

Least square and Empirical Bayes Approaches for Estimating Random Change Points

Yuanjia Wang¹ and Yixin Fang²

¹*Columbia University* and ²*Georgia State University*

Abstract: Here we develop methods for applications where random change points are known to be present a priori and the interest lies in their estimation and investigating risk factors that influence them. A simple least-square method estimating each individual's change point based on one's own observations is first proposed. An easy-to-compute empirical Bayes type shrinkage is then proposed to pool information from separately estimated change points. A method to improve the empirical Bayes estimates is developed. Simulations are conducted to compare least-square estimates and Bayes shrinkage estimates. The proposed methods are applied to the Berkeley Growth Study data to estimate the transition age of the puberty height growth.

Key words: Longitudinal data, mixed effects model, random change points model.

1. Introduction

In longitudinal studies, the change rate over time of the expected values of outcomes may vary after a random time point. For example, the pubertal height growth shows a rapid increasing period followed by a sharp declination of growth rate until the height stops to increase towards the end of puberty (Tuddenham and Snyder 1954, Berkeley Growth Study). The age at which the growth rate changes is usually not observed.

Another example where random change points are present is described in Alzheimer's disease research. In the preclinical phases, declination in cognitive function can be gradual and hard to distinguish from the normal aging. As dementia progresses, cognitive impairment declines more rapidly (Hall *et al.* 2000 and 2003). When the declination in cognitive function begins to accelerate in demented patients is not observed. To estimate the change points in cognitive impairment and to understand risk factors that contribute to their variation are of interest.

A third example involving random change point is a study of patterns of developmental transitions in the emerging adulthood period (Cohen *et al.* 2003). In this study, one of the primary goals was to examine when the young adults transit to independent adulthood. Two hundred and forty subjects were interviewed and scored on several measures of independence (for example, leaving the parental home, being financially independent and so on). Each subject may transit to adulthood at a different age. The interest lies in estimating transition ages and investigating influences of factors such as parental marital status, parental socioeconomic status on the individual differences of the transition ages.

Several methods including both frequentist and Bayesian are proposed to analyze data with change points. See for example, Carlin *et al.* (1992), Hall *et al.* (2000) and (2003), Hinkley (1970), and Jacqmin-Gadda *et al.* (2006). Hall *et al.* (2000) modeled the change point as a fixed value for all subjects and maximized the profile likelihood in a piecewise linear mixed model. However, it is not clear whether the change point of the cognitive function for each Alzheimer patient should be the same. Hall *et al.* (2003) proposed a random change point model with parameters estimated in Bayesian framework. They warned against the sensitivity to choice of prior distribution and identifiability. Jacqmin-Gadda *et al.* (2006) proposed a joint model for time-to-event and repeated measure of a disease marker where the time-to-event data had unknown random change points. The distribution of change point is assumed to be log-normal and the parameters are estimated by MLE. However, these methods are not easy to implement.

Here we develop methods for applications where random change points are known to be present a priori and the interest lies in estimation and investigating risk factors that influence them. A simple least-square method estimating each individual's change point based on one's own observations is first proposed. An easy-to-compute empirical Bayes type shrinkage is then proposed to pool information from separately estimated change points. A method to improve empirical Bayes estimates is developed. Simulations are conducted to compare least-square estimates and Bayes shrinkage estimates. The proposed methods are applied to the Berkeley Growth Study data to estimate the transition age of the puberty height growth.

2. Methods

Let Y_{ij} denote the observed response for the i th subject at the j th occasion and let t_{ij} denote the time at the same occasion. Let τ_i denote the unobserved change point of the i th subject. For the purpose of illustration, we consider the simple case where there are no covariates in the linear model. It is straightforward to add covariates to the model.

Two models are considered here. The first one is a simple piecewise linear

model with one random change point and random intercepts:

$$Y_{ij} = (\alpha_{i1} + \beta_1 t_{ij})1_{\{t_{ij} < \tau_i\}} + (\alpha_{i2} + \beta_2 t_{ij})1_{\{t_{ij} > \tau_i\}} + \varepsilon_{ij}. \quad (2.1)$$

Here α_{i1} , τ_i and ε_{ij} are independent. Since the lines before and after the change point intersects at τ_i , we have the expression

$$\alpha_{i2} = \alpha_{i1} + (\beta_1 - \beta_2)\tau_i,$$

or equivalently,

$$\tau_i = \frac{\alpha_{i1} - \alpha_{i2}}{\beta_2 - \beta_1}.$$

We will exploit this relationship in the estimation. Note that the difference between (2.1) and a traditional piecewise linear mixed effects model is that the change points τ_i are unknown and are different for each subject.

In the second model, it is assumed that each individual has random intercept and random slope:

$$Y_{ij} = (\alpha_{i1} + \beta_{i1} t_{ij})1_{\{t_{ij} < \tau_i\}} + (\alpha_{i2} + \beta_{i2} t_{ij})1_{\{t_{ij} > \tau_i\}} + \varepsilon_{ij}. \quad (2.2)$$

Here we assume β_{i1} and β_{i2} are independent random slopes before and after the change point, respectively.

2.1 Least-square approach

A simple and intuitive approach to estimate change points is to apply least-square estimation on observations from each individual separately. To be specific, least square estimation of τ_i based on each individual's observations is

$$\min_{\alpha_{i1}, \tau_i, \beta_{i1}, \beta_{i2}} \sum_j [y_{ij} - (\alpha_{i1} + \beta_{i1} t_{ij})1_{\{t_{ij} < \tau_i\}} - (\alpha_{i2} + \beta_{i2} t_{ij})1_{\{t_{ij} > \tau_i\}}]^2, \quad (2.3)$$

$i = 1, \dots, n$.

This method is very easy to implement. However, estimating each individual change point separately using only observations on the same subject ignores the information on the distribution among the change points. The mean square error for such least-squares estimates is large especially when there are not many observations per subject. We introduce an empirical Bayes method to pool together information among individual change points which will give smaller mean squared error.

2.2 Empirical Bayes approach

To improve upon the least square estimation of change points in section (2.1) we apply empirical Bayes estimation to combine information across subjects.

Let Y_i denote the vector of the responses from the i th subject. Let $Y_i^{(1)}$ denote the vector of responses that satisfy $t_{ij} < \tau_i$. Let n_{i1} denote the number of observations in $Y_i^{(1)}$. Observe that given the change points, the empirical Bayes estimation of the random intercept is (Laird and Ware 1982)

$$\hat{\alpha}_{i1} = \left(1 - \frac{\hat{\sigma}_{\alpha 1}^2}{\hat{\sigma}_{\alpha 1}^2 + \hat{\sigma}_{\varepsilon}^2/n_{i1}}\right)\hat{\mu}_{\alpha 1} + \sum_{j: t_{ij} < \tau_i} \frac{\hat{\sigma}_{\alpha 1}^2}{\hat{\sigma}_{\alpha 1}^2 + \hat{\sigma}_{\varepsilon}^2/n_{i1}}(Y_{ij} - \hat{\beta}_1 t_{ij}). \quad (2.4)$$

Here $\hat{\mu}_{\alpha 1}$ is the estimated population mean, and $\hat{\sigma}_{\alpha 1}^2$ and $\hat{\sigma}_{\varepsilon}^2$ are the estimated variance of the random effect α_{i1} and the residual ε_{ij} , respectively. The empirical Bayes estimation is shrunk towards the population mean. The estimation of population mean and variance terms involves all subjects and is the source of the strength of pooling. Note that when given τ_i , (2.4) is the ‘‘best linear unbiased predictor’’ or BLUP of the random effect in a linear mixed effects model (Robinson 1991).

We develop procedures to estimate unknown parameters in (2.4) and provide estimation of τ_i and β . The algorithm for model (2.1) is

Step 1. Given β_1 and β_2 , for each i , compute τ_i as the solution of

$$\min_{\alpha_{i1}, \tau_i} \sum_j \|Y_{ij} - (\alpha_{i1} + \beta_1 t_{ij})1_{\{t_{ij} < \tau_i\}} - (\alpha_{i2} + \beta_2 t_{ij})1_{\{t_{ij} > \tau_i\}}\|^2.$$

Step 2. Given τ_i from step 1, compute empirical Bayes estimates of the random intercept α_{i1} and α_{i2} as in (2.4) by fitting a mixed effects model with random intercepts using observations occurring before and after τ_i , respectively. Estimation of the fixed effects β_1 and β_2 are also obtained through the fitting. Iterate the steps 1 and 2 until the algorithm converges.

Step 3. After steps 1 and 2 converges, the empirical Bayes estimation of the change point is

$$\hat{\tau}_i = \frac{\hat{\alpha}_{i1} - \hat{\alpha}_{i2}}{\hat{\beta}_2 - \hat{\beta}_1}. \quad (2.5)$$

From (2.4), the empirical Bayes estimation of α_i is shrunk from the individual estimation towards the population mean. Therefore the estimation of τ_i in the expression (2.5) is also shrunk towards its population mean. The amount of shrinkage depends on the relative magnitude of the variation of the within subject residual, σ_{ε}^2 , and the variation of the random effect, σ_{τ}^2 . When σ_{τ}^2 is small

compared to σ_ε^2 , indicating a more homogeneous population, the empirical Bayes estimation is shrunken more towards the population mean.

To compute the variance of $\hat{\beta}$, we need to take into account of the extra variability introduced by $\hat{\tau}_i$. The variance of $\hat{\beta}$ can be expressed

$$\text{var}(\hat{\beta}) = \text{var}(E(\hat{\beta}|\hat{\tau})) + E(\text{var}(\hat{\beta}|\hat{\tau})). \quad (2.6)$$

The second term on the right hand side is obtained by variance estimates in least-square method. Computation of the first term is less straightforward; we resorted to bootstrap re-sampling. To be specific, for each i , we simulated 200 replications of τ_i^* , where

$$\tau_i^* \sim N(\hat{\tau}_i, \text{var}(\hat{\tau}_i|\tau_i)).$$

Given each replication of τ_i^* , we computed a $\hat{\beta}^*$ as described before. The term $\text{var}(E(\hat{\beta}|\tau))$ is then estimated by the sample variance of $\hat{\beta}^*$.

To get an estimate of the variance of τ_i , observe that from the relationship $\alpha_{i2} = \alpha_{i1} + (\beta_1 - \beta_2)\tau_i$, we have

$$\text{var}(\tau_i) = \frac{\text{var}(\alpha_{i2}) - \text{var}(\alpha_{i1})}{(\beta_1 - \beta_2)^2}. \quad (2.7)$$

The variance of α_{i1} and α_{i2} are estimated when fitting the mixed effects model, and the variance of the change points is then estimated by plugging in estimates of each component on the right hand side of (2.7). The variance of $\hat{\beta}$ is computed as in (2.6).

The algorithm for fitting model (2.2) in which the individual slopes are random is similar. In step 1, the change points τ_i are computed as in (). In step 2, given τ_i from step 1, a mixed effects model with random intercept and random slope is fitted. In step 3, the empirical Bayes estimate of the change point is

$$\hat{\tau}_i = \frac{\hat{\alpha}_{i1} - \hat{\alpha}_{i2}}{\hat{\beta}_{i2} - \hat{\beta}_{i1}}. \quad (2.8)$$

Since the slopes are random in model (2.2), the variance of τ_i has a different form:

$$\text{var}(\tau_i) = \frac{\text{var}(\alpha_{i2}) - \text{var}(\alpha_{i1}) + [E(\beta_1 - \beta_2)E(\tau_i)]^2}{\text{var}(\beta_1 - \beta_2) + [E(\beta_1 - \beta_2)]^2} - (E(\tau_i))^2. \quad (2.9)$$

The estimated variance of τ_i is obtained by plugging in the estimates of each component on the right hand side. The variance of $\hat{\beta}$ is again estimated by (2.6).

After estimating the change points as in (2.5) and (2.8), a regression with estimated change points as outcome can be used to investigate covariates contributing to the variation in individual change points.

2.3 Improved empirical Bayes approach

Because the variability in estimating τ_i affects the estimation of slope and variances (2.7) and (2.9), we develop an intuitive improvement of the empirical Bayes approach to minimize the influence of variability in $\hat{\tau}_i$. Since $\hat{\tau}_i$ are estimated with error, whether the observations around $\hat{\tau}_i$ should contribute to the group before or after the change point cannot be determined with certainty. Therefore based on $\hat{\tau}_i$ and $\hat{\sigma}_{\tau_i}$, we exclude the observations that are within the interval $(\hat{\tau}_i - z^* \hat{\sigma}_{\tau_i}, \hat{\tau}_i + z^* \hat{\sigma}_{\tau_i})$ to remove the influence of these observations around the change points. We then re-estimate slopes, change points, and variances. The values of z^* are chosen a priori by normal quantiles. When the slopes are fixed effects, the influence of estimating τ_i is negligible, therefore in the simulations we only apply this improvement to the random slope model.

3. Simulation

To investigate performances of proposed methods, we conducted two sets of simulations. The simulation settings are closely related to the Berkeley Growth data. The first set of simulations generates growth rate over time as a fixed effect for all individual and the second set simulates the growth rate as a random effect. In each set of the simulations, 100 subjects were generated. Each subject had 31 observations and the time spacing is the same as observed in the Berkeley Growth Study.

In the first setting, the slopes were treated as fixed effects. The true values of the slopes before and after the change point were 2.8 and 0.5, respectively. The random intercepts were simulated from $\alpha_{i1} \sim N(28, 1)$. The change points were simulated from $\tau_i \sim N(12, 0.5^2)$. The residuals had a normal distribution with mean zero and standard deviation one.

In the second setting, the random slopes were simulated from

$$\beta_{i1} \sim N(2.8, 0.1^2), \quad \beta_{i2} \sim N(0.5, 0.1^2).$$

The random change points were simulated from a normal distribution with mean nine and standard deviation 0.5. The distribution of random intercepts and the residuals were the same as the first setting.

Table 1a recorded properties of the slope estimates in the first setting where the slopes were fixed effects. The algorithm described in section (2.2) converges very fast, normally within five steps. The slopes were estimated accurately. The estimate of the variance of $\hat{\beta}$ was close to the empirical variance of $\hat{\beta}$. The mean squared error (*MSE*) was 0.007 for the first slope and 0.016 for the second.

Table 1a: Properties of the slope estimates in 200 simulations: setting 1

EB [†]		
	$\beta_1 = 2.8$	$\beta_2 = 0.5$
Mean($\hat{\beta}$)	2.80	0.50
Empirical S.E. of $\hat{\beta}$ 1	0.007	0.016
Mean $\widehat{SE}(\hat{\beta})$	0.007	0.019
$\sqrt{MSE(\hat{\beta})}$	0.007	0.016

[†]: Empirical Bayes method

Table 1b: Properties of the slope estimates in 200 simulations: setting 2

	LS [†]		EB		EB improved [‡]	
	$\mu(\beta_{i1})$ =2.8	$\mu(\beta_{i2})$ =0.5	$\mu(\beta_{i1})$ =2.8	$\mu(\beta_{i2})$ =0.5	$\mu(\beta_{i1})$ =2.8	$\mu(\beta_{i2})$ =0.5
Mean($\hat{\beta}_i$)	2.78	0.55	2.78	0.56	2.79	0.53
Mean($\hat{\sigma}_{\beta_i}$)	0.10	0.11	0.10	0.11	0.10	0.11
$\sqrt{MSE(\hat{\beta}_i)}$	0.18	0.15	0.074	0.11	0.066	0.098
Improvement in $\sqrt{MSE(\hat{\beta}_i)}$ [*]	-	-	59%	56%	64%	76%

[†]: [†]: Least square method

[‡]: Empirical Bayes method excluding observations close to $\hat{\tau}_i$

^{*}: Proportionate reduction in MSE using EB and EB corrected over least-square method.

Table 1b recorded properties of the slope estimates in the second setting where the slopes were random. All methods provide adequately good estimations for the mean and the variance of the slopes. Note that the improved empirical Bayes method excluding observations close to the $\hat{\tau}_i$ provides slightly more accurate estimate of the mean of β_i , and offers some improvement in MSE . Here we computed $MSE(\hat{\beta}_i)$ by

$$\frac{1}{Nn} \sum_k \sum_i (\hat{\beta}_i^{(k)} - \beta_i)^2,$$

where k indexes simulation, i indexes subject, and N is the total number of simulations. Compared to the empirical Bayes method without improvement, the improved approach lowered the $\sqrt{MSE(\hat{\beta}_i)}$ by 5% and 10% for the first and the second slope, respectively. Note that in this table $\hat{\sigma}_{\beta_i}$ is the estimated standard deviation of the random variable β_i , not the standard error of $\hat{\beta}_i$, so that it is possible that $MSE(\hat{\beta}_i)$ is smaller than $\hat{\sigma}_{\beta_i}^2$.

Table 2: Properties of the change points estimates in 200 simulations

	Setting 1		Setting 2		
	LS $\mu(\tau_i) = 12,$	EB $\sigma(\tau_i) = 0.5$	LS $\mu(\tau_i) = 9,$	EB $\sigma(\tau_i) = 0.5$	EB improved
Mean($\hat{\tau}_i$)	11.13	11.99	8.14	8.88	8.93
$\hat{\sigma}_{\tau_i}$	0.43	0.49	0.48	0.46	0.53
$\sqrt{MSE}(\hat{\tau}_i)$	1.22	0.16	1.26	0.33	0.03
Improvement in $\sqrt{MSE}(\hat{\tau}_i)$	-	87%	-	74%	97%

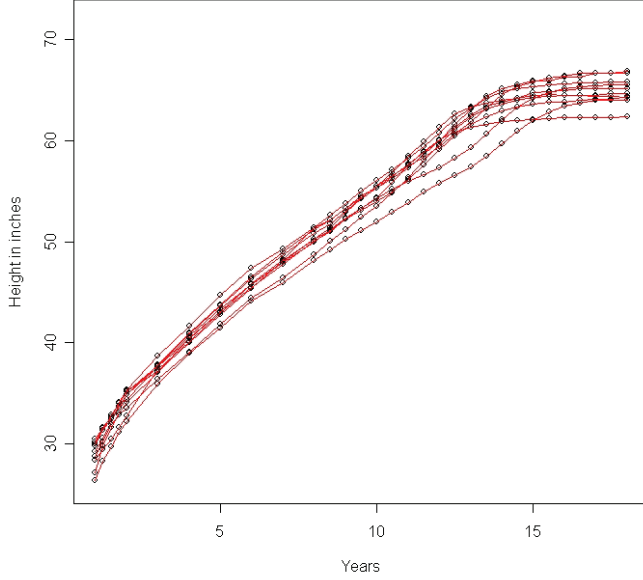


Figure 1: Height growth for ten girls

Table 2 recorded properties of the predicted change points. The mean of the predicted change points computed by the least square method was notably smaller than the true value. In contrast, the mean of $\hat{\tau}_i$ was close to the true value for the empirical Bayes method and its improved version in both settings. The true standard deviation of τ_i was 0.5, where the estimated values were close to 0.5. The MSE of the change points computed by the empirical Bayes based methods was substantially lower than the least square method. The improved empirical Bayes estimates offer more reduction in MSE : the proportionate reduction in $\sqrt{MSE}(\hat{\tau}_i)$ obtained by the empirical Bayes method was 74% in the second setting, while the corresponding reduction computed by the improved method was 97%.

Finally, the residual variance was estimated well by all methods in all settings. The true value of the residual standard deviation was one, where the estimated value was 0.99 in all cases.

4. Data Analysis

In this section the proposed methods are applied to the data in Tuddenham and Snyder (1954). The data recorded heights for 54 girls and 39 boys at 31 ages. The ages were not equally spaced. There were four measurements from age one to two, annual measurement from age two to eight, and biannual measurements from age eight to eighteen. Here we analyze the data on girls. Figure 1 displays height data for ten girls. The height growth can be approximately summarized by two linear growth phase: a sharp increase followed by a slow increase. We are interested in at what age the change of height growth rate happens.

Table 3a: Properties of the estimates in Berkeley Growth Data: fixed slope

	EB		EB Improved	
	β_1	β_2	β_1	β_2
Mean($\hat{\beta}$)	2.72	0.43	2.72	0.38
$\widehat{SE}(\hat{\beta})$	0.02	0.16	0.02	0.14
Mean($\hat{\tau}_i$)	12.86		12.92	
$\hat{\sigma}_{\tau_i}$	0.78		0.76	
$\hat{\sigma}_{\varepsilon}$	1.27		1.27	

Table 3b: Properties of the estimates in Berkeley Growth Data: random slope

	EB		EB Improved	
	β_1	β_2	β_1	β_2
Mean($\hat{\beta}_i$)	2.76	0.62	2.76	0.43
$\hat{\sigma}_{\beta_i}$	0.21	0.32	0.21	0.34
Mean($\hat{\tau}_i$)	12.29		12.63	
$\hat{\sigma}_{\tau_i}$	1.61		1.29	
$\hat{\sigma}_{\varepsilon}$	1.00		1.02	

Since the least square method did not show promising results in the simulation studies, we only applied empirical Bayes method and its improved version to analyze the height data.

Table 3a recorded results from model (2.1) where fixed slopes were assumed. The estimated first slope was 2.72 (SE: 0.02) for both un-adjusted and improved empirical method; the estimated second slope was 0.43 (SE: 0.16) and 0.38 (SE: 0.14) for the two methods, respectively. The mean of the estimated change points

were 12.86 and 12.92, and the estimated standard deviation of the change points were 0.78 and 0.76, respectively for the two methods. The residual standard deviation was estimated at 1.27 by both approaches.

Table 3b recorded results from model (2.2) where random slopes were assumed. The estimated first slope had mean 2.76 for both un-adjusted and improved empirical Bayes method; the second slope had mean 0.62 and 0.43 for the two methods, respectively. The estimated standard deviation of the two slopes were around 0.2 and 0.3. The mean of the estimated change points were 12.29 and 12.63 for the two methods, respectively, and the estimated standard deviation were 1.61 and 1.29. The residual standard deviation was estimated at 1.0 by both approaches.

Noticing that the magnitude of the estimated standard deviations of the slopes are non-negligible, we consider the random slope model to be more appropriate for this data. These results show that height increasing rate for girls changes at around age 12.6: height increases much faster from age zero to age 12.6, at a rate of 2.8 inches per year; compared to a rate of 0.4 inches per year from age 12.6 to age 18.

5. Discussion

An empirical Bayes based method and its improvement are proposed to estimate random change points. The methods are easy to implement and easy to interpret. The empirical Bayes methods shrink the individual change point estimates towards their population mean, therefore reduces the mean squared error. Compared to the least square approach estimating individual change point only using observations on the same subject, the empirical Bayes methods decreased the mean squared error substantially.

The proposed methods explore the first order trend in the data and does not use information on higher derivatives. Exploiting information on the second derivatives was proposed in Ramsay and Silverman (2005) to analyze the same data, which revealed detailed features in the data. Methods developed here provide simple-to-implement and easy-to-interpret alternatives.

It is possible that the relationship between outcome and time is not linear. The proposed methods can be extended to non-linear models. However, the computation burden for fitting these models increases. Another alternative is to project the data onto several basis and use functional data analysis tools to estimate parameters as in Ramsay and Silverman (2005). However, these methods can be hard to interpret.

It is straightforward to extend the current model to handle more than one change point, given that there is prior knowledge that multiple change points exist. However, the methods developed here can not be used to determine how

many change points are present.

Finally, here the outcomes do not encounter censoring. A maximum likelihood type estimation such as in Jacqmin-Gadda *et al.* (2006) can be adapted to account for censoring. However, in those cases the empirical Bayes shrinkage are more complicated and do not have close-form solutions.

References

- Carlin, B. P., Gelfand, A. E., Smith, A. F. M. (1992). Hierarchical Bayesian analysis of changepoint problems. *Applied Statistics*. **41**, 381-405.
- Cohen, P., Kasen, S., and Chen, H. (2003). Variations in patterns of developmental transitions in the emerging adulthood period. *Developmental Psychology* **39** 657-669.
- Hall C. B., Lipton R. B., Sliwinski M., and Stewart W. F. (2000). A change point model for estimating the onset of cognitive decline in preclinical Alzheimer's disease. *Statistics in Medicine* **19**, 1555-1566.
- Hall, C. B., Ying, J., Kuo, L., and Lipton, R. B. (2003). Bayesian and profile likelihood change point methods for modeling cognitive function over time. *Computational Statistics & Data Analysis* **42**, 91-109.
- Hinkley, D. V. (1970). Inference about the change-point in a sequence of random variables. *Biometrika*. **57**, 1-17.
- Jacqmin-Gadda, H., Commenges, D., and Dartigues, J. F. (2006). Random chagepoint model for joint modeling of cognitive decline and dementia. *Biometrics* **62**, 254-260.
- Laird, N., and Ware, J. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis* (2nd Edition). Springer.
- Robinson, G. K. (1991). That BLUP is a good thing: the estimation of random effects. *Statistical Sciences* **6**, 15-51.
- Tuddenham, R. D., and Snyder, M. M. (1954). Physical growth of California boys and girls from birth to eighteen years. *University of California Publications in Child Development* **1**, 183-364.

Received June 4, 2007; accepted August 14, 2007.

Yuanjia Wang
Department of Biostatistics
Mailman School of Public Health
Columbia University
722 W168th St., New York, NY 10032, USA
yw2016@columbia.edu

Yixin Fang
Department of Mathematics and Statistics
Georgia State University
750 COE, 7th floor
30 Pryor Street Atlanta
Georgia 30303, USA
matyxf@langate.gsu.edu