ANNALS OF HUMAN BIOLOGY, 1992, VOL. 19, NO. 4, 337–346

A two-level cross-sectional model using grafted polynomials

H. Q. PAN†, H. GOLDSTEIN‡ and GUO DI†

†WHO Collaborating Centre for Physical Growth and Psychosocial Development of Children, Shanghai Second Medical University, Shanghai, China

‡Department of Mathematics, Statistics and Computing, Institute of Education, University of London

Received December 15 1990; revised May 1 1991

Summary. A new statistical model is proposed for the analysis of hierarchically structured cross-sectional growth data, especially for where measurements are made over long age ranges. The model combines a two-level model with grafted piecewise polynomials, to make efficient use of available data.

1. Introduction

Two approaches to the modelling of longitudinal repeated measures data—polynomial growth curves and fixed-occasion models—were proposed by Goldstein (1986, 1987, 1989). They utilize a two-level framework in which the individual is considered to be a level 2 unit and measurement occasions within an individual are level 1 units. Many growth studies, however, involve cross-sectional data with a hierarchical structure which consists of measurements of children at different ages within geographical or other clusterings. There has been no attempt to use multilevel models to study cross-sectional growth data for clustered populations, although the existence of hierarchical population structures will generally require such models to be used.

The example given in this paper consists of 8971 weight measurements of children from birth to 6 years old in 18 subdistricts in the middle east and southeast of China. A two-level model will be employed where the level 2 unit is a subdistrict and the level 1 unit is a child. If the subdistricts were further clustered within higher-level units then these would constitute a third level of the hierarchy.

First we are interested in the average growth trajectory and the variation in growth trajectories between subdistricts or clusters. Secondly we are interested in the factors which may account for this variation. Thirdly we wish to study the structure of the between-child variation in growth, and the factors which might explain that.

2. The two-level growth model

2.1. *Piecewise polynomials*

The use of a single curve to describe growth over a long time period raises difficult problems. If it is a polynomial it will typically need to be of high order and will often fit badly at some ages, especially at the extremes of the range. If it is a nonlinear curve it will typically require a large number of parameters, which makes estimation difficult, or else may introduce fixed relationships between growth events which are unrealistic (Goldstein 1979). In the present paper we consider the use of piecewise polynomials; that is, the smooth joining of separate low-order polynomials, which are joined together with the property that the derivatives of a desired order are continuous. We shall be concerned here with the following fixed-order piecewise polynomials of degree

p. These are a special case of general spline functions (Cox 1971) and are also known as 'grafted polynomials' (Fuller 1976). We write

$$f(t) = \sum_{s=0}^{p} \beta_s t^s + \sum_{j=1}^{m-1} c_j (t - \zeta_j)_+^p \tag{1}$$

where

$$(t-\zeta_j)_+^p = \begin{cases} 0, & t < \zeta_j \\ (t-\zeta_j)^p, & t \geq \zeta_j \end{cases}$$

The values ζ_j, $\zeta_1 < \zeta_2 < \ldots < \zeta_{m-1}$ are termed join points and $f(t)$ is a polynomial of degree p within the intervals $t \leq \zeta_1$; $\zeta_{j-1} \leq t \leq \zeta_j$, $j = 2,3,\ldots, m-1$, $\zeta_{m-1} \leq t$. Single polynomials are special cases of the above functions when $m = 1$. The value of p is typically 3 or 5 for human growth data. For example, when one join point ζ_1 and polynomials of degree 3 we obtain

$$f(t) = \begin{cases} \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 & t < \zeta_1 \\ \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (t - \zeta_1)^3 & t \geq \zeta_1 \end{cases}$$

Such functions are useful for many purposes on the grounds of simplicity and flexibility, and possess the property of being continuous in as many derivatives as possible without degenerating into simple polynomials. In addition the form of these functions makes them suitable for use in multilevel models.

2.2. *A basic two-level model*

Let person i in subdistrict j be measured on response variable Y at age t. For subdistrict or cluster j, we write the basic grafted polynomial of degree p with $m-1$ join points as follows

$$Y_{ij} = \beta_{0ij} + \beta_{1j} t_{ij} + \beta_{2j} t_{ij}^2 + \ldots + \beta_{pj} t_{ij}^p + \beta_{p+1,j}(t_{ij} - \zeta_1)_+^p + \ldots + \beta_{p+m-1,j}(t_{ij} - \zeta_{m-1})_+^p$$
$$\beta_{0ij} = \beta_{0j} + e_{ij} \tag{2}$$

In this model, the term e_{ij} is referred to as the level 1 'residual' for the ith child in the jth cluster. The coefficients of the intercept β_{0j} and slope β_{1j} etc. can vary across clusters and these coefficients are treated as random variables at level 2. Thus each cluster has its own set of such coefficients. We write

$$\beta_{0j} = \gamma_{00} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + u_{1j}, \text{ etc.} \tag{3}$$

We also have

$$E(e_{ij}) = \text{cov}(e_{ij}, e_{kj}) = E(u_{ij}) = 0$$

We may wish to attempt to account for between-cluster variation in terms of one or more features, Z, of the children or W of the clusters being studied. For example, geographical features of subdistricts or socioeconomic characteristics of the child's family and the model could be extended by including such explanatory variables to give the following

$$\beta_{0ij} = \gamma_{00} + \sum_{k=1} \gamma_{0k} w_{kj} + \sum_{l} \delta_{0l} z_{lij} + u_{0j} + e_{ij} \tag{4}$$

Likewise the slope, etc. coefficients can be modelled as functions of cluster or person level explanatory variables. One important application is where separate coefficients are fitted for different subgroups, for example boys and girls, and typically this would be done only for the low-order polynomial coefficients.

The random variation at level 1 can also be further structured. For example we can fit separate level 1 variances for subgroups or allow the between-child variance with age. Goldstein (1987) gives details of how these can be specified, and also provides an introduction to multilevel modelling.

2.3. *Calculation*

Commonly available statistical packages are not available to provide efficient estimates of the parameters in complex multilevel models. In this paper we use the ML3 software produced by the Multilevel Model Project, Department of Mathematics, Statistics and Computing, Institute of Education (Prosser, Rasbash and Goldstein 1990).

3. An example

The data are derived from 4292 girls and 4697 boys with weight measurements from birth to 72 months in six districts in Shanghai and five provinces in the middle east and southeast of China. The data were measured by the WHO Collaborating Centre for Physical Growth and Psychosocial Development of Children in Shanghai, in 1986. The staff were specially trained to ensure validity under the guidance of experts from the WHO Maternal and Child Health Division. The sample was finally clustered into 18 subdistricts which were classified as rural or urban.

The first step in the analysis was to identify the degree of the polynomial to be fitted to the data and the number of polynomials which should be fitted. A previous analysis (Pan, Goldstein and Yang 1990), suggested that there should be two polynomials joining at 12 months, since the velocity of growth changes markedly at this age.

3.1. *Basic models*

Let y_{ij} denote the weight in kg of the ith child in the jth cluster, t_{ij} denote the age in months. Our basic model is as follows

$$y_{ij} = \beta_{0j} + \beta_{1j} t_{ij} + \beta_{2j} t_{ij}^2 + \beta_{3j} t_{ij}^3 + \beta_{4j} (t_{ij} - 12)_+^3 + \delta_3 z_{3ij} + e_{ij} \tag{5}$$

where Z_3 is a dummy variable for gender. We suppose that the growth pattern varies from district to district, with just the coefficients of the intercept and slope varying. We have

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + u_{0j} \\ \beta_{1j} &= \gamma_{10} + u_{1j} \\ \beta_{2j} &= \gamma_{20} \\ \beta_{3j} &= \gamma_{30} \\ \beta_{4j} &= \gamma_{40} \end{aligned}$$

where γ_{00} is the mean intercept and γ_{10} is the mean linear growth rate. The variances of u_{oj} and u_{1j} and their covariance denoted by σ_{u0}^2, σ_{u1}^2 and σ_{u01} respectively.

The level 1 random variation can also be structured. Suppose we wish to fit separate level 1 variances for boys and girls. If Z_3 is coded 1 for boys and 0 for girls, we specify the level 1 random component as

$$e_{ij} = e_{0ij} + e_{3ij}z_{3ij}$$

Then the level 1 variance is given by

$$\sigma_{e0}^2 + \sigma_{e3}^2 z_{3ij}^2 + 2\sigma_{e03}z_{3ij} \tag{6}$$

By constraining $\sigma_{e3}^2 = 0$ we obtain for boys a level 1 variance $\sigma_{e0}^2 + 2\sigma_{e03}$ and for girls σ_{e0}^2. The ML3 software used in the analysis (Prosser *et al.* 1990) allows such a model to be specified. Note that setting $\sigma_{e3}^2 = 0$ is simply used as a device to avoid the over-parameterization implied by (6).

By calculating weight variances for fixed age groups we find that the weight variance increases with age. This suggests that we should model the level 1 variance as a function of age. If t_{ij} is the age of the ith child in the jth cluster we now write

$$e_{ij} = e_{0ij} + e_{3ij}z_{3ij} + e_{tij}t_{ij} \tag{7}$$

We now obtain the following level 1 variance

$$\sigma_{e0}^2 + 2\sigma_{e03}z_{3ij} + 2\sigma_{e0t}t_{ij} + 2\sigma_{e3t}z_{3ij}t_{ij} + \sigma_{et}^2 t_{ij}^2$$

In fact, in our data σ_{et}^2 is actually estimated as zero and σ_{e3t} is very small and not significant, so that the level 1 variance is modelled as a linear function of age, that is

$$\sigma_{e0}^2 + 2\sigma_{e03}z_{3ij} + 2\sigma_{e0t}t_{ij}$$

Table 1 shows the estimated coefficients for the fixed part where we have fitted grafted polynomials of order 3. The estimates of the fixed part of the model represent the mean growth curve for the entire sample. In both models A and B the fixed coefficients are substantially larger than their standard errors. The estimates in model A and model B are similar, but the standard errors for the fixed coefficients in model B are smaller than in A because the level 1 random structure is more accurately specified. The variance estimates at level 2 are similar in the two models but of course are quite different at level 1. The variance of 1·62 in model A is an overall average. In model B, at a fixed age t in months, the boys' level 1 variance is $0{\cdot}545 + 0{\cdot}046t$, and the girls' is $0{\cdot}38 + 0{\cdot}046t$. We could also allow for an interaction between age and gender so that the rate of increase of variance differs between boys and girls. The difference, in the present sample, is negligible. At level 2 we see that the between-cluster variation changes with age, as a result of both the variance and the mean increasing with age.

3.2. *Adding further explanatory variables*

We now study the effect of including terms for urban/rural, gender, and region (southeast and middle east of China). Let W_1, W_2, Z_3 denote region, urban/rural and gender respectively. We have included gender already in models A and B, but only as an overall boy/girl difference. Now we allow an interaction between gender and the linear and quadratic polynomial coefficients by adding the terms $z_{3ij}t_{ij}$, $z_{3ij}t_{ij}^2$ to the model.

Table 1. Estimated parameters and variance components of models A and B (standard errors in parentheses).

Parameters	A	B
Fixed		
Constant	$3\cdot06\,(9\cdot70\times10^{-2})$	$3\cdot05\,(7\cdot33\times10^{-2})$
t	$1\cdot13\,(2\cdot65\times10^{-2})$	$1\cdot15\,(1\cdot78\times10^{-2})$
t^2	$-8\cdot07\times10^{-2}\,(2\cdot59\times10^{-3})$	$-8\cdot23\times10^{-2}\,(1\cdot85\times10^{-2})$
t^3	$2\cdot24\times10^{-3}\,(7\cdot64\times10^{-5})$	$2\cdot29\times10^{-3}\,(5\cdot59\times10^{-5})$
$(t-12)^3_+$	$-2\cdot24\times10^{-3}\,(7\cdot83\times10^{-5})$	$-2\cdot29\times10^{-3}\,(5\cdot81\times10^{-5})$
Z_3	$0\cdot52\,(2\cdot70\times10^{-2})$	$0\cdot51\,(2\cdot26\times10^{-2})$
Random		
Level 2		
σ^2_{u0}	$7\cdot23\times10^{-2}\,(2\cdot71\times10^{-2})$	$6\cdot14\times10^{-2}\,(2\cdot20\times10^{-2})$
σ^2_{u0t}	$2\cdot55\times10^{-3}\,(9\cdot59\times10^{-4})$	$2\cdot51\times10^{-3}\,(9\cdot15\times10^{-4})$
σ^2_{ut}	$1\cdot28\times10^{-4}\,(4\cdot52\times10^{-5})$	$1\cdot34\times10^{-4}\,(4\cdot76\times10^{-5})$
Level 1		
σ^2_{e0}	$1\cdot62\,(2\cdot43\times10^{-2})$	$0\cdot38\,(2\cdot21\times10^{-2})$
σ_{e03}		$8\cdot25\times10^{-2}\,(1\cdot50\times10^{-2})$
σ_{e0t}		$2\cdot28\times10^{-2}\,(6\cdot91\times10^{-4})$

Z_3 is coded 1 = boy 0 = girl.
Number of boys = 4679, number of girls = 4292.

Table 2. Estimated parameters and variance components of models C and D (standard errors in parentheses).

Parameters	C	D
Fixed		
Constant	$2\cdot85\,(9\cdot27\times10^{-2})$	$2\cdot85\,(8\cdot80\times10^{-2})$
t	$1\cdot13\,(1\cdot88\times10^{-2})$	$1\cdot11\,(1\cdot83\times10^{-2})$
t^2	$-8\cdot27\times10^{-2}\,(1\cdot92\times10^{-3})$	$-8\cdot24\times10^{-2}\,(1\cdot86\times10^{-3})$
t^3	$2\cdot30\times10^{-3}\,(5\cdot80\times10^{-5})$	$2\cdot30\times10^{-3}\,(5\cdot64\times10^{-5})$
$(t-12)^3_+$	$-2\cdot31\times10^{-3}\,(6\cdot02\times10^{-5})$	$-2\cdot30\times10^{-3}\,(5\cdot85\times10^{-5})$
W_1	$0\cdot42\,(8\cdot27\times10^{-2})$	$0\cdot43\,(7\cdot82\times10^{-2})$
$W_1{*}t$	$1\cdot59\times10^{-2}\,(3\cdot99\times10^{-3})$	$1\cdot58\times10^{-2}\,(3\cdot94\times10^{-3})$
W_2	$0\cdot21\,(7\cdot55\times10^{-2})$	$0\cdot20\,(7\cdot14\times10^{-2})$
$W_2{*}t$	$1\cdot27\times10^{-2}\,(3\cdot63\times10^{-3})$	$1\cdot30\times10^{-2}\,(3\cdot59\times10^{-3})$
$Z_3{*}t$		$4\cdot93\times10^{-2}\,(2\cdot80\times10^{-3})$
$Z_3{*}t^2$		$-7\cdot24\times10^{-4}\,(5\cdot67\times10^{-5})$
Random		
Level 2		
σ^2_{u0}	$2\cdot03\times10^{-2}\,(8\cdot40\times10^{-3})$	$1\cdot80\times10^{-2}\,(7\cdot53\times10^{-3})$
σ^2_{u0t}	$5\cdot86\times10^{-4}\,(3\cdot04\times10^{-4})$	$5\cdot89\times10^{-4}\,(2\cdot88\times10^{-4})$
σ^2_{ut}	$4\cdot86\times10^{-5}\,(1\cdot94\times10^{-5})$	$4\cdot76\times10^{-5}\,(1\cdot90\times10^{-5})$
Level 1		
σ^2_{e0}	$0\cdot43\,(2\cdot40\times10^{-2})$	$0\cdot40\,(2\cdot25\times10^{-2})$
σ_{e03}	$9\cdot06\times10^{-2}\,(1\cdot62\times10^{-2})$	$8\cdot45\times10^{-2}\,(1\cdot52\times10^{-2})$
σ_{e0t}	$2\cdot32\times10^{-2}\,(7\cdot24\times10^{-4})$	$2\cdot29\times10^{-2}\,(7\cdot00\times10^{-4})$

W_1 is coded 1 = middle east, 0 = southeast.
W_2 is coded 1 = urban, 0 = rural.
Z_3 is coded 1 = boy, 0 = girl.
Interaction are denoted by the * symbol. Thus, for example, the interaction between age and region is denoted by $W_1{*}t$.
Number of boys = 4679, number of girls = 4292.

The intercept, linear, and quadratic coefficients of age are structured in terms of W_1, W_2 as follows.

$$\beta_{0j} = \gamma_{00} + \gamma_{01} w_{1j} + \gamma_{02} w_{2j} + u_{0j}$$
$$\beta_{1j} = \gamma_{10} + \gamma_{11} w_{1j} + \gamma_{12} w_{2j} + u_{1j}$$
$$\beta_{2j} = \gamma_{20}$$
$$\beta_{3j} = \gamma_{30}$$
$$\beta_{4j} = \gamma_{40}$$

Table 2 shows the estimates for fixed and random effects of models C and D. All the coefficients of the fixed part of models C and D are statistically significant. In model D the coefficient of region is 0·43, which indicates that the average weight at the initial status in the middle east is considerably larger than that in the southeast, and the estimate of γ_{01} of 0·016 indicates tha the average growth rate of weight in the middle east is greater than that in the southeast. In other words the regional difference in weight increases with age, although only by 1%. Likewise, the urban children on average grow just under 2% faster than the rural children. Both in model C and D the level 2 variances are smaller than their counterparts under model B: each variance has decreased more than 60%. These reductions suggest that much of the between-cluster variation is accounted for by region and urban/rural. In model D we see the different growth trajectories for boys and girls. Figure 1 is a plot of average weight curves of boys of urban and rural areas in the middle east and the southeast of China for model D. Figure 2 shows a plot of estimated standardized child level residuals against predicted values for model D. The pattern of residuals shows no trend of variance with the predicted values, and no exceptionally large outliers, which implies that the model D fits adequately.

Table 3 presents results for model E which uses the same explanatory variables as model D but with a logarithmic transformation of weight. We see that in model E the coefficient of $W_1{*}t$ and σ_{e03} are no longer significant at the 5% level and are omitted in

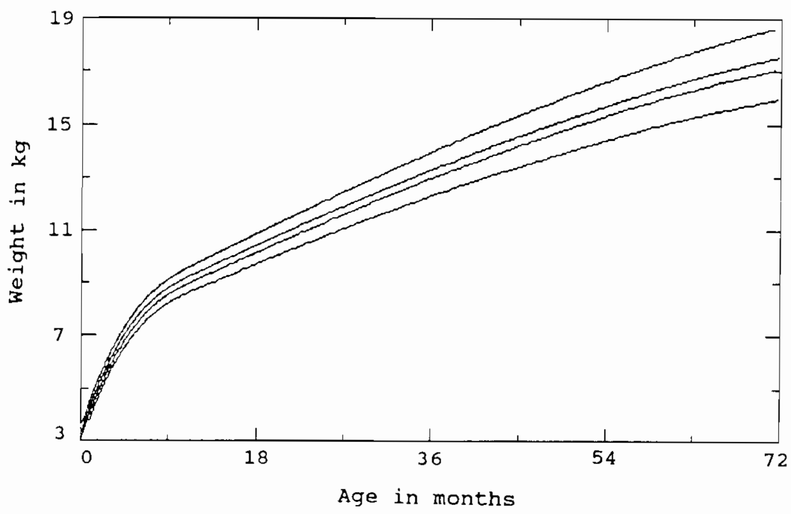

Figure 1. Mean weight curves of boys in urban and rural areas in the middle east and southeast of China for model D.

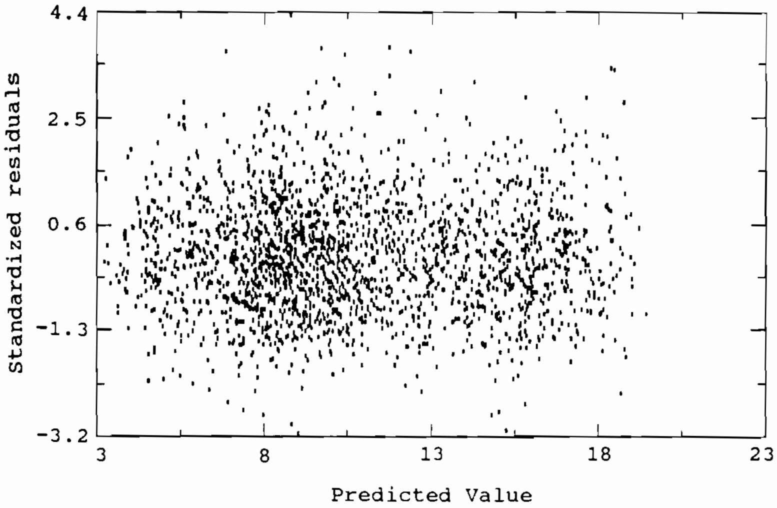

Figure 2. Child (level 1) standardized residuals by predicted values for model D.

Table 3. Estimated parameters and variance components of models E and F (standard errors in parentheses).

Parameters	E	F
Fixed		
Constant	$1\cdot23\,(1\cdot42\times10^{-2})$	$1\cdot22\,(1\cdot42\times10^{-2})$
t	$0\cdot20\,(2\cdot54\times10^{-3})$	$0\cdot20\,(2\cdot54\times10^{-3})$
t^2	$-1\cdot52\times10^{-2}\,(2\cdot47\times10^{-4})$	$-1\cdot53\times10^{-2}\,(2\cdot48\times10^{-4})$
t^3	$4\cdot22\times10^{-4}\,(7\cdot26\times10^{-6})$	$4\cdot23\times10^{-4}\,(7\cdot27\times10^{-6})$
$(t-12)^3_+$	$-4\cdot23\times10^{-4}\,(7\cdot43\times10^{-6})$	$-4\cdot24\times10^{-4}\,(7\cdot44\times10^{-6})$
W_1	$7\cdot20\times10^{-2}\,(1\cdot25\times10^{-2})$	$7\cdot26\times10^{-2}\,(1\cdot25\times10^{-2})$
W_1*t	$1\cdot56\times10^{-4}\,(2\cdot02\times10^{-4})$	
W_2	$3\cdot59\times10^{-2}\,(1\cdot14\times10^{-2})$	$3\cdot58\times10^{-2}\,(1\cdot14\times10^{-2})$
W_2*t	$3\cdot67\times10^{-4}\,(1\cdot87\times10^{-4})$	$3\cdot65\times10^{-4}\,(1\cdot92\times10^{-4})$
Z_3*t	$3\cdot84\times10^{-3}\,(2\cdot59\times10^{-4})$	$3\cdot84\times10^{-3}\,(2\cdot59\times10^{-4})$
Z_3*t^2	$-5\cdot60\times10^{-5}\,(4\cdot68\times10^{-6})$	$-5\cdot60\times10^{-5}\,(4\cdot68\times10^{-6})$
Random		
Level 2		
σ^2_{u0}	$4\cdot92\times10^{-4}\,(1\cdot90\times10^{-4})$	$4\cdot93\times10^{-4}\,(1\cdot91\times10^{-4})$
σ^2_{u0t}	$1\cdot01\times10^{-6}\,(2\cdot22\times10^{-6})$	$9\cdot36\times10^{-7}\,(2\cdot27\times10^{-6})$
σ^2_{ut}	$9\cdot25\times10^{-8}\,(5\cdot12\times10^{-8})$	$1\cdot00\times10^{-7}\,(5\cdot38\times10^{-8})$
Level 1		
σ^2_{e0}	$1\cdot55\times10^{-2}\,(4\cdot02\times10^{-4})$	$1\cdot55\times10^{-2}\,(3\cdot41\times10^{-4})$
σ_{e03}	$4\cdot69\times10^{-6}\,(2\cdot03\times10^{-4})$	
σ_{e0t}	$-3\cdot46\times10^{-5}\,(4\cdot53\times10^{-6})$	$-3\cdot46\times10^{-5}\,(4\cdot53\times10^{-6})$

W_1 is coded 1 = middle east, 0 = southeast.
W_2 is coded 1 = urban, 0 = rural.
Z_3 is coded 1 = boy, 0 = girl.
Interaction are denoted by the * symbol. Thus, for example, the interaction between age and region is denoted by W_1*t.
Number of boys = 4679, number of girls = 4292.

model F. We see that the variance on the log scale is a decreasing function of age. Since the standard deviation of log (weight) is approximately equal to the standard deviation of weight divided by the mean weight, this simply means that relative weight is a decreasing function of age.

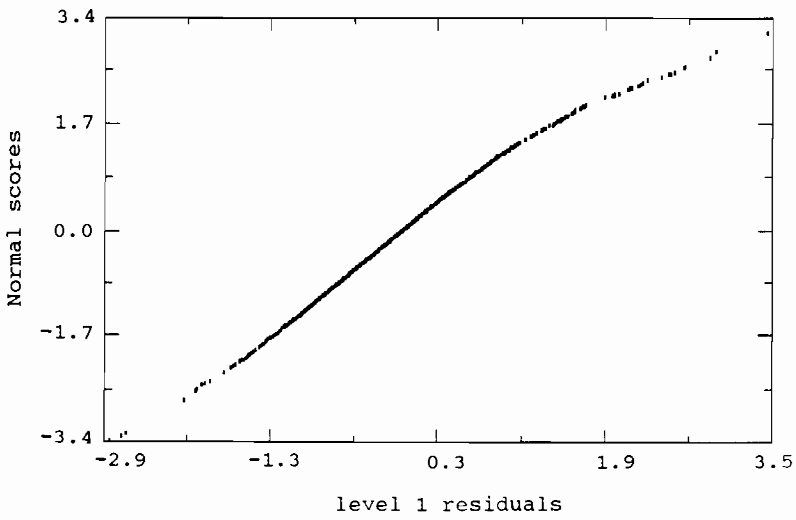

Figure 3. Normal probability plot for level 1 residuals for model D.

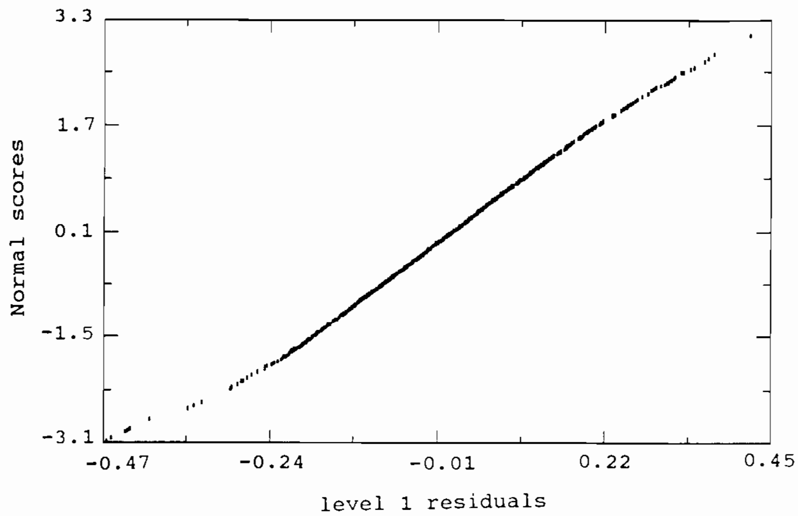

Figure 4. Normal probability plot for level 1 residuals for model F.

The advantage of using model F is that we need to specify fewer parameters. Furthermore, the distribution of the residuals is less skew, as is shown in figures 3 and 4 which compare normal probability plots for the level 1 residuals from analyses D and F. Substantively, the omission of the interaction between region and age in analysis F implies that the ratio of weight for the regions is an additive function of region and age. The use of F rather than D, however, does not change our general conclusions in any important way.

4. Discussion

The main feature of a two-level model for cross-sectional data on growth is that we can study between-cluster variation and at the same time fit heterogeneous variation at level 1 to allow for different gender variances and changing variability with age. The present paper has shown the feasibility of fitting grafted polynomials to a wide age range with a multilevel structure which also allows for covariates.

In our paper only t^p in equation (1) is permitted to be different for the separate pieces of the polynomial. A more general form, known as weak splines, can allow t^{p-1} etc. to have different coefficients in the various segments of the piecewise polynomials (Cox 1971). Further study of this will be presented in another paper. In practice we have found that the grafted polynomials can cope with multilevel growth data successfully. More than one join point would be necessary to fit curves for longer age ranges, and in special situations, such as the age range including adulthood, a more complicated method is needed and further study of this problem is planned.

In model A we found that the estimates of the coefficients of t^3 and $(t-12)_+^3$ are almost equal in absolute value, which implies that the polynomial in the second subrange is effectively quadratic, and we could have incorporated a constraint to make these coefficients equal in absolute value.

In our examples only the linear coefficient of age is assumed to be random at level 2. In some cases we may wish to make the coefficients of higher-order terms random, and that involves a straightforward extension. If, however, the coefficient of the highest-order term, in our case the cubic, is made random we require both the coefficients of t^3, $(t-12)_+^3$ to be random and their covariance to be fitted as well as their variances.

Finally, other factors, for example family circumstances, can be incorporated readily into the fixed part of the model to see how far they might explain between-cluster or between-child variation in growth.

Acknowledgements

We are most grateful to Mr Robert Prosser and Jon Rasbash for their comments and assistance. This work was carried out while the first author was a WHO visiting fellow in London. It was partly supported by a research grant from the Economic and Social Research Council (UK). We are grateful to the staff of the WHO Collaborating Centre in Shanghai for the data, and to the WHO Collaborating Centre in Growth and Psychosocial Development at the Institute of Education and Child Health, University of London for providing facilities for this research.

References

Cox, M. G., 1971, Curve fitting with piecewise polynomials, *Journal of the Institute of Mathematics and its Applications*. **8**, 36–52.

Fuller, W. A., (1976). *Introduction to statistical time series* (New York: John Wiley & Sons), pp. 393–398.

Goldstein, H., 1979, *The Design and Analysis of Longitudinal Studies* (London: Academic Press).

Goldstein, H., 1986, Efficient statistical modelling of longitudinal data. *Annals of Human Biology*, **13**, 129–141.

Goldstein, H., 1987, *Multilevel Model in Educational and Social Research* (London: Griffin; New York: Oxford University Press).

Goldstein, H., 1989, Models for multilevel response variables with an application to growth curves. In *Multilevel Analysis of Educational Data*, edited by D. Bock (New York: Academic Press), pp. 108–125.

Pan, H., Goldstein, H., and Yang, Q., 1990, Non-parametric estimation of age-related centiles over long age ranges. *Annals of Human Biology*, **17**, 475–482.

PROSSER, R., RASBASH, J., and GOLDSTEIN, H., 1990, ML3, *Software for Three-level Analysis: Users' Guide* (London: Institute of Education, University of London).

Address for correspondence: Huiqi Pan, Department of Mathematics, Statistics and Computing, Institute of Education, 20 Bedford Way, London WC1H 0AL, UK.

Zusammenfassung. Es wird ein neues statistisches Modell zur Analyse hierarchisch strukturierter Querschnittsdaten zum Wachstum vorgestellt, das insbesondere geeignet ist, wenn Messungen für eine große Altersspanne vorliegen. Das Modelle kombiniert ein Zwei-Stufen Modell mit "grafted piecewise polynomials", um aus den verflügbaren Daten effizient Nutzen zu ziehen.

Résumé. Un nouveau modèle statistique est proposé pour l'analyse de données de croissance transversale hiérarchiquement structurées, en particulier pour les cas où les mesures sont faites en englobant de vastes gammes d'âge. Le modèle combine un modèle à deux niveaux avec assemblages polynomiaux construits pas à pas, afin de faire un usage efficace des données disponibles.