



Percentile Smoothing Using Piecewise Polynomials, with Covariates

Harvey Goldstein; Huiqi Pan

Biometrics, Vol. 48, No. 4. (Dec., 1992), pp. 1057-1068.

Stable URL:

<http://links.jstor.org/sici?sici=0006-341X%28199212%2948%3A4%3C1057%3APSUPPW%3E2.0.CO%3B2-F>

Biometrics is currently published by International Biometric Society.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ibs.html>.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

Percentile Smoothing Using Piecewise Polynomials, with Covariates

Harvey Goldstein and Huiqi Pan

Institute of Education, University of London, 20 Bedford Way,
London WC1H 0AL, England

SUMMARY

A nonparametric procedure for the joint smoothing of a series of percentile curves is presented. It allows separate curves to be fitted over contiguous age ranges, constraining them to join smoothly. Covariates can be incorporated, thus allowing efficient estimation simultaneously for subgroups.

1. Introduction

The construction of norms for growth measurements is an important activity with several uses—for example, in screening for growth disorders. The data used for the construction of norms usually comprise a random sample of measurements taken cross-sectionally on individuals and covering the age range of interest. Until recently the actual process of constructing the norms typically has been carried out using ad hoc smoothing procedures applied to raw percentile estimates. The latter would be obtained by ordering measurements within narrow age categories.

Apart from a few measurements such as stature, the distribution at any age is difficult to characterise with a simple function, and usually is far from Gaussian. One approach to providing efficient estimation procedures for percentile norms has therefore centered around finding transformations of the measurements to produce a standard distributional form. Cole (1988) uses a smoothed power transformation to make the distributions at each age conform as closely as possible to a Gaussian distribution. Assuming this distributional form, he then estimates the percentiles in a straightforward fashion.

A different approach has been taken by Healy, Rasbash, and Yang (1988), who present a nonparametric procedure for estimating a set of smooth percentile norms for continuous measurements over time. The procedure fits high-order polynomials to preliminary estimates of the chosen percentiles within contiguous age bands. The constant, linear, etc. coefficients of these polynomials are constrained to be smoothly varying, in turn, by regressing them on the equivalent standard normal deviates of the percentiles. A very flexible class of curves is obtained whose shape can be manipulated by altering the order of polynomials fitted to the percentiles and the form of the regression on the normal deviates. The use of the normal distribution in this context is largely for convenience, and other, possibly skew, distributions could be used. The procedure is as follows.

First a set of percentiles of interest is decided upon. Typically for growth data these will be the 3rd, 10th, 25th, 50th, 75th, 90th, and 97th percentiles. It is assumed that the data consist of measurements covering an age range, preferably uniformly distributed.

Key words: Constrained estimation; Growth; Nonparametric methods; Percentile estimation; Piecewise polynomial; Smoothing.

The second stage is to carry out an initial smoothing of the data to produce "raw" percentile estimates at each age point. The total number of such estimates is N and in general we can allow unequal numbers of estimates for each percentile. To these estimates we can envisage fitting polynomials, one for each percentile. Suppose that polynomials of degree q are fitted, and let j index the percentiles and i index the coefficients. We would thus obtain a set of polynomial coefficients, say

$$\boldsymbol{\beta} = \{\beta_j^{(i)}\}, \quad i = 0, \dots, q, \quad j = 1, \dots, t,$$

where the set of elements for each percentile is stacked into the column vector $\boldsymbol{\beta}$. If we denote the set of raw percentile estimates by the vector \mathbf{Y} and the polynomial "design" matrix for the polynomial coefficients containing powers of age, by \mathbf{X}^* , the predicted values of \mathbf{Y} are

$$\mathbf{Y}^* = \mathbf{X}^*\boldsymbol{\beta}, \quad (1)$$

where \mathbf{Y}^* is of order $(N \times 1)$ and $\boldsymbol{\beta}$ is of order $(t(q + 1) \times 1)$.

We now envisage carrying out polynomial regressions of order p , of these coefficients, for each i separately, on z_j , the normal equivalent deviate (NED) corresponding to the j th percentile. This would give a new set of predicted coefficients, say

$$\hat{\beta}_j^{(i)} = \sum_{k=0}^p \alpha_k^{(i)} z_j^k \quad (2)$$

or

$$\hat{\boldsymbol{\beta}} = \mathbf{Z}\mathbf{A}.$$

Substituting $\hat{\boldsymbol{\beta}}$ for $\boldsymbol{\beta}$ in (1) gives new predicted values which can be written as

$$\hat{\mathbf{Y}} = \mathbf{X}^*\mathbf{Z}\mathbf{A} = \mathbf{X}\mathbf{A}, \quad (3)$$

where \mathbf{A} is of order $((q + 1)(p + 1) \times 1)$ and consists of the set of p parameter values for each coefficient. The matrix \mathbf{Z} is block-diagonal of order $(t(q + 1) \times (q + 1)(p + 1))$ and has one block for each coefficient. The matrix \mathbf{X} is known and of order $(N \times t(q + 1))$. In the next section we allow different-order polynomial regressions and the detailed structures of these matrices are given. The linear model estimates $\hat{\mathbf{A}}$ are used to construct the final percentile curves. Healy et al. (1988) use ordinary least squares (OLS) estimation and the present paper considers generalised least squares estimators.

In practice, three difficulties arise with this procedure. The first problem occurs when estimates are required over a wide age range, such as typically is found in studies of child growth, where a single polynomial is inadequate. In such a case an obvious extension is to construct two or more polynomials joined in a smooth fashion. This is incorporated into the procedure proposed below.

The second problem occurs when separate estimates are required for population subgroups. To carry out completely separate estimation procedures in general will make inefficient use of the data, since we would expect smooth relationships between the percentiles for the subgroups. For example, in the simple case all the percentiles may differ by a constant from one subgroup to another. By incorporating the possibility of such relationships into a joint estimation, the data can be analysed more efficiently.

The third problem lies in the original choice of percentiles to estimate. The procedure recommended by Healy et al. when smoothing the polynomial coefficients is to carry out an ordinary least squares regression of the polynomial coefficients on the normal equivalent deviates (z). A particular choice of percentiles, however, determines the set of normal deviate values used in this regression and hence an implicit weighting that will determine

the relative accuracy of different percentiles. Thus, for example, if a relatively large proportion of the percentiles that are fitted to the raw data are chosen to lie near the 50th percentile, then the OLS coefficients will be relatively heavily weighted by the position of these percentiles. In the extreme case when, say, all the chosen percentiles are between the 25th and 75th, predictions for percentiles outside this range will be based on an extrapolation of polynomials fitted to data points only within the range. Nevertheless, since most of the original data points occur around the median, it would seem appropriate to sample more percentiles from that region in order to utilise the observed data more efficiently. To deal with this problem, a more flexible procedure is to allow the data analyst to be able to specify a weight to be attached to each chosen percentile so that the effect of varying these weights can be explored.

The present paper presents a comprehensive model that extends the Healy et al. (1988) procedure to take account of these difficulties.

2. Estimation in Contiguous Age Ranges

Suppose that the total age range is divided into m contiguous subranges, with n_{lj} values for the j th raw percentile in the l th subrange, with $N_j = \sum_l n_{lj}$. We can fit separate polynomials in each range, and for range l and percentile j we obtain a set of observed coefficients $\beta_{lj}^{(i)}$. The subranges need not be the same width, nor do the polynomials fitted within each range need to be of the same order. Equation (2) now becomes, for the i th coefficient of the j th percentile in the l th subrange,

$$\beta_{lj}^{(i)} = \sum_{k=0}^{p_{li}} \alpha_k^{(i)} z_j^k, \quad i = 0, \dots, q_l. \tag{4}$$

Note that in (2) we imposed a degree of smoothness by assuming that the coefficients $\alpha_k^{(i)}$ were the same for all percentiles. Equation (4) can be rewritten as $\beta = \mathbf{Z}\mathbf{A}$, where

$$\begin{aligned} \beta^T &= (\beta_1^T, \dots, \beta_m^T), \quad \beta_l^T = (\beta_l^{(0)T}, \dots, \beta_l^{(q_l)T}), \quad \beta^{(i)T} = (\beta_{1l}^{(i)}, \dots, \beta_{ml}^{(i)}), \\ \mathbf{A}^T &= (\alpha_1^T, \dots, \alpha_m^T), \quad \alpha_l^T = (\alpha_l^{(0)T}, \dots, \alpha_l^{(q_l)T}), \quad \alpha^{(i)T} = (\alpha_{1l}^{(i)}, \dots, \alpha_{ml}^{(i)}), \\ \mathbf{Z} &= \bigoplus_{l=1}^m \mathbf{Z}_l, \quad \mathbf{Z}_l = \bigoplus_{i=1}^{q_l} \mathbf{Z}_l^{(i)}, \\ \mathbf{Z}_l^{(i)} &= (\mathbf{Z}_{1l}^{(i)}, \dots, \mathbf{Z}_{ml}^{(i)})^T, \quad \mathbf{Z}_l^{(i)T} = (1, z_j, \dots, z_j^{p_{li}}), \\ j &= 1, \dots, t, \quad l = 1, \dots, m, \quad i = 0, \dots, q_l, \end{aligned} \tag{5}$$

where \bigoplus is the direct sum operator. We refer to this model as a piecewise polynomial model and note that the OLS estimates of (3) are simply those that would be obtained from separate regressions in each subrange.

2.1 Smoothing the Join Points

Because growth is a continuous process, we require all the curves to join “smoothly.” The degree of smoothness can be expressed in terms of setting derivatives of the curves equal at the join points. Thus, if $f_{lj}(x_k)$ is the value predicted for the curve in the l th subrange for percentile j at the k th join point ($k = 1, \dots, m - 1$), we require at least

$$f_{lj}(x_k) = f_{l+1,j}(x_k), \tag{6}$$

$$f'_{lj}(x_k) = f'_{l+1,j}(x_k). \tag{7}$$

In addition, we might also require

$$f''_{ij}(x_k) = f''_{i+1,j}(x_k). \quad (8)$$

Condition (7) states that growth velocity changes smoothly at a join point. In practice, the inherent smoothness of growth data allows us to dispense with (8). The estimation problem now becomes the restricted least squares problem of fitting (3) subject to the constraints (6), (7), and possibly (8).

In the Appendix, we show how the smoothness constraints can be incorporated into the model. We also show how a weighted analysis can be carried out where the raw percentile data may be differentially weighted. The response vector contains the "raw" percentile values, and is of length N , where, for the i th percentile, $N = \sum_i N_i$. It is ordered by age or time within percentile. In the most general case we can consider a set of weights, w_{ij} for the j th data point for the i th percentile, where there is no restriction on the structure of $W = \{w_{ij}\}$. In many applications, however, it is sufficient to write w_{ij} in the form $w_i w_j$, where the w_i specify common weights for the i th percentile and the w_j specify common weights for the j th age. Thus the matrix V^{-1} in the Appendix is diagonal with elements w_{ij} .

3. Estimation for Subgroups

Suppose that the population is divided into r subgroups, indexed by g . These might, for example, be defined in terms of social class, gender, or say, parental height. Given sufficient data, a separate estimation could be carried out for each group. Typically this is done for the growth of boys and girls. In general, however, we would expect that the subgroup differences are similar for each percentile or at least vary across percentiles in a simple fashion. If this variation can be modelled then it will be possible to carry out the estimation for all subgroups simultaneously, with allowance being made for the variation. The procedure described in the preceding sections can be modified in straightforward fashion to do this, as follows.

Write $\beta_{ij}^{(l)}(g)$ for the polynomial coefficient for the j th percentile for the l th subrange for the i th coefficient for the g th subgroup. Equation (2) now becomes

$$\beta_{ij}^{(l)}(g) = \sum_{k=0}^{p_i} \alpha_{gk}^{(i)} z_{ij}^k w_g^{(i)}. \quad (9)$$

In the simplest case the $w_g^{(i)}$ are r dummy variables defining the groups. In general, we might expect the groups to differ in their regressions on z in terms of the lower-order polynomial coefficients only. Thus, for large values of i , only one of the $w_g^{(i)}$, say for $g = 1$, will be nonzero. If the grouping is based on a continuous variable, then we might wish to give w the value of the midpoint of the appropriate interval.

Clearly equation (9) has the same general form as (2) and we can apply the same estimation procedure.

4. Extensions to the Basic Model

The model described above will be adequate for many applications. In some circumstances, however, it may be desirable to include further constraints or to alter the functional form.

For example, in a particular age subrange, some but not all percentiles may effectively be changing linearly, so that we would wish to constrain their higher-order coefficients to be zero. Such further constraints can be added as additional rows of the matrix C , defined in the Appendix.

The procedure described for polynomial models can be adapted straightforwardly for any curves linear in their parameters, simply by using (6) and (7). Different functional forms can be mixed either within age groups or across age groups.

We can also relax the condition that the join points are the same for each percentile. This again involves straightforward modifications to (6) and (7). This may be important at the extremes of the age distribution (see Section 7).

5. Examples

The data for these examples come from the area of Shanghai and five provinces in Southeastern China. They comprise weight measurements on 4,690 male and 4,305 female children between birth and 6 years of age. They were measured by trained staff at the World Health Organisation (WHO) Collaborating Centre for Growth and Development in Shanghai.

In the first example, all the percentile points are equally weighted. The GROSTAT II program developed for the WHO (Rasbash and Pan, 1990) was used in the computations. The first example fits a set of three percentiles to the whole age range using a single fourth-degree polynomial. The intercept terms are assumed to have a cubic regression on the NEDs of the percentiles; the first- to third-order polynomial coefficients are assumed to have a quadratic regression; the fourth-order coefficients are assumed to have a linear relationship. We refer to this as a (4, 3, 2, 2, 2, 1) model. For illustrative purposes just the 3rd, 50th, and 97th percentiles for females are used. In practice, it is desirable generally to fit more than three, typically seven or more, percentiles. Figure 1 shows the fitted smoothed percentiles together with the raw percentiles calculated by ordering measurements within successive age bands. We show just the first 19 months of age.

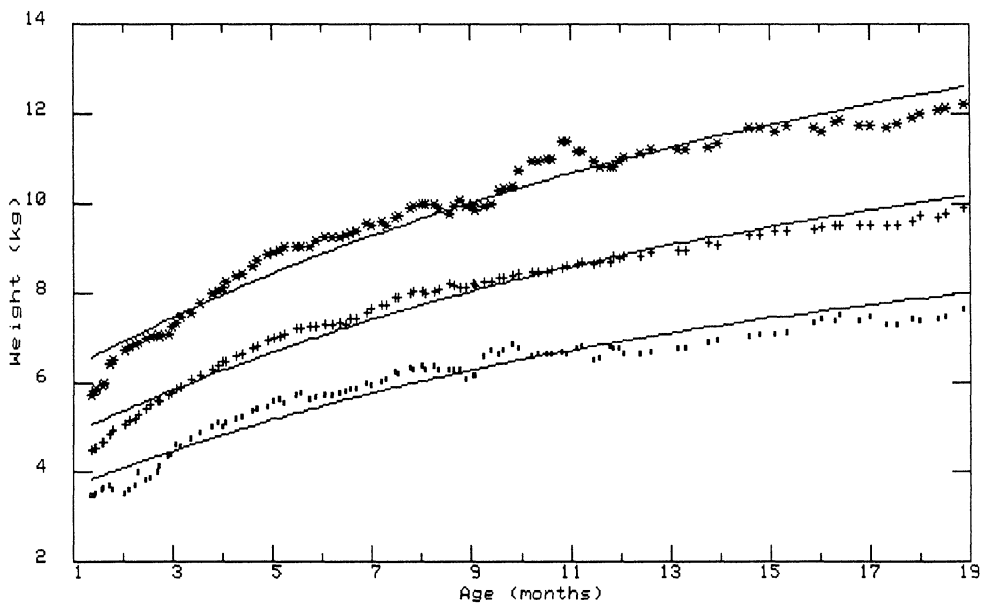


Figure 1. 3rd, 50th, 97th centiles of model (4, 3, 2, 2, 2, 1) for females, superimposed on raw percentile estimates.

While the raw percentiles are reasonably well fitted in the central part of the age range, the fit is relatively poor at the extremes. In particular, the higher percentiles appear to have asymptotes, which is clearly incorrect as will be seen below. Using higher-order polynomials will not in general solve this problem; often local fluctuations will be more pronounced.

In the next example two sets of curves, following an exploratory analysis, are joined at the age of 17 months using the model $(3, 2, 2, 2, 1)$ and $(2, 2, 2, 2)$ for the first and second sets. As Figure 2 shows, this produces smoothly joining curves that fit the data well. As was

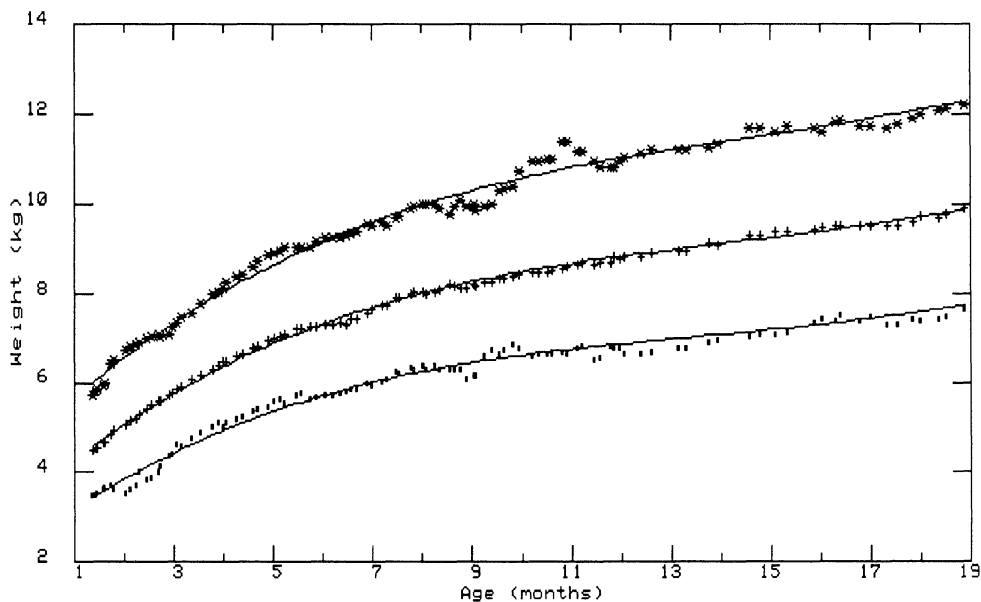


Figure 2. 3rd, 50th, 97th centiles of model $(3, 2, 2, 2, 1)$, $(2, 2, 2, 2)$ for females, superimposed on raw percentile estimates.

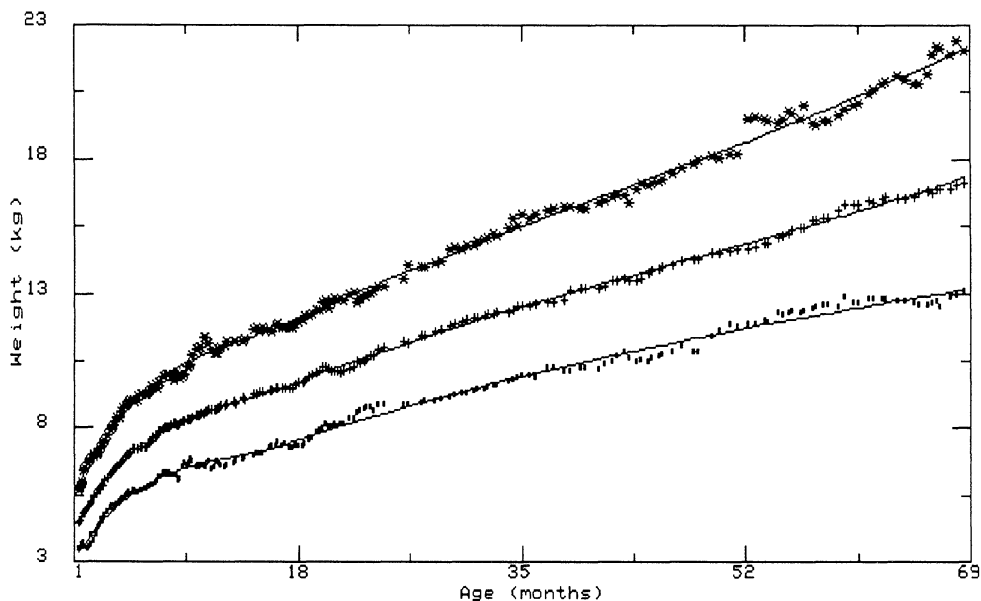


Figure 3. 3rd, 50th, 97th centiles of model $(3, 2, 2, 2, 2)$, $(3, 2, 2, 0, 0)$ for females, superimposed on raw percentile estimates.

the case in this example, considerable experimentation is often required to find an appropriate combination of parameters and a suitable choice of join points. The model $(3, 2, 2, 2, 2), (3, 2, 2, 2, 2)$ as shown in Figure 3 is overparameterised so that the 3rd percentile curve in particular follows local data patterns too closely. Figure 4 uses the model $(3, 2, 1, 1, 0), (2, 1, 1, 0)$ and illustrates what happens when dependencies occur among constraints. Because the coefficients of age, age-squared, and age-cubed have at most a linear regression on the NEDs (z), the constraints on the slopes of the two sets of curves at

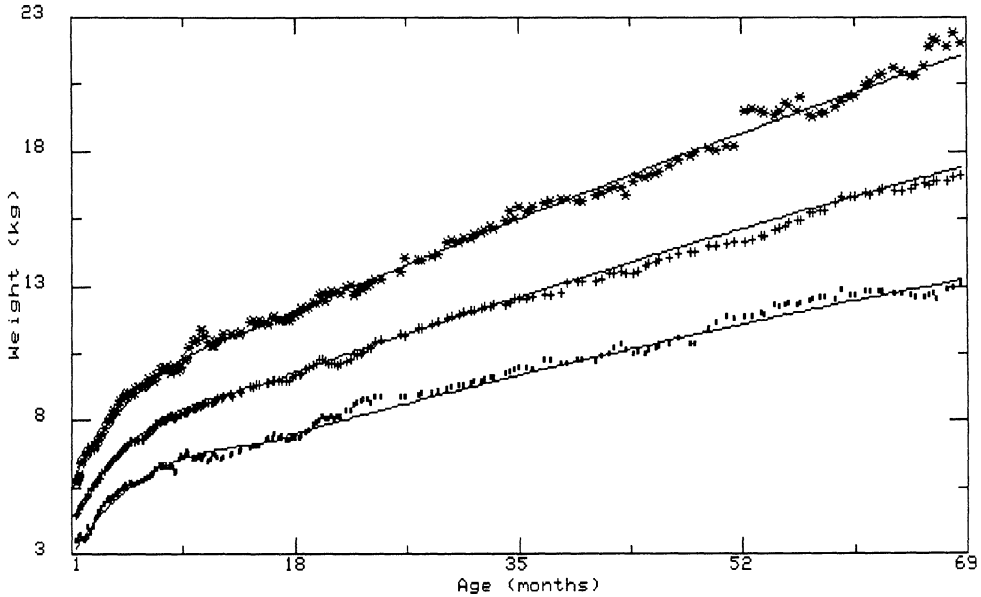


Figure 4. 3rd, 50th, 97th centiles of model $(3, 2, 1, 1, 0), (2, 1, 1, 0)$ for females, superimposed on raw percentile estimates.

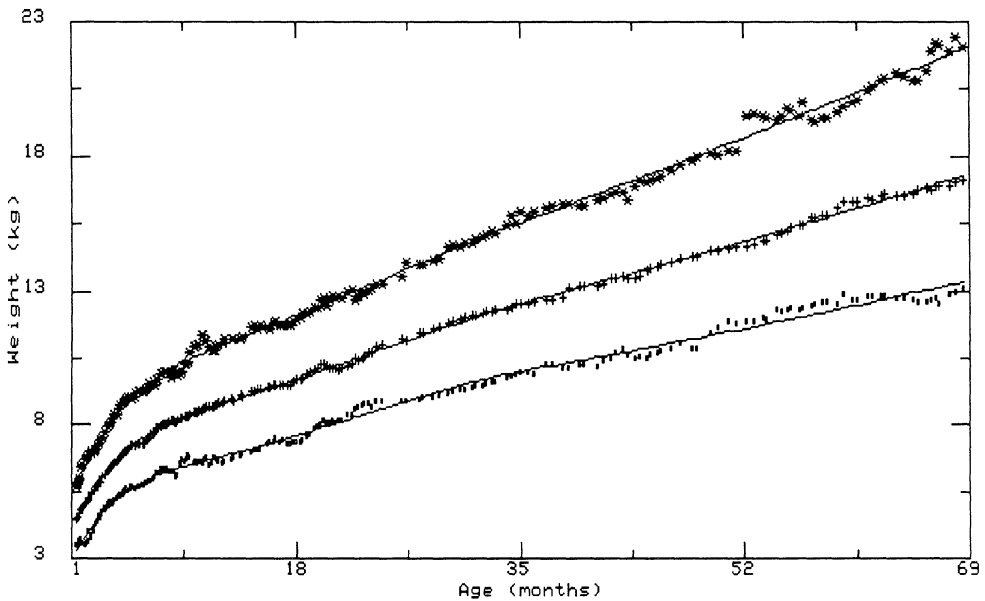


Figure 5. 3rd, 50th, 97th centiles of model $(3, 2, 2, 1, 1), (4, 2, 2, 1, 1, 0)$ for females, superimposed on raw percentile estimates.

the join point are linear functions of the z . This implies that there are at most two independent constraints, whereas we require three—one for each percentile. In general we can ensure that sufficient independent constraints are present by requiring the regression on z to be of order one less than the number of percentiles, for both the intercept and at least one of the polynomial coefficients other than the intercept. This will be required for at least one and preferably both of the curves at each join point.

Figure 5 shows the results of fitting three sets of curves, joining at ages 10 and 36 months

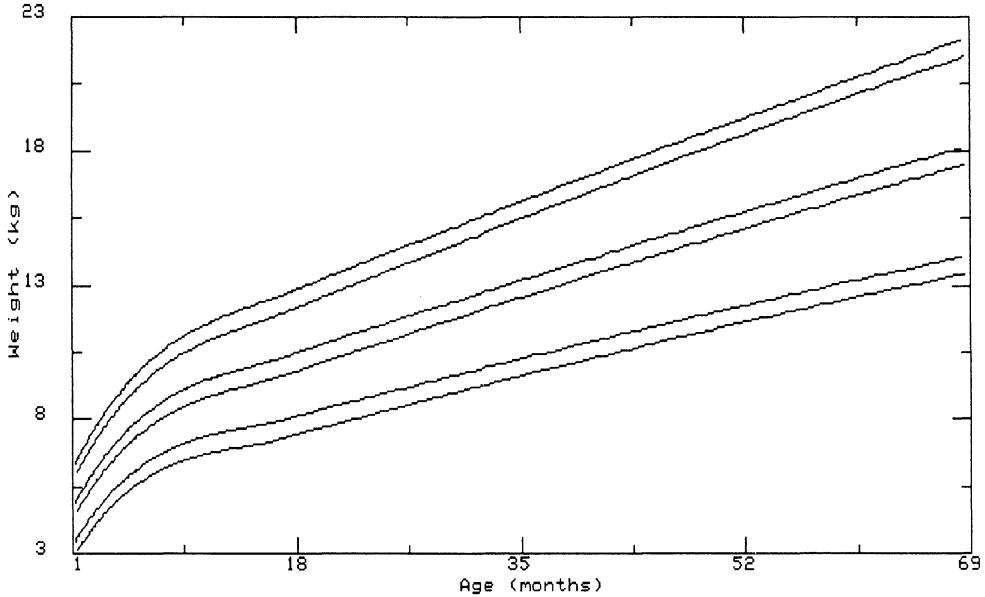


Figure 6. 3rd, 50th, 97th centiles of model $(3, 2, 2, 1, 0)$, $(2, 1, 1, 0)$ for males and females.

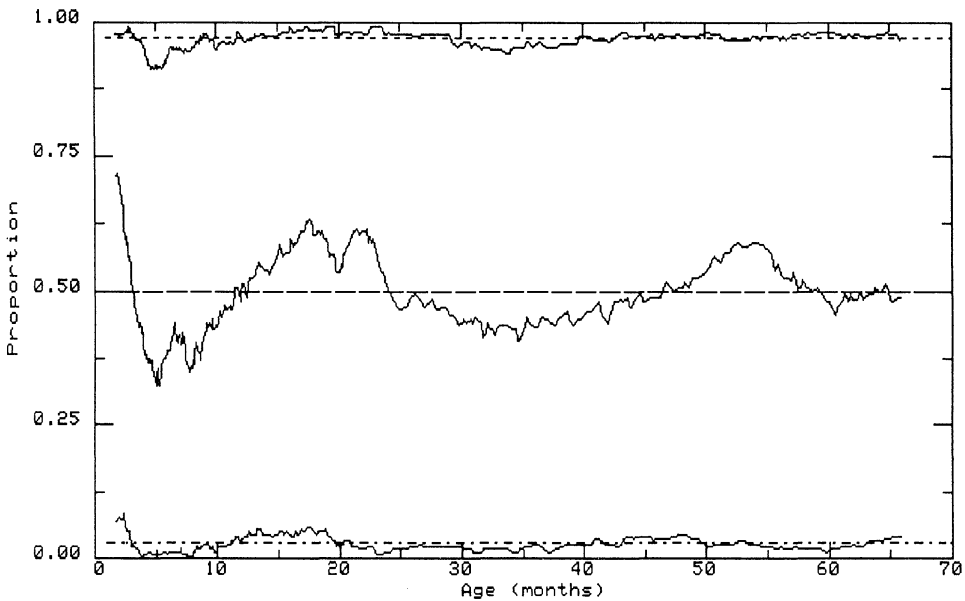


Figure 7. Percentage of individuals below estimated 3rd, 50th, 97th centiles of model $(4, 3, 2, 2, 2, 1)$ for females.

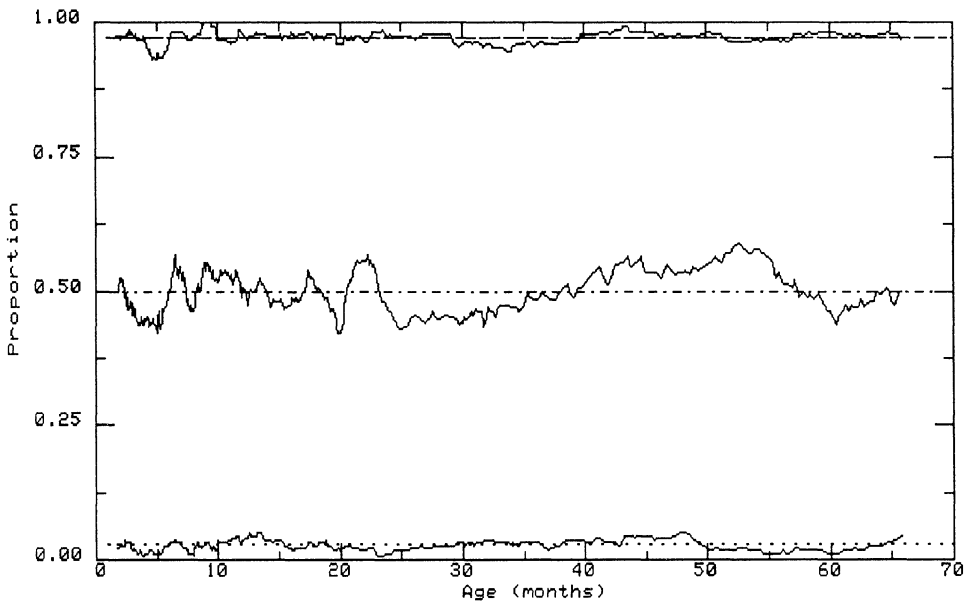


Figure 8. Percentage of individuals below estimated 3rd, 50th, 97th centiles of model (3, 2, 2, 2, 1), (2, 2, 2, 2) for females.

using the model (3, 2, 2, 1, 1), (4, 2, 2, 1, 1, 0), (2, 2, 2, 0). The fit to the raw percentiles is now somewhat improved over that in Figure 2.

Finally, we illustrate the use of covariates by fitting jointly to males and females. The model uses two sets of curves (3, 2, 2, 1, 0), (2, 1, 1, 0), again joined at 17 months. It is assumed that the groups differ in terms of the intercept and “slope” coefficients of the polynomials for both sets of curves. Figure 6 shows the jointly fitted smoothed polynomials.

5.1 Checking for Fit

We can study the fit of the estimated percentiles to the data graphically as follows. Starting with the youngest ages, an interval is chosen to include a fixed number of measurements—typically about 100. The numbers falling below each estimated percentile are counted and plotted at the midpoint of the interval. This is repeated, moving the starting age of the interval so as to omit and include the same number of measurements at the ends of the interval. To obtain a smooth plot this number should be small. The resulting plot then can be examined for trends or other irregularities. This procedure is implemented in the program GROSTAT II.

Figures 7 and 8 show plots of the actual percentages of individuals below the 3rd, 50th, and 97th percentiles by age corresponding to the models displayed in Figures 1 and 2. The improvement in fit is clear, with no obvious trends apparent in Figure 8.

6. Asymptotic Curves

There is a particular difficulty associated with the age range where percentiles reach asymptotic values. With height, for example, this occurs at the end of the adolescent growth period and each percentile reaches the asymptotic adult value at a different age. Another case is that of bone maturity scores, where all the percentiles eventually reach the same final mature value. This will therefore imply an equality constraint for the asymptotic values.

To see how such cases may be modelled, consider first the smoothing of a single percentile. If the age at which the adult value is achieved were known, we could choose this as an age boundary and fit a horizontal straight line for the asymptotic value. If the boundary point were unknown, then a set of different, successive, boundary points would be chosen and models fitted for each one. A suitable age range for these boundary points could be chosen by inspection. The model that produced the minimum residual sum of squares could then be chosen. The optimum boundary points will in general vary across percentiles so that the search for an overall minimum would involve a great deal of computing.

An alternative procedure is to use curves with a horizontal asymptote for the final age ranges. One possibility is the general inverse polynomial, which has the form

$$y = \sum_{i=0}^p \beta_i x^{-i}. \quad (10)$$

This function is linear in its coefficients and so can be incorporated within the general framework of our models. It is also asymptotic and so satisfies our requirement. In practice, a judicious choice of age origin will result in only a small value of p being necessary, but this is not crucial. It is also worth noting that (10) can be used for curves that are asymptotic with a negative slope (that is, concave functions), and also for curves with a lower asymptote. In the latter case, a suitable origin would be chosen that was greater than the range of ages defining the asymptote.

7. Discussion

A major advantage of a nonparametric approach to percentile estimation is that distributional assumptions are unnecessary. This is important for variables such as skinfold measurements and many biochemical measurements, where there may be no simple procedure for transforming, for example, to normality. The power transformation advocated by Cole (1988) may be adequate for some measurements such as weight. Nevertheless, it needs to be demonstrated in the case of each new measurement whether it in fact produces a close enough approximation to normality.

We have followed Healy et al. (1988) in using polynomials as smoothing functions since they are relatively easy to handle and interpret. For describing the regression of the coefficients on the NEDs, however, caution is needed if we wish to make any extrapolations to percentiles beyond those estimated from the raw data. There is no guarantee that the extrapolated values would even be smoothly increasing or decreasing. In general, we would not recommend such extrapolations. Some care also needs to be taken in specifying the order of the regression of the polynomial coefficients on the NEDs in relation to the constraints. Thus, for example, if this regression is of insufficiently high order, the constraints may not be linearly independent. This results in too few constraints to ensure that the curves join smoothly. This difficulty can usually be avoided by ensuring that, at each join point, sufficient of the $\beta^{(i)}$ have at least a quadratic regression on z . In practice, different combinations of model components will need to be tried, as illustrated in the examples. In general, considerable empirical exploration may be necessary to establish the appropriate degree of polynomials to be used, the number of subranges, and the join points. The GROSTAT II program allows such interaction with the data. Further research into this issue clearly would be useful.

A more general cubic spline with knots at the N_j observation points—that is, the raw percentile ages—could be considered instead of the piecewise model, but has important drawbacks. First, there is no straightforward way to incorporate other explanatory covariates or grouping factors into the model (see below). Second, while individual percentiles can be

smoothed, the joint smoothing of the percentiles would be very complicated. For these reasons we have not considered this possibility further.

Whereas this paper has been concerned with the joint estimation of population percentiles, the same procedure can be used to fit growth curves to individual series of longitudinal measurements. Thus, growth in height from birth could readily be fitted with a series of smoothly joining piecewise polynomials, with a final inverse polynomial for the approach to adulthood. Such a procedure would have several advantages over the more common nonlinear models. First, they can be used for the whole growth period or parts of it as required. Second, they can graduate “local” events accurately, given the availability of data points. Third, they are easy to compute, and fourth, they do not suffer the drawback of imposing rigid internal relationships between growth events [see, for example, Goldstein (1986)].

ACKNOWLEDGEMENTS

Thanks are due to Michael Healy and Jon Rasbash for their helpful comments. This work was partly supported by a WHO fellowship to the second author.

RÉSUMÉ

Une procédure non paramétrique de lissage conjoint de séries de courbes de percentiles est présentée. Elle permet le lissage de courbes distinctes sur des étendues d'âge contigües en forçant un lissage conjoint. Des covariables peuvent être introduites permettant ainsi des estimations simultanées efficaces par sous-groupes.

REFERENCES

Cole, T. J. (1988). Fitting smoothed centile curves to reference data (with Discussion). *Journal of the Royal Statistical Society, Series A* **151**, 385–418.
 Goldstein, H. (1986). Efficient statistical modelling of longitudinal data. *Annals of Human Biology* **13**, 129–141.
 Healy, M. J. R., Rasbash, J., and Yang, M. (1988). Distribution-free estimation of age-related centiles. *Annals of Human Biology* **15**, 17–22.
 Rasbash, J. and Pan, H. (1990). GROSTAT II: A program for estimating age-related centiles. London: World Health Organisation Collaborating Centre for Growth and Development, University of London.

Received January 1990; reviewed January and July 1991; accepted August 1991.

APPENDIX

The complete set of equations for the general model discussed in Section 2 can be written

$$\beta = ZA.$$

For the curves to join smoothly we shall require constraints (6) and (7), which lead to

$$\sum_{i=0}^{q_l} \beta_{ij}^{(l)} x_i^j = \sum_{i=0}^{q_{l-1}} \beta_{i-1,j}^{(l-1)} x_i^j \quad \text{or} \quad (\mathbf{X}^* \boldsymbol{\beta})_{x_l}^{(l)} = (\mathbf{X}^* \boldsymbol{\beta})_{x_l}^{(l-1)},$$

$$\sum_{i=1}^{q_l} i \beta_{ij}^{(l)} x_i^{j-1} = \sum_{i=1}^{q_{l-1}} i \beta_{i-1,j}^{(l-1)} x_i^{j-1} \quad \text{or} \quad \left(\frac{\partial \mathbf{X}^* \boldsymbol{\beta}}{\partial \mathbf{x}} \right)_{x_l}^{(l)} = \left(\frac{\partial \mathbf{X}^* \boldsymbol{\beta}}{\partial \mathbf{x}} \right)_{x_l}^{(l-1)}, \quad l = 2, \dots, m,$$

where x_l is the age at the $(l - 1)$ th join point. These constraints can be written in the general form $\mathbf{DB} = \mathbf{0}$.

The Z are known and this gives the constraint system

$$\mathbf{DZA} = \mathbf{F}^T \mathbf{A} = \mathbf{0}.$$

The OLS solution to (3) subject to these constraints is given by

$$\hat{\mathbf{A}} = \hat{\mathbf{A}}^* - (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{F} (\mathbf{F}^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{F})^{-1} \mathbf{F}^T \hat{\mathbf{A}}^*,$$

where

$$\hat{\mathbf{A}}^* = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y}$$

is the unconstrained solution.

We can also introduce a matrix \mathbf{V} of weights and obtain the corresponding generalised least squares (GLS) solution given by

$$\hat{\mathbf{A}} = \hat{\mathbf{A}}^* - (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{F} (\mathbf{F}^T (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{F})^{-1} \mathbf{F}^T \hat{\mathbf{A}}^*,$$

with

$$\hat{\mathbf{A}}^* = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{Y}.$$

We note that \mathbf{X} , \mathbf{F} have a block-diagonal structure so that the matrices to be inverted and stored are of order no larger than

$$\max[2(m-1)t, G], \quad G = \sum_{i=1}^m \sum_{j=0}^{q_i} (p_{ij} + 1).$$

The generalised least squares solution subject to linear constraints is derived as follows. We wish to estimate \mathbf{A} in

$$\mathbf{Y} = \mathbf{XA} + \mathbf{E},$$

where \mathbf{E} is a vector of residuals, $E(EE^T) = \mathbf{V}$.

We minimise

$$(\mathbf{Y} - \mathbf{XA})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{XA})$$

subject to

$$\mathbf{A}^T \mathbf{C} = \mathbf{k}^T.$$

Thus we minimise

$$(\mathbf{Y} - \mathbf{XA})^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{XA}) + (\mathbf{A}^T \mathbf{C} - \mathbf{k}^T) \boldsymbol{\lambda},$$

where $\boldsymbol{\lambda}$ is a vector of Lagrange multipliers. Differentiating with respect to \mathbf{A} , we obtain

$$2\mathbf{X}^T \mathbf{V}^{-1} (\mathbf{Y} - \mathbf{XA}) + \mathbf{C} \boldsymbol{\lambda} = \mathbf{0}.$$

Ignoring the factor 2, and multiplying from the left by $\mathbf{L} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1}$ gives an estimate, say

$$\hat{\mathbf{A}}^{(c)} = \hat{\mathbf{A}} - \mathbf{LC} \boldsymbol{\lambda}.$$

Utilising the constraint $\mathbf{C}^T \mathbf{A} = \mathbf{k}$, this gives

$$\mathbf{k} = \mathbf{C}^T \hat{\mathbf{A}} - (\mathbf{C}^T \mathbf{L} \mathbf{C}) \boldsymbol{\lambda}, \quad \boldsymbol{\lambda} = (\mathbf{C}^T \mathbf{L} \mathbf{C})^{-1} (\mathbf{C}^T \hat{\mathbf{A}} - \mathbf{k}),$$

which leads directly to

$$\hat{\mathbf{A}}^{(c)} = \hat{\mathbf{A}} - \mathbf{LC} (\mathbf{C}^T \mathbf{L} \mathbf{C})^{-1} (\mathbf{C}^T \hat{\mathbf{A}} - \mathbf{k}).$$

If $(\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})$ is singular, then \mathbf{L} is replaced by a generalised inverse and likewise if $(\mathbf{C}^T \mathbf{L} \mathbf{C})$ is singular a generalised inverse is used.