Flexible Statistical Models for Growth Fragments: a Study of Bone Mineral Acquisition

Trevor J. Hastie [*] Laura K. Bachrach [†]

Balasubramanian Narasimhan [‡] May Choo Wang [§]

December 31, 1999

Abstract

We describe some statistical modeling of longitudinal data from a pediatric study of bone mineral acquisition in 423 adolescents in the San Francisco Bay area (Bachrach et al. 1999). Several different measurements of bone mineral density are available for each subject, taken roughly one year apart, and between one and four measurements per

[*]Department of Statistics and Division of Biostatistics, Stanford, University, Stanford CA 94305; hastie@Stat.Stanford.Edu This work was partially supported by grant DMS-9803645 from the National Science Foundation, and grant ROI-CA-72028-01 from the National Institutes of Health.

[†]Departments of Pediatrics, Stanford Medical Center, S302, Stanford University, Stanford University, CA94305; lkbach@Leland.Stanford.Edu

[‡]Department of Statistics, Stanford University; naras@Stat.Stanford.Edu

[§]Department of Nutrition, San Jose State University

subject. In this paper we discuss models for representing the growth curves for sub-populations (mean curves and percentiles), and present a simple model for the tempo of growth. Our approach makes use of mixed-effects models for longitudinal data, regression splines, generalized additive models and the bootstrap. We show Splus code for fitting our models, and have created a web page for interacting with the growth curves.

Keywords: Growth curves, Splines, Local Regression, Mixed-Effects, Bootstrap

1 Introduction

Osteoporosis has reached epidemic proportions in the United States with yearly expenditures of \$13.8 billion to care for pathologic fractures from bone fragility (Ray et al. 1997). Since there is no cure for osteoporosis once it has developed, there has been increasing focus on means to prevent the disorder. One key determinant of the lifetime risk of osteoporosis is the amount of bone mineral acquired by early adulthood, termed peak bone mass. Osteoporosis prevention begins by fostering early gains in bone mineral in order to maximize peak bone mass. This has led to increased interest in the process of bone mineral acquisition during childhood and adolescence. Several studies have reported age-related gains in bone mineral using a non-invasive technique for measuring bone mineral called dual energy x-ray absorptiometry (DXA). To date, most of these investigations have been cross-sectional studies in predominantly white and female cohorts (Bonjour 1996). The goals of this study were to define the longitudinal gains in bone mineral in an ethnically diverse cohort of healthy youth and to determine if there were gender and/or ethnic differences in the magnitude and tempo of bone mineral acquisition.

Measurement of bone mineral in children and adolescents is more complex than in adults because of changes in bone size and puberty. The marked increases in bone size during adolescent growth influence DXA measurements of bone mineral. Bone mass measurements using DXA are reported either as bone mineral content (BMC), the grams of bone mineral within region studied or as bone mineral density (BMD), which normalizes BMC for the projected area (gm/cm^2). Both BMC and BMD overestimate the bone mineral of large bones and underestimate the bone mineral of smaller bones. To address this problem, the Pediatrics Department at Stanford University has developed a model to estimate volumetric bone density which takes into account the thickness of bone (Carter et al. 1992). Although these estimates are only approximations of true bone volume, they represent a reasonable approach when comparing the bone mineral status in children of different size. Rates of bone growth and mineral acquisition are more closely linked to pubertal development than to chronologic age (Bonjour 1996), gender, ethnicity and individual differences determine the timing of pubertal development which influences growth rates for bone size and mineral content. These differences in tempo add to the complexity of modeling bone growth curves.

While the data are sufficient to model the mean growth curves in subpopulations defined by gender and ethnicity, care needs to be taken in modeling the growth velocity. In this paper we develop a growth curve model for each of the BMD measurements as a function of age, ethnicity and gender. The mean curves differ in shape for males and females, and within gender classes, there are vertical shifts due to ethnic classes. We perform tests of ethnic differences that take into account the correlation in the repeated measurements. The end product for each group is a growth curve with pointwise standard error curves, which allow one to calculate Z-scores for any future measurements. We also present a simple model for representing the differences in the growth velocity, broken down by the various groups. This paper describes in detail the methodlogy behind the results reported in Bachrach et al. (1999).

Figures 1 (females) and 2 (males) show the final models selected for Spinal BMD. The percentile curves in each case are labeled with the appropriate normal score (Z-score).

A more comprehensive collection of these growth curve models are available online at `http://www-stat.stanford.edu/pediatric-bones`, where similar percentile curves are shown for measurements taken at different sites, as well as volumetric measurements. This is an interactive website, that allows users to include their own data on the plot, extract Z-scores, and print the personalized growth curves.

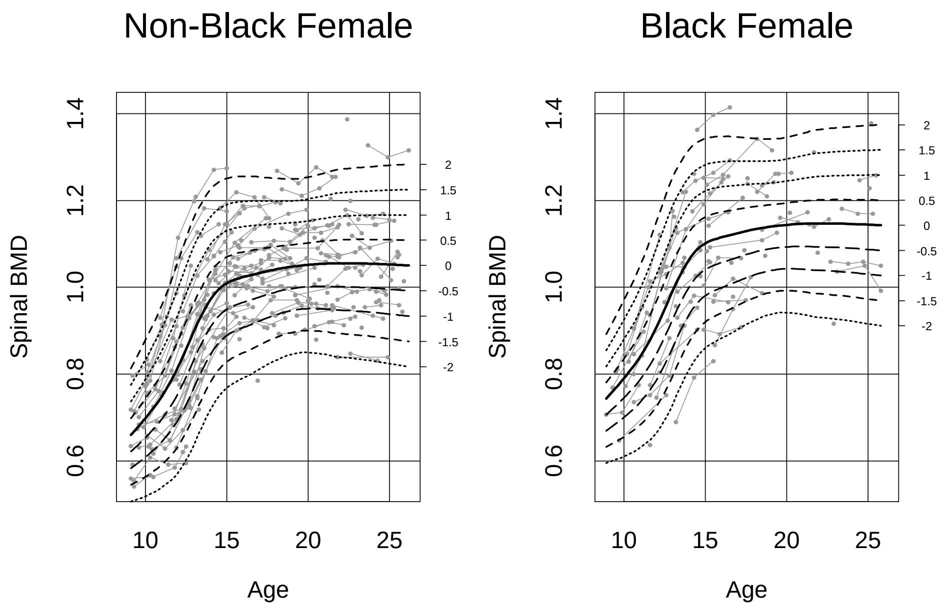

Figure 1: The mean curve and several selected quantiles for females, separately for Blacks and non-Blacks. The mean curves differ ($p < 0.05$) by a level shift, and the quantile curves are computed from the pooled residuals for the two groups. The quantiles are labeled in terms of Z-scores. The data for each subject are connected by line segments.

Figure 3 lists all the measurements modeled, and shows the gender and ethnic effects.

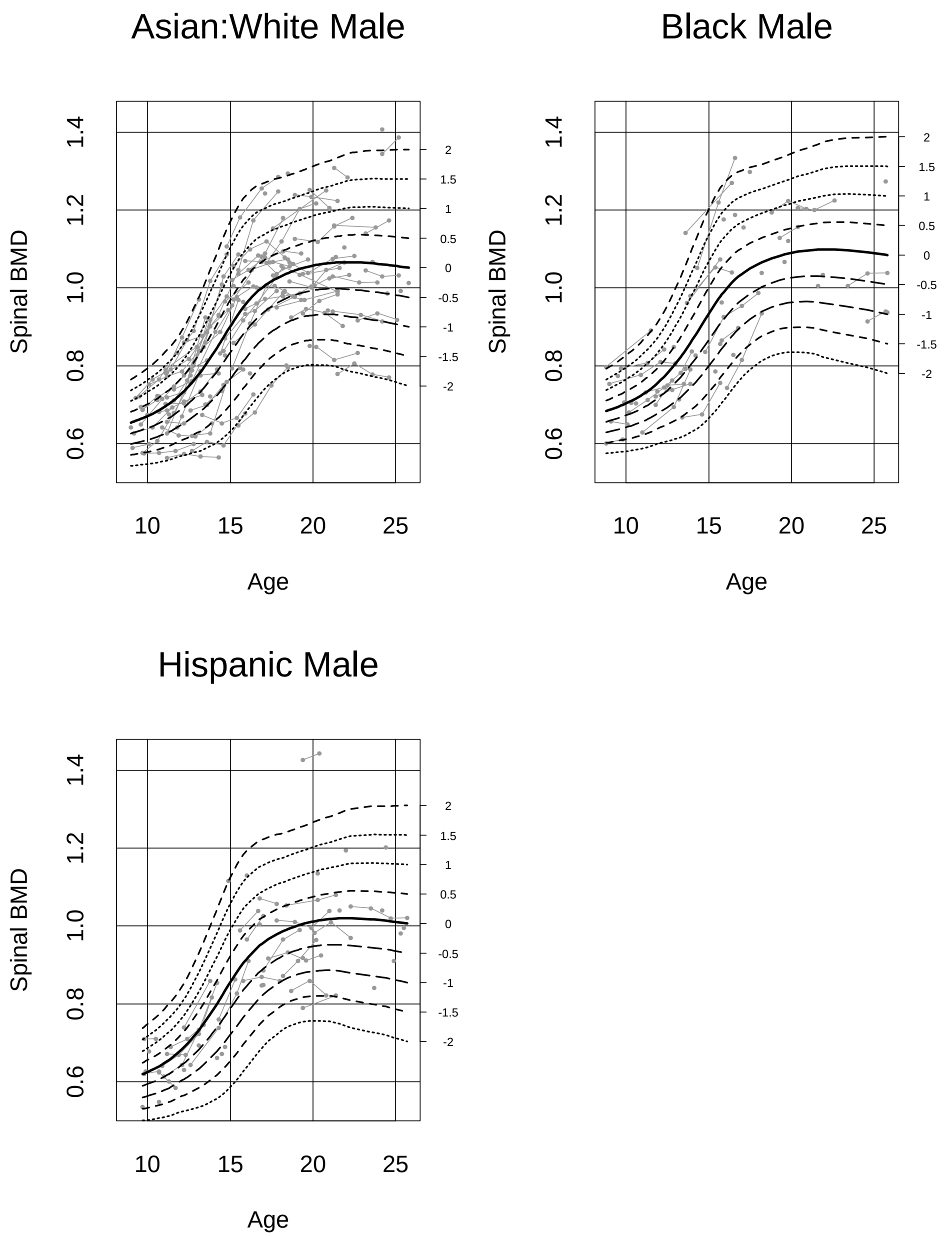

Figure 2: The mean curve and several selected quantiles for males, separately for Blacks, Whites/Hispanics and Asians. The mean curves differ ($p < 0.05$) by a level shift, and the quantile curves are computed from the pooled residuals for the three groups. The quantiles are labeled in terms of Z-scores.

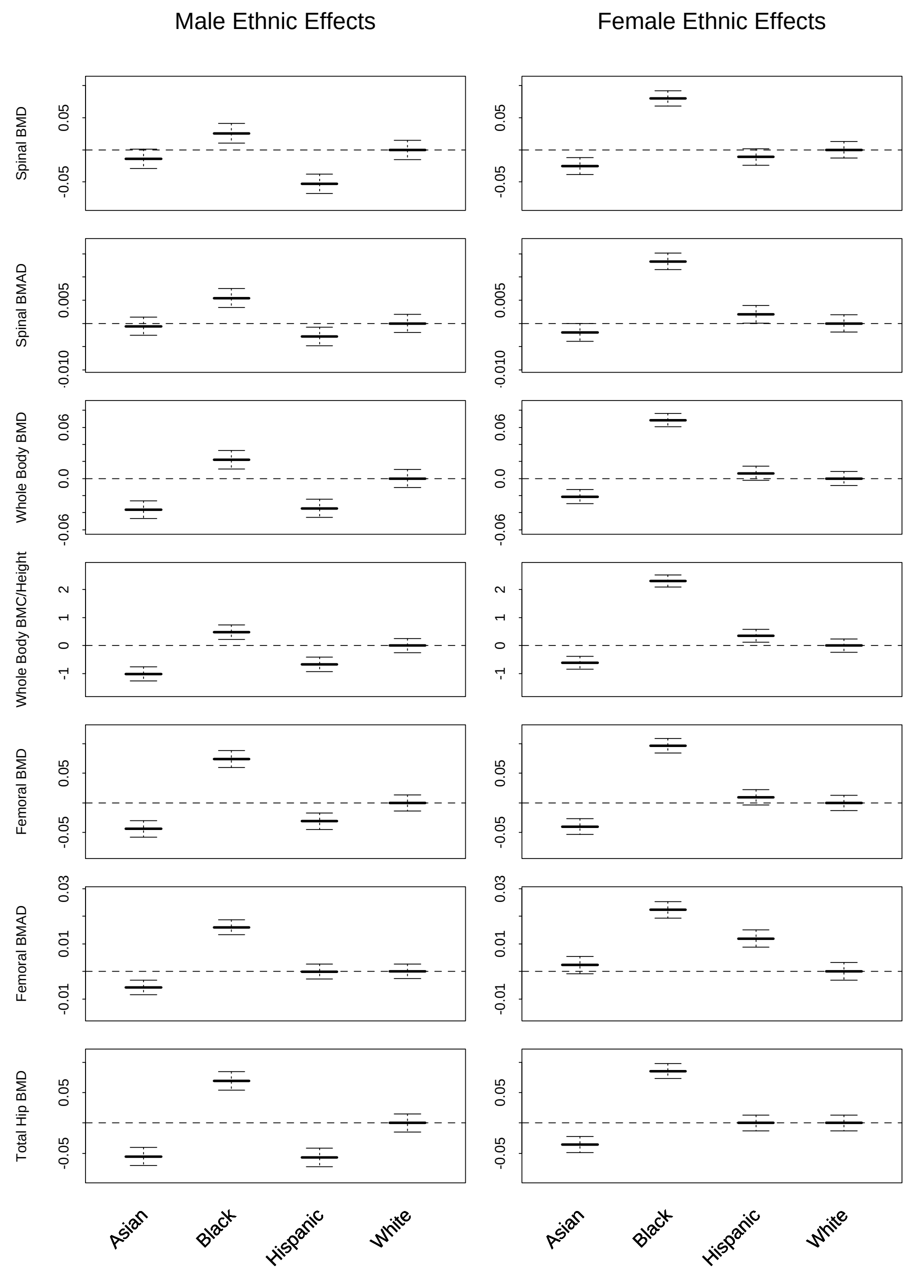

Figure 3: Ethnic effects for all the BMD measurements. For each panel, the shape of the mean growth curve is the same, but the level differs by ethnicity. We see the estimated effects, with a standard error bar included for comparison. This allows us to see at a glance the level shifts in each of the curves.

2 Statistical Models and Methods

The complete set of measurements for Spinal BMD for females are shown in figure 1 and for males in figure 2. The ethnic groupings were determined using the models we describe below. The repeated measurements by age per subject are shown as connected fragments. The breakdown of repeats by gender are given in the table below.

Gender/Repeats	1	2	3	4
Female	77	48	46	59
Male	66	44	30	53

The methods we have used to model the growth curves and produce the percentile curves are a marriage of several popular classes of techniques in applied biostatistics:

- *Mixed Effects* models allow one to properly account for repeated measurements on an individual, such as those arising in a longitudinal growth study. In particular, they allow for accurate estimation of the standard errors of the fixed effects in the model.

- *Semi-parametric regression models* allow for flexible modeling of the mean growth curves using splines.

- *Robust non-parametric smoothing* techniques allow us to estimate the standard deviation as a function of age. We give a brief description of each of these in turn.

2.1 Mixed Effects Models

Mixed-effects models have gained popularity in the last 10 years as a simple way to account for repeated and hence correlated measurements in a longitudinal study. Ignoring these correlations might have little or no effect on the estimated fixed effects such as ethnic or gender differences, but can have disastrous consequences on their estimated standard errors and the subsequent inferences (Diggle et al. 1994). Our model for any one of the bone mineral measures, separately for males and females, has the form

$$y_{ij} = f(t_{ij}) + \beta_{E_i} + U_i + \epsilon_{ij} \tag{1}$$

where

- y_{ij} is the jth measurement on subject i made at time or age t_{ij}.
- $f(t)$ is the *fixed effect*, which in this case is the mean growth curve for all subjects. We use a parametric representation $f(t) = \sum_k \eta_k(t)\beta_k$ to represent $f(t)$, where the $\eta_k(t)$ are fixed basis function in age; see the next section for details.
- β_{E_i} is a shift due to ethnicity. One can also model a separate curve for each ethnic group. We considered such models as well, but found that in no cases were their sufficient data to find significant differences with so many parameters involved. This shift model is a parsimonious compromise, and in section 3 we consider a different parsimonious model for representing tempo and onset of the growth spurt.

- U_i is a random effect associated with the ith subject, and is common and fixed for all measurements on this subject. It may just happen that subject i is much larger than all the other subjects, and *all* the measurements made at different times on subject i will be larger. U_i is intended to account for such an effect.

- ϵ_{ij} is the error term, which we assume to be serially-correlated within subjects. Since the measurements are made at irregularly spaced time points, we assume a *continuous AR* correlation function, which assumes

$$\operatorname{cov}(\epsilon_{ij}, \epsilon_{ik}) = \sigma^2 \rho^{|t_{ij} - t_{ik}|} \tag{2}$$

  where ρ is the correlation of measurements one unit apart.

In this model we assume that each individual follows the average growth curve shifted up or down by a constant depending on their ethnic class and their particular U_i. In addition the repeated measurements are subject to measurement error and other random fluctuations; these are assumed to be correlated within a subject, and the correlations die off the further apart in age the measurements occur.

We can think of the error term in the model as being the compound expression $U_i + \epsilon_{ij}$, and thus providing a way of capturing the correlation in the measurements. Typically we assume that $U_i \sim N(0, \sigma_R^2)$ i.i.d. across subjects, and $\epsilon_{ij} \sim N(0, \sigma^2)$ for all measurements, with correlations as specified above. So for example, if subject i has three measurements at t_{i1}, t_{i2}

and t_{i3}, then the three-component covariance matrix block for subject i has the form

$$\Sigma_i = \sigma_R^2 \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} + \sigma^2 \begin{pmatrix} 1 & \rho^{|t_{i2}-t_{i1}|} & \rho^{|t_{i3}-t_{i1}|} \\ \rho^{|t_{i2}-t_{i1}|} & 1 & \rho^{|t_{i3}-t_{i2}|} \\ \rho^{|t_{i3}-t_{i1}|} & \rho^{|t_{i3}-t_{i2}|} & 1 \end{pmatrix} \tag{3}$$

It may seem that there is redundancy in having both a random subject specific intercept U_i and a serially correlated error term ϵ_{ij}. Both represent correlation in repeated measurements, but in different ways. Having them both tends to damp the correlation decay of the AR error covariance. We found that in general this model seemed to fit better than if one of the terms were omitted, and the only down side is a single extra parameter (since we do not try and interpret the variance parameters).

The models are fit by maximum likelihood, which includes both the parameters of the fixed effects and of the variance functions.

There are good reasons for going to all this trouble to model the errors accurately. We wish to test for ethnic and other differences in the fixed effects. By properly modeling the correlation in the observations, we get more accurate tests. For example, many highly correlated observations on an individual should carry diminished weight, since together they do not tell us much more than just one of them. We used likelihood ratio tests for these purposes.

Mixed-effects models are described in detail and clearly in Diggle et al. (1994), which contains many useful references to original sources.

2.2 Cubic Regression Splines

Our model for $f(t)$ has the form

$$f(t) = \sum_{k=1}^{K} \eta_k(t)\beta_k \tag{4}$$

where the $\eta_k(t)$ are *natural cubic-spline basis functions* with pre-chosen fixed *knots*. Perhaps a more familiar basis expansion has $\eta_k(t) = t^k$ as in polynomial regression. Our spline bases allows for more local behavior than polynomials by specifying a number of knots or breakpoints. Between any adjacent pair of knots the model is a cubic-polynomial, with constraints built in to force the separate cubic polynomial to join up smoothly. For example, our models have knots in age at 12, 14, 16 and 18 years. The constraints help control the number of parameters, in our case six altogether (in addition we assume the "natural" boundary conditions; the model is linear beyond the range of the data). For further details see Hastie & Tibshirani (1990) or Alzola & Harrell (1999). We chose the knots subjectively to cover the region where the curves visually changed the most. Once the basis functions are selected, the model can be fit just like any other linear model, since it is linear in the parameters. Her the fitting is done on the context of the mixed-effects model. Figure 4 illustrates the idea.

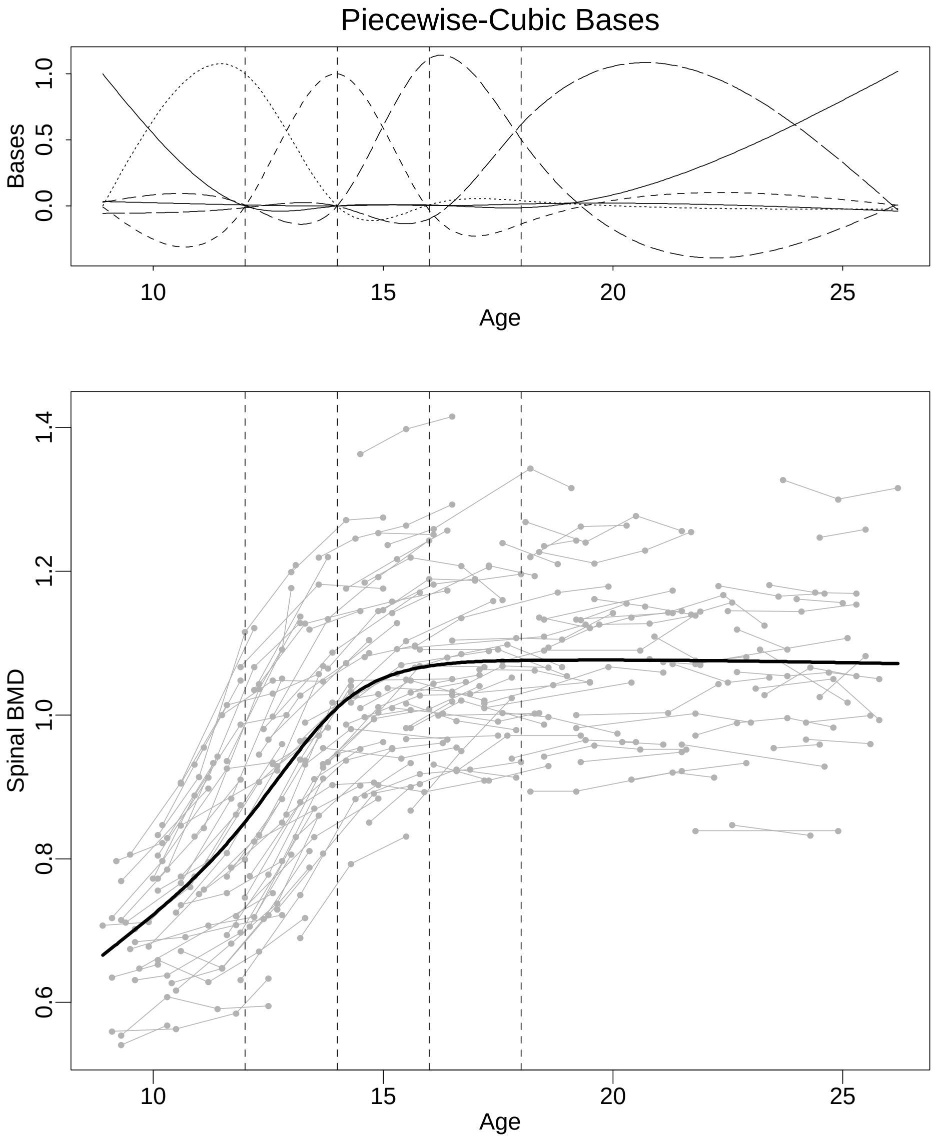

Figure 4: The top panel shows a basis of 6 natural cubic splines with knot locations indicated by the vertical dashed lines. The lower panel shows the regression fit to the female Spinal BMD data using this bases. Notice that in the region around the growth spurt we have placed several knots, allowing for more flexible behavior in the fitted model

2.3 Robust non-parametric smoothing for standard deviations

While it is possible to model the standard deviation (changing with age) and the mean simultaneously, we found this too cumbersome. Our standard deviation curves are based on the squared residuals from the mean model. We used a robust locally weighted linear regression estimate, as described in Cleveland (1979). This procedure allows one to estimate the variance in smoothly overlapping windows in age, and omit outlying observation which would otherwise exaggerate any estimate of variance.

Since variances are positive, and squared residuals tend to behave like Chi-squared quantities, we avoided standard least squares estimation. Instead we fit a *Generalized Additive Model* to the squared residuals as a function of age. We used a *Gamma* error model (Chi-square is a member of the Gamma family) with *log* link, and local regression was chosen as the method of smoothing. In addition, a *robust* version of maximum-likelihood was used to fit the model (Hastie & Tibshirani 1990, Chambers & Hastie 1991). More details of this implementation follow below.

2.4 Fitting the Models

The mixed-effects models are fit by *maximum likelihood* estimation, the details of which are beyond the scope of this article. Essentially the (marginal) gaussian likelihood is used both to estimate the fixed-effect coefficients β and the variance parameters ρ and σ^2. The random-effects part of the model implies a covariance matrix for the entire response vector, which will be block diagonal to account for the correlations among the repeated measurements per subject. For any given estimate of the covariance matrix, the fixed effects can be estimated by generalized least squares. The EM algorithm is used to alternate between updating the covariance matrix and estimating the parameters of the fixed effects. We used the linear mixed effects (LME) software in Splus (Pinheiro & Bates 1995), which is convenient for this type of modeling. Here follows a fragment of Splus code for fitting and comparing models leading up to figure 1:

```
fit1 <- lme(fixed=spnbmd ~  ns(age, knots = c(12, 14, 16, 18)),
     random=~1,  cluster=~idnum, data=BMA.females,
      serial.structure = "ar1.continuous",
      serial.covariate = ~age, est.method="ML")

fit2 <- update(fit1,
     fixed=spnbmd ~ ethnic + ns(age, knots = c(12, 14, 16, 18)))

fit3 <- update(fit1,
     fixed=spnbmd ~ ethnic * ns(age, knots = c(12, 14, 16, 18)))

anova(fit1, fit2, fit3)
```

Here we compare no ethnic effects (`fit1`), separate ethnic levels (`fit2`) as in (1), and separate ethnic curves (`fit3`) in models for Spinal BMD for females. The `random=~1` term calls for a random constant U_i per subject, and the individual subjects are identified by `cluster=~idnum`.

The (abbreviated) output of the **anova** follows:

```
     Model Df     AIC     BIC Loglik    Test Lik.Ratio  P value
fit1     1  9 -1536.0 -1497.2 776.99
fit2     2 12 -1561.6 -1509.9 792.79 1 vs. 2    31.599 0.000001
fit3     3 27 -1554.9 -1438.7 804.47 2 vs. 3    23.351 0.076973
```

Reported are the maximized marginal likelihoods. Although model `fit3` has the highest likelihood, it uses 15 parameters more than `fit2`. The model `fit2` (separate ethnic levels) is chosen by the minimum BIC (Bayesian Information Criterion), AIC (Akaike information criterion) and likelihood-ratio tests.

We fit similar models to all the measurements, separately for males and females. For each measurement we tested the effect of ethnicity (Asian, Black, Hispanic or White), and found that for different responses certain ethnic groupings were reasonable (typically Black versus non-Black for females). Figure 3 shows simultaneously all the ethnic level effects with standard-error bars. These effects and standard errors are easily derived from the estimated fixed effects and their covariance matrix, available from the fitted LME models.

The standard deviation curves were computed using the following Splus code

```
Resids <- residuals(fit2)$population
sd.fit <- gam(I(Resids^2) ~ lo(age, span = 0.4),
      family = robust(Gamma(link = log)), data = BMA.females)
sd.curve <- sqrt(fitted(sd.fit))
```

The `family = robust(Gamma(link = log))` argument to `gam()` specifies three things: the Gamma family is to be used, the mean (of the squared residuals) is to be modelled on the log scale, and a robust version of the likelihood is to be used in the estimation. The latter ensures that excessively extreme residuals are downweighted in the estimation. The model formula `~ lo(age, span = 0.4)` specifies that local regression should be used to fit the log-mean-squared residuals as a function of age, and that at each point the support of the kernel should be 40% of the data. Note that a single standard-deviation curve is fit for all ethnic groups, since there is insufficient data in some subclasses to estimate these separately. We did entertain simple modifications; for example

```
I(Resids^2) ~ lo(age, span = 0.4) + ethnic
```

calls for the same shape curve for each ethnic group, but each scaled by a different positive factor. In all cases the `ethnic` term did not make a significant difference.

The Z-curves are simply given by

$$f_\alpha(t) = \hat{f}(t) + \alpha\hat{\sigma}(t) \tag{5}$$

where $f(t)$ is the estimated mean curve, $\sigma(t)$ the standard deviation curve (`sd.curve`), and α the appropriate quantile of a standard normal distribution.

3 Modeling Tempo of Growth

There is some evidence in the data that differences exist in the *tempo* of bone growth when stratified by gender and ethnicity. This will be exhibited in the rate of bone acquisition, and the age at which bone growth ceases. Although one can imagine rather exotic models to investigate these phenomena, we are limited by having short fragments of data for each individual. We use the following simple model to represent this growth in each ethnic/gender modality:

$$y_{ij} = \beta_0 + \beta_1(t_{ij} - \theta)_- + U_i + \epsilon_{ij} \tag{6}$$

where $(t-\theta)_-$ is the *negative part* of $t-\theta$ ($t-\theta$ if $t-\theta < 0$, else 0). This model is linear in age up to θ, and then constant after θ. All three parameters are of interest, and in the context of this rather crude model, measure:

β_0: the final level achieved; we expect this to differ across ethnic groups and sex based on the previous models.

β_1: the rate of bone acquisition.

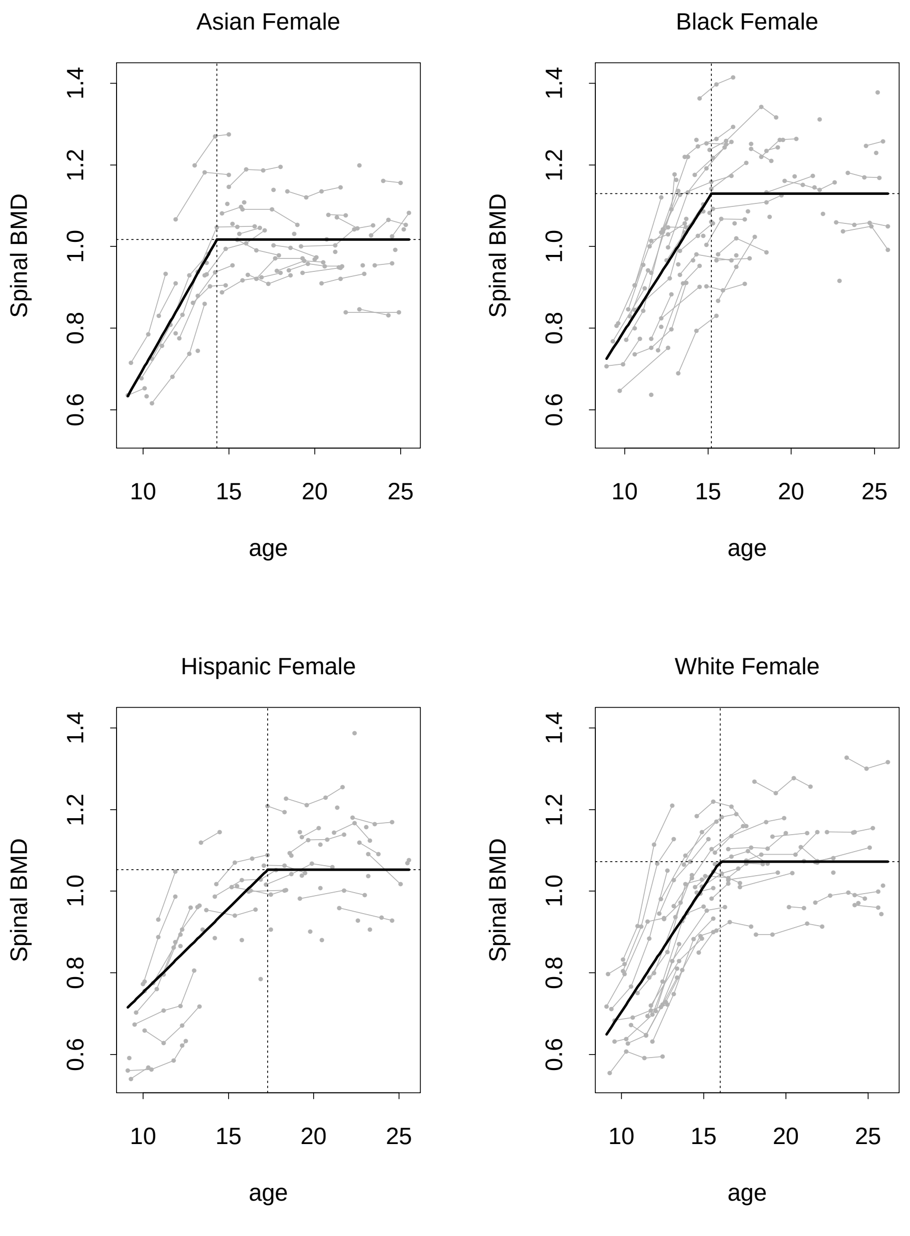

Figure 5: Piecewise linear model to capture growth velocity and stage for female Spinal BMD

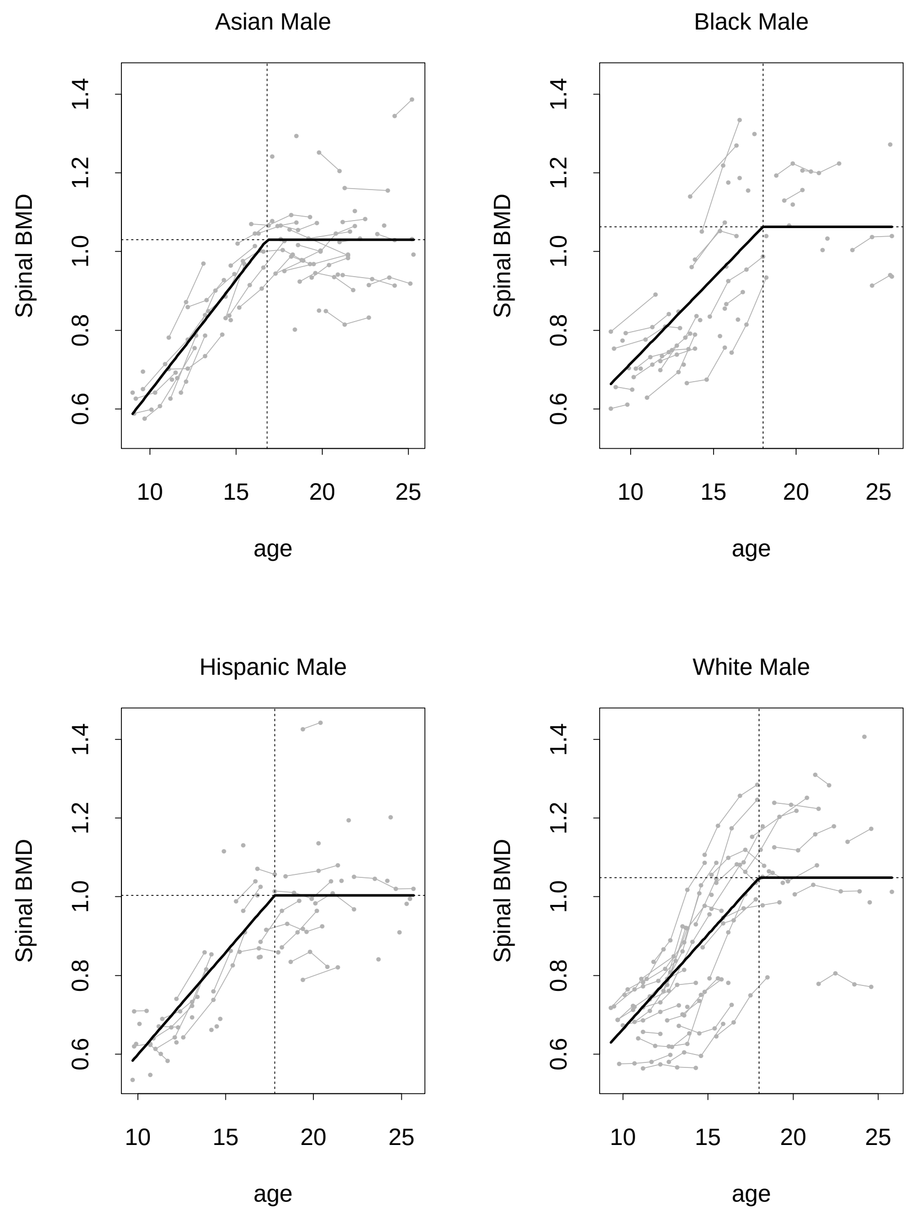

Figure 6: Piecewise linear model to capture growth velocity and stage for male Spinal BMD

θ: the age at which bone acquisition stops.

This piecewise-linear model is also a spline (with a single knot θ), but a degree-1 rather than the degree-3 splines used earlier. Figures 5 and 6 show these models fit to the spinal BMD data. For fixed θ, the model (6) is a special case of (1) using a single very simple basis function $\eta_\theta(t) = (t - \theta)_-$. So for each fixed value of θ, the model can be fit as before, by maximum likelihood using LME. The resulting *profile* log-likelihood is maximized by a grid search for the optimum value of θ. Because of the adaptive nature of the search for θ, the distribution theory used in linear mixed effects models no longer applies to this situation. We used the *bootstrap* (Efron & Tibshirani 1993) to asses the sampling variability of our three parameter estimates.

Separately in each ethnic/sex modality, we created $B = 200$ bootstrap datasets by sampling the individuals *with replacement* (to preserve the within subject correlation). For each of these bootstrap samples we fit the model (6) and recorded the parameters. Note that we preserve the correlation structure by sampling individuals rather than observations, since we keep the repeated measurements for each subject intact. We thus get a bootstrap distribution for each parameter, which can then be used to compare groups, produce confidence intervals and so on. We provide a display of the parameter θ for each of the ethnic groups in figure 7, and include it's bootstrap standard error for pairwise comparisons. From the figure it can be seen that Asians tend to terminate growth earlier than other ethnic groups, especially for Spinal

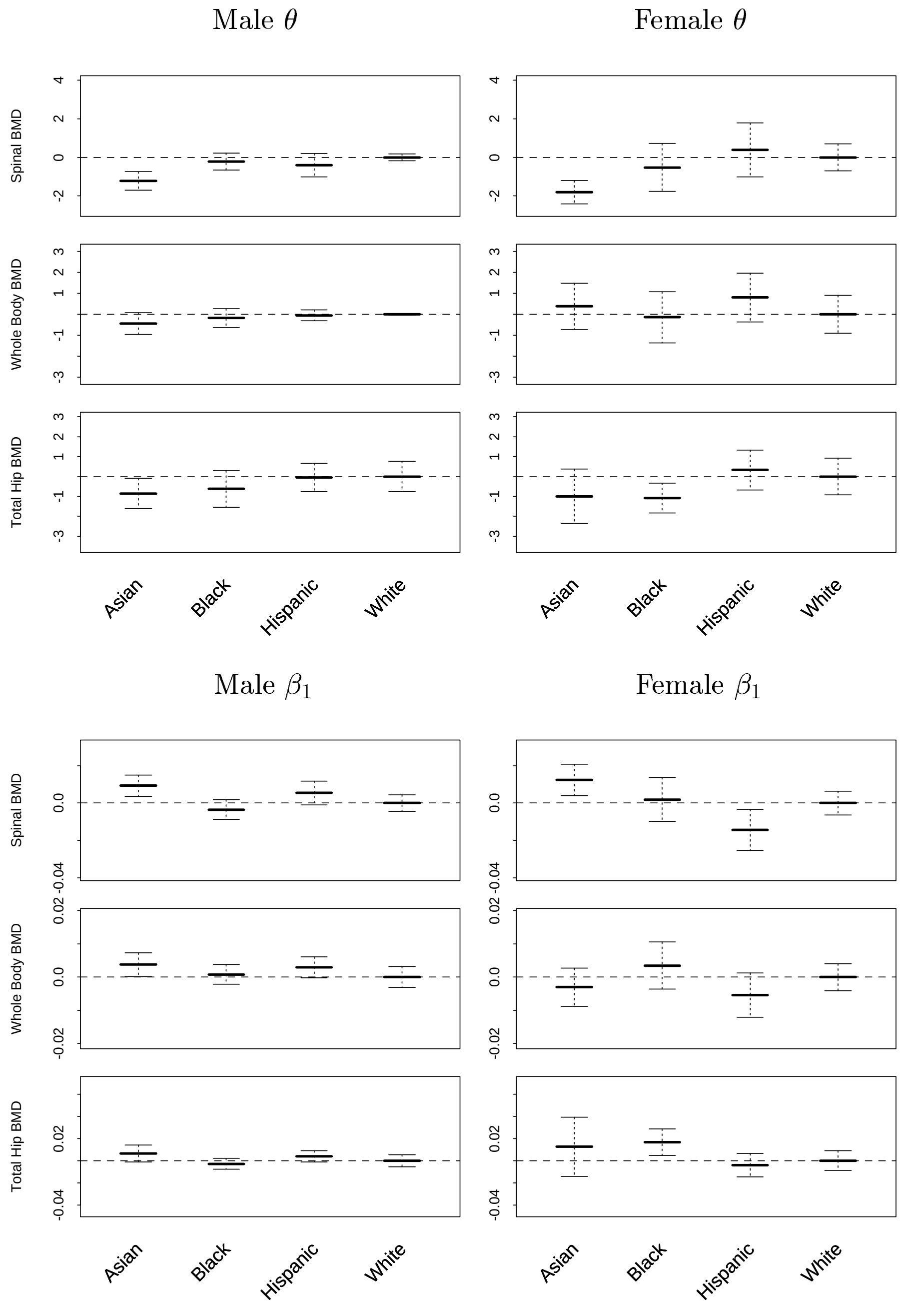

Figure 7: Top panel: A summary of the bootstrap distribution of the location of the *knot* θ in the piecewise linear curve. The position of the knot provides an estimate for the subgroup of the point at which bone-growth stops. We show the mean value $\pm$ one bootstrap standard error. Lower Panel: The same for the slope β_1

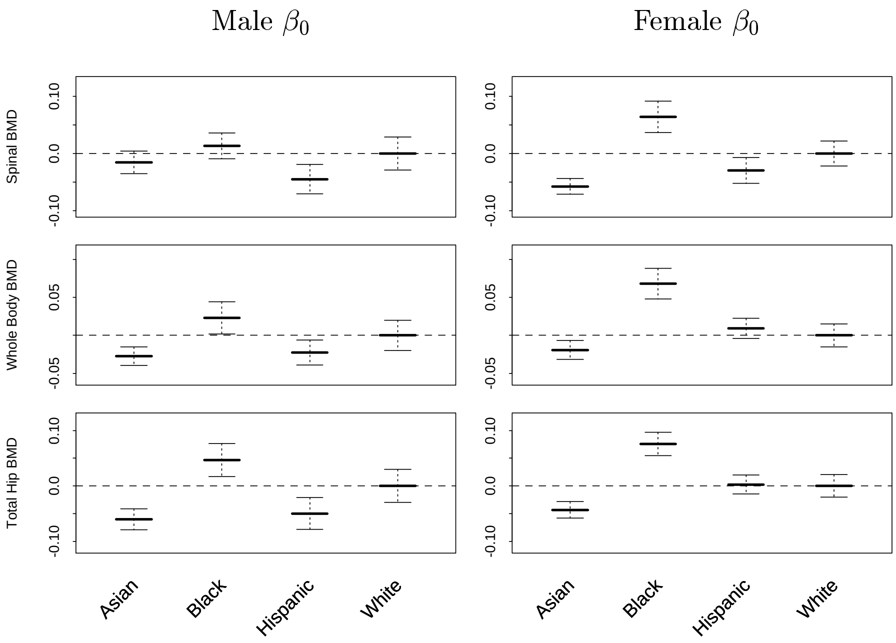

Figure 8: *Bootstrap distribution of the level* β_0 *in model (6).*

BMD. The lower panel in figure 7 shows the slope β_1, which is negatively correlated with θ.

4 Discussion

In this paper we have combined some modern techniques to produce a flexible model for representing growth curves. The tools we combined were mixed effects models for longitudinal data, regression splines for flexible function fitting, and generalized additive models and local regression for flexibly estimating the standard deviations. In addition we used the bootstrap for inference. These models can be fit rather simply using existing tools in Splus.

The bootstrap can be used even in cases where the theory does apply, as in the inference for the ethnicity effects in model (1). As an experiment (not shown), we used the bootstrap to compute the standard errors for the ethnic effects for Spinal BMD, and got very similar results to those reported in figure 3. The piecewise linear model (6) has a level effect, β_0. Figure 8 shows the bootstrap mean and standard error for three different measurements. These are to be compared with the corresponding panels in figure 3; the estimates and standard errors are very similar.

References

Alzola, C. & Harrell, F. (1999), *An Introduction to S-Plus and the Hmisc and Design Libraries*, Department of Biostatistics and Epidemiology, University of Virginia. Preprint.

Bachrach, L., Hastie, T., Wang, M.-C., Narasimhan, B. & Marcus, R. (1999), 'Bone mineral acquisition in healthy asian, hispanic, black and caucasian youth. a longitudinal study.', *J. Clin. Endocrinol. Metab.* **84**, 4702–4712.

Bonjour, J. (1996), Bone mineral acquisition in adolescence., *in* M. R., F. D. & K. J., eds, 'Osteoporosis', Academic Press, San Diego, pp. 465–476.

Carter, D., Bouxsein, M. & Marcus, R. (1992), 'New approaches for interpreting projected bone densitometry data.', *Journal of Bone Mineral Research* **7**, 137–145.

Chambers, J. & Hastie, T. (1991), *Statistical Models in S*, Wadsworth/Brooks Cole, Pacific Grove, California.

Cleveland, W. (1979), 'Robust locally-weighted regression and smoothing scatterplots', *Journal of the American Statistical Society* **74**, 829–836.

Diggle, P., Liang, K. & Zeger, S. (1994), *Analysis of Longitudinal Data*, Oxford Science Publications.

Efron, B. & Tibshirani, R. (1993), *An Introduction to the Bootstrap*, Chapman and Hall.

Hastie, T. & Tibshirani, R. (1990), *Generalized Additive Models*, Chapman and Hall.

Pinheiro, J. & Bates, D. (1995), *Mixed Effects Models Methods and Classes for S and Splus*, Department of Statistics, University of Wisconsin.

Ray, N., Chan, J., Thamer, M. & Melton, L. (1997), 'Medical expenditures for the treatment of osteoporotic fractures in the united states in 1995: Report from the national osteoporosis foundation.', *Journal of Bone Mineral Research* **12**, 24–35.