Diss. ETH 6039

Estimation and Parameterization of Growth Curves

A dissertation submitted to the

Swiss Federal Institute of Technology Zürich
for the degree of
Doctor of Mathematics

Presented by

WERNER STÜTZLE
Dipl. Math. ETH
born on September 14, 1950
Citizen of Western Germany

Accepted on the recommendation of

Prof. P. J. Huber, referee
Prof. A. Prader, co-referee

Juris Druck + Verlag Zürich
1977

ISBN 3 260 04357 8

CONTENTS

ACKNOWLEDGEMENTS

I would like to express my gratitude to the following persons for stimulation and help:

PD Dr. T. Gasser
Prof. Dr. P.J. Huber
Dr. R. Largo
Prof. Dr. A. Prader

Work was partially supported by the Fritz Hoffmann-La Roche foundation.

SUMMARY

Two methods are proposed to characterize the individual height growth of a child by a set of parameter values:

1) Estimation of the growth curve by smoothing splines.
   Calculation of the parameters directly from the estimated curve. Ways are proposed to spot gross errors in the data, to complete the measurement series by filling in missing observations, to determine the smoothing parameter and to reduce the bias caused by smoothing. The necessary programs have been implemented on a CDC-6000.

2) Parameterization by fitting a nonlinear model.
   This approach has been tried often before, but common to all papers published so far are two shortcomings:
   a) The models do not take into account that puberty stops growth by closing the epiphyses.
   b) The regression procedures do not make use of the fact, that usually one has observed many children and so has obtained many realizations of the same model and not only one.

   A model and a fitting algorithm without these shortcomings are proposed. An interactive program for this algorithm has been implemented on a DEC-10.

Finally, the advantages and disadvantages of the two methods are discussed.

ZUSAMMENFASSUNG

Zwei Methoden werden vorgeschlagen, um das individuelle Grössenwachstum eines Kindes durch einen Satz von Parameterwerten zu charakterisieren:

1) Schätzung der Wachstumskurve durch glättende Splines. Berechnung der Parameter direkt aus den geschätzten Kurven. Verfahren zur Entdeckung von gross errors, zur Vervollständigung von Messreihen durch Auffüllen fehlender Beobachtungen, zur Bestimmung des Glättungsparameters und zur Reduzierung des durch das Glätten verursachten Bias werden vorgeschlagen. Die notwendigen Programme sind auf einer CDC-6000 implementiert worden.

2) Parametrisierung durch Anpassung eines nichtlinearen Modells. Dieses Vorgehen ist schon oft versucht worden, aber allen bis jetzt veröffentlichten Arbeiten gemeinsam sind zwei Nachteile:

   a) Die Modelle berücksichtigen nicht, dass die Pubertät durch Schliessung der Epiphysenfugen das Wachstum beendet.

   b) Die Regressionverfahren nützen die Tatsache nicht aus, dass man gewöhnlich viele Kinder beobachtet hat und so über viele Realisierungen des gleichen Modells und nicht nur eine verfügt.

   Ein Modell und ein Fit-Algorithmus ohne diese Nachteile werden vorgeschlagen. Ein interaktives Program für diesen Algorithmus ist auf einer DEC-10 implementiert worden.

Schliesslich werden die Vor - und Nachteile der beiden Methoden diskutiert.

1. INTRODUCTION

During the years 1955-1976 a longitudinal study of human growth was carried out by the Kinderspital Zürich in cooperation with the "Centre International de l'Enfance" in Paris. 400 new-born Zürich children were selected. They were measured and interviewed at the Kinderspital, half-yearly in early childhood and during puberty, yearly, otherwise, until maturity. So for each child and each variable a longitudinal series of observations was obtained. (See [1] for the complete collection of questionnaires.)

The aim of this study was on one hand to provide data for a description of the normal growth process, thus offering the possibility to detect abnormal growth patterns. This is of practical use, as such abnormalities often are indicators for diseases like chromosome defects or disturbances in the hormonal system. On the other hand the study is helpful in answering mainly theorectical questions in the fields of endocrinology and developmental psychology.

How should one analyse such longitudinal data statistically ? I shall treat the case of height growth, but most of the ideas immediately generalize to other variables with values on an interval scale. The straightforward approach, considering each individual series of measurements as an observation vector and directly applying multivariate methods, for most questions make no sense: The dimension is too high (36 observations between birth and 20 years) and the results are hardly interpretable. The first step has to be a data reduction: Define a set of parameters (as few as possible), which summarize the important properties of the growth process, and find a method, which assigns to each measurement sequence the corresponding set of parameter values. Multivariate Analysis is then applied to the parameter vectors.

Two ways for performing this parameterization are suggested:

1) Curve estimation approach: Estimate the growth curve by a smoothing procedure. Calculate parameters directly from the estimated curve.

   This will be treated in Chapter 2. Results gained by the analysis of the so obtained parameters can be found in [2].

2) Regression approach: Fit a parametric model by least squares.

   This will be the subject of Chapter 3.

Finally, in Chapter 4 the two methods are compared.

2. THE CURVE ESTIMATION APPROACH

2.1. Characteristics of Human Height Growth

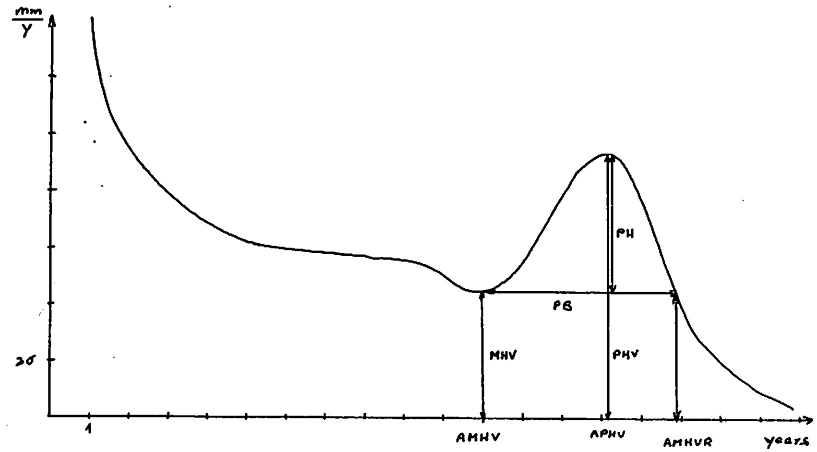

Figure - 1

To provide a better understanding of the following, in figure 1 a typical human height growth velocity curve and some of the parameters of interest are shown.

Growth velocity is highest after birth. It then decreases, first very fast, later more slowly, until the onset of puberty, which is associated with the adolescent growth spurt. After the end of puberty the epiphyses are closed and no further growth is possible. The basic parameters we estimated are parameters of the velocity curve, namely:

1) Peak height velocity PHV = maximum of the velocity curve between 9 and 16 years.

2) Age at peak height velocity APH = abscissa of maximum.

3) Minimal prespurt height velocity MHV = last local minimum before APH.

4) Age at minimal prespurt height velocity AMHV = abscissa of minimum.

5) Age at minimal prespurt height velocity return AMHVR = age at which, for the first time after APH, MHV is attained again.

6) Peak height PH = PHV - MHV

7) Peak basis PB = AMHVR - AMHV

2.2. The Smoothing Procedure

As smoothing procedure we use smoothing spline functions.
Let f be the unknown true function, which we want to estimate, $x_1 \dots x_n$ the abscissas, were we have observations, and $u_1 \dots u_n$ the observations.
We assume that

$$u_i = f(x_i) + \varepsilon_i$$

$$E(\varepsilon_i) = 0 \quad i = 1, \dots, n \qquad cov(\underline{\varepsilon}) = \Sigma$$

Def: The smoothing spline function s for $\underline{x}$, $\underline{u}$, Σ with smoothing parameter S is the solution of the minimum problem

$$\left.\begin{array}{l} \int s''^2 = min! \\ \text{under the constraint} \quad (\underline{u} - s(\underline{x}))^T \Sigma^{-1} (\underline{u} - s(\underline{x})) \le S \end{array}\right\} \text{problem } 1$$

Motivation: $\int s''^2$ is a measure for the smoothness of s, whereas $(\underline{u}-s(\underline{x}))^T\Sigma^{-1}(\underline{u}-s(\underline{x}))$ measures the distance of s from the given points. That means: s is the smoothest function with distance from the observations less or equal to S.

The properties of the smoothing spline functions are well known. See, for example, [3]. They are listed here only for the sake of completeness.

1) Problem 1 is equivalent to problem 2:

$$\int s''^2 + \lambda (\underline{u} - s(\underline{x}))^T \Sigma^{-1} (\underline{u} - s(\underline{x})) = min!$$

in the following sense:

Let l(x) be the least-squares straight line through ($\underline{x}$,$\underline{u}$) and R its residual sum of squares. Then for each S < R there exists a unique λ such that the solutions of problem 1 and of problem 2 are the same. λ is called Lagrange parameter.

2) In the intervals $[x_i, x_{i+1}]$, i=1...n-1, s coincides with a polynominal of degree 3. In the points x_i, called knots, these segments are fitted together two times differentiably.

3) For fixed λ the coefficients of the polynomials depend linearly on the observations u_i. This especially means that the smoothed ordinates $y_i = s(x_i)$ depend linearly on the observations:

$$\underline{y} = W(\lambda)\underline{u}$$

An algorithm for the computation of s is described in Chapter 2.4.

The Role of the Smoothing Parameter S

The smoothing parameter steers the smoothness of s. Let us first consider two examples:

1) S = 0. The spline function passes exactly through the points (x_i, u_i). Thus s might be rather wiggly.

2) S $\geq$ R. The spline function is the least-squares straight line.

The influence of S may also be discussed in terms of bias and variance:

1) S = 0. $s(x_i)$ is an unbiased estimate of $f(x_i)$, but it has the same variance as ε_i.

2) S $\geq$ R. s coincides with the least-squares straight line. So $s(x_i)$ estimates $f(x_i)$ with a small variance, but, depending on f, the bias may become very large.

As a criterion for the quality of the estimate, we use as usual the average expected squared error

$$AESE = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(s(x_i)-f(x_i)\right)^2\right)$$

Because $\underline{y}$ depends linearly on $\underline{u}$ the AESE is the sum of a bias term b and a variance term v, where for given λ resp. S the bias term depends only on f and the variance term only on the ε_i :

$$AESE = b+v$$
$$b = \frac{1}{n}\,\|\,W(\lambda)\underline{f}(x)-\underline{f}(x)\,\|^2$$
$$v = \frac{1}{n}\,trace\,\left(W(\lambda)\,\Sigma\,W^T(\lambda)\right)$$

For S increasing (λ decreasing) the bias component of AESE increases and the variance component decreases. If Σ and f were known, one could analytically determine the optimal λ . But of course this has no practical meaning.

In the following we shall always assume that $\Sigma = \sigma^2 C$, where only σ^2 is unknown, but C is a known matrix.

How to Determine the Smoothing Parameter ?

There are a few possibilities:

1) For each individual series of measurements plot the points (x_i,u_i) and the spline for several S. Choose that S for which the spline looks smooth enough and for which the bias is not too large. This may for example be done using a graphics terminal as an output device.

   This procedure has several disadvantages: It is time-consuming (think of 200 - 400 series of measurements) and not reproducible.

2) Use an asymptotic result (see [4])

$$S_{opt} = c(f)\cdot\sigma^2\cdot n$$

   where $c(f) \rightarrow 1$ as the sampling rate increases. (The sampling rate is defined as $1/\sup(x_{i+1}-x_i)$. In our context "asymptotically" will always mean "for increasing sampling rate").

   Good choices of c(f) are between 0.8 and 1 (see [5]).
   The only remaining problem is to find a reasonable estimate of σ^2.

3) Estimate the optimal λ or S for each individual by cross-validation (see [5]). The cross-validation procedure may be described as follows:

   a) Choose a λ.
   b) For i = 1...n define weights

$$W^*_{ij}(\lambda) = \frac{W_{ij}(\lambda)}{1-W_{ii}(\lambda)} \qquad j \neq i$$

      Define the i-th cross-validation residual d_i by

$$d_i(\lambda) = \sum_{j\neq i} W^*_{ij}(\lambda)\,u_j - u_i$$

      Loosely speaking, d_i is the difference between u_i and the ordinate at x_i, estimated without using (x_i,u_i).
   c) Define the cross-validation sum of squares CV(λ) by

$$CV(\lambda) = \underline{d}^T\Sigma^{-1}\underline{d}$$

      This procedure is repeated in order to find the λ^* for which CV(λ) is minimized. Now there are two possibilities for proceeding:
   d) Use λ^* as Lagrange parameter for smoothing the particular series of measurements

      or
   e) Define the smoothing parameter S as $c(f)\cdot CV(\lambda)$

We use 3d with some modifications (See Chapter 2.3).

Reducing the Bias ("Prewhitening")

It suggests itself, to try to estimate the bias introduced by smoothing and (partially) eliminate it. Heuristically such a bias correction would look as follows:

1) Assume S to be given. Determine λ, get $W(\lambda)$ and the smoothed ordinates

$$y_i = \sum_j w_{ij}(\lambda)\, u_j$$

as well as the other polynomial coefficients of the smoothing spline s.

2) Smooth the (x_i, y_i) once more with the same weight matrix. Define

$$y_i^* = \sum_j w_{ij}(\lambda)\, y_j$$

and call the obtained spline function s*. Use s-s* as an estimate of the bias. Define the corrected spline function s^c as

$$s^c = s + (s - s^*) = 2s - s^*$$

The smoothed ordinates of s^c are $y^c = 2y - y^*$

In Chapter 2.6 an asymptotic argument is given, that this correction really eliminates the highest terms (in λ) of the bias.

The bias correction may also be considered as "prewhitening" :
First the large effects are found by smoothing the measurements.
Then these effects are taken out by subtraction and the residuals are smoothed again.

2.3. Data Checking, Treatment of Missing Data, Estimation of Parameters

Data Checking

Before performing any analysis of the data, they have to be checked for gross errors. The reasons for such errors are mainly interchange of rows on the questionnaires and punching errors. For error detection, it is necessary to make use of the longitudinal structure of the data and not only use cross-sectional methods. This can be seen looking at a simple example:
The (cross-sectional) standard deviation of standing height in boys is
63 mm at 16 years, mean 1738 mm
62 mm at 17 years, mean 1759 mm
Now think of a boy, who was measured 1700 mm at 16 years and who was punched 1791 instead of 1719 at 17 years. Of course, no cross-sectional rejection rule would reveal this error, whereas a longitudinal one should.

We suggest cross-validation as a tool for outlier detection.
The procedure works in the following way:
Given a series of measurements (x_i, u_i).

1) Perform cross-validation and obtain λ^*, the Lagrange parameter, for which $CV(\lambda)$ is minimal.
2) For i=1...n compute $d_i(\lambda)$, the i-th cross-validation residual.
3) Apply a normal rejection rule to the d_i, for example consider a point as an outlier if the corresponding d_i is more than 3 median deviations away from the median of the d_i.

Filling in Missing Observations

We found it useful to complete the measurement series if possible. This leads to the same design for all individuals when smoothing the velocity curves, which considerably simplifies programs and reduces computing costs.

Measurement series were not completed if

1) more than 3 observations were missing (out of 24)

2) 2 consecutive observations were missing

The completion procedure works as follows:
Assume (x_i, u_i) missing.

1) Choose a smoothing parameter S. This may be done separately for each series by cross-validation. Another way is, to use a smoothing parameter according to average error variance over the sample. It does not seem to make a great difference.

2) Compute the smoothing spline s for parameter S and the given points. Set $u_i^* = s(x_i)$

3) Compute a smoothing spline for parameter S though $(x_1,u_1)\ldots(x_i,u_i^*)\ldots(x_n,u_n)$. Define $u_i^{**} = s(x_i)$. If $|(u_i^* - u_i^{**})|$ is less than a selected bound, goto 4, else set $u_i^* = u_i^{**}$ and goto 3.

4) Perform bias correction. Use u_i^c as a pseudo-observation.

In simulations this procedure showed a good performance (see Chapter 2.5). Of course, the "error" of the pseudo-observation is correlated with the errors of near-by observations. We neglect this effect, as otherwise the advantages of completion would get lost.

Estimation of the Parameters

We now describe the complete procedure which led to the parameters finally

1) Check the individual series of measurements for outliers.

2) Complete the series. Series which could not be completed, were not further used.

3) For each series compute the raw velocities

$$v_i\left(\frac{t_K + t_{K+1}}{2}\right) = \frac{h_i(t_{K+1}) - h_i(t_K)}{t_{K+1} - t_K} \qquad K = 1 \cdots n-1$$

In the following observe that

$$cor(v_i, v_{i+1}) = cor(v_i, v_{i-1}) = -\frac{1}{2}$$

$$var(v_i) = 2\sigma^2/(t_{i+1} - t_i)^2 \quad if \quad var(h(t_i)) = \sigma^2 \ \forall i$$

4) Find λ_i^*, the Lagrange parameter which minimizes the cross-validation sum of squares for individual i. Set $S_i = S(\lambda_i^*)$.

5) Compute μ = med S_i, δ = meddev S_i

6) Set

$$S_i^* = \begin{cases} \mu+\delta & \mu+\delta < S_i \\ S_i & \mu-\delta \le S_i \le \mu+\delta \\ \mu-\delta & \mu-\delta > S_i \end{cases}$$

7) Smooth the raw velocities of individual i with smoothing parameter S_i, perform bias correction and obtain spline s_i. Two examples for smoothed curves can be found on photos 1 and 2 (Appendix 1)

8) Estimate the parameters for individual i by the corresponding parameters of s_i.

The only thing special here is step 6, the winsorizing of smoothing parameters. It is the result of a compromise: If one would assume error variances equal for all individuals, it would be appropriate to use $\overline{\lambda}^*$ as a smoothing parameter for all individuals. This is not the case. We do not know the reason, but we assume that seasonal effects play a role here (see [6]). Such effects would lead to a higher variance in half-year measurements for children born in spring or autumn compared to those born in summer or winter. On the other hand, the sampling rate we have seems to lie at the lower limit, where cross-validation gets usable. Using the unmodified smoothing parameters and then visually inspecting the smoothed curves, some of them seemed us strongly over-smoothed, others were not smoothed at all. However, the effect was in the right direction. So we decided to use the compromise described above.

2.4. Algorithms and Routines for Spline-Smoothing and Cross-Validation

By definition, the smoothing spline function for data points $(x_1,u_1)\ldots(x_n,u_n)$, error covariance matrix Σ and smoothing parameter S_0 is the solution of the minimum problem

$$\int s''^2 = \min !$$

under the constraint

$$(\underline{u} - s(\underline{x}))^T \Sigma^{-1} (\underline{u} - s(\underline{x})) \leq S_0$$

Using variational calculus, this leads to the equation for $\underline{s}(\underline{x}) = \underline{y}$:

$$\underline{y} = - \Sigma B^T (\lambda A + B \Sigma B^T)^{-1} B \underline{u} + \underline{u}$$

$$A = \frac{1}{3} \begin{array}{c} \left[\begin{matrix} 2(h_1+h_2) & h_2 & 0 & 0 \\ h_2 & 2(h_2+h_3) & h_3 & 0 \\ 0 & h_3 & 2(h_3+h_4) & h_4 \\ & & \ddots & \ddots \end{matrix}\right] n-2 \\ n-2 \end{array} \qquad B = \begin{array}{c} \left[\begin{matrix} -\frac{1}{h_1} & \frac{1}{h_1}+\frac{1}{h_2} & -\frac{1}{h_2} & 0 & 0 \\ 0 & -\frac{1}{h_2} & \frac{1}{h_2}+\frac{1}{h_3} & -\frac{1}{h_3} & 0 \\ 0 & 0 & -\frac{1}{h_3} & \frac{1}{h_3}+\frac{1}{h_4} & -\frac{1}{h_4} \\ & & & \ddots & \ddots \end{matrix}\right] n-2 \\ n \end{array}$$

$$h_i = x_{i+1} - x_i$$

$\lambda = 0$ gives the least-squares straight line. For given S_0, λ is determined by a Newton procedure.. With

$$\Sigma = R^T R \qquad \Sigma^{-1} = R^{-1} R^{-T}$$

we get

$$S^2(\lambda) = \| R B^T (\lambda A + B \Sigma B^T)^{-1} B \underline{u} \|^2$$

We will show later, that $1/S(\lambda)$ is a concave function of λ. The Newton procedure is used to find a solution of

$$\frac{1}{S(\lambda)} - \frac{1}{\sqrt{S_0}} = 0$$

Cross-Validation

If

$$\underline{y} = W(\lambda) \underline{u}$$

the cross-validation sum of squares $CV(\lambda)$ was defined as follows:

1) For $i=1\ldots n$ compute

$$W_{ij}^*(\lambda) = \frac{W_{ij}(\lambda)}{1 - W_{ii}(\lambda)} \qquad j \neq i$$

Define d_i, the ith cross-validation residual, by

$$d_i(\lambda) = \sum_{j \neq i} W_{ij}^*(\lambda) u_j - u_i$$

2) Define

$$CV(\lambda) = \underline{d}^T \Sigma^{-1} \underline{d}$$

With $\alpha = 1 - W_{ii}$ this yields

$$d_i(\lambda) = \frac{1}{\alpha} \left(W_{i1}(\lambda) u_1 + \ldots + (W_{ii}(\lambda) - 1) u_i + \ldots + W_{in}(\lambda) u_n \right)$$

For smoothing splines the weight matrix is

$$W(\lambda) = I - \Sigma B^T (\lambda A + B \Sigma B^T)^{-1} B$$

So we obtain for the cross-validation residuals

$$d_i(\lambda) = \frac{-\Sigma_i B^T (\lambda A + B \Sigma B^T)^{-1} B \underline{u}}{\Sigma_i B^T (\lambda A + B \Sigma B^T)^{-1} B^i}$$

For $\Sigma = \operatorname{diag}(\sigma_1^2 \ldots \sigma_n^2)$ this formula has a nice interpretation:

Theorem: $d_i(\lambda) = \lim_{\sigma_i \to \infty} (y_i - u_i)$

So $d_i(\lambda)$ is in this case really the difference between the observed ordinate u_i and the value of the spline at x_i, when the spline has been computed omitting the point (x_i,u_i)

Proof:

Set $D = \operatorname{diag}(\sigma_1 \ldots\ldots \rho\sigma_i \ldots\ldots \sigma_n)$ and $Z = BD$. Then we have

$$y_i = -\rho\sigma_i z_i^T (\lambda A + Z Z^T)^{-1} B \underline{u} + u_i$$

$$= -\rho\sigma_i z_i^T (\underbrace{\lambda A + \sum_{k \neq i} z^k z^{kT}}_{C} + \rho z^i z^{iT})^{-1} B \underline{u} + \underline{u}_i$$

Using the Sherman-Morrison identity

$$(A + \underline{u}\underline{v}^T)^{-1} = A^{-1} - \frac{A^{-1} \underline{u}\, \underline{v}^T A^{-1}}{1 + \underline{v}^T A^{-1} \underline{u}}$$

we get

$$y_i = -\rho\,\sigma_i\, z^{iT}\left(C^{-1} - \frac{C^{-1} z^i z^{iT} C^{-1}}{\frac{1}{\rho} + z^{iT} C^{-1} z^i} \right) B\underline{u} + u_i$$

$$= -\rho\,\sigma_i \frac{\frac{1}{\rho} z^{iT} C^{-1} B\underline{u} + z^{iT} C^{-1} z^i z^{iT} C^{-1} B\underline{u} - z^{iT} C^{-1} z^i z^{iT} C^{-1} B\underline{u}}{\frac{1}{\rho} + z^{iT} C^{-1} z^i} + u_i$$

$$= \frac{-\sigma_i z^{iT} C^{-1} B\underline{u}}{\frac{1}{\rho} + z^{iT} C^{-1} z^i} + u_i$$

$$\lim_{\rho\to\infty} (u_i - y_i) = \frac{\sigma_i z^{iT} C^{-1} B\underline{u}}{z^{iT} C^{-1} z^i}$$

$$= \frac{B^{iT}(\lambda A + z^{(i)} z^{(i)T})^{-1} B\underline{u}}{B^{iT}(\lambda A + z^{(i)} z^{(i)T})^{-1} B^i}.$$

$$z^{(i)} = (z^1 \ldots z^{i-1}\; z^{i+1} \ldots z^n)$$

Using the Sherman-Morrison identity once more, this gives for the numerator

$$B^{iT}(\lambda A + z^{(i)} z^{(i)T})^{-1} B\underline{u} = B^{iT}((\underbrace{\lambda A + zz^T}_{C}) - z^i z^{iT})^{-1} B\underline{u}$$

$$= B^{iT}\left(C^{-1} + \frac{C^{-1} z^i z^{iT} C^{-1}}{1 - z^{iT} C^{-1} z^i} \right) B\underline{u}$$

$$= \frac{B^{iT} C^{-1} B\underline{u}}{1 - z^{iT} C^{-1} z^i}$$

Analog we get for the denominator

$$B^{iT}(\lambda A + z^{(i)} z^{(i)T})^{-1} B^i = \frac{B^{iT} C^{-1} B^i}{1 - z^{iT} C^{-1} z^i}$$

So we finally obtain

$$\lim_{\rho\to\infty} (u_i - y_i) = \frac{B^{iT} C^{-1} B\underline{u}}{B^{iT} C^{-1} B^i} = \frac{B^{iT}(\lambda A + zz^T)^{-1} B\underline{u}}{B^{iT}(\lambda A + zz^T)^{-1} B^i}$$

Algorithm for Spline Computation

This algorithm was proposed to me by Gene Golub. The aim is, to be able to evaluate

$$\Sigma B^T(\lambda A + B\Sigma B^T)^{-1} B$$

for many different values of λ without much cost.

Defining

$$A = R^T R \qquad \Sigma = C^T C \qquad Z = BC^T$$

we get

$$\Sigma B^T(\lambda A + B\Sigma B^T)^{-1} B = C^T Z^T(\lambda R^T R + ZZ^T)^{-1} Z C^{-T}$$

$$= C^T Z^T R^{-1}(\lambda I + R^{-T} Z Z^T R^{-1}) R^{-T} Z C^{-T} \qquad (I)$$

Set

$$G = R^{-T} Z = R^{-T} B C^T$$

This yields

$$(I) = C^T G^T (\lambda I + GG^T) G C^{-T}$$

Using the singular value decomposition of G^T

$$G^T = HDV^T \qquad H^T H = I \qquad V^T V = I \qquad D \text{ diagonal}$$

we get

$$(I) = C^T H D (\lambda I + D^2)^{-1} B H^T C^{-T}$$

and, setting

$$\varphi_\lambda = \mathrm{diag}\,(D_{ii}^2 / (D_{ii}^2 + \lambda))$$

$$\Sigma B^T(\lambda A + B\Sigma B^T)^{-1} B = C^T H \varphi_\lambda H^T C^{-T}$$

So if abscissas and covariance matrix remain the same, we only once perform the operations listed here and compute $C^T H$ and $H^T C^{-T}$. This is done by the routine SPLINIT.

Further on, we easily obtain

$$S^2(\lambda) = \| H \varphi_\lambda H^T C^{-T} \underline{u} \|^2 = \| \varphi_\lambda H^T C^{-T} \underline{u} \|^2$$

or, with $\underline{v} = H^T C^{-T} \underline{u}$

$$S^2(\lambda) = \| \varphi_\lambda \underline{v} \|^2$$

The argument, that $1/S(\lambda)$ is concave, is the same as in [7].

The Newton Procedure for Determining λ .

We want to find a solution of

$$\frac{1}{S(\lambda)} - \frac{1}{\sqrt{S_0}} = S^*(\lambda) = \sigma$$

We start with $\lambda_0 = 0$.

$$\Delta\lambda = -\frac{S^*(\lambda_0)}{S^{*\prime}(\lambda_0)} = \frac{S(\lambda_0)}{S'(\lambda_0)}\left[1 - \frac{S(\lambda_0)}{\sqrt{S_0}}\right]$$

$$S^2(\lambda) = \sum_i (\varphi_{\lambda ii} v_i)^2$$

$$S'(\lambda) = \frac{1}{2}\left[\sum (\varphi_{\lambda ii} v_i)^2\right]^{1/2} \cdot 2 \sum v_i^2 \varphi_{\lambda ii} \varphi'_{\lambda ii}$$

$$= -\left[\sum_i \frac{D_{ii}^4}{(D_{ii}^2+\lambda)^2} v_i^2\right]^{-1/2} \sum_i \frac{D_{ii}^4}{(D_{ii}^2+\lambda)^3} v_i^2$$

After having determined λ, we easily obtain the smoothed ordinates $\underline{y}$. The polynomial coefficients are then determined by computing the interpolating spline function for the points $(x_1,y_1)\ldots(x_n,y_n)$. This requires essentially a Choleski decomposition and a forward-backward substitution. All this is done by the routine SPLINE.

Algorithm for Cross-Validation

Using the same notation as before, we get

$$d_i(\lambda) = \frac{(C^T)_i H \varphi_\lambda H^T C^{-T} \underline{u}}{(C^T)_i H \varphi_\lambda H^T (C^{-T})^i}$$

and

$$CV(\lambda) = \| C^{-1} \underline{d} \|^2$$

Before calling the function CV, initialisation has to be done by calling SPLINIT. Then, if S is supplied, the corresponding λ is computed as in SPLINE. Afterwards, the cross-validation sum of squares is evaluated straightforwardly.

There are two more subroutines in the package:
MINCV, a minimization routine for minimizing CV, and SPLINT, a function for spline interpolation. Of course, before calling SPLINT, the spline coefficients have to be computed by calling SPLINE.

2.5. Monte-Carlo Results

As true underlying functions for our simulations we use two curves out of a parametric family used in [8] to parameterize human growth curves.

The family is given by the equation

$$h(t, a_1, b_1, c_1, a_2, b_2, c_2) = \frac{a_1}{1 + e^{-b_1(t-c_1)}} + \frac{a_2}{1 + e^{-b_2(t-c_2)}}$$

We choose the parameter combinations

	Model 1	Model 2
a_1	156.7	156.7
b_1	0.24	0.28
c_1	1.18	0.48
a_2	35.30	27.00
b_2	1.07	0.73
c_2	12.23	12.23

Model 1 corresponds to a growth velocity curve with very high peak (PHV = 117.8 mm/year, PH = 63.9 mm/year), Model 2 to one with very low peak (PHV = 65.1 mm/year, PH = 15.9 mm/year).

Filling in Missing Observations

Three cases were considered in the simulations: Observation missing at age 9, 11, or 13.5 years. As both curves have an APH of 12.1 years, the cases where an observation is missing at 11 or 13.5 years are critical, because one expects a bias here. 9 years is just the age, where the spurt starts.

Standard deviations of simulated measurement error were 2, 4 and 7 mm. 30 runs per combination of model, missing observation and measurement error were simulated. The smoothing in the completion routine was done according to a variance of measurement error of 16 mm^2 in all cases. The results are summarized in table 1.

Model	Error Std (mm/year)	Age (Years)	Bias (mm)	MSE (mm^2)
1	2.0	9	3.0	10.8
1	2.0	11	-3.4	13.2
1	2.0	13.5	2.4	8.0
1	4.0	9	0.8	9.5
1	4.0	11	1.2	11.7
1	4.0	13.5	-0.8	14.9
1	7.0	9	0.2	87.4
1	7.0	11	2.9	43.6
1	7.0	13.5	-3.5	53.6
2	2.0	9	0.6	1.2
2	2.0	11	-1.8	3.8
2	2.0	13.5	1.6	3.6
2	4.0	9	1.8	12.8
2	4.0	11	0.5	3.4
2	4.0	13.5	-0.5	6.1
2	7.0	9	0.4	46.3
2	7.0	11	2.6	56.2
2	7.0	13.5	-0.9	85.3

TABLE - 1

The bias of the filled-in observation does not exceed 3.5 mm in absolute value. For error variance 16 mm^2 we get an excellent behaviour. For error variance 49 mm^2 the behaviour is not quite as good, but still MSE is within two times the error variance. The only case, where the completion procedure is bad, is : Model 1 (with the high peak), error variance 4 mm^2. Here oversmoothing spoils the results : large bias of up to 3.5 mm, MSE of two to three times the error variance. So one might produce points identifiable as outliers. But of course this worst case is not likely to appear often in the data. We decided to run the risk of adding an outlier in some cases in order to make use of the large advantages offered by having only complete series of observations.

Parameter Estimation

We again simulated height measurements with standard deviations of 2, 4 and 7 mm for each of the two parameter sets. 20 runs per combination were generated. From the simulated height measurements raw velocities were computed. Lagrange parameters for smoothing the velocities were determined by cross-validation. No winsorizing of smoothing parameters was performed.

The question we want to answer by means of these simulations is: Are the estimated parameters "valid", i.e. do they tell us something meaningful about the population. This may be done by comparing the mean squared errors of the parameters with the variance of the parameters over the population. Of course the MSEs depend on the underlying true curve, in our case mainly on its peak height. That is why we chose as true underlying curves for our simulation one with a very high peak, the other with a low peak.

In table 2 we give for our parameters the average mean squared errors for the two parameter combinations and the variance over the sample.

If one accepts as a rule of thumb, that for meaningful parameters the standard deviation over the population should be about 3 times the standard error (square root of MSE), one can draw the following conclusions:

For 2 mm error standard deviation we have no problems.

For 7 mm error standard deviation the parameters are useless. This is not astonishing, as the error variance of half-year velocity measurements is 392 mm^2, which makes a standard deviation of nearly 20 mm/year.

4 mm standard deviation is the most interesting case, for the true standard deviation of height measurements is about 3-4 mm. Here the parameters PHV, APH and AMHVR are well determined, AMHV and PH are tolerable. As was to be expected, PB and MHV are critical: PB because AMHV enters, which itself is not very well determined, and MHV, because the variance over the population is small compared for example with PHV and PH. So the results based on these two parameters have to be regarded with some caution.

	2 mm		4 mm		7 mm		Pop Var
	Par 1	Par 2	Par 1	Par 2	Par 1	Par 2	
PHV	4.5	1.15	14.6	3.15	31.7	21.4	120
APH	0.009	0.02	0.03	0.06	0.03	0.21	0.69
MHV	1.2	0.63	14.7	4.12	12.4	18.3	35
AMHV	0.17	0.05	0.19	0.14	0.27	0.55	1.34
AMHVR	0.005	0.01	0.02	0.04	0.04	0.19	0.75
PH	6.3	2.25	28.2	9.68	23.4	72.1	157
PB	0.16	0.09	0.22	0.14	0.39	0.48	0.86

<u>TABLE - 2</u>

How well the smoothing procedure picks out the true curve, is illustrated by photo 3...6 (Appendix 1). Each photo shows a simulated measurement series (height measurement error std = 4 mm), the true underlying curve (dashed) and the estimated curve (solid).

2.6. Asymptotics

To provide some insights into the asymptotics of curve estimation, we first set up the problem in general. Then for a simple type of smoothing procedure some asymptotic properties are derived and the asymptotic effect of the bias correction described in Chapter 2.2 is evaluated. Finally, we reproduce results on the asymptotics of spline smoothing and cross-validation given by G. Wahba in [4], [10].

General Situation

Given an unknown function f, the properties of which we will specify from case to case. At the points

$$x_i = i\Delta \qquad i = 0 \ldots n \qquad \Delta = \frac{1}{n}$$

we observe

$$u_i = f(x_i) + \varepsilon_i \qquad E(\varepsilon_i) = 0 \qquad \operatorname{cov}(\varepsilon_i, \varepsilon_j) = \sigma^2 \delta_{ij}$$

We want to estimate $f(x_i)$, i=1...n.
The criterion for the quality of the estimate will be the average expected squared error

$$AESE = \frac{1}{n} \sum_{i=0}^{n} \left(f(x_i) - \hat{f}(x_i) \right)^2$$

It is often convenient to assume, that one has observations all over the real line, not only in the interval [0,1]. This simplifies the theory by avoiding boundary effects, and in most cases has no influence on the asymptotic behaviour of the AESE.

Most asymptotic theory developed so far deals with linear estimates of the moving average type. Such an estimate is defined by a family $\underline{w}(n, \lambda)$ of sequences with

$$\sum_{-\infty}^{\infty} w_i(n,\lambda) = 1$$

at least asymptotically (for $n \to \infty$, $\lambda \to \infty$).

For fixed n the sequence $\hat{\underline{f}}(n, \lambda)$ of estimated ordinates is given by

$$\hat{\underline{f}}(n,\lambda) = \underline{u} * \underline{w}(n,\lambda)$$

(* denotes the convolution)

Let us demonstrate this at the example of rectangular weights.

Define

$$w_i(n,\lambda) = \begin{cases} \frac{\lambda}{n} & -\frac{n}{2\lambda} \le i \le \frac{n}{2\lambda} \\ 0 & \text{elsewhere} \end{cases}$$

Then

$$\hat{f}_i(n,\lambda) = \sum_{j=-\infty}^{\infty} u_{i-j}\, w_j(n,\lambda) = \frac{\lambda}{n} \sum_{-\frac{n}{2\lambda}}^{\frac{n}{2\lambda}} f_{i-j} + \frac{\lambda}{n} \sum_{-\frac{n}{2\lambda}}^{\frac{n}{2\lambda}} \varepsilon_{i-j}$$

Assume f to be continuous. Then we immediately see that $\hat{f}_i(n, \lambda)$ is a consistent estimate of f_i if

$$\frac{\lambda}{n} \to 0 \quad \text{because} \quad \frac{\lambda}{n} \to 0 \implies \frac{\lambda}{n} \sum \varepsilon_{i-j} \to 0 \text{ a.s.}$$

$$\lambda \to \infty \quad \text{because} \quad \lambda \to \infty \implies \frac{\lambda}{n} \sum f_{i-j} \to f_i$$

Neglecting boundary effects, smoothing splines are linear estimates of moving average type. This may easily be seen by considering spline smoothing in a situation, where the abscissas are equidistant points on the circle. λ corresponds to the Lagrange parameter.

Moving Averages Derived From a Kernel

This type of estimate has been treated often before in the context of density or spectrum estimation (see, for example, [9]).

Let K be a Lebesque integrable, two times continuously differentiable symmetric kernel with compact support and let K and its first two derivatives be bounded. Define $K_\lambda(x) = \lambda K(\lambda x)$. For $\lambda \to \infty$ K_λ becomes more and more concentrated around 0. $\int_{-\infty}^{\infty} K(x)\,dx = 1$

We define the weights

$$w_i(n,\lambda) = \Delta K_\lambda(i\Delta) = \lambda\Delta K(i\lambda\Delta)$$

For

$$\frac{\lambda}{n} \to 0 \qquad \sum w_i(\lambda,n) \to 1$$

So we get

$$\hat{f}_i(n,\lambda) = \lambda\Delta \sum_j K(j\lambda\Delta)\, u_{i-j}$$

This gives for the AESE

$$AESE = \frac{1}{n} E \sum_i (\hat{f}_i(n,\lambda) - f_i)^2 = \frac{1}{n} E \sum_i (\lambda\Delta \sum_j K(j\lambda\Delta) u_{i-j} - f_i)^2$$

$$= \Delta \sum_i [\lambda\Delta \sum_j K(j\lambda\Delta) f_{i-j} - f_i]^2 + \sigma^2\lambda^2\Delta^2 \sum_j K^2(j\lambda\Delta)$$

We approximate this by

$$IESE = \int_0^1 [\int_{-\infty}^{\infty} \lambda K(\lambda t)(f(x-t) - f(x))\,dt]^2 dx + \sigma^2\lambda\Delta \int_{-\infty}^{\infty} K^2(t)\,dt$$

Lemma

The approximation error is of the order of magnitude $o(\lambda^2\Delta^2)$

Proof:

Let supp(K) $\subset [-a, a]$

Then supp(K_λ) $\subset [-\frac{a}{\lambda}, \frac{a}{\lambda}]$

We are first interested in approximating

$$\lambda \int_{-\frac{a}{\lambda}}^{\frac{a}{\lambda}} K(\lambda t) f(x_i - t)\,dt \quad \text{by} \quad \lambda\Delta \sum K(j\lambda\Delta) f_{i-j}$$

Without loss of generality we assume $x_i = 0$.

Set

$$g_\lambda(t) = \lambda K(\lambda t) f(t)$$

Then the error in the interval $[-\frac{\Delta}{2}, \frac{\Delta}{2}]$ is

$$\left|\int_{-\frac{\Delta}{2}}^{\frac{\Delta}{2}} g_\lambda(t)\,dt - \Delta g_\lambda(\sigma)\right| \le -\frac{1}{2}\left|\int_{-\frac{\Delta}{2}}^{\Delta/2} x^2 g''_\lambda(\xi(x)x)\right| \le \sup|g''_\lambda(x)|\frac{\Delta^3}{24}$$

This inequality is obtained by simply expanding g_λ in a Taylor series around 0.

The number of such intervals of length Δ grows with $2a/\lambda\Delta$.
So the total approximation error becomes

$$\left|\int_{-\frac{a}{\lambda}}^{\frac{a}{\lambda}} g_\lambda(t)\,dt - \Delta\sum_{-\frac{a}{\lambda\Delta}}^{\frac{a}{\lambda\Delta}} g_\lambda(i\Delta)\right| \le \sup|g''_\lambda(x)|\frac{2a}{\lambda}\frac{\Delta^2}{24}$$

The dominating term in $g_\lambda''(x)$ is $\lambda^3 K''(\lambda x)f(x)$. So finally have

$$\left|\int_{-\frac{a}{\lambda}}^{\frac{a}{\lambda}} g_\lambda(t) - \Delta\sum_{-\frac{a}{\lambda\Delta}}^{\frac{a}{\lambda\Delta}} g_\lambda(i\Delta)\right| \le \frac{a}{12}\sup|K''(\lambda x)f(x)|(\lambda\Delta)^2 = O(\lambda^2\Delta^2)$$

Now we are prepared to look at one term we are really interested in:
What is the error when approximating

$$\Delta^2\lambda\sum_k\Big[\sum_i K(i\lambda\Delta)f_{k-i} - f_k\Big]^2 \text{ by } \int_0^1\Big[\lambda\int_{-\frac{a}{\lambda}}^{\frac{a}{\lambda}} K(\lambda t)f(x-t) - f(x)\,dt\Big]^2dx$$

Let us use the following nomenclature:

$$f_\lambda(x) = \int\lambda K(\lambda t)f(x-t)\,dt$$

$$f^*_\lambda(x) = \lambda\Delta\sum K(i\lambda\Delta)f(x_k - i\Delta)$$

$$\delta(x) = [f_\lambda(x) - f(x)]^2$$

$$\delta^*(x_k) = [f^*_\lambda(x_k) - f(x_k)]^2$$

Then we have

$$\left|\int_0^1\Big[\int\lambda K(\lambda t)(f(x-t)-f(x))\,dt\Big]^2dx - \Delta\sum_k\Big[\lambda\Delta\sum_i K(i\lambda\Delta)f(x_k - i\Delta) - f(x_k)\Big]^2\right|$$

$$= \left|\int_0^1\delta(x)dx - \Delta\sum_{k=1}^n\delta^*(x_k)\right| = \left|\int_0^1\delta(x)dx - \Delta\sum\delta^*(x_k) + \Delta\sum\delta(x_k) - \Delta\sum\delta(x_k)\right|$$

$$\le \left|\int_0^1\delta(x)dx - \Delta\sum\delta(x_k)\right| + \left|\Delta\sum(\delta^*(x_k) - \delta(x_k))\right|$$

$$\le \left|\int_0^1\delta(x)dx - \Delta\sum\delta(x_k)\right| + \sup|\delta^*(x_k) - \delta(x_k)|$$

The first term is bounded by $\frac{\Delta^2}{24}\sup\delta''(x)$. By simply differentiating it can easily be seen that $\delta''(x)$ has a bound which only depends on the suprema of the first two derivatives of f. If we assume f and its first two derivatives to be bounded, we get

$$\left|\int_0^1\delta(x) - \Delta\sum\delta(x_k)\right| = O(\Delta^2)$$

Let us now look at the second term:

$$|\delta^* - \delta| \le |f_\lambda^{*2} - f_\lambda^2| + 2|f||f_\lambda - f^*_\lambda|$$

$$\le |f_\lambda - f^*_\lambda|\,[\,|f_\lambda + f^*_\lambda| + 2|f|\,]$$

$$\sup_x|\delta^*(x) - \delta(x)| \le \sup_x|f_\lambda(x) - f^*_\lambda(x)|\,\sup_x[\,|f_\lambda(x) + f^*_\lambda(x)| + 2|f(x)|\,]$$

The first factor is $O(\lambda^2\Delta^2)$, the second is bounded, so that we finally have

$$\left|\Delta\sum_i\Big[\lambda\Delta\sum_j K(j\lambda\Delta)f_{i-j} - f_i\Big]^2 - \int_0^1\Big[\int\lambda K(\lambda t)(f(x-t) - f(x))\,dt\Big]^2dx\right| = O(\lambda^2\Delta^2)$$

We further need the error when approximating

$$\sigma^2\Delta^2\lambda^2\sum K^2(i\lambda\Delta) \quad \text{by} \quad \lambda^2\sigma^2\Delta\int K^2(\lambda x)\,dx$$

Set $h_\lambda = K^2(\lambda x)$. With the same argument as before we have

$$\left|\int_{-\infty}^{\infty} h_\lambda(x) - \Delta\sum K^2(i\lambda\Delta)\right| \le \frac{2a}{\lambda\Delta}\frac{\Delta^3}{24}\sup|h''_\lambda(x)|$$

$$|h''_\lambda(x)| \le c\lambda^2 \Rightarrow \left|\int_{-\infty}^{\infty} h_\lambda(x) - \Delta\sum K^2(i\lambda\Delta)\right| \le \frac{2}{24}a\lambda\Delta^2c$$

So we have proved, that the total approximation error is $O(\lambda^2\Delta^2)$

We shall from now on work with IESE, but we have to keep in mind that, because of the approximation error, terms smaller than $O(\lambda^2\Delta^2)$ have no asymptotic meaning.

The first question we shall now deal with is: How do we have to choose λ as a function of n in order to obtain convergence of IESE to 0 at an optimal rate and under which conditions do we really obtain this rate. Questions of this type have been studied often before (see, for example, [9]).

Def

We call $\int K(t)t^m dt$ for m = 0.... the m-th moment of K.
We get the following

Theorem

Let K be a symmetric, two times continuously differentiable kernel with compact support and $\int K(t)dt = 1$. Assume the first l-1 moments of K vanish and the l-th moment exists. Let f be l times continuously differentiable with f and all l derivatives bounded and $f^{(l)}$ square-integrable. Then to obtain optimal convergence rate of IESE one has to choose

$$\lambda_n = \left[\frac{\int_0^1 f^{(l)^2}(x)\,dx\,[\int K(t)t^l]^2 \frac{1}{(l!)^2}}{\sigma^2 \int K^2(t)\,dt}\right]^{\frac{1}{2l+1}} n^{\frac{1}{2l+1}}$$

Then the IESE is

$$\mathrm{IESE} = 2\Big[\int_0^1 f^{(l)^2}(x)\,dx\,\big(\int K(t)t^l dt\big)^2 \frac{1}{(l!)^2}\Big]^{\frac{1}{2l+1}} \Big[\sigma^2 \int K^2(t)\,dt\Big]^{\frac{2l}{2l+1}} \cdot n^{-\frac{2l}{2l+1}} + o\big(n^{-\frac{2l}{2l+1}}\big)$$

Proof:

$$\mathrm{IESE} = \int_0^1 \Big[\int_{-\infty}^{\infty} \lambda_n K(\lambda_n t)\big(f(x-t)-f(x)\big)\,dt\Big]^2 dx + \sigma^2 \lambda_n \Delta \int K^2(t)\,dt$$

Let us first look at the bias term

$$b_\lambda = \int_0^1 \Big[\int_{-\infty}^{\infty} \lambda_n K(\lambda_n t)\big(f(x-t)-f(x)\big)\,dt\Big]^2 dx = \int_0^1 \Big[\int_{-\infty}^{\infty} K(t)\big(f(x-\tfrac{t}{\lambda})-f(x)\big)\,dt\Big]^2 dx$$

By simply expanding f in a Taylor series around x and observing, that the first l-1 moments of K vanish, one obtains

$$b_{\lambda_n} = \frac{1}{\lambda_n^{2l}(l!)^2} \int_0^1 \Big[\int_{-\infty}^{\infty} K(t)t^l f^{(l)}\big(x+\frac{\delta(t)\,t}{\lambda_n}\big)\,dt\Big]^2 dx \qquad \delta(t)\in[0,1] \quad \forall t$$

As $\lambda_n \longrightarrow \infty$

$$\int_0^1 \Big[\int_{-\infty}^{\infty} K(t)t^l f^{(l)}\big(x+\frac{\delta(t)t}{\lambda_n}\big)\,dt\Big]^2 dx \longrightarrow \int_0^1 f^{(l)^2}(x)\,dx\,\Big[\int K(t)t^l dt\Big]^2$$

This will be proved later.

So we have

$$\mathrm{IESE} \doteq \frac{1}{\lambda_n^{2l}(l!)^2} \int_0^1 f^{(l)^2}(x)\,dx\,\big[K(t)t^{(l)}dt\big]^2 + \sigma^2 \lambda_n \Delta \int K^2(x)\,dx + o(\lambda_n^{-2l})$$

Setting

$$c_1 = \int_0^1 f^{(l)^2}(x)\,dx\,\Big[\int K(t)t^l dt\Big]^2 \frac{1}{(l!)^2}$$

$$c_2 = \sigma^2 \int K^2(t)\,dt$$

and making the ansatz

$$\lambda_n = c\,n^{\alpha}$$

we obtain optimal convergence rate for the IESE if

$$\frac{c_1}{c^{2l}\,n^{2\alpha l}} = c_2\,c\,n^{\alpha-1}$$

This gives

$$\alpha = \frac{1}{2l+1}$$

$$c = \Big[\frac{c_1}{c_2}\Big]^{\frac{1}{2l+1}}$$

That means for the IESE

$$\mathrm{IESE} = 2c_1^{\frac{1}{2l+1}}\, c_2^{\frac{2l}{2l+1}}\, n^{-\frac{2l}{2l+1}} + o\big(n^{-\frac{2l}{2l+1}}\big)$$

That is what was stated.

It remains to show that

$$\int_0^1 \Big[\int_{-\infty}^{\infty} K(t)t^l f^{(l)}\big(x+\frac{\delta(t)t}{\lambda_n}\big)\,dt\Big]^2 dx \xrightarrow[\lambda_n\to\infty]{} \int_0^1 f^{(l)^2}(x)\,dx\,\Big[\int K(t)t^l dt\Big]^2$$

We have

$$\int_0^1 \Big[\int_{-\infty}^{\infty} K(t)\,t^l f^{(l)}\big(x+\frac{\delta(t)t}{\lambda_n}\big)\,dt\Big]^2 dx = \int_0^1 \Big[\underbrace{\int_{-\infty}^{\infty} K(t)t^l\Big(f^{(l)}\big(x+\frac{\delta(t)t}{\lambda_n}\big)-f^{(l)}(x)\Big)\,dt}_{\mathrm{II}}$$

$$+\underbrace{\int_{-\infty}^{\infty} K(t)t^l f^{(l)}(x)\,dt}_{\mathrm{I}}\Big]^2 dx = \int_0^1 \mathrm{I}^2 + 2\int_0^1 \mathrm{I}\,\mathrm{II} + \int_0^1 \mathrm{II}^2$$

$$\int \mathrm{I}^2\,dx = \Big[\int K(t)t^l dt\Big]^2 \int f^{(l)^2}(x)\,dx$$

II is Lebesgue-integrable. By Lebesgue's theorem it suffices to show that II $\longrightarrow$ 0 pointwise. Show:

$$\int K(t)\, t^{l} \left(f^{(l)}\left(x + \frac{\delta(t)\, t}{\lambda_n}\right) - f^{(l)}(x)\right) dt \xrightarrow[\lambda_n \to \infty]{} 0 \quad x\text{-a.e.}$$

For t fixed, the integral goes to 0 for $\lambda_n \to \infty$ ($f^{(l)}$ assumed to be continuous). The result is now obtained by applying Lebesgue's theorem once more.

Two comments on the theorem:

1) The approximation error is $\mathcal{O}(\lambda^2 \Delta^2)$. With $\lambda_n = c\, n^{\frac{1}{2l+1}}$ this gives $\mathcal{O}(n^{-\frac{4l}{2l+1}})$. IESE = $\mathcal{O}(n^{-\frac{2l}{2l+1}})$. So the approximation error is of smaller order of magnitude and we have AESE = $\mathcal{O}(n^{-\frac{2l}{2l+1}})$, too.

2) The theorem has no practical use, even if the sampling rate is large enough to make the asymptotics valid, for $\int f^{(l)^2}(x)\, dx$ as well as σ^2 are unknown.

Correcting the Bias

The smoothing procedure with bias correction may be described as follows:

1) Compute the smoothed ordinates

$$\hat{\underline{f}}(n,\lambda) = \underline{K}_\lambda * \underline{u}$$

2) Estimate the bias by smoothing $\hat{\underline{f}}(n, \lambda)$ once more:

$$\underline{f}^*(n,\lambda) = \underline{K}_\lambda * \hat{\underline{f}}(n,\lambda)$$

$$\hat{\underline{b}}(n,\lambda) = \hat{\underline{f}}(n,\lambda) - \underline{f}^*(n,\lambda)$$

3) Define the corrected ordinates $\hat{\underline{f}}^c(n, \lambda)$ by

$$\hat{\underline{f}}^c(n,\lambda) = \hat{\underline{f}}(n,\lambda) + \hat{\underline{b}}(n,\lambda) = (2\underline{K}_\lambda - \underline{K}_\lambda * \underline{K}_\lambda) * \underline{u}$$

Does this procedure really reduce the bias asymptotically ?
The answer lies in the following

Lemma:

If K has vanishing first $l-1$ moments and existing l-th to $2l$-th moment = 0, then K* = (2K-K*K) has vanishing $2l-1$ moments and existing $2l$-th moment.

Proof:

Let m_i denote the i-th moment of K and m_i^* the i-th moment of K*.
We have

$$m_{2l}^* = \int K^*(t)\, t^{2l}\, dt = 2m_{2l} - \iint K(t)\, K(x-t)\, x^{2l}\, dt\, dx$$

$$= 2m_{2l} - \iint K(t)\, K(x)\, (x+t)^{2l}\, dt\, dx$$

$$(x+t)^{2l} = \sum_{k=0}^{2l} \binom{2l}{k} x^k t^{2l-k}$$

Because of $m_1 \ldots m_{l-1} = 0$ and the symmetrie in (x,t) this yields

$$m_{2l}^* = 2m_{2l} - \binom{2l}{l} \int K(t)\, K(x)\, x^l t^l\, dt\, dx - 2m_{2l}$$

$$= -\binom{2l}{l} m_l^2$$

In the same way we get $m_1^* \ldots m_{2l-1}^* = 0$.

If K has moments m_i and we smooth with $K_\lambda(x) = \lambda K(\lambda x)$, the bias b_λ at x_0 approximately is

$$b_\lambda(x_0) = \sum_{1}^{\infty} \frac{1}{\lambda^k k!} f^{(k)}(x_0)\, m_k$$

So the lemma immediately yields, that the bias correction really eliminates the highest terms in λ of the bias.

Asymptotics of Spline Smoothing and Cross-Validation

The theory of spline asymptotics is somewhat involved. We will give here only the essentials and not bother about assumptions which are made only for technical reasons.

a) Optimal choice of the Lagrange parameter λ and the corresponding AESE (See [4]).

Let f, the unknown true function, be in $W_2^{(2)}$, the vector space of functions on [0,1] with absolutely continuous 3-rd and square-integrable 4-th derivative. Then the AESE is minimized by choosing

$$\lambda_n = c_1 \left[\frac{\sigma^2}{\int f^{(4)2}} \right]^{4/9} n^{-\frac{5}{9}}$$

For this choice one obtains for the AESE

$$AESE = c_2 \, [\sigma^2]^{8/9} \, [\int f^{(4)2}]^{1/9} \, n^{-8/9} + o(n^{-8/9})$$

There is a clear analogy to the formulas for kernel estimates. The dependence on $\int f^{(4)2}$ suggests, that the spline asymptotically coincides with a kernel with vanishing first 3 moments.

b) Cross-validation (see [10]).

In Chapter 2.2 it was pointed out, that for fixed λ the ordinates $y_i = s(x_i)$ of the smoothing spline depend linearly on the observations u_i:

$$\underline{y} = W(\lambda)\underline{u}$$

The cross-validation sum of squares for homoscedastic uncorrelated errors was defined as

$$CV(\lambda) = \sum_i \left(\frac{((I - W(\lambda))\underline{u})_i}{(I - W(\lambda))_{ii}} \right)^2$$

In [10], G. Wahba derives her asymptotic results for a procedure she calls "generalized cross-validation". The generalized cross-validation sum of squares is defined as

$$GCV(\lambda) = \frac{\| (I - W(\lambda))\underline{u} \|^2}{[\mathrm{trace}\,(I - W(\lambda))]^2}$$

If one looks at the formulas for $W(\lambda)$, one easily sees, that for equi-spaced data and n large, the diagonal elements are nearly equal and so

$$GCV(\lambda) \approx \frac{1}{n^2} CV(\lambda)$$

For the generalized cross-validation, the following asymptotic result can be stated:

Let f and its first 3 derivatives be absolutely continuous and $f^{(4)}$ square-integrable. Let $\tilde{\lambda}$ minimize the AESE of the smoothing spline and λ^* minimize $E(GCV(\lambda))$, the expected value of its generalized cross-validation sum of squares. Then:

If f is a polynomial of degree 1 or less, $\lambda^* = \tilde{\lambda} = \infty$

If f is not a polynomial of degree 1 or less, $\lambda^* = \tilde{\lambda}(1 + o(1))$.

3. THE REGRESSION APPROACH

3.1. Models for Human Height Growth

Using the word "model" in the title gives a somewhat optimistic picture of what seems possible in the analysis of the human growth process. Making a model for a dynamic system usually consists of the following essential steps:

1) Structure identification: Find a set of differential equations (which may contain unknown parameters) describing the mechanisms of the system.

2) Parameter identification: Find the solution of this set of differential equations, which best approximates the observed behaviour of the system.

Step 1 at the moment seems to be impossible, as the mechanisms steering the growth process are to a great extent unknown. All approaches tried so far omitted step 1: A parametric family of curves was chosen, not because it was thought to be the family of solutions for differential equations describing the growth mechanisms, but only because it was thought to fit well to the observed data.

Two curves play a central role in the analysis of growth:

1) The Gompertz curve $g(t) = e^{-e^{-t}}$

2) The logistic curve $l(t) = (1+e^{-t})^{-1}$

The work so far done is summarized in the following review:

1) Burt [11]

Age range : Conception to maturity

Proposed model

$$h(t) = \sum_{i=1}^{3} a_i \, l\left(\frac{t-b_i}{c_i}\right) \quad h : \text{standing height}$$

The first term in the sum accounts for the very fast growth during pregnancy, the second for growth during childhood, the third for the adolescent growth spurt.

Burt's model has never been used in full generality.

2) Deming [12]

Age range: Early childhood to maturity

Proposed model:

$$h(t) = \begin{cases} a_1 + b_1 t + c_1 \log t & \text{for early childhood} \\ a_2 \, g\left(\frac{t-b_2}{c_2}\right) & \text{for the adolescent growth spurt} \end{cases}$$

The question is, how and where these two components should be fitted together.

3) Bock et al [8]

Age range: 1 year to maturity

Proposed model:

$$h(t) = a_1 \, l\left(\frac{t-b_1}{c_1}\right) + (h_a - a_1)\, l\left(\frac{t-b_2}{c_2}\right) \quad h_a : \text{adult height}$$

There is a clear lack of fit : average squared residual is 2 cm^2 for boys and 0.9cm^2 for girls compared to a measurement error variance of about 0.2 cm^2.

The authors also make a step on the slippery path of interpretation: They call the first component "prepubertal" and the second "adolescent". However, there is the somewhat uncomfortable fact that they find a large sex difference in the "prepubertal" parameters.

The same authors in a later paper [13] compare 4 additive two component models (logistic + logistic, logistic + Gompertz....) on the basis of average squared residual and find double logistic best.

4) Marubini et al [14]

Age range: Adolescent growth spurt

Proposed models:

$$h(t) = p + a \, g\left(\frac{t-b}{c}\right)$$

$$h(t) = p + a \, l\left(\frac{t-b}{c}\right)$$

The authors compare these two models and find logistic best. There seems to be a lack of fit (average squared residual is 0.9 cm^2 in girls)

One has to decide where the adolescent growth spurt starts before running the fit algorithm.

For most of the individuals there are 11 or less measurements. The authors use run tests to judge the quality of the fit. Of course, a run test with 11 observations has no power at all and so will hardly detect model bias.

5) Marubini et al [15]
   Age range : Adolescent growth spurt
   Proposed models

$$h(t) = p + a\,l\left(\frac{t-b}{c}\right)$$

$$h(t) = p + a\,g\left(\frac{t-b}{c}\right)$$

The authors fit these two models to the London data and find logistic best. They also study the influence of the choice of "start of growth spurt" (take-off point), which turns out to be small.

On the London data the average residual standard deviation for the logistic model is only 4 mm in boys and 3.5 mm in girls, which is as good as can be expected. Good fit is also obtained for leg length, sitting height and biacromial diameter.

We are interested in a model for height growth, which covers about the same age range as that proposed by Bock et al in [8], [13]. Reading these papers, two questions arise:

1) Does the biological knowledge about the growth process justify a <u>two</u> component <u>additive</u> model ?
2) Searching for the best model, the authors restrict themselves to all possible combinations of logistic and Gompertz components. Is this restriction really necessary, or is it possible to reduce the bias and thus get a better fit by just allowing for larger classes of component functions ?

These two questions will be the subject of the rest of Chapter 3.

3.2. The Question of Additivity

Biological knowledge about the mechanisms of human growth is not sufficient for a structure identification of the growth process. Nevertheless, a model should take into account the qualitative knowledge available. The following facts can be found in [17], [18].

Human growth can be thought to consist of two components : Non-pubertal and pubertal. The non-pubertal component is that, which is observed in childhood before the occurence of puberty. Later, only the total effect of the two components is observable.

Effects caused by the pubertal component are the pubertal growth spurt, full development of the primary and secondary sex characteristics, accelerated bone maturation and epiphyse closure. The latter is especially important for models of human height growth, as after epiphyse closure no further growth is possible.

That the pubertal component via acceleration of bone maturation and epiphyse closure really "switches off" the non-pubertal component, is demonstrated by the effect of some disturbances of the hormonal system:

a) Eunuchoidism, prepubertal castration
   The pubertal component is missing, no signs of puberty are observed. Non-pubertal growth continues till long after the age of 20. Full ossification of the bones is usually never reached. Although the individuals lack of pubertal growth spurt, they grow on the average taller than normal individuals (eunuchoid gigantism).
b) Precocious puberty
   We use the term "precocious puberty", if some or all effects of puberty turn up at an earlier stage of development (measured in bone age) as usual. This may either be idiopatic or caused by diseases like gonadotropin-producing tumors, gonadal tumors or adrenogenital

syndrome. Cases where puberty starts at an age of less than 2 years have been reported. These individuals are taller than normal during their puberty, because of the pubertal growth spurt. They are smaller than normal as adults, because non-pubertal growth has been switched off at an early age and so has supplied only a reduced contribution to adult height.

With this in view, let us have a look at the model proposed by Bock et al in [8] and their results:

The model is

$$h(t) = a_1\, l(b_1(t-c_1)) + (h_a - a_1)\, l(b_2(t-c_2)) \qquad h_a: \text{adult height}$$

They neglect the switch-off effect. The following table gives the means and standard deviations for the parameters:

	a1 (mm)	b1 (1|years)	c1 (years)	a2 (mm)	b2 (1|years)	c2 (years)	
mean	1497	0.26	0.83	312	0.90	13.0	♂ (N=56)
std	71	0.02	0.35	41	0.17	0.79	
mean	1380	0.30	0.39	298	0.88	11.0	♀ (N=51)
std	71	0.03	0.36	62	0.19	0.87	

There is a large sex difference in the "prepubertal" parameters a_1, c_1. It is not caused by real differences between growth of boys and girls in childhood (these differences are known to be small), but is due to the fact, that puberty in girls occurs earlier than in boys and so non-pubertal growth is switched off at an earlier age. So, due to the neglection of the switch-off effect, in the "prepubertal" parameters properties of grwoth in childhood and the timing of puberty are confounded. This greatly limits the value of the parameterization.

A natural way to incorporate the switch-off effect in a model for height growth velocity is the following: Set

$$v(t) = a_1 s_1\left(\frac{t-b_1}{c_1}\right) \varphi\left(\frac{t-b_2}{c_2}\right) + a_2 s_2\left(\frac{t-b_2}{c_2}\right)$$

where φ is a function decreasing from 1 to 0, modelling the switch-off effect. It is coupled in location and scale to the second component modelling the growth spurt. s_1 will be a function which first decreases rapidly and then is approximately constant, s_2 will be bell-shaped.

In the following we will work with models of this type.

3.3. Shape-Invariant Models

All models for human height growth detailed in Chapter 3.1 with the exception of Deming's are shape-invariant models in the sense of the following definition:

Def:

Let $t_1(\underline{p}_1,t)\ldots t_k(\underline{p}_k,t)$ be time transformations depending on the parameters $\underline{p}_1\ldots\underline{p}_k$. Let $s_1\ldots s_k$ be given functions. Let $\circ$ denote addition or multiplication. Then the parametric family of curves

$$y(t, a_1\ldots a_k, p_1\ldots p_k) = a_1 s_1(t_1(p_1,t)) \circ \ldots \circ a_k s_k(t_k(p_k,t))$$

is said to be shape-invariant. The s_i are called shape-functions.

Instead of "shape-invariant family of curves" we also use the term "shape-invariant model" (abbreviated SIM).

For example, Burt's model is a 3 component additive SIM. s_1, s_2 and s_3 are logistic functions, $\underline{p}_i = (b_i,c_i)$, $t_i(\underline{p}_i,t) = b_i(t-c_i)$.

Shape-invariant models were first introduced by Lawton et al [19].

We shall restrict ourselves to linear time transformations as used in Burt's model. In principle this is not necessary.

3.4. Residual Alignment

a) The case of 1 component SIM

Think we have observations generated following the model

$$y_{ij} = a_i\, s\left(\frac{t_j - b_i}{c_i}\right) + \varepsilon_{ij} \qquad i = 1 \dots N,\; j = 1 \dots n$$

s as well as the parameters $\underline{p}_i = (a_i, b_i, c_i)$ are unknown. We start with a shape-function s' selected after visual inspection of the data.

$$s'(x) = s(x) + \Delta(x)$$

If we knew the $\underline{p}_i$, there would be a straightforward way to estimate Δ (x):

Def: Set

$$x_{ij} = \frac{t_j - b_i}{c_i}$$

$$r_{ij} = y_{ij} - a_i\, s'(x_{ij})$$

The pairs (x_{ij}, r_{ij}) are called "aligned residuals".

With this notation we have

$$\frac{r_{ij}}{a_i} = -\Delta(x_{ij}) + \frac{\varepsilon_{ij}}{a_i}$$

So Δ (x) may be estimated by applying one of the standard curve estimation procedures to the aligned residuals.

In practice we do not have the true $\underline{p}_i$, but instead use the estimated $\hat{\underline{p}}_i$. This process (for given s' estimate the $\underline{p}_i$ by least squares, then estimate a correction Δ (x) for s' by applying a curve estimation procedure to the aligned residuals) may be repeated if necessary. Such a procedure has first been suggested in [19]. There is a clear analogy to our population bias correction for splines (see [2]).

b) 2 component additive SIM

Here we have

$$y_{ij} = a_{1i}\, s_1(x^1_{ij}) + a_{2i}\, s_2(x^2_{ij}) + \varepsilon_{ij}$$

$$x^1_{ij} = \frac{t_j - b_{1i}}{c_{1i}} \qquad x^2_{ij} = \frac{t_j - b_{2i}}{c_{2i}}$$

There are two sets of aligned residuals: Aligned according to the first or according to the second time scale. We also want to estimate two corrections Δ_1, Δ_2. Again assuming for a moment that we know the true parameters, we get

$$r_{ij} = -(a_{1i}\,\Delta_1(x^1_{ij}) + a_{2i}\,\Delta_2(x^2_{ij})) + \varepsilon_{ij}$$

The simple way of estimating the bias functions Δ_1, Δ_2 for example by smoothing is not applicable here.

3.5. Finding the Best Shape-Invariant Model

As a criterion for the goodness of fit of a SIM we use the overall residual sum of squares

$$S = \sum_{i,j} r_{ij}^2$$

Let us continue with the 2 component additive case and again assume for the moment the true parameters $\underline{p}_i = (a_{1i}, b_{1i}, c_{1i}, a_{2i}, b_{2i}, c_{2i})$ to be known.

We want to find the best Δ_1, Δ_2. This of course makes no sense.

We have to specify a class functions and find the best Δ_1, Δ_2 out of this class. We can for example find the best Δ_1, Δ_2 out of the class of polynomials of degree at most k, or out of the class of spline functions of a given degree with pre-specified knots. It makes life simple, if the classes are vector spaces.

Let Z_1, Z_2 be vector spaces and let $z_{11} \dots z_{1k}$ be a basis for Z_1, $z_{21} \dots z_{2m}$ be a basis for Z_2. Now finding the best SIM means solving the least squares problem

$$\sum_{i,j} \left(r_{ij} - a_{1i} \sum_{s=1}^{k} d_{1s}\, z_{1s}(x^1_{ij}) - a_{2i} \sum_{t=1}^{m} d_{2t}\, z_{2t}(x^2_{ij}) \right)^2 = \min! \qquad (*)$$

The fact that Z_1, Z_2 are vector spaces is essential for making the problem an unrestricted least squares problem. We also clearly see the advantage of additive SIMs: In the case of multiplicative effects in the above equation terms with $d_{1i}d_{2j}$ turn up and the least squares problem is no longer linear.

The minimum in (*) is unique, although the solution parameters may be not. If the solution parameters are ill determined, this means, that we can not decide whether a bias is attributable to an error in s_1' or to an error in s_2'.

In practice the true parameters p_i are unknown. We then have to proceed iteratively:

1) Start with some s_1^0, s_2^0 and starting parameters p_i^0, $i = 1...N$.

2) For fixed s_1^k, s_2^k correct the nonlinear parameters by performing some steps of a nonlinear least squares algorithm.

$$p_i^k \longrightarrow p_i^{k+1} \quad i = 1 \ldots N$$

(Nonlinear step)

3) For p_i^{k+1}, $i=1...N$ solve the linear least squares problem (*) and set

$$s_1^{k+1} = s_1^k + \Delta_1 , \quad s_2^{k+1} = s_2^k + \Delta_2$$

(Linear step)

Steps 2 and 3 are repeated.

3.6. Some Comments on Our Model

In Chapter 3.2. we suggested the model

$$v_{ij} = a_{1i}\, s_1\left(\frac{t_j - b_{1i}}{c_{1i}}\right) \varphi\left(\frac{t_j - b_{2i}}{c_{2i}}\right) + a_{2i}\, s_2\left(\frac{t_j - b_{2i}}{c_{2i}}\right)$$

where φ is the switch-off function. It seems to be totally hopeless to estimate a correction for φ with the data available to us. So φ is not corrected, but held fixed. In this case we can use the correction procedure for a 2 component additive SIM. As starting functions we use

$$s_1^0(x) = e^{-x} + 1$$
$$s_2^0(x) = e^{-x^2}$$

$$\varphi(x) = 1 - \frac{\int_{-\infty}^{x} s_2^0(t)\,dt}{\int_{-\infty}^{\infty} s_2^0(t)\,dt}$$

As vector spaces Z_1, Z_2 we use spaces of cubic splines. To be precise : for each component we may specify

1) The location FKNOT of the first knot of the spline

2) The distance DELTA between the knots. The knots have to be equi-spaced.

3) The number NKNOT of knots

All what is said in the following about the first component applies as well to the second:

Let $x_1 \ldots x_k$ be the prespecified knots. Z_1 is the vector space of cubic splines with support $[x_1, x_k]$ and knots $x_1, \ldots, x_k$. This assures, that the sum of the starting shape function s_1^0 and the correction is a smooth curve. Z_1 has dimension k-4. As a basis for Z_1 we use the k-4 B-splines B_i, $i=1...k-4$. The B-splines are the minimal-support splines of degree 3. B_i has support $[x_i, x_{i+4}]$. It is defined as

$$B_i(x) = \sum_{j=i}^{i+4} \frac{4\,(x - x_j)_+^3}{\prod_{k \neq j} (x_j - x_k)}$$

$$(x - x_j)_+ = \begin{cases} 0 & x < x_j \\ x - x_j & x \geq x_j \end{cases}$$

The restriction, that the knots for each component have to be equi-spaced, has been introduced for convenience. It is not a severe one, however, for the following reason: The first component essentially accounts for the growth in childhood, where the measurements have been taken in one-year intervals. The second component accounts for the growth spurt, where the measurements occured in half-year intervals.

3.7. The Algorithm

1) Nonlinear Step

We use an algorithm given by Nagel & Wolf in [21]. It is a Marquardt derivate, very simple, and worked well in tests. The convergence could certainly be improved by splitting off the parameters a_1, a_2 and then minimizing the variable projection functional (see [22]). This was not done, because no program was available and there was no time left to write one.

2) Linear Step

We have to deal with rather large linear least squares problems: 10 knots/component, 30 individuals, 30 measurements/individual make a least squares problem with 12 variables and 900 observations. To solve it, we use sequential Householder transformations (see [23]). This procedure has the following advantages:

a) It is numerically sound

b) Only a small part of the design matrix has to be held in core at a time (at least NVAR+2 rows, where NVAR is the number of unknown variables).

c) The design matrix has to be read only once.

d) The number of operations increases only slightly compared to non-sequential Householder transformations.

3.8. The Fitting Program

The most important features of the program are given below:

Computer : DEC-10 + Tektronix 4014 terminal
Language : Fortran-10
Mode : Interactive
Length : 4000 lines without comments
Development Time : 2.5 man-months
Computer costs for development : 70 h hook-up time, 1.5 h run time ≈ 3000 Fr.
Utility software used : Tektronix Terminal Control System
Structure : Tree. The root corresponds to the main program, which is only a calling program. 10 branches originating from the root correspond to 10 subroutines doing the work (description see below). Communication between different branches is done via labeled COMMONs.

Dimensions : Up to 30 individuals can be handled at a time. (This restriction would be easy to remove)

Interactivity is absolutely necessary to prevent waste of computer- and user-time and nonsense results.

The tree structure of the program and the communication between subroutines via COMMONs is very convenient for program development, testing and modification. However it is less convenient for use: To come from a node (activity) in one main branch to a node in a different branch, one has first to climb up the tree to the root and then descend again. So it may be necessary to enter 3 or 4 commands for this purpose.

The names of the 10 main subroutines and the tasks performed by them are:

SELCHI:

Select a random sample of individuals from the data file. To offer the possibility of drawing disjoint samples, this is done via random permutations: If there are N boys and N' girls in the data file, a random permutation of $\{1\ldots N\}$ and a random permutation of $\{1\ldots N'\}$ are generated. Let $\{p_1\ldots p_N\}$ and $\{p'_1\ldots p'_{N'}\}$ denote these two permutations. The user now specifies starting indices i, i' and sample sizes n,n'. The data file is read sequentially and the number of boys resp. girls already found is counted in c resp c'. If the individual last read is a boy, it is checked whether $c \in \{p_i\ldots p_{i+n}\}$. If yes, the boy is selected for the sample, if no, he is discarded. The same applies to the girls.

START :

Find starting values for the nonlinear least squares. The subroutine can find starting values for two models:

model 1 (without switch-off)

$$y(t) = a_1 s_1\left(\frac{t}{c_1}\right) + a_2 s_2\left(\frac{t-b_2}{c_2}\right)$$

$$s_1(x) = e^{-x}$$

$$s_2(x) = e^{-x^2}$$

So there is no location parameter for the first component. It would not be identifiable. However, after a linear step the shape-functions have changed and this is no longer true. The location parameter may be estimated and 0 is taken as a starting value.

model 2 (with switch-off) see Chapter 3.6.

NLSTEP:

Perform nonlinear step (see Chapter 3.7)

COMRES:

Compute residuals (see Chapter 3.9)

LSTEP:

Perform linear step (see Chapter 3.7)

MAPLO:

MAPLO is a calling routine for plotting. It calls the plot routines

PLOCHI	plot individual curves and measurements
PLORES	plot aligned residuals (one-dimensional)
PLOTTA	plot aligned residuals (two-dimensional)
PLSHA	plot corrections of shape-functions.

SAVE:

Save current status of the program in a file with user-defined name.

REVIVE:

Read file with program status saved before calling SAVE.

CHAMOD:

Select Model. Code for model is a two-digit integer. The rightmost digit stands for switch-off effect. 0:no, 1:yes. The second digit indicates, whether the correlation structure of the residuals shall be taken into account for nonlinear and linear least squares. 0:no, 1:yes.

SELECT:

Children in the sample can be set active or passive. For computations using data from all individuals in the sample (for example linear step, residual plots) the non-active children are omitted.

3.9. Residual Plots

In relation with the residual plots a small difficulty arises: As our velocity measurements are really divided differences of measured heights, half-year velocities are expected to have twice the random error as whole-year velocities. This would make comparision of the residuals difficult. So in years with two measurements the two residuals are averaged. Furtheron, the residuals of the four measurements in the first year of life are averaged.

1) The one-dimensional aligned residuals plot (see photo 8-13 in Appendix 1)

   The plot consists of two parts: At the top residuals aligned according to the first time scale, at the bottom aligned according to the second time scale.

   Scales on the age-axis correspond to the average child. Let, for example, be $\bar{b}_1$, $\bar{c}_1$ be the averages of b_{1i}, c_{1i} over the sample. Then the mark '5' on the age-axis corresponds to age 5 of a child with parameters $\bar{b}_1$, $\bar{c}_1$. The three lines (the solid one in the middle of the two dashed ones) are running (median $\pm$ standard deviation of the median).

2) The two-dimensional aligned residuals plot (see photo 20)

   This plot is created in the following way:

   a) Define new, rescaled shape functions

   $$s_1'(x^1) = s_1\left(\frac{x^1-\beta_1}{\gamma_1}\right)$$

   $$s_2'(x^2) = s_2\left(\frac{x^2-\beta_2}{\gamma_2}\right)$$

   Let $b_{1i}, c_{1i}, b_{2i}, c_{2i}$ resp. $b_{1i}', c_{1i}', b_{2i}', c_{2i}'$ be the scale parameters for individual i before resp. after rescaling the shape functions. We choose $\beta_1, \gamma_1, \beta_2, \gamma_2$ such that

   $$\overline{b_1'} = \overline{b_2'} = \sigma$$

   $$\overline{c_1'} = \overline{c_2'} = 1$$

   This yields

   $$\beta_1 = \bar{b}_1, \quad \gamma_1 = \bar{c}_1$$

   $$b_{1i}' = \frac{\bar{c}_1 b_{1i} - \bar{b}_1 c_{1i}}{\bar{c}_1} \qquad c_{1i}' = \frac{c_{1i}}{\bar{c}_1}$$

   and analog for the second component.

   b) We have

   $$x_i^1(t_j) = x_{ij}^1 = \frac{t_j - b_1'}{c_1'} \qquad x_i^2(t_j) = x_{ij}^2 = \frac{t_j - b_2'}{c_2'}$$

   The points (x_{ij}^1, x_{ij}^2) lie on a straight line in the x^1-x^2 plane. The line is given by

   $$x^2 = \frac{c_{1i}'}{c_{2i}'} x^1 + \frac{b_{1i}' - b_{2i}'}{c_{2i}'}$$

   So the line for the average child ($b_1' = b_2' = 0$, $c_1' = c_2' = 1$) is just the main diagonal. On the line for each child, ages 0 and 13 are marked by long dashes, ages 11, 12, 14, 15 are marked by short dashes.

   c) Positive residuals correspond to $\vartriangle$ triangles, negative ones to $\triangledown$ triangles. The sidelength of a triangle indicates the absolute value of a residual. Residuals with absolute values larger than three times the median absolute value are indicated by $\blacktriangle$ resp. $\blacktriangledown$.

An example for a two-dimensional aligned residuals plot can be found on photo 20 (Appendix 1). The origin of the lower left coordinate cross is $(x^1,x^2) = (0,0)$, the origin of the second cross is $(x^1,x^2) = (\bar{b}_2,\bar{b}_2)$. One unit on the x^1 (x^2) -axis corresponds to one year for a child with parameter $c_1 = 1$ ($c_2 = 1$).

Interpretation of the Plot:

Let us have a look at the line for child 83.

a) The line crosses the x^1-axis of the upper coordinate cross at an age of about 13.5 years. As

   $$\frac{t - b_2'}{c_2'} = \bar{b}_2 \implies t = b_2$$

   this means $b_2 = 13.5$ years.

b) The length of the projection of the line onto the x^1-axis is 35 units. This means $c_1 = 20/35 \approx 0.6$.

c) The length of the projection of the line onto the x^2-axis is 17 units, which yields $c_2 = 20/17 \approx 1.2$.

The parameter b_1 is of small importance (see Chapter 3.11).
So the two-dimensional aligned residuals plot for a child contains all the information about the scale parameters and the quality of the fit. Furtheron it is possible to mark the appearance of sex characteristics or other development indicators on the line. By drawing several lines on one plot, one can produce a condensed picture of the development of a group of children.

3.10. A Comment on Regression with Non-Diagonal Error Covariance Matrix

Our model for the observed height growth velocities is (see Chapter 3.5):

$$v_{ij} = f(x_i, \underline{p}_i, \underline{d}_1, \underline{d}_2) + \varepsilon_{ij} \qquad i = 1 \dots N, \; j = 1 \dots n$$

where

$$f(x, \underline{p}, \underline{d}_1, \underline{d}_2) = a_1 s_1\left(\frac{x-b_1}{c_1}\right) \varphi\left(\frac{x-b_2}{c_2}\right) + a_2 s_2\left(\frac{x-b_2}{c_2}\right)$$

$$\underline{p} = (a_1, b_1, c_1, a_2, b_2, c_2)$$

$$s_1 = s_1^0 + \sum_{s=1}^{K} d_{1s} z_{1s}$$

$$s_2 = s_2^0 + \sum_{t=1}^{m} d_{2t} z_{2t}$$

Using the vector notation

$$\underline{v} = (v_{11} \dots v_{1n} \dots\dots v_{N1} \dots v_{Nn})^T$$

$$\underline{\varepsilon} = (\varepsilon_{11} \dots \varepsilon_{1n} \dots\dots \varepsilon_{N1} \dots \varepsilon_{Nn})^T$$

$$\underline{g}(\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2) = (f(x_1, \underline{p}_1, \underline{d}_1, \underline{d}_2) \dots f(x_n, \underline{p}_1, \underline{d}_1, \underline{d}_2) \dots\dots f(x_1, \underline{p}_N, \underline{d}_1, \underline{d}_2) \dots f(x_n, \underline{p}_N, \underline{d}_1, \underline{d}_2))^T$$

the model can be written as

$$\underline{v} = \underline{g}(\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2) + \underline{\varepsilon}$$

Due to the definition of the "observed" velocities as divided differences we have, under the hypotheses of homoscedastic, uncorrelated errors in

the height measurements

$$\mathrm{cov}(\underline{\varepsilon}) = \sigma^2 \underbrace{\begin{bmatrix} \Sigma & & \\ & \Sigma & \\ & & \Sigma \end{bmatrix}_{nN}}_{nN} = \tilde{\Sigma} \sigma^2 \qquad \Sigma = D \varphi D$$

$$D = \mathrm{diag}\left(\frac{1}{x_2 - x_1} \dots \frac{1}{x_{n+1} - x_n}\right)$$

$$\varphi = \begin{bmatrix} 1 & -\frac{1}{2} & 0 & 0 & \dots \\ -\frac{1}{2} & 1 & -\frac{1}{2} & 0 & \dots \\ 0 & -\frac{1}{2} & 1 & -\frac{1}{2} & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{bmatrix}$$

and $\sigma^2 = 2$ times the height measurement error variance.

Gauss-Markov theorem suggests to estimate the unknown parameters $\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2$ by minimizing

$$RSS_w = (\underline{v} - \underline{g}(\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2))^T \tilde{\Sigma}^{-1} (\underline{v} - \underline{g}(\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2))$$

A computationally simpler estimate is obtained by forgetting the negative correlation in the residuals and minimizing

$$RSS_n = (\underline{v} - \underline{g}(\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2))^T \tilde{D}^{-2} (\underline{v} - \underline{g}(\underline{p}_1 \dots \underline{p}_N, \underline{d}_1, \underline{d}_2))$$

$$\tilde{D} = \begin{bmatrix} D & & & \\ & D & & \\ & & D & \\ & & & \ddots \; D \end{bmatrix}_{nN}$$

The Gauss-Markov estimate should give a better fit and smaller variance of the estimated parameters compared to the naive estimate neglecting the negative correlations. In practice, however, the Gauss-Markov estimate showed some strange properties. In particular, the estimate for σ^2 based on RSS_w was in some cases twice as high as the estimate based on RSS_n. The reason apparently is, that the Gauss-Markov estimate with this error covariance matrix is very sensitive to low-frequency deviations of the true individual growth velocity curves from the stated model. The following simple example delivers the basis for this conjecture:

Think we have measurements $x_1 \dots x_n$ at times $t_1 \dots t_n$ and we set up the model

$$x_i = \mu + \varepsilon_i \qquad \underline{\varepsilon} \sim N(\underline{0}, \sigma^2 \varphi)$$

We want to estimate μ and σ^2.

φ has the eigenvalues (see [24])

$$\lambda_k = 1 - \cos\left(\frac{k\pi}{n+1}\right) \qquad k=1\dots n$$

The corresponding eigenvectors are

$$\underline{e}_k = \left(\sin\left(\frac{k\pi}{n+1}\right), \sin\left(\frac{2k\pi}{n+1}\right) \dots\dots \sin\left(\frac{nk\pi}{n+1}\right)\right)^T$$

We will later need the second eigenvector $\underline{e}_2$. We call it $\underline{b}$ and its components $b_1 \dots b_n$.

Applying the Gauss-Markov theorem, one obtains by straight-forward calculations:

$$\hat{\mu} = \frac{\sum_i (\varphi^{-1}\underline{x})_i}{\sum_{i,j} \varphi^{-1}_{ij}}$$

is the minimum variance linear unbiased estimate for μ.

$$\hat{\sigma}_w^2 = \frac{1}{n-1} (\underline{x}-\hat{\underline{\mu}})^T \varphi^{-1} (\underline{x}-\hat{\underline{\mu}}) \qquad \hat{\underline{\mu}} = (\hat{\mu} \dots \hat{\mu})^T$$

is an unbiased estimate of σ^2.

The naive variance estimate

$$\hat{\sigma}_n^2 = \frac{1}{n} \sum_i (x_i-\hat{\mu})^2$$

is biased, but the bias is only O(1/n).

Now, think the assumed model is not true, and the data have been generated following the model

$$x_i = \mu + \alpha\sigma b_i + \varepsilon_i$$

α is called amplitude factor of the bias.

What do we get in this case using the Gauss-Markov estimate for our assumed model ? The bias $\underline{b}$ does not influence $\hat{\mu}$, as

$$\sum_i (\varphi^{-1}\underline{b})_i = \frac{1}{\lambda_2}\sum_i b_i = 0$$

for symmetry reasons. Let us now look at E($\hat{\sigma}_w^2$)

$$E((\underline{x} - \hat{\underline{\mu}})^T\varphi^{-1}(\underline{x}-\hat{\underline{\mu}})) = (n-1)\sigma^2 + \sigma^2\alpha^2 \underline{b}^T \varphi^{-1}\underline{b}$$

This gives

$$E(\hat{\sigma}_w^2) = \sigma^2 + \frac{1}{(n-1)\lambda_2}\alpha^2\sigma^2\|b\|^2 = \sigma^2\left(1+\frac{1}{(n-1)\lambda_2}\alpha^2\|b\|^2\right)$$

We need an approximation for

$$\|b\|^2 = \sum_{k=1}^{n} \sin^2\left(\frac{2k\pi}{n+1}\right)$$

Application of the rectangle rule yields

$$\sum_{k=1}^{n} \sin^2\left(\frac{2k\pi}{n+1}\right) = \frac{n+1}{2} + O\left(\frac{1}{n}\right)$$

where $O(\frac{1}{n}) = c_n \frac{1}{n} \quad |c_n| \le 20 \ \forall n$

So:

$$E(\hat{\sigma}_w^2) = \sigma^2\left[1 + \frac{1}{(n-1)\lambda_2}\alpha^2\left(\frac{n+1}{2} + O\left(\frac{1}{n}\right)\right)\right]$$

$$= \sigma^2\left[1 + \frac{\alpha^2}{2\lambda_2}\left(\frac{n+1}{n-1} + O\left(\frac{1}{n^2}\right)\right)\right]$$

where $O(\frac{1}{n^2}) = c_n\frac{1}{n(n-1)} \qquad |c_n| \le 20 \ \forall n$

For n=100 we have $\lambda_2 \approx \frac{2\pi}{n^2} \approx 0.002$

That means, that an amplitude factor of only 0.06 (which makes the bias practically undetectable), is still enough to overestimate the variance by a factor 2. The naive estimate is not significantly influenced by this small bias.

3.11. Results

As we needed measurements from birth to 20 years, we had to use recumbent length in early childhood, standing height later. Children were excluded from the analysis, if the measurement series between birth and 3 years was incomplete, or if later totally more than 3 or two consecutive observations were missing. This left us 120 usable measurement series. To simplify computations, missing observations were completed using the procedure described in Chapter 2.3.

Model (See Chapter 3.6) :

$$v(t) = a_1 s_1\left(\frac{t-b_1}{c_1}\right) \varphi\left(\frac{t-b_2}{c_2}\right) + a_2 s_2\left(\frac{t-b_2}{c_2}\right)$$

Specification of knots for the linear step (in absolute scale) :

	FKNOT	DELTA	NKNOT
component 1	-4.33	2.68	7
component 2	-4.72	1.07	11

For the average child this corresponds to

	FKNOT	DELTA	
component 1	-3	3	(years)
component 2	7	1.5	

3 disjoint random samples, each with 15 boys and 15 girls, were drawn out of the 120 children available. The fitting procedure specified in the following was applied to each of the 3 samples independently:

1) Determination of starting values for the nonlinear parameters
2) Nonlinear step (5 iterations of the NLLS-algorithm)
3) Linear step
4) Nonlinear step (5 iterations of the NLLS-algorithm)
5) Linear step

Between the steps, curves were visually inspected on the terminal screen, mainly to be sure that the NLLS-algorithm had not converged towards a dègenerate solution. If such a case was detected, starting values were entered manually and the NLLS-algorithm was run again. Performing steps 1...5 for 30 children takes about 10 minutes CPU-time on a DEC-10.

3 children, one out of each sample, had to be discarded, because their raw velocity series were totally wild. Photo 7 (Appendix 1) illustrates, what that means.

Photos 8...13 show the one-dimensional aligned residuals plots, first after step 2, then after step 5. Qualitatively there is no difference between the three samples. Let us first have a look at the aligned residuals after step 2:

a) Alignment according to the first time scale and alignment according to the second time scale yield totally different patterns. This indicates, that it is well determined, which part of the bias is attributable to an error in s_1^0 and which part is attributable to an error in s_2^0.

b) The shape function s_1^0 is too low between 2 and 6 years and too high between 11 and 14 years.

c) There is a dip before the spurt starts (s_2^0 is too high between 10 and 12 years). Furtheron, s_2^0 descends too rapidly at the end of the spurt.

Remember that the scales on the age-axis correspond to the average child.

After step 5, the residuals show no pattern at all. They are nicely distributed in narrow bands around 0.

As a quantitative measure to judge the goodness of fit we use ASR0 and ASR1: ASR0 is the average squared residual based on all observations between birth and 20 years, ASR1 is based only on the observations between 1 and 20 years. ASR0 is much larger than ASR1. This has two reasons:

1) The recumbent length measurement error in the first year of life is larger than the standing height measurement error in later ages. (Think of the technical problems when measuring a two year old baby).

2) The bias in the shape-function for the first year of life is not corrected, because the distance between the knots of the spline is too large. This is a limitation of our program, which only allows for equi-spaced knots in each component.

Table 3 gives ASR0 and ASR1 for the 3 samples after the different steps of the fit procedure:

		Sample 1	Sample 2	Sample 3
after	ASR0	30.5	28.3	37.1
step 2	ASR1	23.7	21.5	27.0
after	ASR0	27.1	25.7	33.9
step 3	ASR1	19.9	18.4	23.4
after	ASR0	26.1	24.5	32.1
step 4	ASR1	19.7	18.2	22.6
after	ASR0	25.8	24.1	31.8
step 5	ASR1	19.2	17.6	22.1

TABLE - 3

For the interpretation of ASR0, ASR1 it is important to know, that the observations were weighted according to their theoretical variances: If the variance of the height measurement error is σ^2, then the raw velocity $v_i((t_{j+1}+t_j)/2)$ has error variance $2\sigma^2/(t_{j+1}-t_j)^2$ and was therefore weighted with $(t_{j+1}-t_j)/2$. So, if the velocity curves were estimated bias-free, we would get $ASR \approx \sigma^2$. σ is known to be close to 4 mm. The values of ASR1 in table 3 indicate, that the fit between 1 year and maturity is really good. This is confirmed by visual inspection of the fitted curves. As illustration, photos 14...18 (Appendix 1) show the velocity measurements and the fitted curves for the first 5 children of the first sample. The curves on each picture are

1) The non-pubertal component $a_1 \varphi_1(\frac{t-b_1}{c_1})\Phi(\frac{t-b_2}{c_2})$ (dotted)
2) The pubertal component $a_2 \varphi_2(\frac{t-b_2}{c_2})$ (dashed)
3) The sum of the two components

Photo 19 shows the estimated corrections for the shape-functions: At the top $s_1^2-s_1^0$, at the bottom $s_2^2 - s_2^0$, both for the 3 samples. The scales on the age-axis correspond to the average child. The pattern is the same for the 3 samples. The small quantitative differences may indicate slight overfitting.

For further use, the shape-functions estimated for the 3 samples were averaged. Tables 4 and 5 gives s_1^0 and s_1^2 resp. s_2^0 and s_2^2.

x^1	$s_1^0(x^1)$	$s_1^2(x^1)$
-2.0	8.39	8.40
-1.0	3.72	3.74
0.0	2.00	2.05
1.0	1.37	1.44
2.0	1.14	1.23
3.0	1.05	1.15
4.0	1.02	1.10
5.0	1.01	1.04
6.0	1.00	1.00
7.0	1.00	0.98
8.0	1.00	0.98
9.0	1.00	0.99
10.0	1.00	1.00

<u>TABLE - 4</u>

x^2	$s_2^0(x^2)$	$s_2^2(x^2)$
-4.0	0.00	-0.01
-3.5	0.00	-0.06
-3.0	0.00	-0.13
-2.5	0.00	-0.20
-2.0	0.02	-0.21
-1.5	0.11	-0.07
-1.0	0.37	0.33
-0.5	0.78	0.84
0.0	1.00	1.02
0.5	0.78	0.72
1.0	0.37	0.36
1.5	0.11	0.22
2.0	0.02	0.18
2.5	0.00	0.11
3.0	0.00	0.06
3.5	0.00	0.05
4.0	0.00	0.04
4.5	0.00	0.02
5.0	0.00	0.01

<u>TABLE - 5</u>

A Look at the Estimated Parameters

Table 6 gives means and standard errors (standard deviation of the mean) for the estimated parameters:

		a_1	b_1	c_1	a_2	b_2	c_2
boys	mean	58.2	1.67	0.96	56.1	14.5	1.39
(N=45	stderr	0.79	0.06	0.04	1.47	0.14	0.03
girls	mean	57.8	1.69	1.03	38.1	13.1	1.49
(N=42)	stderr	0.95	0.08	0.05	1.80	0.13	0.05

TABLE - 6

There is a perfect coincidence between boys and girls in the non-pubertal parameters a_1, b_1, c_1. This supports the assertion of Chapter 3.2, that the sex differences in the prepubertal parameters found by Bock et al [8] are due to improper choice of their model (neglection of the switch-off effect).

The sex difference in the timing of the spurt (b_2) is slightly smaller than that found in [2] (1.4 years instead of 1.7 years). Moreover, the sample average of b_2 is considerably larger than the average APH. The difference is 0.6 years in boys and 0.9 years in girls. This is due to the different definition of the two parameters: The velocity curve is the sum of the non-pubertal component, which decreases during the spurt due to the switch-off effect, and the pubertal component. b_2 is the age, where the second component reaches its maximum, whereas APH is the age, where the sum of the two components reaches its maximum. So it is clear, that APH will always be smaller than b_2.

Table 7 gives the correlation between the parameters for boys. For girls, the pattern is the same. HA is the adult height, PAR is defined as $a_2 * c_2$ and measures the intensity of the spurt.

	a_1	b_1	c_1	a_2	b_2	c_2	HA	PAR
a_1	1.00	-0.70	-0.61	0.03	-0.48	-0.02	0.39	0.00
b_1		1.00	0.95	-0.18	0.19	0.03	0.19	-0.14
c_1			1.00	-0.22	0.12	0.01	0.25	-0.18
a_2				1.00	-0.27	-0.34	-0.03	0.63
b_2					1.00	-0.11	0.03	-0.35
c_2						1.00	-0.02	0.51
HA							1.00	-0.06

TABLE - 7

Comments

a) The most surprising result obtained already by the spline approach (See [2]) is confirmed: Adult height is independent of the duration of growth (b_2) and of the duration, height and intensity (c_2,a_2,PAR) of the spurt.

b) The correlation between b_1 and c_1 is extremely high in both sexes. In good approximation we can set

$$b_{1i} = \alpha c_{1i} + \beta \qquad \alpha = 1.4 \qquad \beta = 0.3$$

The fitted curve for child i is then

$$v_i(t) = a_{1i}\, s_1\left(\frac{t-\beta}{c_{1i}} - \alpha\right) \varphi \ldots\ldots$$

When seen from the biological point of view, this could be interpreted as follows: For small ages, $x^1(t)$ depends only weakly on the individual (see, for example, the children on the two-dimensional aligned residuals plot shown on photo 20). For $t = \beta$ there is no such dependence at all. Or, the other way: x^1 measures the state of development of the non-pubertal component. The development starts at $x^1 = -\alpha$, which corresponds to the same age for all individuals. Whether this interpretation is supported by biological knowledge has not yet been checked. In the case, it would be difficult to assign a meaning to x^1, if $x^1(0)$ for one child would coincide with $x^1(-5)$ for an other child.

4. CONCLUSIONS

What are the advantages and the shortcomings of the two methods proposed in Chapter 2 resp. 3 ? Let us first have a look at the curve estimation approach:

1) It is methodologically simple.

2) The program length is only 20% and the computer costs (measured in CPU-time/individual) are only 10% compared to SIMs.

3) It is objective in the sense, that only a simple qualitative assumption is necessary (the true underlying curve has to be smooth) and that only one parameter (S resp. λ) has to be chosen. Furtheron, it is clearly visible, how the choice of this parameter influences the results.

4) As the parameters do not depend on model assumptions, they can be used to judge, whether a proposed model will possibly give a good fit or not.

The advantages of the regression approach are:

1) It offers the possibility of really finding the curve for the "average individual". Using smoothing splines, this was only approximately possible: Alignment could be done according to the location of the growth spurt (APH), (see the definition of "individual-type standards" in [2]), whereas, using SIMs, alignment can be done according to location and scale of the peak as well as location and scale of the first component. The so obtained curve shows more details: When looking at the spline-smoothed individual curves, we had conjectured that there was systematically a dip in the curve before the adolescent spurt started. However, this dip did not show up in the individual-type mean curve, clearly because no alignment according to duration of spurt was possible. When, using SIMs, alignment was done according to location and scale, the dip became clearly visible.

2) When choosing the type of the model, qualitative knowledge about the structure of the growth process was used (See Chapter 3.2). So there is a coincidence between parameters and characteristics of the growth process thought to be important. This is not entirely true for the spline approach. There is no natural way to measure the duration of the spurt. PB was defined arbitrarily and is not well determined.

3) By fitting SIMs, we get for each child two individual time scales. It is not yet clear, whether these time scales are useful. This will turn out when comparing them with the time scale defined by bone age and when setting them in relation to the timing of the different stages of puberty.

APPENDIX - 1 : Photos

The numbers in brackets below each photo are the pages, where the photo is referred to.

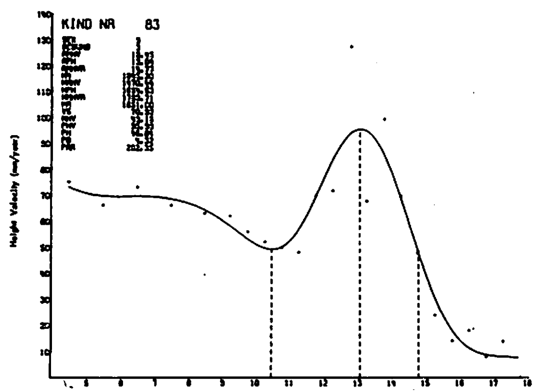

Photo 1

Smoothed curve child 83 (17)

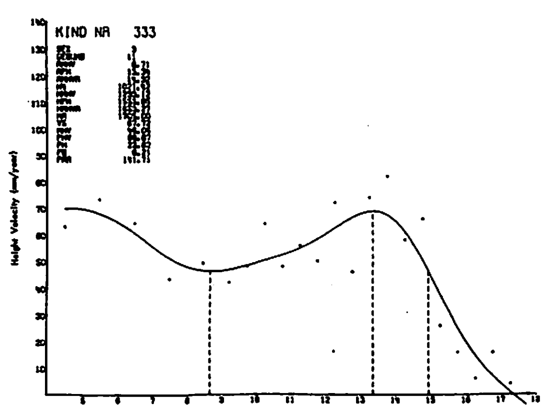

Photo 2

Smoothed curve child 333 (17)

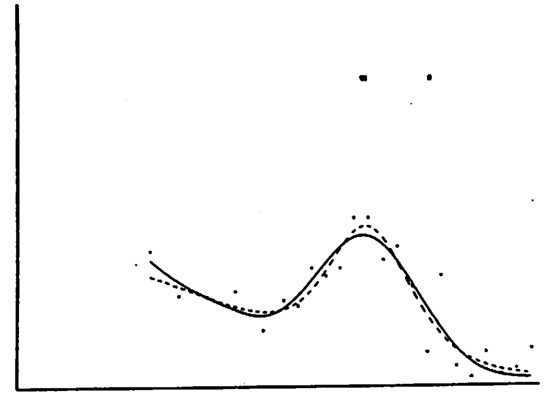

Photo 3

Simulated measurements, true and estimated curve model 1 (27)

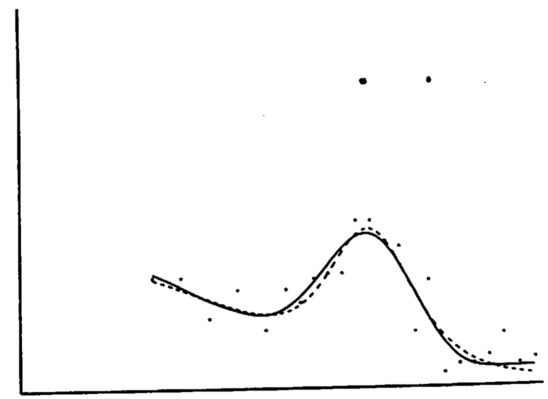

Photo 4

Simulated measurements, true and estimated curve model 1 (27)

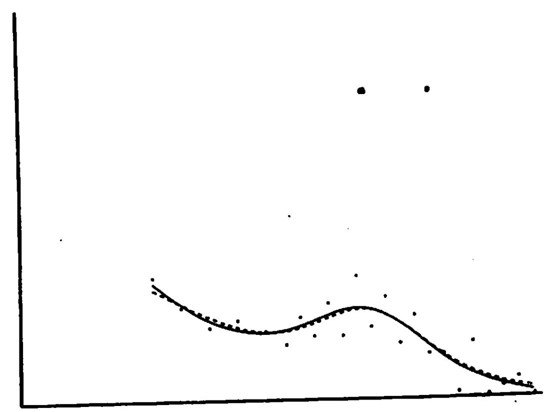

Photo 5

Simulated measurements, true and estimated curve model 2 (27)

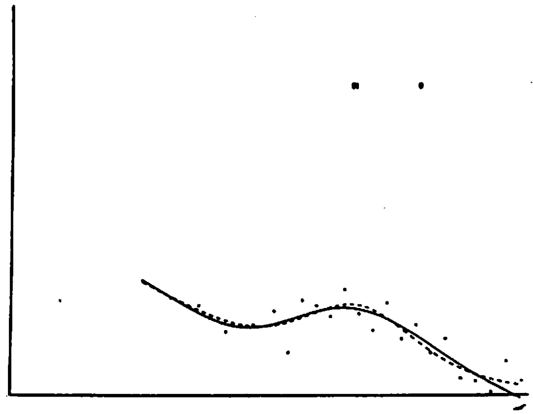

Photo 6

Simulated measurements
true and estimated
curve model 2
(27)

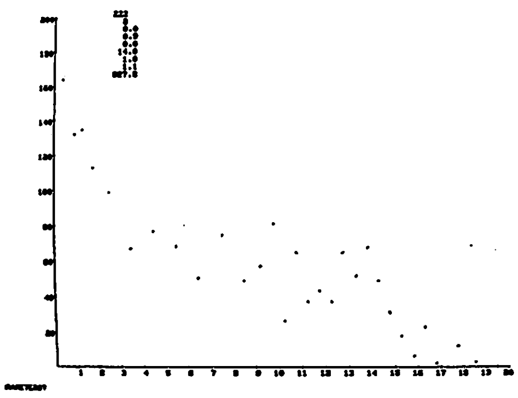

Photo 7

Raw velocities
child 222
(58)

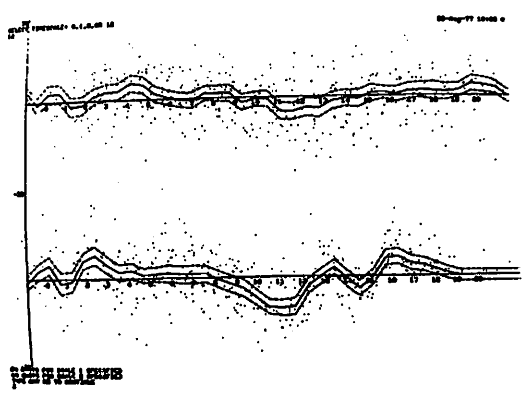

Photo 8

Aligned residuals
sample 1
after step 2
(52, 58)

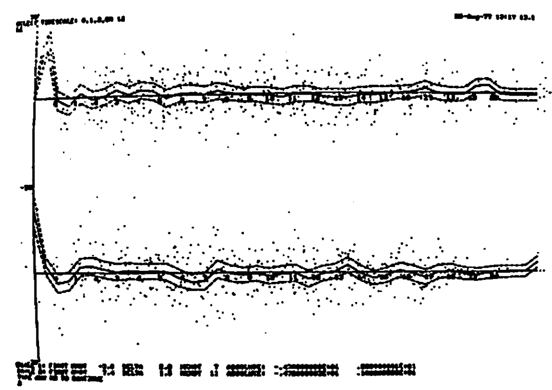

Photo 9

Aligned residuals
sample 1
after step 5
(52, 58)

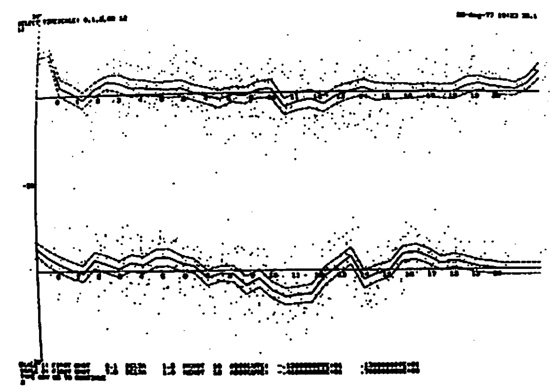

Photo 10

Aligned residuals
sample 2
after step 2
(52, 58)

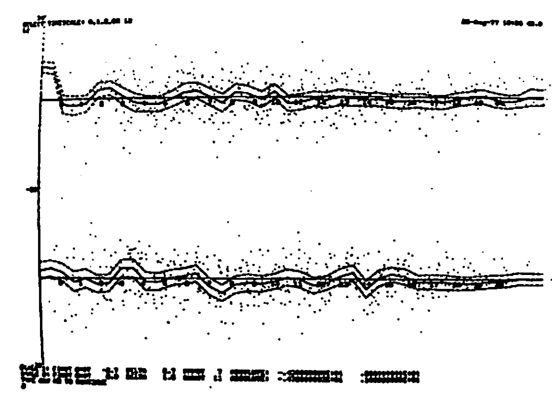

Photo 11

Aligned residuals
sample 2
after step 5
(52, 58)

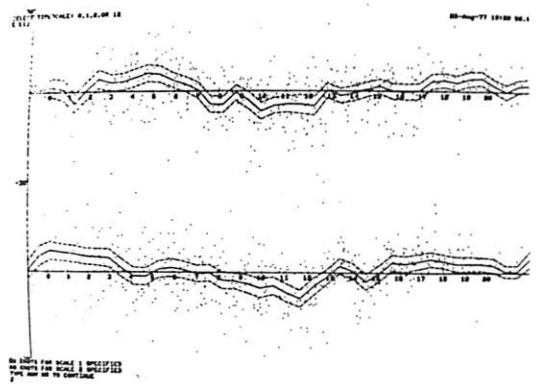

Photo 12

Aligned residuals
sample 3
after step 2
(52, 58)

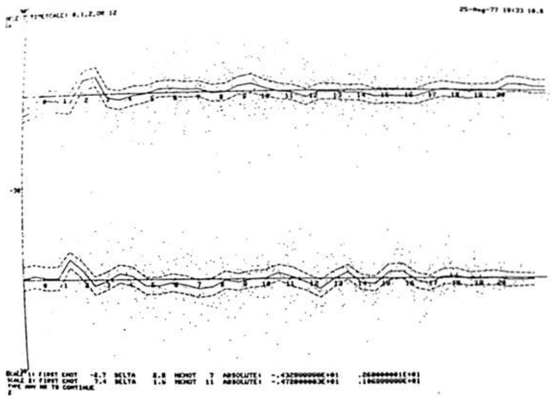

Photo 13

Aligned residuals
sample 3
after step 5
(52, 58)

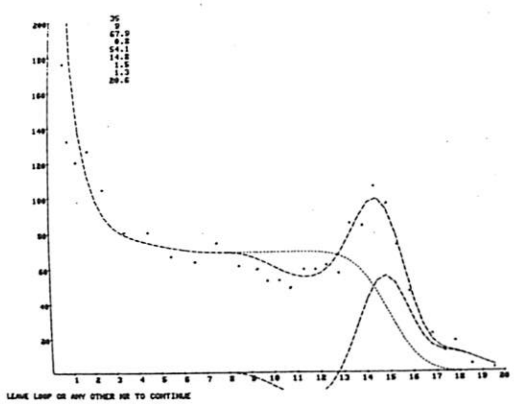

Photo 14

Fitted curve
child 35
(60)

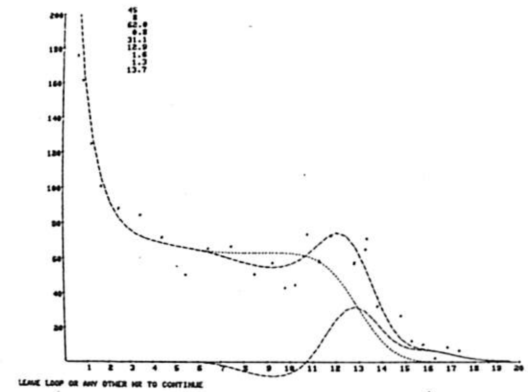

Photo 15

Fitted curve
child 45
(60)

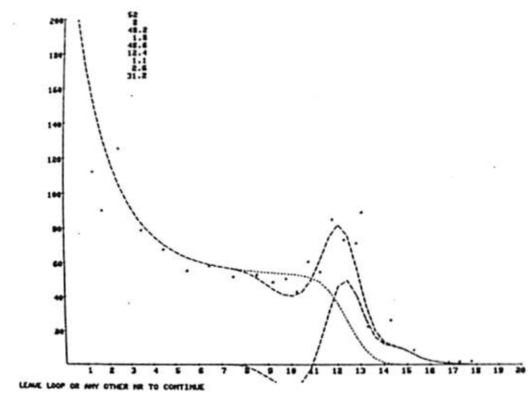

Photo 16

Fitted curve
child 52
(60)

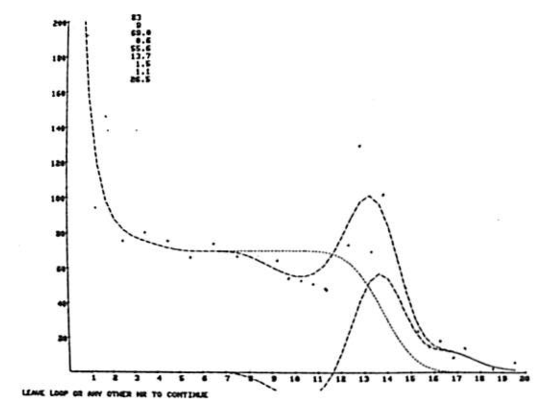

Photo 17

Fitted curve
child 83
(60)

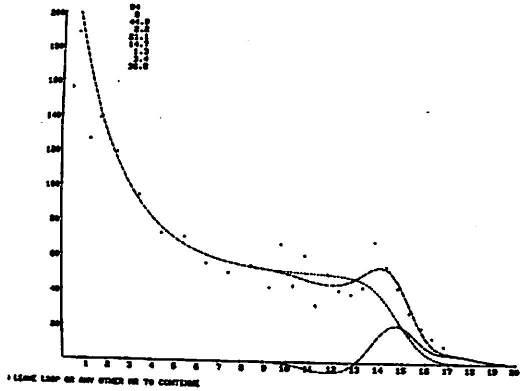

Photo 18

Fitted curve
child 94
(60)

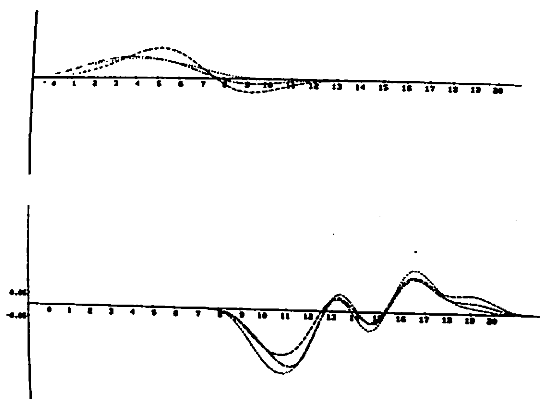

Photo 19

Estimated corrections
for shape-functions
(60)

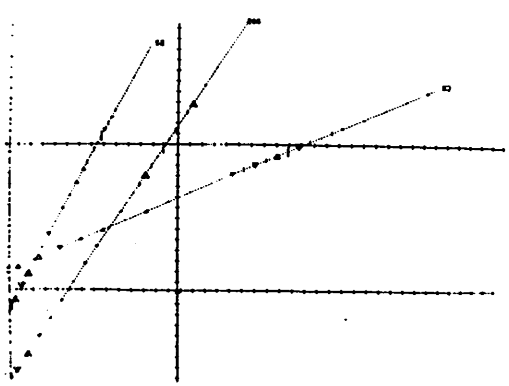

Photo 20

Two-dimensional
aligned residuals
plot
(52, 63)

[1] Falkner, F., Croissance et développement de l'enfant normal, Masson et Cie, Paris 1961

[2] Largo, R.H., Stützle, W., Gasser, T., Huber, P.J., Prader, A., A description of the adolescent growth spurt using smoothing spline functions, to be published in Annals of Human Biology

[3] Reinsch, C.H., Smoothing by spline functions, Numerische Mathematik 10 (1967), pp 177-183

[4] Wahba, G., Smoothing noisy data by spline functions II Technical report 380, August 1974, University of Wisconsin, Madison, Wiscon

[5] Wahba, G., Wold, S., A completely automatic french curve: Fitting spline functions by cross-validation, Communications in Statistics 4 (1975), pp 1-17

[6] Marshall, W.A., Evaluation of growth rates over periods of less t one year, Archives of Disease in Childhood 46 (1971), pp 414-420

[7] Reinsch, C.H., Smoothing by spline functions II, Numerische Mathematik 16 (1971), pp 451-454

[8] Bock, R.D., Wainer, H., Petersen, A., Thissen, D., Murray, J., Roche, A., A parameterization for individual human growth curves, Human Biology 45 (1973), pp 63-80

[9] Rosenblatt, M., Curve estimates, Annals of Mathematical Statistic 42 (1971), pp 1815-1842

[10] Wahba, G., A survey of some smoothing problems and the method of generalized cross-validation for solving them, Technical report 457, July 1976, University of Wisconsin, Madison, Wisconsin

[11] Burt, C., The backward child, Appleton-Century, New York 1937

[12] Deming, J., Application of the gompertz curve to observed patterr of growth in length of 48 individual boys and girls during the adolescent cycle of growth, Human Biology 29 (1957), pp 83-122

[13] Thissen, D., Bock, R.D., Wainer, H., Roche, A.F., Individual growth in stature: a comparison of four growth studies in the U.S.A., Annals of Human Biology 3 (1976), pp 529-542

[14] Marubini, E., Resele, F., Barghini, G., A comparative fitting of gompertz and logistic functions to longitudinal height data during adolescence in girls, Human Biology 43 (1971), pp 237-252

[15] Marubini, E., Resele, L.F., Tanner, J.M., Whitehouse, R.H., A fit of gompertz and logistic curves to longitudinal data during adolescence on height, sitting height and biacromial diameter in boys and girls of the Harpenden growth study, Human Biology 44 (1972), pp 511-524

[16] Tanner, J.M.,Whitehouse, R.H., Marubini, E., Resele, L.F. The adolescent growth spurt of boys and girls of the Harpenden growtl study, Annals of Human Biology 3 (1976), pp 109-126

[17] Zachmann, M., Ansley-Green, A., Prader, A., Interrelations of the effects of growth hormone and testosterone on hypopituitarism, in: Growth Hormone and Related Peptides, Excerpta Medica International Congress Series No. 381

[18] Labhart, A., Clinical endocrinology, Springer 1971

[19] Lawton, W.H., Sylvestre, E.A., Maggio, M.S., Self modeling nonlinear regression, Technometrics 14 (1972), pp 513-532

[20] Curry, H.B., Schoenberg, I.J., On Polya frequency functions. IV: The fundamental spline functions and their limits, J. Analyse Math. 17 (1966), pp 71-107

[21] Nagel, G., Wolff, W., Ein Verfahren zur Minimierung einer Quadratsumme nichtlinearer Funktionen, Biometrische Zeitschrift 16 (1974), pp 431-439

[22] Golub, G.H., Pereyra, V., The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate, SIAM J. Numer Anal. 10 (1973), pp 413-432

[23] Lawson, C.L., Hanson, R.J., Solving least squares problems, Prentice Hall 1974

ABSTRACT

Two methods are proposed to characterize the individual height growth of a child by a set of parameters:

1) Estimation of the growth curve by smoothing splines. Calculation of the parameters directly from the estimated curve.

2) Parameterization by fitting a nonlinear model. A type of model is proposed, which takes into account the present knowledge about the biological mechanisms of human growth. It is shown how the fact, that one has observed many children and so has obtained many realizations of the same model, can be used to reduce model bias.

The necessary algorithms are given. Finally, the two methods are compared.

CURRICULUM VITAE

Ich wurde am 14.9.1950 in Ravensburg (BRD) geboren. Nach dem Abitur begann ich im Wintersemester 1969/70 mit dem Studium der Mathematik an der Universität Heidelberg. Auf das Wintersemester 1971/72 wechselte ich an die ETH Zürich. Dort schloss ich mein Studium im Herbst 1973 mit dem Diplom ab. Meine Diplomarbeit schrieb ich im Sommer 1973 unter der Leitung von Professor P.J. Huber über ein Thema aus der mathematischen Statistik. Seit Frühjahr 1973 bin ich an der Fachgruppe für mathematische Statistik der ETH angestellt. Von Herbst 1975 bis Herbst 1977 arbeitete ich unter Leitung von Professor P.J. Huber und Professor A. Prader (Kinderspital Zürich) an dem interdisziplinären Forschungsprojekt "Auswertung der Zürcher longitudinalen Wachstumsstudie". Während dieser Zeit entstand auch die vorliegende Dissertation.