World Academy of Science, Engineering and Technology
International Journal of Mathematical and Computational Sciences
Vol:6, No:7, 2012

Human Growth Curve Estimation through a Combination of Longitudinal and Cross-sectional Data

Sedigheh Mirzaei S., Debasis Sengupta

***Abstract*—**Parametric models have been quite popular for studying human growth, particularly in relation to biological parameters such as peak size velocity and age at peak size velocity. Longitudinal data are generally considered to be vital for fittinga parametric model to individual-specific data, and for studying the distribution of these biological parameters in a human population. However, cross-sectional data are easier to obtain than longitudinal data. In this paper, we present a method of combining longitudinal and cross-sectional data for the purpose of estimating the distribution of the biological parameters. We demonstrate, through simulations in the special case ofthePreece Baines model, how estimates based on longitudinal data can be improved upon by harnessing the information contained in cross-sectional data.We study the extent of improvement for different mixes of the two types of data, and finally illustrate the use of the method through data collected by the Indian Statistical Institute.

***Keywords*—**Preece-Baines growth model, MCMC method, Mixed effect model

I. Introduction

GROWTH curves arise naturally in a wide variety of applied areas, including biology, psychology, economics and sociology. In a broad sense, a 'growth curve' represents the way a physical or conceptual variable grows over time. In the context of human growth, the physical variable can be the height (stature) of a person, or some other body dimension. Ideally, these variables may be measured at different points of time for many individuals, leading to a longitudinal data set. The time, expense and effort associated with longitudinal studies may be substantial. As a result, there are also cross-sectional studies, where one individual is measured only once.

An individual growth curve is generally a smooth function that represents the central tendency in the size vs. age graph of a particular individual. Longitudinal growth data are often of the form $(t_{i1}, y_{i1}), (t_{i2,} y_{i2}), \dots, (t_{in}, y_{in}), i = 1,2, \dots, n,$ where n is the number of individuals,n_i is the number of observations for the ith individual, $1 \leq i \leq n$,and y_{ij} is the observed size variable at age t_{ij} regarded without loss of generality as the 'height' variable in this paper.

S. Mirzaei and D. Sengupta are with the Applied Statistical Unit, Indian Statistical Institute, 203 B.T. Road, Kolkata 700108, India. Email: sedigheh_r@isical.ac.in and sdebasis@isical.ac.in

A general model for such a data set is :

$$y_{ij} = h(t_{ij}) + \varepsilon_{ij}, \quad j = 1,2, \dots, n_i, \quad i = 1,2, \dots, n, \qquad (1)$$

where h is the growth function and ε_{ij}'s are additive random errors. Apart from the function$h$, one is typically interested in estimating its derivative (referred to as the 'velocity' function) and various biological parameters, such as the points of inflection of the velocity function (referred to as the age at takeoff and the age at peak height velocity), the velocities at these ages (referred to as takeoff velocity and peak height velocity, respectively) and the limiting value of *h* for large age (referred to as final height). One may also seek to estimate the age-specific quantiles of the height variable, which depends on the value of the function *h* at that age, as well as the distribution of the error. When covariates are present, all these estimation problems become regression problems.

If the number of measurements per individual is large, one can use nonparametric smoothing or regression [10], [16] with age as the explanatory variable, to estimate the function$h$. Staniswalis and Lee [20] provided a nonparametric method that can even handle covariates. However, it is rather unusual to find a longitudinal data set with a large number of height measurements per individual, let alone height data with covariates. Consequently, there has been emphasis on parametric models of the form:

$$y_{ij} = h(t_{ij}; \boldsymbol{\tau}_i) + \varepsilon_{ij}, \quad j = 1,2, \dots, n_i, \quad i = 1,2, \dots, n \qquad (2)$$

where the function $h$has a known functional form, with an unknown vector parameter $\boldsymbol{\tau}_i$controlling its shape. Potthoff and Roy [17]considered a model of $h$that is possibly nonlinear (e.g., polynomial) in the age variable, but is linear in the parameter τ_i. Since this model lies in the framework of multivariate linear models, Pothoff and Roy's (1964) work was followed up by many other researchers. However, when there are only a handful of observations per individual, a parsimonious model can be fitted to individual growth data only if the model is allowed to be nonlinear in the parameters. Simplest examples of such models include the exponential growth model and the logistic growth model, while more complex models with larger number of parameters have also been considered [9], [12]. One of the most popular has been model I of Preece and Baines [18], given by

$$h(t; s_0, s_1, \theta, h_{max}, h_\theta) = h_{max} - \frac{2(h_{max} - h_\theta)}{e^{s_0(t-\theta)} + e^{s_1(t-\theta)}} \quad (3)$$

where$s_0$and s_1 are parameters controlling the rates of growth at different stages, $h_{max}$is the final height, and $h_\theta$is the height at a threshold age θ. One can use nonlinear regression [2] to fit any of these models to individual growth curves. A range of nonparametric and parametric methods for longitudinal data can be found in [6], [22].

One can also treat the vector parameter τ_i appearing in the model (2) as a random vector, which has a probability distribution over the population. This amounts to assuming a random effects model. Estimation of this model can give rise to estimation of the population distribution of various biological parameters that are functions of τ_i, e.g., the final height, the takeoff and peak height velocities and the corresponding ages, and so on. Work in this area began with the Pothoff-Roy model [19] and has continued ever since [3], [7], [8], [13], [14].

Because of the above mentioned difficulties of obtaining longitudinal data, some studies are designed to track different individuals over different age ranges. The different age ranges used in the study may have only partial overlap. This way, the duration of the study can be shorter. Huggins and Loesch[13] considered analysis of this type of data.

However, many large scale studies on human growth happen to be entirely cross-sectional. With these data, one seeks to obtain age-specific height quantiles for the population, which can be used as reference for comparing the growth of an individual subject. The trajectories of a specific quantile of height over different ages are sometimes referred to as centile growth curves or reference centile charts. A combination of these charts for different quantiles is used as a reference for the growth trajectory of growing children. The shapes of these curves can be very different from the shape of an individual growth curve.

The estimation of centile growth curves has traditionally relied on normal distribution theory. Typically, quantiles for an age are computed from the mean and the standard deviation estimated from cross-sectional data for that age, and the quantiles for different ages are smoothed to obtain centile growth curve. In this context, the LMS method[5] has become a classical work on growth chart construction. If the normal distribution assumption holds, then extreme quantiles computed from the mean and standard deviation may be fine even for short data. In large scale studies, such as the one used to develop the National Center of Health Statistics (NCHS) growth chart [11],quantiles are computed from empirically observed fractions. More recently, semiparametricquantile regression models have been used for this purpose [21], so that conditional quantiles can be estimated without any distributional assumption.

Even if one assumes a model such as (2) and a distribution for τ_i's, the observed data can be seen as samples from the convolution of the distributions of ε_{ij}'s and τ_i's. The dependence of the location parameter of this distribution on age can have a very different form than the shape of the function h for any specific individual. In particular, it is generally a smoother function. Zemel and Johnston [23] have reported issues of interpretability and model validity that can arise, when one attempts to fit the Preece Baines model (3) to cross sectional data. Similar problems arise in the case of spline models that are linear in the parameters [13], particularly while determining the population distribution of biological parameters.

On the other hand, cross sectional data are generally much more abundant than longitudinal data. Thus, one might seek to combine the strengths of the two types of data, for solving longitudinal data problems.

In this paper, we present a method of estimating the population distribution of the biological parameters from a combination of longitudinal and cross-sectional data. The work is motivated by an anthropometric study conducted by the Indian Statistical Institute under the leadership of Professor S.R. Das during the 1950's and 1960's, from the Sarshuna-Barisha (S-B) region of Kolkata. The data set of male subjects obtained from this study reflects 298 individuals, many of whom were tracked over the said period for different durations. The variables include age, stature and a few other anthropometric characteristics. The number of observations per individual ranges from 1 to 21, for the age interval 0.5 to 21. Lack of samples in the 19-or-more and 10-or-less age ranges come in the way of fitting a reasonable parametric model in most of the cases. Excluding another 5 cases due to convergence problems, only 36 cases are found to be amenable to fitting a reasonable parametric model. On the other hand, there is arelatively healthy count of total number of observations, which may be tapped for improved estimation.

II. Estimating Population Distribution of Parameters

Consider the random effects model defined by (2) and the additional assumptions:

(a) $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_n$ are samples from a common population distribution f, and

(b)$\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ are samples from a distribution φ.

For fully cross-sectional data, $n_i = 1$ for $i = 1,2, \dots, n$ Here, we permit a part of the data to be cross-sectional.We assume a functional form of f subject to an unspecified vector parameter$\boldsymbol{\theta}$, and a functional form ofφ subject to an unspecified scale parameter σ. The problem of estimating the population distribution of biological parameters is then reduced to the problem of estimating $\boldsymbol{\theta}$. Since the distribution of any function of $\boldsymbol{\tau}_i$ can be derived from $f(\boldsymbol{\tau}_i; \boldsymbol{\theta})$, this distribution can always be estimated by substitution, once $\boldsymbol{\theta}$ is estimated.

The likelihood for $\boldsymbol{\theta}$ and σ is

$$\prod_{i=1}^{n} f(\tau_i; \boldsymbol{\theta}) \prod_{j=1}^{n_i} \frac{1}{\sigma} \varphi\left(\frac{y_{ij} - h(t_{ij}; \boldsymbol{\tau}_i)}{\sigma}\right). \quad (4)$$

Since $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_n$ are unobserved, these can be treated as nuisance parameters. Maximizing the likelihood in the presence of the nuisance parameters is usually rather difficult.

A standard approach to this problem is to maximize the likelihood (4) with respect to $\boldsymbol{\theta}$, σ and $\boldsymbol{\tau}_1, \boldsymbol{\tau}_2, \dots, \boldsymbol{\tau}_n$. The other approach is to integrate the likelihood with respect to the nuisance parameters, i.e., to maximize the integrated likelihood

$$\prod_{i=1}^{n} \int f(\tau_i; \boldsymbol{\theta}) \prod_{j=1}^{n_i} \frac{1}{\sigma} \varphi\left(\frac{y_{ij} - h(t_{ij}; \boldsymbol{\tau}_i)}{\sigma}\right) d\tau_i \qquad (5)$$

The EM algorithm and Gibb's sampling [15] provide computational methods for solving such problems. Even so, the nonlinear nature of the function *h*complicates the optimization problem that needs to be solved at each step of an iterative procedure.

If f is treated as prior density for $\boldsymbol{\tau}_i$ then the corresponding posterior is

$$g_i(\boldsymbol{\tau}_i | y_{i1}, \dots, y_{in_i}; \boldsymbol{\theta}, \sigma) \propto f(\boldsymbol{\tau}_i; \boldsymbol{\theta}) \prod_{j=1}^{n_i} \frac{1}{\sigma} \varphi\left(\frac{y_{ij} - h(t_{ij}; \boldsymbol{\tau}_i)}{\sigma}\right). \qquad (6)$$

If the number of individuals is large, the average of these posterior distributions should resemble the correct distribution of the $\boldsymbol{\tau}_i$'s, even if the prior distribution is not the same as that distribution. Instead of using a prior density, we can use the longitudinal part of the data set to estimate $\boldsymbol{\tau}_i$ for the corresponding individuals, use these estimated$\boldsymbol{\tau}_i$'s to estimate $\boldsymbol{\theta}$, and substitute the latter in $f$to get an empirical version of the prior density. The use of an empirically determined density in place of the prior is in the spirit of the empirical Bayes approach [4]. We can then iterate over this entire process, by treating the average posterior distribution at a particular step as the prior distribution at the next step, until the 'prior' and the average of the posteriors come sufficiently close.

The Markov Chain Monte Carlo technique is a convenient tool for implementing this method. The crux of the problem is to avoid computing the proportionality constant of (6), even though the samples need to be drawn from an average of the posterior densities (and not the posterior densities themselves). In order to make this possible, the average of the posterior densities is viewed as a mixture distribution, so that the samples from the targeted density can be obtained by judiciously pooling samples from the posterior densities of the individuals.

The steps to be used, adapted from the Metropolis-Hastings algorithm [1], are as follows.

- Estimate $\boldsymbol{\tau}_i$ for each individual i correspondingto the *longitudinal part* of the data, through nonlinear least squares [2]. Estimate $\boldsymbol{\theta}$ by using these estimates as observed data, and denote the estimator by $\boldsymbol{\theta}^{(0)}$. Also estimate σ from the pooled data, and denote the estimator by $\hat{\sigma}$. Set the index of iteration$k = 0$.
- For each individual i (for which n_i can be 1 orgreater than 1), generate samples from the posterior density of τ_i, defined by (6) with $\boldsymbol{\theta}$and σreplaced by $\boldsymbol{\theta}^{(k)}$ and $\hat{\sigma}$, respectively, as follows. Generate$M$samples from a proposal distribution, say$\boldsymbol{\tau}_1^*, \boldsymbol{\tau}_2^*, \dots, \boldsymbol{\tau}_M^*$. For$j = 1,2, \dots, M$, compute
$$r_{ij} = \frac{g_i(\boldsymbol{\tau}_j^* | y_{i1}, \dots, y_{in_i}; \boldsymbol{\theta}^{(k)}, \hat{\sigma})}{g_i(\boldsymbol{\tau}_0 | y_{i1}, \dots, y_{in_i}; \boldsymbol{\theta}^{(k)}, \hat{\sigma})}$$
whereτ_0 is the mean of the distribution f for $\boldsymbol{\theta} = \boldsymbol{\theta}^{(k)}$. If$r_{ij} > 1$ accept the sample $\boldsymbol{\tau}_j^*$; else, accept it with probability r_{ij}. Let M_i be the number of selected samples.
- Draw N samples from the average posterior asfollows. Let
$$p_i = \frac{n_i I(M_i > 0)}{\sum_{j=1}^{M} n_j I(M_j > 0)},$$
for$i = 1,2, \dots, n$ and$m_1, m_2, \dots, m_n$ be multinomial with parameters $N, p_1, p_2, \dots, p_n$. Then, for$i = 1,2, \dots, n$ the desired sample would consist of m_i samples selected with replacement from the M_i samples generated from the posterior density of τ_i, as mentioned in Step II.
- Define the updated estimate$\boldsymbol{\theta}^{(k+1)}$as thatobtained from the sample of size N generated in Step III.

Steps II to IV are iterated until the estimates of $\boldsymbol{\theta}$ fromsuccessive steps come sufficiently close. The population distribution of any function of τ_i can be obtained from f evaluated at the converged value of$\boldsymbol{\theta}$.

III. Simulation Results

For the purpose of simulation, we assume that the densityφ of the measurement errors is normal with mean 0 and variance σ^2. As for the growth function h,we work with the Preece-Baines model (3) having five parameters. For the sake of identifiability, it is assumed that $s_0 < s_1$.It can be shown, by analyzing the derivative of the growth function, that the ages at takeoff, t_{to} and peak height velocity,t_{phv}, are defined in terms of the model parameters as

$$t_{to} = \frac{2}{s_1 - s_0} \log\left(\frac{(s_1 - s_0) - \sqrt{(s_1 - s_0)^2 - 4 s_1 s_0}}{2 s_1}\right) + \theta,$$

$$t_{phv} = \frac{2}{s_1 - s_0} \log\left(\frac{(s_1 - s_0) - \sqrt{(s_1 - s_0)^2 - 4 s_1 s_0}}{2 s_1}\right) + \theta$$

A necessary condition for the existence of these two distinct ages is that

$$\frac{s_1}{s_0} > 3 + 2\sqrt{2} \qquad (7)$$

Other biological parameters of interest can be expressed as the final height, h_{max}, the takeoff velocity,$h'(t_{to})$ and the peak height velocity,$h'(t_{phv})$.

The conditions for the simulation study are mostlydetermined by the characteristics of the S-B data mentioned in Section 1. Any computational method for the

model parameters may be affected by the different orders of theirmagnitude and the various constraints. It follows from a preliminary analysis of the S-B data that an unconstrained set of parameters of somewhat comparable magnitude is

$$\alpha_1 = \log\left(\frac{3s_0}{1-3s_0}\right), \alpha_2 = \log\left(\frac{\frac{(3+2\sqrt{2})s_0}{s_1}}{1-\frac{(3+2\sqrt{2})s_0}{s_1}}\right),$$

$$\alpha_3 = \frac{\theta}{10}, \qquad \alpha_4 = \log(h_{max} - h_\theta), \qquad \alpha_5 = \log(h_\theta).$$

The distribution of these transformed parameters, obtained from the nonlinear least squares fit of the longitudinal part of the S-B data, appeared to be normal, with a rank 2 variance-covariance matrix. Accordingly, the first two principal components, with empirically determined coefficients, were used for the simulations. Thus, the random parameter used here is a vector$\boldsymbol{\tau}$with two components, such that the parameters $\alpha_1, \dots, \alpha_5$are linear functions of these, and the model parameters $s_0, s_1, \theta, h_{max}$ and $h_\theta$are nonlinear functions thereof. Independent normal distributions for the two components of $\boldsymbol{\tau}$are assumed.

As for estimation of parameters of the distribution ofτ, we assume that the distribution is bivariate normal, and estimate its parameters through the sample mean and the sample variance-covariance matrix. The parameter $\sigma^2$is estimated from the longitudinal part of the data by averaging over the error sum of squares, after the parameter $\boldsymbol{\tau}$has been estimated through nonlinear regression, separately for each individual. The proposal distribution is considered to be the bivariate normal distributions, with mean and dispersion matrix given by the current mean and dispersion matrix of the components of $\boldsymbol{\tau}$.

While evaluating the proposed method for a mixture oflongitudinal and cross-sectional data, the important questions are as follows. (a)Is there any value addition to the estimate from the cross-sectional part of the data? (b) Would the performance be substantially better if the cross-sectional part of the data are replaced by equivalent amount of longitudinal data?

In order to answer these questions, we repeatedly (100times) generate three types of data. The first type of data consists of 50 individuals each with 10 data points (height at ages 4-21).The second type comprises 10 individuals each with 10 data points and 400 individuals with only one data point (i.e., cross-sectional data). The third type of data is a subset of the second one, where only the longitudinal part of the data is included.

By using the above method, we can find the estimateddistribution of individual specific biological parameters and compare them with the true mean and standard deviation. Table I gives a comparison of the bias and the standard deviation of the biological parameters estimated from the three types of data. Table 2 gives a similar comparison for the parameters of the mathematical model. As expected, the bias and the standard deviation for the second type of data (mixture of 20% longitudinal and 80% cross-sectional data) generally lies in between those of the other two types.

Even where there is an exception (e.g., in the case of peak height velocity), the mean squared error follows this order.

In Figure 1, histograms of biological parameters of the three data sets are compared with their respectively true means. In Figure 2, standard errors are compared likewise. The estimated values seem to be mostly in line with the theoretical values.

TABLE I
BIAS AND STANDARD ERROR FOR BIOLOGICAL PARAMETERS

Biological parameter	Data type 1 Bias (Std err)	Data type 2 Bias (Std err)	Data type 3 Bias (Std err)
Age at takeoff (year)	0.09 (0.38)	0.15 (0.65)	0.230 (0.7)
Takeoff velocity (cm/year)	-0.12 (0.09)	-0.099 (0.31)	-0.11 (0.37)
Age at PHV (year)	0.24 (0.19)	0.27 (0.49)	0.27 (0.52)
PHV (cm/year)	-0.56 (0.33)	-0.15 (0.69)	-0.31 1.1
Final height (cm)	0.93 (1.35)	0.91 (1.99)	1.07 (2.18)

TABLE II
BIAS AND STANDARD ERROR FOR MATHEMATICAL PARAMETERS

Mathematical parameter	Data type 1 Bias (Std err)	Data type 2 Bias (Std err)	Data type 3 Bias (Std err)
s_0 (cm/year)	–0.002 (0.006)	–0.002 (0.013)	–0.002 (0.012)
s_1 (cm/year)	-0.07 (0.07)	-0.03 (0.13)	0.019 (0.16)
θ (year)	0.29 (0.17)	0.3 (0.46)	0.36 (0.4)
h_{maz} (cm)	0.93 (1.35)	0.91 (1.99)	1.07 (2.18)
h_θ (cm)	1.07 (0.94)	0.76 (1.95)	0.98 (2.37)

IV. DATA ANALYSIS

Turning to the Sarshuna-Barisha (boys) data, we note thatthere are 36 cases with 10 to 18 data points in the range7 to 18 years, where estimation of the model parameterssubject to the constraint (7) is possible. For the remaining 262 cases, there are many with only with a few observations,including 16 cases with only one observation. This makes itimpossible to estimate the population distribution of individualspecific biological parameters, using conventional methods.

Application of the method proposed in this paper gives riseto the summary of mean and standard error of biological andmathematical parameters, reported in Table 3. For comparison,the summary from the longitudinal part of the data are alsoreported alongside. The standard deviations show substantialimprovement when the additional 262 cases are included inthe analysis.

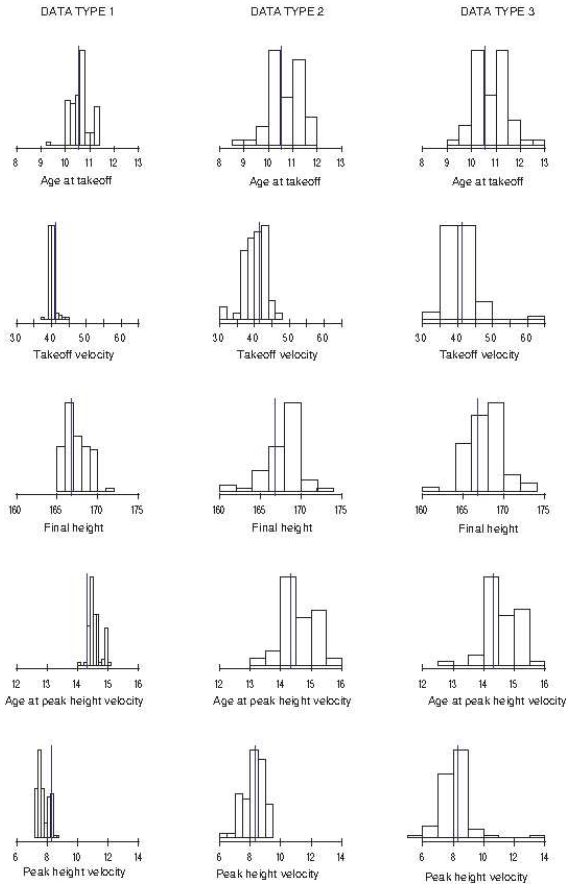

Fig. 1 Histogram of estimated biological parameters compared with true mean, for three types of simulated data

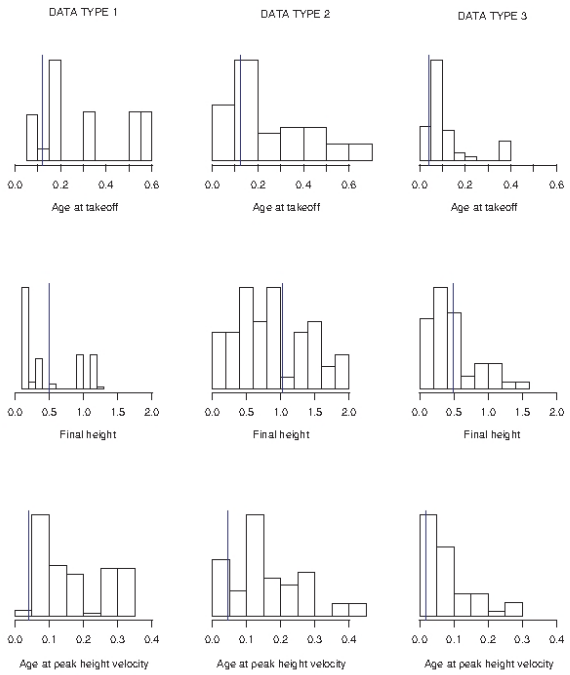

Fig. 2 Histogram of standard errors of estimated biologicalparameters compared with true standard deviation, for types ofsimulated data

TABLE III
MEAN AND STANDARD ERROR OF ESTIMATED BIOLOGICAL AND MATHEMATICAL PARAMETERS, USING FULL DATA AND A PART OF DATA

	Whole data set	Part of data set
Mathematical parameter	Mean (Std err)	Mean (Std err)
s_0 (cm/year)	0.0926 (0.0027)	0.10 (0.012)
s_1 (cm/year)	1.089 (0.029)	1.17 (0.14)
θ (year)	14.72 (0.040)	14.73 (0.15)
h_{maz} (cm)	166.30 (0.125)	166.5 (0.46)
h_θ (cm)	154.36 (0.323)	154.22 (1.34)
Biological parameter	Mean (Std err)	Mean (Std err)
Age at takeoff (year)	10.20 (0.12)	10.49 (0.55)
Takeoff velocity (cm/year)	4.02 (0.052)	4.11 (0.18)
Age at PHV (year)	14.3 (0.042)	14.33 (0.18)
PHV (cm/year)	7.9 (0.07)	8.04 (0.24)
Final height (cm)	166.30 (0.125)	166.5 (0.46)

REFERENCES

[1] Albert, J. *Bayesian Computation with R*, Springer, New York. 2007.
[2] Bates, D.M. and Watts, D.G. *Nonlinear Regression Analysis and Its Applications*. Wiley, New York. 2007.
[3] Cnaan A., Laird, N.M. Laird and Slasor, P. "Using the general linear mixed model to analyseunblanced repeated measures and longitudinal data."*Statist. Med.*, vol. 16, 2349-2380. 1997.
[4] Carlin, B.P., and Louis, T.A. *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman and Hall, Boca Raton, FL. 2000.
[5] Cole, T.J. "Fitting smoothed centile curves to reference data." *J. Roy. Statist. Soc. Ser.A*, vol. 151, 385-418. 1988.
[6] Diggle, P., Heagerty, P., Liang. K.-Y. and Zeger, S.L. *Analysis of Longitudinal Data*. Oxford University Press, Oxford. 2002.
[7] Donnet, S., Foulley, J. L., Samson, A. "A Bayesian analysis of growth curves using mixed models defined by stochastic differential equations." *Biometrics*, vol. 66, 733-741. 2010.
[8] Duncan, T.E., Duncan, S.C., Strycker, L.A., Li, F. and Alpert, A. *An Introduction to Latent Variable Growth Curve Modeling: Concepts, Issues and Applications*, Lawrence Erlbaum Associates, Mahwah, NJ. 2006
[9] Falkner, F. and Tanner, J.M. *Human Growth: A Comprehensive Treatise*, Vol. 3, second edition, Plenum Press, New York. 1986.
[10] Gasser, T., Muller, H.G., Kohler, W., Molinari, L. and Prader, A. "Nonparametric regression analysis of growth curves", *Ann. Statist.*, vol. 12, 210-229. 1984.
[11] Hamill , P.V., Drizd , T.A., Johnson , C.L., Reed, R.B., Roche, A.F. and Moore, W.M. "Physical growth: National Center for Health Statistics percentiles."*Amer. J. Clinical Nutrition*, vol. 32, 607-629. 1979.
[12] Hauspie, R.C., Cameron, N. and Molinari, L. *Methods in Human Growth Research*. Cambridge University Press, Cambridge. 2004.
[13] Huggins R.M. and Loesch D.Z. "On the analysis of mixed longitudinal growth data."*Biometrics*, vol. 54, 583-595. 1998.
[14] Laird, N.M. and Ware, J.H. "Random effects models for longitudinal data: an overview of recent results."*Biometrics*, vol. 38, 963-974. 1982.
[15] McLachlan, G. and Krishnan, T. *The EM Algorithm and Extensions*, Wiley, New York. 1997.
[16] Muller, H.G. *Nonparametric Regression Analysis of Longitudinal Data*. Sringer-Verlag, Berlin. 1988.

[17] Potthoff, F. and Roy, S.N. "A generalized multivariate analysis of variance model useful specially for growth curve problems." *Biometrika*, vol. 51, 313-326. 1964.
[18] Preece M.A. and Baines M.J. "A new family of mathematical models describing the human growth curve."*Annals of Human Biology*, vol. 5, 1-24. 1978.
[19] Rao, C.R. "The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves."*Biometrika*, vol. 52, 447-458. 1965.
[20] Staniswalis, J.G. and Lee, J.J. "Nonparametric regression analysis of longitudinal data."*J. Amer. Statist. Assoc*., vol. 93, 1403-1418. 1998.
[21] Wei, Y., Pere, A., Koenker, R. and He, X. "Quantile regression methods for reference growth charts."*Statist. Medicine*, vol. 25, 1369-1382. 2006.
[22] Wulfsohn, M.S. Tsiatis, A.A. "A joint model for survival and longitudinal data measured with error."*Biometrics*, vol. 53, 330-339. 1997.
[23] Zemel, B.S. and Johnston, F.E. "Application of the Preece-Baines growth model to cross-sectional data: Problems of validity and interpretation."*Amer. J. Human Biol.*, vol. 6, 563-570. 1994.

Debasis Sengupta is a Professor in the Applied Statistical Unitof the Indian Statistical Institute, Kolkata. He did his Ph.D. degreesin Electrical and Computer Engineering and in Statistics from theUniversity of California, Santa Barbara. His research interests includemodeling and analysis of lifetime data, linear models and regression,multivariate analysis, statistical signal processing, and applications ofstatistics in various fields.

Sedigheh Mirzaei S. is a Senior Research Fellow in the AppliedStatistical Unit of Indian Statistical Institute, Kolkata 700108, INDIA.She did her M.Sc. (Statistics) degrees from the Isfahan Universityof Technology, Iran. Her research interests include growth models,regression and linear models, survival analysis, and computational statistics.