing can be carried out, e.g., a functional confidence interval for $\beta_j(t)$ can be determined and a hypothesis test of $\beta_j(t) \equiv 0$ can be carried out.

Estimators of varying coefficients are usually obtained by kernel smoothing method (see, Hoover, Rice, Wu and Yang (1998), Guo (2002), Satoh and Ohtaki (2006)). The smoothing method is an essensially linear regression using data around fixed time point and it is difficult to construct a confidence interval as a function of time. In fact, they showed the pointwise confidence intervals.

This paper is organized as follows: In Section 2 an estimate for the varying coefficient is discussed and its asymptotic confidence interval as a function of time is derived theoretically. Moreover, we propose a procedure to test for the significance of the effects of a covariate. Although $\mathbf{\Sigma}_i$ is a nuisance parameter, it needs to be estimated for obtaining the estimators of the varying coefficients. Therefore we illustrated two types of growth curve models and those specific structures of $\mathbf{\Sigma}_i$ in Section 3. Two numerical examples are presented in Section 4. Section 5 contains a discussion and our conclusions.

2. ESTIMATION OF VARYING COEFFICIENTS AND THEIR EVALUATION.

Now we are interested in estimating $\boldsymbol{\beta}(t)$ in (3). In fact, this is equivalent to estimating $\mathbf{\Theta}$ in (2). The estimator $\hat{\mathbf{\Theta}} = (\hat{\boldsymbol{\theta}}_1, \ldots, \hat{\boldsymbol{\theta}}_k)'$ for $\mathbf{\Theta}$ can be obtained as explicit forms from the results by Potthoff and Roy (1964) and Rao (1965) for the balanced case or Vonesh and Carter (1987) for the unbalanced case. Note that they also showed the asymptotic normality of those estimators. Scince the growth curve model can be considered as a special model of mixed effect model developed by Laird and Ware (1982), each elements of $\mathbf{\Theta}$ can be estimated using restricted maximum likelihood method (see, Patterson and Thompson (1971)), based on the mean having interactions between covariates and time related variables, $E[y_i(t)] = \sum_{l=1}^{k} \sum_{m=1}^{q} \theta_{lm}\{a_{il}x_m(t)\}$ here $a_{il}x_m(t)$'s are kq known covariates. Thus, we can also have the estimator under the condition that some elements of $\mathbf{\Theta}$ are zero using mixed effect model. The restriction on the unknown parameters can be considered as a variable selection problem (see, Satoh, Kobayashi and Fujikoshi (1997)).

Let the asymptotic variance of $\hat{\boldsymbol{\theta}}_j$ $(j = 1, \ldots, k)$ be denoted by $\mathbf{\Omega}_j$, where $\mathbf{\Omega}_j = O(n^{-1})$. For both the balanced and unbalanced cases, we can obtain the estimators of the regression parameters which satisfy $\sqrt{n}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) \xrightarrow{d} N_q(\mathbf{0}, n\mathbf{\Omega}_j)$ $(n \to \infty)$ and an estimator $\hat{\mathbf{\Omega}}_j$ can also be derived for $\mathbf{\Omega}_j$. The asymptotic normality of the estimator has been discussed by Vonesh and Carter (1987), Srivastava and von Rosen (1999). Let $\omega_j(t) = \boldsymbol{x}(t)'\mathbf{\Omega}_j\boldsymbol{x}(t)$ and $\hat{\omega}_j(t) = \boldsymbol{x}(t)'\hat{\mathbf{\Omega}}_j\boldsymbol{x}(t)$. From the result of Rao (1973, p.60), for any time point $t \in \mathbb{R}$, we have

$$\frac{\{\hat{\beta}_j(t) - \beta_j(t)\}^2}{\hat{\omega}_j(t)} = \frac{\{(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)'\boldsymbol{x}(t)\}^2}{\boldsymbol{x}(t)'\hat{\mathbf{\Omega}}_j\boldsymbol{x}(t)} \leq (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)'\hat{\mathbf{\Omega}}_j^{-1}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j). \quad (4)$$

Note that $n\hat{\mathbf{\Omega}}_j$ is a consistent estimator of $n\mathbf{\Omega}_j$. Therefore, the right side of (3) converges to χ_q^2 when n tends to infinity because $\hat{\mathbf{\Omega}}_j^{-1/2}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) \xrightarrow{d} N_q(\mathbf{0}, \boldsymbol{I}_q)$ $(n \to \infty)$. Hence, we obtain the following equation:

$$\max_{t\in\mathbb{R}} \frac{\{\hat{\beta}_j(t) - \beta_j(t)\}^2}{\hat{\omega}_j(t)} \leq (\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j)'\hat{\mathbf{\Omega}}_j^{-1}(\hat{\boldsymbol{\theta}}_j - \boldsymbol{\theta}_j) \xrightarrow{d} \chi_q^2. \tag{5}$$

Since the right side of (3) does not depend on t, we obtain a functional confidence interval for $\beta_j(t)$ as follows: Let $c_{q,\alpha}$ be the upper $\alpha \times 100$ percentage point of χ_q^2, i.e., $\Pr(\chi_q^2 \geq c_{q,\alpha}) = \alpha$. A confidence interval for the varying coefficient $\beta_j(t)$ for $j = 1, \ldots, k$ can be obtained by

$$\mathcal{I}_{j,\alpha}(t) = \left[\hat{\beta}_j(t) - \sqrt{\hat{\omega}_j(t)c_{q,\alpha}}, \quad \hat{\beta}_j(t) + \sqrt{\hat{\omega}_j(t)c_{q,\alpha}}\right]. \tag{6}$$

For any time point t, the coverage probability of the interval $\mathcal{I}_{j,\alpha}(t)$ is asymptotically given by $\Pr(\beta_j(t) \in \mathcal{I}_{\alpha,j}(t) : {}^{\forall}t \in \mathbb{R}) \geq 1 - \alpha$. Note that the confidence interval at a fixed time point t can also be obtained from $\{\hat{\beta}_j(t) - \beta_j(t)\}^2/\omega_j(t) \xrightarrow{d} \chi_1^2$, and this interval is narrower than the functional confidence interval given in (6).

Directly, we have a test statistic for the null hypothesis that a varying coefficient is zero for any time point, i.e.,

$$H_0: \ \beta_j(t) = 0 \quad ({}^{\forall}t \in \mathbb{R}). \tag{7}$$

This implies the corresponding covariate has no effect on observations. Let us define $T_j(t) = \hat{\beta}_j(t)^2/\omega_j(t)$ $(j = 1, \ldots, k)$. In order to test the hypothesis H_0 in (7), we have to consider the probability $\Pr(T_j(t) \leq x : {}^{\forall}t \in \mathbb{R})$. Note that

$$\Pr(T_j(t) \leq x : {}^{\forall}t \in \mathbb{R}) = \Pr\left(\max_{t\in\mathbb{R}} T_j(t) \leq x\right).$$

Now we define $T_j = \hat{\boldsymbol{\theta}}_j'\hat{\mathbf{\Omega}}_j^{-1}\hat{\boldsymbol{\theta}}_j$. The result in (4) implies that $\max_{t\in\mathbb{R}} T_j(t) \leq T_j$. Hence, T_j is the test statistic for the null hypothesis H_0. Recall that T_j is asymptotically null distributed as χ_q^2. Therefore, the null hypothesis H_0 is rejected when $T_j > c_{q,\alpha}$, or the p-value can be calculated from the chi-square approximation $\Pr(\chi_q^2 > T_j)$.

3. COVARIANCE STRUCTURES OF ERROR DISTRIBUTIONS

Varying coefficients can be estimated by applying the methodology discussed in Section 2, but the methodology requires the estimators of $\boldsymbol{\Theta}$ and those asymptotic covariance matrices in advance. The estimator, $\hat{\boldsymbol{\Theta}}$ usually depends on the covariance matrix of the error distribution, $\boldsymbol{\Sigma}_i$, which is regarded as a nuisance parameter in Section 1 and Section 2. Therefore we can not obtain the estimators of varying coefficients without giving the structure of $\boldsymbol{\Sigma}_i$ in actual situations. In this section, we illustrate two covariance structures of the error distributions, and describe $\hat{\boldsymbol{\Theta}}$ and estimated asymptotic covariance matrices in closed form.

Since the covariance structure of the error distribution is often decided from the design of observed time points or prior informations, two types of growth curve models are presented here, which correspond to balanced and unbalanced designs. Several covariance structures can be considered under the balanced design. As one of the most general and popular structures, an unstructured covariance is given for example. For unbalanced design, we express a random effect model of which regression coefficients have random effects and the covariance structure is related to the individual design matrix $\boldsymbol{X}_i$.

Balanced Case: Note that $\boldsymbol{X}_i$ can be rewritten as $\boldsymbol{X}$ in the balanced case. The GCM proposed by Potthoff and Roy (1964) is defined by

$$\boldsymbol{y}_i = \boldsymbol{X}\boldsymbol{\Theta}'\boldsymbol{a}_i + \boldsymbol{\varepsilon}_i, \quad (i = 1, \ldots, n),$$

where $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n \sim i.i.d.\ N_p(\boldsymbol{0}, \boldsymbol{\Sigma})$. Note that $\boldsymbol{\Sigma}$ is unstructured and it has $p(p+1)/2$ parameters. Let $\boldsymbol{Y} = (\boldsymbol{y}_1, \cdots, \boldsymbol{y}_n)'$ and $\boldsymbol{A} = (\boldsymbol{a}_1, \cdots, \boldsymbol{a}_n)'$, then the estimators of $\boldsymbol{\Theta}$ and $\boldsymbol{\Sigma}$ are obtained by

$$\hat{\boldsymbol{\Theta}} = (\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{Y}\boldsymbol{S}^{-1}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X})^{-1}, \qquad \hat{\boldsymbol{\Sigma}} = \frac{1}{n}(\boldsymbol{Y} - \boldsymbol{A}\hat{\boldsymbol{\Theta}}X)'(\boldsymbol{Y} - \boldsymbol{A}\hat{\boldsymbol{\Theta}}X),$$

respectively, where $\boldsymbol{S} = \boldsymbol{Y}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{Y}/m$, $m = n - k$ and $\boldsymbol{P}_{\boldsymbol{A}} = \boldsymbol{A}(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{A}'$. The covariance matrix of $\hat{\boldsymbol{\theta}}_j$ is estimated by

$$\hat{\boldsymbol{\Omega}}_j = \frac{m(m-1)}{\{m-(p-q)-1\}\{m-(p-q)\}}\boldsymbol{e}_j'(\boldsymbol{A}'\boldsymbol{A})^{-1}\boldsymbol{e}_j \otimes (\boldsymbol{X}'\boldsymbol{S}^{-1}\boldsymbol{X})^{-1},$$

where $\boldsymbol{e}_j$ is a k-dimensional basis vector with jth entry equal to 1 and the others equal to 0.

Unbalanced Case: The GCM proposed by Vonesh and Carter (1987) is defined by

$$\boldsymbol{y}_i = \boldsymbol{X}_i \boldsymbol{\Theta}' \boldsymbol{a}_i + \boldsymbol{\varepsilon}_i, \quad (i = 1, \ldots, n),$$

where $\boldsymbol{\varepsilon}_1, \ldots, \boldsymbol{\varepsilon}_n$ are independent random vectors, $\boldsymbol{\varepsilon}_i \sim N_{p_i}(\boldsymbol{0}, \boldsymbol{\Sigma}_i)$ and $\boldsymbol{\Sigma}_i = \sigma^2 \boldsymbol{I}_{p_i} + \boldsymbol{X}_i \boldsymbol{\Delta} \boldsymbol{X}_i'$. Note that the covariance matrix of individual error distribution depends on the design matrix $\boldsymbol{X}_i$ and it is quite different than the unstructured covariance matrix introduced in balanced case. Let $\boldsymbol{u}_i = (\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}\boldsymbol{X}_i'\boldsymbol{y}_i$, $\boldsymbol{U} = (\boldsymbol{u}_1, \cdots, \boldsymbol{u}_n)'$, $\boldsymbol{W} = \boldsymbol{U}'(\boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{A}})\boldsymbol{U}$ and $s^2 = \sum_{i=1}^n \boldsymbol{y}_i'$ $(\boldsymbol{I}_{p_i} - \boldsymbol{P}_{\boldsymbol{X}_i})\boldsymbol{y}$, then we have the variance of the observations and covariance matrix of the random effect,

$$\hat{\sigma}^2 = s^2/(\sum_{i=1}^n p_i - nq), \qquad \hat{\boldsymbol{\Delta}} = \boldsymbol{W}/m - \hat{\sigma}^2 \sum_{i=1}^n r_{i,i}(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}/m,$$

respectively, here $\boldsymbol{R} = (r_{i,j}) = \boldsymbol{I}_n - \boldsymbol{P}_{\boldsymbol{A}}$. By using $\hat{\boldsymbol{\Gamma}}_i = \hat{\boldsymbol{\Delta}} + \hat{\sigma}^2(\boldsymbol{X}_i'\boldsymbol{X}_i)^{-1}$, estimators of $\boldsymbol{\Theta}'$ and the estimated covariance matrix of $\hat{\boldsymbol{\theta}}_j$ are given by

$$\mathrm{vec}(\hat{\boldsymbol{\Theta}}') = \sum_{i=1}^n \left(\boldsymbol{a}_i \boldsymbol{a}_i' \otimes \hat{\boldsymbol{\Gamma}}_i^{-1} \right)^{-1} \left(\sum_{i=1}^n \boldsymbol{a}_i \otimes \hat{\boldsymbol{\Gamma}}_i^{-1} \boldsymbol{u}_i \right),$$

and

$$\hat{\boldsymbol{\Omega}}_j = (\boldsymbol{e}_j' \otimes \boldsymbol{I}_q) \left\{ \sum_{i=1}^n \boldsymbol{a}_i \boldsymbol{a}_i' \otimes \hat{\boldsymbol{\Gamma}}_i^{-1} \right\}^{-1} (\boldsymbol{e}_j \otimes \boldsymbol{I}_q),$$

respectively, where $\mathrm{vec}(\hat{\boldsymbol{\Theta}}') = (\hat{\boldsymbol{\theta}}_1', \cdots, \hat{\boldsymbol{\theta}}_k')'$.

4. NUMERICAL EXAMPLES.

4.1. GIRLS AND BOYS DATA.

We apply our estimation method to Table 1 considered by Potthoff and Roy (1964), consisting of measurements of the distance (mm) from the center of the pituitary to the pteryomaxillary fissure for 11 girls and 16 boys at 4 different ages (8, 10, 12, 14 years old).

TABLE 1. Girls and boys data.

Girls					Boys				
ID	8	10	12	14	ID	8	10	12	14
1	21	20	21.5	23	1	26	25	29	31
2	21	21.5	24	25.5	2	21.5	22.5	23	26.5
3	20.5	24	24.5	26	3	23	22.5	24	27.5
4	23.5	24.5	25	26.5	4	25.5	27.5	26.5	27
5	21.5	23	22.5	23.5	5	20	23.5	22.5	26
6	20	21	21	22.5	6	24.5	25.5	27	28.5
7	21.5	22.5	23	25	7	22	22	24.5	26.5
8	23	23	23.5	24	8	24	21.5	24.5	25.5
9	20	21	22	21.5	9	23	20.5	31	26
10	16.5	19	19	19.5	10	27.5	28	31	31.5
11	24.5	25	28	28	11	23	23	23.5	25
					12	21.5	23.5	24	28
					13	17.5	24.5	26	29.5
					14	22.5	25.5	25.5	26
					15	23	24.5	26	30
					16	22	21.5	23.5	25

Here we fit the model (1) to the data by letting $\boldsymbol{a}_i = (1,0)'$ for the girls and $(1,1)'$ for the boys and taking $\boldsymbol{x}(t) = (1,t,t^2)'$ for $t \in \{8,10,12,14\}$. The first covariate expresses the distance for girls and the second covariate expresses the sex effect, which is determined by the additional distance for the boys beyond that of the girls. Since the design is balanced with respect to the time points, we have estimators of $\boldsymbol{\Theta}$, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ from the result of Potthoff and Roy (1964), which are given by

$$\hat{\boldsymbol{\Theta}} = \begin{pmatrix} 17.096 & 0.537 & -0.003 \\ 4.946 & -0.852 & 0.053 \end{pmatrix},$$
$$\hat{\boldsymbol{\Omega}}_1 = \begin{pmatrix} 26.065 & -4.634 & 0.199 \\ -4.634 & 0.843 & -0.037 \\ 0.199 & -0.037 & 0.002 \end{pmatrix},$$
$$\hat{\boldsymbol{\Omega}}_2 = \begin{pmatrix} 43.985 & -7.820 & 0.335 \\ -7.820 & 1.423 & -0.062 \\ 0.335 & -0.062 & 0.003 \end{pmatrix},$$

respectively. Here we focus on the varying coefficient function for the sex effect. Estimated varying coefficient curve $\hat{\beta}_2(t)$ was shown in Figure 1. The broken lines represent 95% confidence intervals derived using Theorem 1. (a) The jointed gray lines show individual observations; the solid lines are fitted curves. (b) Sex effects: $\boldsymbol{\beta}_2(t)$ versus time. The significance level was calculated as $p = 0.002$.

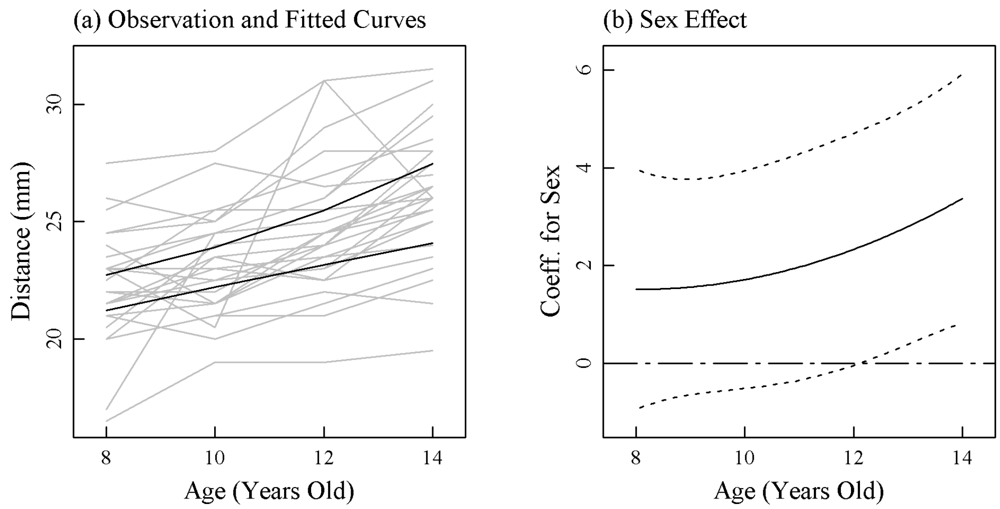

FIGURE 1. Varying coefficient for male and female data.

4.2. MICE BODY WEIGHT DATA.

Here we apply the proposed method to data about the body weights of mice, which was investigated by Watanabe, Takahashi, Lee, Ohtaki, Roy, Ando, Yamada, Gotoh, Kurisu, Fujimoto, Satow and Ito (1996). One of our main concerns in these experiments was to determine whether neutron-induced genetic damage in parental germline cells can affect the F_1 offspring growth or their body weight, which is an important index for measuring the health of mice. The parental mice received a single whole-body exposure to ^{252}Cf neutrons at doses of 0, 50, 100, or 200 cGy. There were 1,232 body-weight observations of 124 mice. Body weights were measured at the same calendar time, but the birth dates for each mouse varied over two weeks. The observation ages for the mice are therefore not balanced, with age values varying between 2.03 and 13.55. The number of pairs of (male, female) mice having doses of 0, 50, 100, and 200 cGy were (4,4), (19,17), (22,25), and (19,14), respectively.

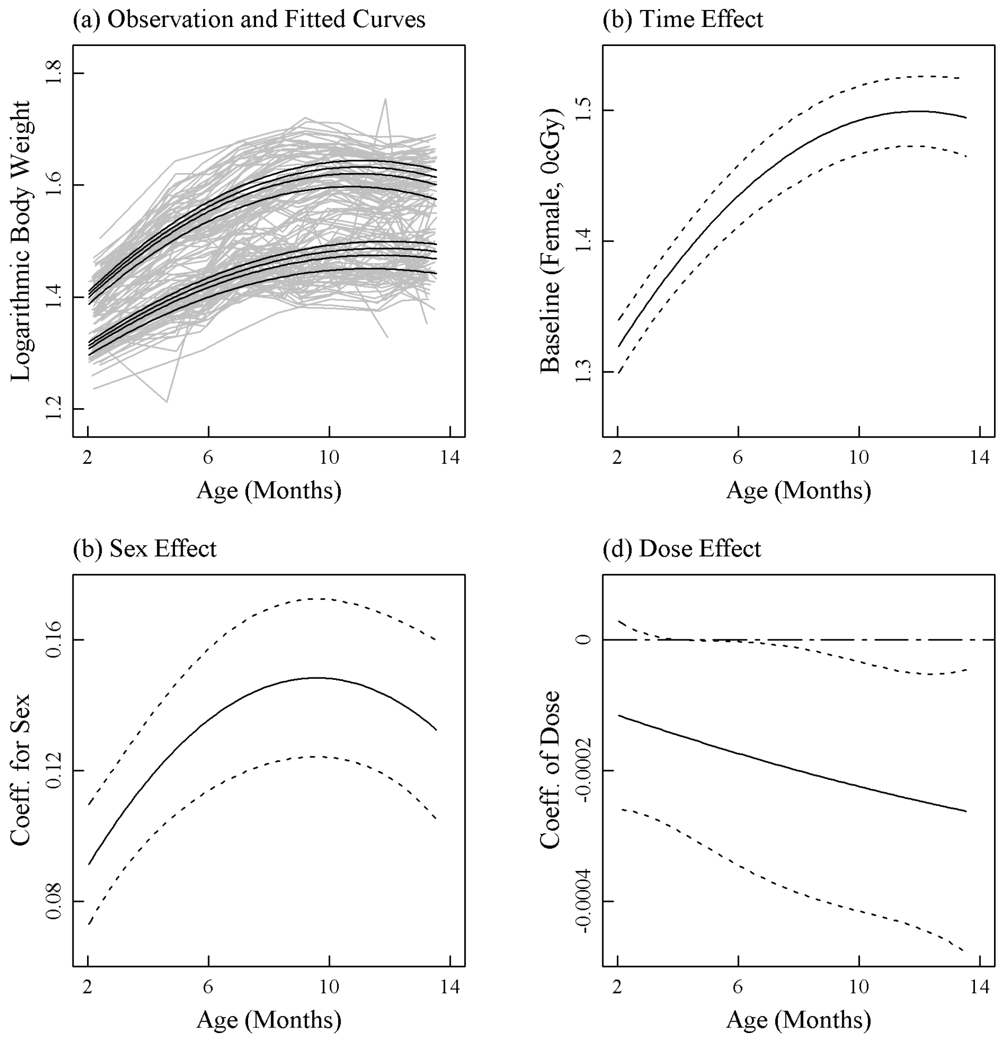

FIGURE 2. Varying coefficient for mice body weight data.

Because the variances of male body weights were greater than those of female body weights, the original weights were transformed by common logarithms as indicated in Figure 2. We thus take y_{ij} to denote the logarithmic body weight (g) of the jth measurement of the ith individual at an age of t_{ij} months old for $i = 1, \cdots, 124$. The covariates of the ith individual are expressed as $\boldsymbol{a}_i = (a_{i1}, a_{i2}, a_{i3})'$ where $a_{i1} \equiv 1$, $a_{i2} = 1$ for a male ith mouse,

$a_{i2} = 0$ for a female mouse, a_{i3} is the parental dose of the ith mouse, and and $\boldsymbol{x}(t) = (1, t, t^2)'$ for t_{ij}.

Applying the estimation method proposed by Vonesh and Carter (1987), estimators of $\boldsymbol{\Theta}$, $\boldsymbol{\Omega}_j$, $j = 1, 2$ and 3 are obtained. Thus, estimated varying coefficient curves $\hat{\beta}_j(t)$, $j = 1, 2$ and 3 are shown in Figure 2. The dotted curves represent the ± 1.96 standard error bands. The broken lines represent 95% confidence intervals derived using Theorem 1. (a) The jointed gray lines show individual observations; the solid lines are fitted curves. (b) Time effects: $\boldsymbol{\beta}_1(t)$ versus time. (c) Sex effects: $\boldsymbol{\beta}_2(t)$ versus time. (d) Dose effects: $\boldsymbol{\beta}_3(t)$ versus time. The sex effect was significantly positive over the whole observation period, and it increased nearly monotonically. On the other hand, the dose effect was almost negative and seemed to be constant with respect to age. The significance level was calculated as $p = 0.004$. Therefore, the fitted curves in Fig 2(a) seem to be parallel with respect to dose level and are clustered by sex.

5. CONCLUSION AND DISCUSSION.

In this paper, we have considered a new view of GCM, regarding $\boldsymbol{\theta}_j'\boldsymbol{x}(t)$ as a varying coefficient $\beta_j(t)$. Since elements of $\boldsymbol{\Theta}$ in the GCM are essentially coefficients of a polynomial function expressing the time trend, it is meaningless to focus only on a single element of $\boldsymbol{\Theta}$. Nevertheless, in the ordinary GCM, interval estimation and testing a hypothesis about only a single element of $\boldsymbol{\Theta}$ are usually considered, i.e., interval estimation for coefficients of polynomial function is carried out, or a test is performed to determine whether a coefficient of the polynomial function is zero. In real data analysis, we are interested in the effects of covariates on the amount of growth. Even if we find that, up to statistical significance, a coefficient of a polynomial function is 0, then, needless to say, it is not necessarily the case that the effect of the covariate is 0. Utilizing this new approach, that is, by varying the coefficient $\beta_j(t)$ over time, we can examine the effect over time of the covariate. Such an effect can be easily visualized by the proposed functional confidence interval for $\beta_j(t)$. Furthermore, by means of the proposed test procedure, we can determine the significance or otherwise of the effect of the covariate. Although it can be difficult to understand and interpret the results that are obtained, there is no doubt that GCM is a very useful model for describing a time trend, and our method may help in understanding results obtained from an actual GCM data analysis. Our estimation method is very easy, i.e., estimates can be obtained from the results of the GCM. Although we introduce only two models in Section 3, explicit estimators have been derived on some other growth curve models, e.g., Fujikoshi and Satoh (1996), Fujikoshi, Kanda and Ohtaki (1999), Srivastava (2002). So it is unnecessary to use a nonlinear optimization algorithm to compute the functional confidence interval and the p-value of the test statistic. Therefore, our result may also be valuable in the sense of providing an easy estimation method for a varying coefficient model.

ACKNOWLEDGEMENT

We really appreciated helpful and valuable referee's coments. A part of this research was supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Young Scientists (B), #19700265 (2007-2010), #21700306 (2009-2010) and the Fujii Fund of Hiroshima University (2008-2009).

REFERENCES

Fujikoshi, Y. and Satoh, K. (1996). Estimation and model selection in an extended growth curve model. Hiroshima Mathematical Journal, 26, 635-647.

Fujikoshi, Y., Kanda, T. and Ohtaki, M. (1999). Growth curve models with hierarchical within-individuals design. Annals of the Institute of Statistical Mathematics, 51, 707-721.

Guo, W. (2002). Functional mixed effects model. Biometrics, 58, 121-128.

Hastie, T. and Tibshirani, R. (1993). Varying-coefficient models. Journals of the Royal Statistical Society B, 55, 757-796.

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

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, Biometrika, 38, 963-974.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. Biometrika, 58, 545-554.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems. Biometrika, 51, 313-326.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to analysis of growth curves. Biometrika, 52, 447-458.

Rao, C. R. (1973). Linear Statistical Inference and Its Applications. John Wiley, Inc., New York.

Ruppert, D., Wand, M. and Carroll, R. (2003) Semiparametric regression. Cambridge University Press, Inc., Cambridge.

Satoh, K., Kobayashi, M. and Fujikoshi, Y. (1997). Variable selection for the growth curve model. Journal of Multivariate Analysis, 60, 277-292.

Satoh, K., Yanagihara, H. and Ohtaki, M. (2003). Bridging the gap between B-spline and polynomial regression model. Communications in Statistics - Computation and Simulation, 32, 179-190.

Satoh, K. and Ohtaki, M. (2006). Nonparametric growth curve model with local linear approximation. Communications in Statistics - Theory and Methods, 35, 641-648.

Srivastava, M. S. (2002). Nested growth curve models. Sankhya A, 64, 379-408.

Srivastava, M. S. and von Rosen, D. (1999). Growth curve models. Multivariate Analysis, Design of Experiments, and Survey Sampling (Ghosh, S. ed.), 547-578, Marcel Dekker, New York.

Vonesh, E. F. and Carter, R. L. (1987). Efficients inference for random-coefficient growth curve models with unbalanced data. Biometrics, 43, 617-628.

Watanabe, H., Takahashi, T., Lee, J., Ohtaki, M., Roy, G., Ando, Y., Yamada, K., Gotoh, T., Kurisu, K., Fujimoto, N., Satow, Y. and Ito, A. (1996). Influence of paternal ^{252}Cf Neutron exposure on abnormal Sperm, embryonal Lethality, and Liver tumorigenesis in F_1 offspring of mice. Japanese Journal of Cancer Research, 87, 51-57.

West, M., Harison, P. J. and Migon, H. S. (1985). Dynamic generalized linear models and Bayesian forecasting (with discussion). Journal of the American Statistical Association, 80, 73-97.