The Annals of Statistics
1995, Vol. 23, No. 2, 551–570

MODEL ESTIMATION IN NONLINEAR REGRESSION UNDER SHAPE INVARIANCE[1]

By Alois Kneip and Joachim Engel

Université Catholique de Louvain and Universität Bonn

Given data from a sample of noisy curves, we consider a nonlinear parametric regression model with unknown model function. An iterative algorithm for estimating individual parameters as well as the model function is introduced under the assumption of a certain shape invariance: the individual regression curves are obtained from a common shape function by linear transformations of the axes. Our algorithm is based on least-squares methods for parameter estimation and on nonparametric kernel methods for curve estimation. Asymptotic distributions are derived for the individual parameter estimators as well as for the estimator of the shape function. An application to human growth data illustrates the method.

1. Introduction. Selection of a suitable model constitutes a central problem in nonlinear regression. In classical nonlinear regression analysis of the model function is specified a priori. Methods that allow estimation of the model as well as the parameters have been introduced only recently by Lawton, Sylvestre and Maggion (1972) and by Kneip and Gasser (1988). These SElf-MOdeling Regression (SEMOR) methods are motivated by the situation encountered in many experiments in biomedicine and the physical sciences. They are based on the availability of observations Y_{ij}, $j = 1, \dots, n_i$, $i = 1, \dots, N$, from a sample of N individuals or experimental units according to some experimental design t_{ij}. While in many situations a nonlinear regression model is an appropriate framework to describe the data, there is no a priori specification for the model available that provides a good fit to the data. SEMOR then offers more flexibility than the classical approach. In SEMOR the data are allowed to speak for themselves in choosing the model function.

Given data from a sample of noisy curves we assume the model

$$Y_{ij} = f(t_{ij}, \theta_i) + \varepsilon_{ij}, \qquad j = 1, \dots, n_i,\ i = 1, \dots, N. \tag{1.1}$$

However, without further restricting assumptions it is not possible to estimate at the same time both parameters $\{\theta_i\}$ and model function f. This is a

Received March 1992; revised April 1994.

[1]This work was performed as part of the research program of the Sonderforschungsbereich 123 at the University of Heidelberg and the Sonderforschungsbereich 303 at the University of Bonn, and was made possible by financial support from the Deutsche Forschungemeinschaft.

AMS 1991 *subject classifications*. 62J02, 62G07.

Key words and phrases. Model selection, samples of curves, nonparametric smoothing, semiparametric methods, kernel estimators, human growth analysis.

consequence of the problem of identifiability. Specification of some class of functions that contains the true model is necessary. SEMOR is a semiparametric method in the sense that it employs methods from nonparametric curve estimation in a parametric nonlinear regression context.

One important class of model functions that allows self-modeling regression estimation to work is the shape invariant model (SIM). Under SIM the individual regression curves f_i are obtained from a common shape function ϕ via some parametric transformation. In the simplest case these transformations are linear, and we have

$$f(t,\theta_i) = f_i(t) = \theta_{i1}\phi\left(\frac{t-\theta_{i3}}{\theta_{i2}}\right) + \theta_{i4}. \tag{1.2}$$

Then the individual regression curves differ only by referring to different vertical and horizontal scales. As special cases the shape invariant model includes some important parametric models that have been thoroughly studied in the literature, such as the Gompertz model and the logistic model [compare Ratkowsky (1983)]. However, different from the classical approach, no particular specification for the model function ϕ is required.

In this paper we study the shape invariant model (1.1) and (1.2) and propose an iterative algorithm for the simultaneous estimation of the parameters θ_i and the shape function ϕ. Our estimation procedure is based on iteratively employing least-squares methods for parameter estimators alternated with kernel methods for nonparametric regression estimation. Kernel estimators are thoroughly studied and enjoy good analytic properties. We derive asymptotic distributions for the estimator of the individual parameters $\hat{\theta}_i$ as well as asymptotic bias and variances for the estimator of the shape function $\hat{\phi}$. If the number N of experimental units is large, the asymptotic distributions of our parameter estimators practically coincide with those resulting from least-squares estimators when ϕ is known. Our work is related to Härdle and Marron (1990), who, in a somewhat different setup, study nonparametric comparison of two regression curves f_1, f_2 which differ only by parametric shift transformations of the axes. They are concerned with testing and estimating the parameters transforming one curve into the other for which they obtain $\sqrt{n}$-convergence. Further they consider the important problem of testing the underlying assumption of shape invariance. Our focus is different, as we are concerned with estimating a model function ϕ and the individual parameters θ_i, in line with Lawton, Sylvestre and Maggio (1972) and Kneip and Gasser (1988). Though related, our approach seems to be simpler and more straightforward to apply than the one proposed in the latter papers. Furthermore, Lawton, Sylvestre and Maggio (1972) gave no theoretical results, while Kneip and Gasser (1988) only proved consistency.

In Section 2 we describe our regression model and introduce an iterative algorithm for estimation. Asymptotic results are presented in Section 3. The efficacy of our method is demonstrated in Section 4, where we apply our algorithm to real data. Based on the data of the Zürich human growth study, we estimate the shape function for the pubertal growth spurt separately for girls and boys.

2. Estimation. We assume data (Y_{ij}, t_{ij}) satisfying the regression model

$$Y_{ij} = f_i(t_{ij}) + \varepsilon_{ij}, \qquad j = 1, \dots, n_i, i = 1, \dots, N.$$

Here $f_1, f_2, \dots$ denote unknown smooth regression functions while ε_{ij} are unknown independent zero mean error terms with variance σ_i^2. The known design points t_{ij} are elements of an interval $J = [a_0, a_1] \subset \mathbb{R}$. We now postulate a shape-invariant model represented as follows:

$$f_i(\theta_{i2}t + \theta_{i3}) = \theta_{i1}\phi(t) + \theta_{i4} \quad \text{for } t \in \bar{J}, i = 1, \dots, N. \tag{2.1}$$

Here θ_i are unknown true parameter vectors, that is, $\theta_i := (\theta_{i1}, \dots, \theta_{i4}) \in \mathbb{R}_+^2 \times \mathbb{R}^2$ and ϕ is an unknown real function; $\bar{J} = [\bar{a}_0, \bar{a}_1]$ is a known interval fulfilling $\bar{J} \subset J$, that is, $a_0 < \bar{a}_0 \le \bar{a}_1 < a_1$. Furthermore, we have to introduce some normalizing condition:

We impose that the true parameters average to the vector (1, 1, 0, 0), that is,

$$\frac{1}{N}\sum_{i=1}^{N} \theta_{i1} = 1, \qquad \frac{1}{N}\sum_{i=1}^{N} \theta_{i2} = 1, \qquad \frac{1}{N}\sum_{i=1}^{N} \theta_{i3} = 0, \qquad \frac{1}{N}\sum_{i=1}^{N} \theta_{i4} = 0. \tag{2.2}$$

Normalizing conditions are necessary since the shape invariant model has some inherent unidentifiability: for given parameters ϑ_i and shape function ϕ we can always find other parameters ϑ_i^* and another shape function ϕ^* such that $\vartheta_{i1}\phi(t - \vartheta_{i3})/\vartheta_{i2}) + \vartheta_{i4} = \vartheta_{i1}^*\phi^*((t - \vartheta_{i3}^*)/\vartheta_{i2}^*) + \vartheta_{i4}^*$ holds for all t.

Conditions (2.2) are no restriction and serve as an anchor needed for the uniqueness of definition. Other choices of normalizing conditions are possible, and estimates based on (2.2) can always be renormalized accordingly. For example, in Härdle and Marron (1990) normalization is done by using the first curve f_1 as reference curve, that is, by requiring that $(\theta_{11}, \theta_{12}, \theta_{13}, \theta_{14}) = (1, 1, 0, 0)$. Our choice is motivated as follows: Condition (2.2) implies that the shape function ϕ allows representation as

$$\phi(t) = \frac{1}{N}\sum_{i=1}^{N} f_i(\theta_{i2}t + \theta_{i3}), \tag{2.3}$$

that is, $\phi = \phi_{f_1, \dots, f_N}$ is an average over shifted individual regression curves. The shifts are such that the individual curves are transformed to an average time scale. Amplitudes and locations of, say, minima and maxima of ϕ are then averages over amplitudes and locations of the corresponding extrema in the individual curves. In Kneip and Gasser (1992) a function ϕ with this property is called "structural average" of the f_i because ϕ summarizes structural information about the f_i in the form of an average. Model (2.1) refers to a subinterval $\bar{J} \subset J$. This is quite natural in applications where SIM is used to model only specific structural features of the f_i, and does not hold on the whole domain J (see our application below). However, even if (1.2) holds for all $t \in J$, the observations from the ith individuum contain information about ϕ only in the interval $J_i = [(a_0 - \theta_{i3})/\theta_{i2}, (a_1 - \theta_{i3})/\theta_{i2}]$. Model (2.1) then refers to an "overlapping" interval $\bar{J} = \bigcap_{i=1}^{N} J_i = [\bar{a}_0, \bar{a}_1]$ containing information from all individual units to be utilized for estimating ϕ. In many cases this will not impose a severe restriction. As illustration in this paper, we consider the problem of modeling human growth velocity. The data are from the Zürich longitudinal growth study consisting of measurements of many somatic variables taken regularly from birth to adulthood. Figure 1 shows kernel estimated growth velocity curves for four boys. Note that here we want to analyze the derivatives f_i' and not the regression function itself. The curves in Figure 1 were obtained by applying kernel estimators for derivatives. The small technical differences when dealing with derivatives will be discussed in Section 4.

The well-known, accentuated pubertal growth spurt is clearly visible in each curve. As a paradigm for the ideas developed in this paper, we consider modeling pubertal growth velocity. An inspection of estimated growth curves suggests the validity of the shape invariant model. This corresponds to the results of Stützle, Gasser, Molinari, Largo, Prader and Huber (1980), who proposed a model for human growth from early childhood to adulthood consisting essentially of two shape invariant components describing prepubertal and pubertal growth. On heuristic reasoning they propose a very complicated algorithm for estimating model and parameters and they obtain a satisfactory fit. Our approach provides a first step toward a simpler

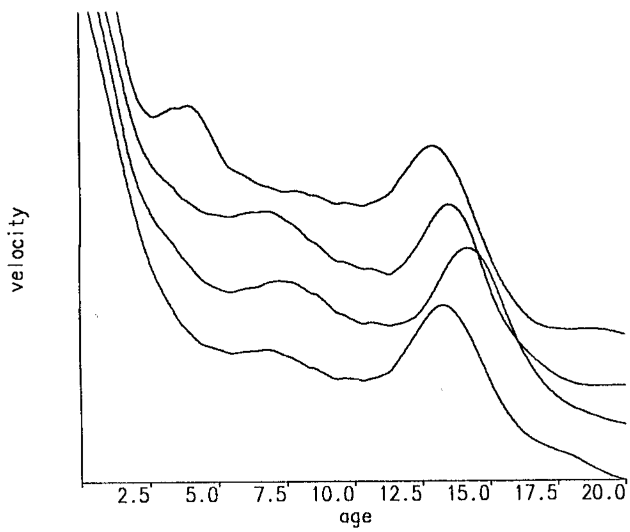

FIG. 1. *Velocity curves for growth of four boys from the Zürich growth study, estimated with nonparametric kernel estimators. For better visual separation, constants have been added to the curves.*

computational structure and better theoretical understanding of such procedures. To reduce complexity, we restrict ourselves to modeling the pubertal component. The age intervals to be considered extend from puberty entrance age (the minimum just before the pubertal peak) to the end of growth at, approximately, age 20. We thus assume that model (2.1) holds with $\mathscr{J} = [\bar{\tau}, 20]$, where $\bar{\tau}$ denotes average puberty entrance. It has to be emphasized that the SIM approach to the pubertal growth process generalizes previous modeling attempts, based on prespecified model functions. References are Deming (1957), Marubini, Resele and Barghini (1971), Bock, Warner, Petersen, Thissen, Murray and Roche (1973) and Preece and Baines (1978), who basically all refer to models of the form (1.2) for describing pubertal growth.

Let us now return to the general model (2.1) and (2.2), and consider the problem of how to estimate ϕ and the individual parameters. The algorithm we propose is based on the existence of initial estimates $\hat{\theta}_{i2}^0$ and $\hat{\theta}_{i3}^0$ for the horizontal shift parameters θ_{i2} and θ_{i3}. Such prior estimates can be obtained in different ways. Following Kneip and Gasser (1992), a simple and practicable method applies if we can identify at least two typical extrema and/or inflection points that each individual curve f_i and, hence, ϕ assumes. In what follows these points are called "structural points." When considering our growth example, identification of structural points is straightforward. Obviously, each individual growth velocity curve possesses a well defined pubertal maximum. Typical inflection points are maximal acceleration or deceleration during puberty.

Now assume that we are able to specify two structural points, and let τ_{1i}, τ_{2i} denote the individual locations of these points for the regression curve f_i. Then if we denote the locations of the corresponding structural points of ϕ by x_1 and x_2, equations (1.2) or (2.1) imply that for any i,

$$\tau_{1i} = \theta_{i2} x_1 + \theta_{i3},$$
$$\tau_{2i} = \theta_{i2} x_2 + \theta_{i3}.$$

As a consequence of the normalizing condition (2.2) it follows that the location of the two structural points of ϕ is given by $x_1 = \bar{\tau}_1 = (1/N)\Sigma_{l=1}^N \tau_{1l}$ and $x_2 = \bar{\tau}_2 = 1/N\Sigma_{l=1}^N \tau_{2l}$, that is, x_1 and x_2 are obtained as averages over the individual locations.

The above equations can be used to determine initial estimators for θ_{i2} and θ_{i3}. Estimates $\hat{\tau}_{1i}, \hat{\tau}_{2i}$ of the locations τ_{1i}, τ_{2i} of the two structural points can be obtained from nonparametric estimates of the regression functions and their derivatives [Müller (1985), (1989); Kneip and Gasser (1992)]. Then we obtain initial estimators for the individual parameters by

$$\hat{\theta}_{i2}^0 = \frac{\hat{\tau}_{i2} - \hat{\tau}_{i1}}{(1/N)\Sigma_{l=1}^N(\hat{\tau}_{l2} - \hat{\tau}_{l1})},$$
$$\hat{\theta}_{i3}^0 = \hat{\tau}_{i1} - \hat{\theta}_{i2}^0 \frac{1}{N}\sum_{l=1}^N \hat{\tau}_{l1}. \tag{2.4}$$

There are, of course, alternative ways for determining initial estimators: for example, one might choose one of the f_i, $i = 1, \dots, N$, as a "reference curve" and then estimate the parameters transforming all other curves into this reference curve, using methods of Härdle and Marron (1990). Suitable normalization, following (2.2), then leads to initial estimates.

Note that if ϕ were known, estimates of the parameters could be obtained by least-squares methods. On the other hand, if the parameters were known, an estimate of ϕ could be based on (2.3). Replacing in (2.3) the unknown regression function f_i with nonparametric estimates, the availability of initial estimators for θ_{i2}, θ_{i3} and relation (2.3) suggests the algorithm below. As nonparametric function estimates, we use kernel estimators of the convolution type [see Gasser and Müller (1984)]. Other estimators, such as splines, local polynomials or nearest neighbor methods, could be used alternatively. For mathematical convenience, we assume the parameters to be obtained from a compact subspace $D \subset \mathbb{R}_+^2 \times \mathbb{R}^2$, that is, we do not allow arbitrary large shifting and scaling.

ALGORITHM.

Step 1. For $i = 1, \dots, N$ estimate the regression function f_i by a nonparametric kernel estimate $\hat{f}_i$. More precisely,

$$\hat{f}_i(t) = \sum_{j=1}^{n_i} \frac{1}{b} \int_{s_{i(j-1)}}^{s_{ij}} W\left(\frac{t-v}{b}\right) dv\, Y_{ij}, \tag{2.5}$$

where $s_{ij} = (t_{ij} + t_{i(j+1)})/2$, W is a prespecified kernel function and $b > 0$ denotes the bandwidth. Then compute an initial estimate $\hat{\phi}^0$ of ϕ by

$$\hat{\phi}^0(t) = \frac{1}{N} \sum_{i=1}^{N} \hat{f}_i\left(\hat{\theta}_{i2}^0 t + \hat{\theta}_{i3}^0\right) \quad \text{for } t \in \bar{J}. \tag{2.6}$$

Step 2. Iterate $h = 1, 2, \dots, h^*$:

(i) For each $i = 1, \dots, N$ determine new estimates $\tilde{\theta}_i^h \in D$ by solving

$$\int_{\bar{J}} \left[\hat{f}_i\left(\tilde{\theta}_{i2}^h t + \tilde{\theta}_{i3}^h\right) - \tilde{\theta}_{i1}^h \hat{\phi}^{h-1}(t) - \tilde{\theta}_{i4}^h \right]^2 dt$$

$$= \min_{\vartheta \in D} \int_{\bar{J}} \left[\hat{f}_i(\vartheta_2 t + \vartheta_3) - \vartheta_1 \hat{\phi}^{h-1}(t) - \vartheta_4 \right]^2 dt.$$

(ii) Normalize these estimates, that is, for $(k, j) \in \{(1,4),(2,3)\}$ set

$$\hat{\theta}_{ik}^h = \frac{\tilde{\theta}_{ik}^h}{(1/N)\sum_{l=1}^N \tilde{\theta}_{lk}^h}, \qquad \hat{\theta}_{ij}^h = \tilde{\theta}_{ij}^h - \frac{\hat{\theta}_{ik}^h}{N} \sum_{l=1}^N \tilde{\theta}_{lj}^h, \qquad i = 1, \dots, N.$$

(iii) Set

$$\hat{\phi}^h(t) = \frac{1}{N} \sum_{i=1}^N \hat{f}_i\left(\hat{\theta}_{i2}^h t + \hat{\theta}_{i3}^h\right) \quad \text{for } t \in \bar{J}.$$

Step 3. Determine final estimators:

$$\hat{\theta}_i = \hat{\theta}_i^{h^*}, \qquad i = 1, \dots, N, \tag{2.7}$$

$$\hat{\phi}(t) = \frac{1}{N} \sum_{i=1}^{N} \hat{f}_i^* \left(\hat{\theta}_{i2} t + \hat{\theta}_{i3} \right) \quad \text{for } t \in \bar{J}, \tag{2.8}$$

where $\hat{f}_i^*$ denotes a nonparametric kernel estimator with bandwidth b^* [replace b by b^* in (2.5)].

For a reasonable implementation and a correct understanding of this algorithm, the following considerations are important:

1. The number h^* of iterations to be done depends on the quality of the initial estimators. Typically, a few iterations suffice to guarantee $\sqrt{n}$-convergence of the parameter estimators $\hat{\theta}_i$. For example, if the initial estimators are derived via extrema as structural points, then only $h^* = 2$ is necessary.
2. The procedure requires that the same bandwidth b is used for all estimates $\hat{f}_i$ and $\hat{\phi}^h$ throughout the iterations. This is of theoretical importance since then the leading bias terms for $\hat{f}_i$ and $\hat{\phi}_i^h$ cancel out (compare the proof of Theorem 1 in the Appendix) leading to approximately unbiased parameter estimates even if $\hat{\phi}^h$ is strongly biased. Note, however, that in Step 3 we allow $\hat{\phi} \neq \tilde{\phi}^{h^*}$.
3. Asymptotic theory (see Theorems 1 and 2 in Section 3) yields information on the choice of b and b^*. It turns out that the bandwidth b used to determine the estimates $\hat{f}_i$, required in Steps 1 and 2, is of minor importance. The leading terms in the asymptotic expansions for $\hat{\theta}_i$ and $\hat{\phi}$ do not depend on b. In particular, this bandwidth does not influence the resulting asymptotic distributions of the parameter estimates, unless b is extremely large or small. One might work with some "reasonable" bandwidth b according to prior experience or inspection by eye. Other possibilities for selecting b include averaging optimal bandwidths obtained by cross-validatory or plug-in methods.
4. In contrast, the choice of the bandwidth b^* is crucial when estimating ϕ. An undersmoothing bandwidth is advisable when determining the final estimate $\hat{\phi}$. Undersmoothing means to choose a smaller bandwidth than the optimal one for estimating an individual curve f_i. Note that $\hat{\phi}$ is defined as the average of N estimated individual curves. By undersmoothing these curves, we reduce the bias, while averaging curves itself diminishes variance. A procedure for determining b^* is sketched at the end of Section 3.
5. The restriction of the domain of estimation to the overlapping interval $\bar{J}$ has been introduced to derive the asymptotic results below. A slight modification of the algorithm allows consistent estimation of the model over the entire range of data as follows: under the above model (2.1) and (2.2), the observations from the ith individual carry information on $\phi(t)$

for $t \in J_i = [(a_0 - \theta_{i3})/\theta_{i2}, (a_1 - \theta_{i3})/\theta_{i2}]$. Therefore,

$$\phi(t) \sum_{\{i: t \in J_i\}} \theta_{i1} + \sum_{\{i: t \in J_i\}} \theta_{i4} = \sum_{\{i: t \in J_i\}} f_i(\theta_{i2} t + \theta_{i3}).$$

Based on this equation we could estimate ϕ over the entire data range using initial estimates of the θ_{ij}'s and of the f_i's. The algorithm described above is the special case of $\{i: t \in J_i\} = \{1, \dots, N\}$. An appropriate modification in Step 3 and integration over J_i in Step 2 then leads to an estimator of ϕ over the whole data range. However, the rate of convergency of this modified estimator would be slower outside the overlapping interval, depending on the size of the set $\{i: t \in J_i\}$.

3. Asymptotic results. Our asymptotic theory is based on the following assumptions on error terms and design:

ASSUMPTION 1. (i) For all i, j the random variables ε_{ij} are independent, and for fixed i the errors $\varepsilon_{i1}, \varepsilon_{i2}, \dots$ are i.i.d. zero mean random variables with variance $\sigma_i^2 < \infty$. Furthermore, for each $\alpha \in \mathbb{N}$ there exists a constant $C_\alpha < \infty$ such that $E\varepsilon_{ij}^\alpha < C_\alpha$ for all i.

(ii) For each individual we have the same number of observations, taken at the same set of design points, that is, $n := n_1 = n_2 = \cdots = n_N$, and for all n and each j we have $t_j := t_{1j} = \cdots = t_{Nj}$. Moreover, $t_{j+1} - t_j = (a_1 - a_0)/n$ for all j.

Condition (ii) has been introduced for mathematical convenience only. It can easily be generalized. Some further assumptions concerning model (2.1) are required:

ASSUMPTION 2. (i) $\theta_i \in \operatorname{int}(D)$ for all $i = 1, \dots, N$, where $\operatorname{int}(D)$ denotes the interior of the set D. For all $\vartheta \in D$ it holds that $\vartheta_2 \bar{a}_0 + \vartheta_3 \in (a_0, a_1)$ and $\vartheta_2 \bar{a}_1 + \vartheta_3 \in (a_0, a_1)$.

(ii) The parameters θ_i are identifiable, that is, for any i there exists a unique vector $\theta_i \in D$ such that (2.1) holds. Moreover, for any i there exists a $\delta_i > 0$ such that (2.1) generalizes to all $t \in [\bar{a}_0 - \delta_i, \bar{a}_1 + \delta_i]$.

A final assumption refers to the initial estimators of θ_{i2} and θ_{i3}, the number of iterations to be done, the smoothness of the curves f_i and ϕ and to the kernel functions.

ASSUMPTION 3. (i) The initial estimators $\hat{\theta}_{i2}^0$ and $\hat{\theta}_{i3}^0$ fulfill

$$|\hat{\theta}_{ik}^0 - \theta_{ik}| = O_P(n^{-\beta}), \qquad k = 2, 3 \text{ and for some } \beta > 0.$$

(ii) $h^* \geq -\log(\beta)/\log(2)$.

(iii) W is a kernel of some even order $k \in \mathbb{N}$, $k \geq 2$. More precisely, $\int W(x) x^q \, dx = 1$ for $q = 0$, $= 0$ for $q = 1, \dots, k-1$, $\neq 0$ for $q = k$.

(iv) W has support $[-1, 1]$, is continuously differentiable on $\mathbb{R}$ and twice continuously differentiable on $[-1, 1]$.

(v) For some $\mu \geq k + 1$ the functions ϕ and $f_1, \dots, f_N$ are μ-times continuously differentiable.

When using kernel methods for estimating the structural points as described by Müller [(1985), (1989)] one can specify the rate of convergence for the initial estimators: For example, if the structural points are based on extrema only and kernels of order greater than or equal to 2 are employed, then $\beta \geq 2/7$ and $h^* \geq 2$ is required by Assumption 3. If inflection points are involved the rates deteriorate to $\beta \geq 2/9$ (if $\mu \geq 4$), and $h^* \geq 3$ is necessary. To formulate asymptotic results we introduce the following notation ($1 \leq i, j \leq N$):

$$G_i(t, \vartheta) = f_i(\vartheta_2 t + \vartheta_3) - \vartheta_2 \phi(t) - \vartheta_4, \qquad Q_i = \operatorname{diag}(\theta_{i1}, \theta_{i2}, \theta_{i2}, \theta_{i1}),$$

$$\Gamma_{ji} = \left\{ \int_J G_{i,\vartheta}(\theta_i, t) G_{j,\vartheta}(\theta_j, t)^T \, dt \right\}^{-1}, \qquad \Gamma_i = \Gamma_{ii},$$

$$\zeta_{ij} = \frac{\theta_{i1}\theta_{j1}}{N} \sum_{l=1}^{N} \frac{1}{\theta_{l2}} - \frac{\theta_{j1}}{\theta_{i2}} - \frac{\theta_{i1}}{\theta_{j2}}, \qquad \zeta_i = \zeta_{ii},$$

where $G_{i,\vartheta}(\cdot, \cdot)$ denotes the vector of the derivatives of G_i with respect to ϑ.

THEOREM 1. *Under the above assumptions let $b \to 0$, $b = O(n^{-\gamma})$ for some $\gamma > 0$, $(nb^4)^{-1} \to 0$ as $n \to \infty$. For $i \in \{1, \dots, N\}$ for the parameter estimators we obtain*

$$\sqrt{n}\left(\hat{\theta} - \theta_i\right) \text{ is } AN(0, \Sigma_i), \tag{3.1}$$

where the covariance matrix can be represented as sum of $\Sigma_i = A_i + B_i + C_i$, defined by

$$A_i = \frac{1}{\theta_{i2}} \Gamma_i,$$

$$B_i = \frac{1}{N}\left(\frac{1}{N} Q_i \sum_{l=1}^{N} \frac{1}{\theta_{l2}} \Gamma_l Q_i - \frac{1}{\theta_{i2}} Q_i \Gamma_i - \frac{1}{\theta_{i2}} \Gamma_i Q_i \right),$$

$$C_i = \frac{1}{N}\left(\zeta_i \Gamma_i - \frac{1}{N} \sum_l \zeta_{il} \Gamma_{li} Q_i - \frac{1}{N} \sum_l Q_i \zeta_{li} \Gamma_{il} + \frac{1}{N^2} \sum_{l,k} Q_i \zeta_{lk} \Gamma_{kl} Q_i \right).$$

The proof is deferred to the Appendix. Let us consider the terms in the expression for the asymptotic covariance matrix in Theorem 1. First note that (2.1) implies that $f_i(t) = \theta_{i1}\phi((t - \theta_{i3})/\theta_{i2}) + \theta_{i4}$ for all $t \in J_{\theta_i} := [(\bar{a}_0 - \theta_{i3})/\theta_{i2}, (\bar{a}_1 - \theta_{i3})/\theta_{i2}]$. If ϕ were known, we could estimate θ_i by least-squares methods, minimizing $\sum_{t_{ij} \in J_{\theta_i}} \{Y_{ij} - \vartheta_1 \phi((t_{ij} - \vartheta_3)/\vartheta_2) - \vartheta_4\}^2$ with respect to $\vartheta \in D$. The term A_i equals the asymptotic covariance matrix of this least-squares estimator [compare Jennrich (1969)]. Subsequent normalization [see Step 2 (ii)] of these estimators would lead to the covariance matrix $A_i + B_i$. The matrix C_i is that part of Σ_i which is due to the fact that

ϕ is unknown. It is immediately seen that the matrices B_i and C_i are of order $1/N$. Thus, for large N, Σ_i approximately equals A_i.

Unfortunately, matrices B_i and C_i are rather complex, and for small N we were unable to analyze their behavior in general. There is, however, a remarkable exception: if $\theta_{i1} = \theta_{i2}$ for all $i = 1, \ldots, N$, then $C_i = 0$ which means we do not have to pay anything for not knowing ϕ. Furthermore, if all parameters θ_{i1}, θ_{j1} and θ_{i2}, θ_{j2} are equal, which by (2.2) implies that $\theta_{i1} = \theta_{i2} = 1$ for all i, then $B_i = -\Gamma_i/N$ is negative definite. We conclude that if $\theta_{i1} \approx \theta_{i2} \approx 1$ for all i, then the matrix $B_i + C_i$ is negative definite. Hence, in this case our parameter estimators are asymptotically more efficient than (nonnormalized) least-squares estimators under the known model ϕ. For the estimator of the shape function $\phi(t)$ we obtain the following result:

THEOREM 2. *In addition to the assumptions of Theorem* 1, *let* $b^* \to 0$, $(nb^{*2+\eta})^{-1} \to 0$ *as* $n \to \infty$ (*for some* $\eta > 0$). *We then obtain for the estimator of the shape function the following asymptotic representation*:

$$\hat{\phi}(t) - \phi(t) = \frac{1}{N}\sum_{i=1}^{N}\left\{\hat{f}_i^*(\theta_{i2}t + \theta_{i3}) - f_i(\theta_{i2}t + \theta_{i3})\right\}$$
$$+ O_p\Big([Nn]^{-1/2} + [nb^2]^{-1} + b^k n^{-1/2} + [nb^{*1.5}]^{-1}\Big).$$

Again, the proof is found in the Appendix. It should be noted that the term $[nb^{*1.5}]^{-1}$ is due to some rather crude approximations of some remainder terms. A more sophisticated analysis reduces this term to $[nb^{*1}]^{-1}$, and it might be possible to derive a still lower bound. The leading term on the right-hand side in Theorem 2 can be analyzed using standard expansions [compare Müller (1988)]:

$$\mathrm{Bias}^2\left[\frac{1}{N}\sum_{i=1}^{N}\left\{\hat{f}_i^*(\theta_{i2}t + \theta_{i3}) - f_i(\theta_{i2}t + \theta_{i3})\right\}\right]$$
$$= \frac{b^{*2k}M_k^2(W)}{k!^2}\phi^{(k)2}(t) - o(b^{*2k}),$$

$$\mathrm{Var}\left[\frac{1}{N}\sum_{i=1}^{N}\left\{\hat{f}_i^*(\theta_{i2}t + \theta_{i3}) - f_i(\theta_{i2}t + \theta_{i3})\right\}\right] = \frac{\sigma^2 V(W)}{Nnb^*} + o\big([nb^*]^{-1}\big).$$

Here $M_k(W) = \int u^k W(u)\,du$, $V(W) = \int W(u)^2\,du$, and $\sigma^2 = (1/N)\Sigma_{j=1}^N \sigma_j^2$. Considering AMISE $= \int(\mathrm{Bias}^2 + \mathrm{Var})$, we obtain as the asymptotically optimal bandwidth

$$b^*_{\mathrm{AMISE}} = \left\{\frac{1}{Nn}\,\frac{V(W)\sigma^2(\bar{a}_1 - \bar{a}_0)k!^2}{2kM_k(W)^2\int_{\mathcal{J}}\phi^{(k)2}(t)\,dt}\right\}^{-1/(2k+1)}$$

A data-dependent bandwidth $\hat{b}^*$ might be chosen with plug-in methods [e.g., Gasser, Kneip and Köhler (1991)]. However, a bandwidth $b^* = N^{-1/(2k+1)} \cdot \bar{b}$ is also asymptotically of the correct order of magnitude if $\bar{b}$ denotes the average of the optimal individual bandwidths. This results in a gain for the asymptotic MISE by the factor of $N^{-2k/(2k+1)}$ as compared to nonparametric curve estimates of a single curve.

4. Application to human growth data. The sample consists of measurements on $N = 112$ girls and $N = 120$ boys, taken yearly or half-yearly (in puberty), from birth to adulthood, following the design of an international prospective study. This resulted in a 32 dimensional data vector per subject for each somatic variable measured (height, leg and trunk length, width of shoulder, pelvis, etc.; the focus in most studies has been on height).

As outlined in Section 2, we apply our methodology in order to model pubertal growth velocity for height, separately for boys and girls. A slight modification has to be done due to the fact that we are dealing with derivatives and not with the regression functions themselves. The basic estimation method remains unchanged, only when determining $\hat{f}_i$ and $\hat{f}_i^*$ we use kernels tailored for estimating derivatives [Gasser and Müller (1984); Müller (1988)]. Asymptotic results generalize to this situation with, however, a slower rate of convergence for $\hat{\phi}$.

We assume that model (2.1) and (2.2) holds with $\bar{J} = [\bar{\tau}, 20]$, where $\bar{\tau}$ denotes average puberty entrance age. The puberty entrance age for each child is estimated as that time point where the nonparametrically estimated velocity curve has a local minimum before the pubertal spurt (PS). For details, see Gasser, Müller, Köhler, Prader, Largo and Molinari (1985) and Gasser, Köhler, Müller, Largo, Molinari and Prader (1985).

To apply the above algorithm initial estimators are needed for the shift parameters along the time axis, that is, for $\theta_i^{(2)}$ and $\theta_i^{(3)}$. These were obtained via the locations of two structural points τ_{1i} and τ_{2i} as the age of maximal acceleration of the ith child in PS and the age of maximum velocity in PS of the ith child. These functionals are determined from nonparametric estimates of growth acceleration curves. The initial estimators then are obtained by (2.4). Further, in this particular application, all vertical shift parameters are set $\theta_{i4} = 0$ since growth ends after puberty. The bandwidth b used in Steps 1 and 2 of the algorithm was chosen by prior experience as $b = 1.5$. This value is close to the average of the estimated optimal bandwidths for the individual growth velocity curves if measurements in early childhood are excluded. When taking different bandwidths $b = 1$ or $b = 2$ the resulting parameter estimates differed by at most 2%. For the bandwidth b^* a value of approximately $b^* = bN^{-1/5}$ was taken. Figure 2 shows a comparison of two methods for estimating growth curves: the estimates relying on the SIM structure of the data are visually clearly superior to kernel estimates based on estimated optimal bandwidths. This also confirms the validity of the shape invariant model.

Finally, Figure 3 compares the estimated model functions $\hat{\phi}_{\text{boys}}, \hat{\phi}_{\text{girls}}$ for boys and girls. The picture on the left shows the estimated model functions after three iteration steps representing the average growth during PS for girls and for boys. The picture on the right shows $\hat{\phi}_{\text{boys}}$ and $\vartheta_1 \hat{\phi}_{\text{girls}}((t -$

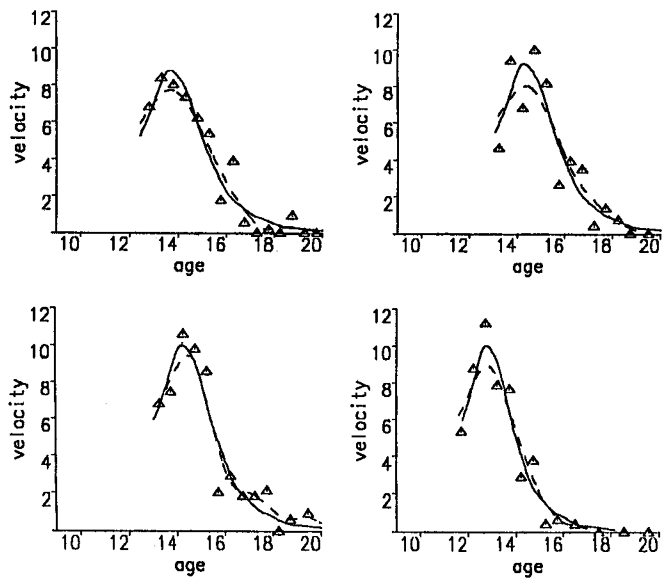

FIG. 2. *Comparison of two different estimates of the growth curve for four boys: a kernel estimate $\hat{f}_i$ based on measurements for the ith child only (dashed line) and a SIM based estimate $\hat{\theta}_{i1}\hat{\phi}((t - \hat{\theta}_{i3})/\hat{\theta}_{i2})$ (solid line). The triangles represent the divided differences of the data, that is, the height and time measurements.*

$\vartheta_3)/\vartheta_2)$, where the parameters ϑ are obtained by a least-square fit to the boys model. It suggests that the pubertal growth of boys is structurally not different from the girl's growth.

APPENDIX

In order to prove the two theorems, we introduce further notation: for $h \in \mathbb{N}$, $i \in \{1, \dots, N\}$, $\vartheta \in D$, $t \in \bar{J}$, we define (E denoting expectation)

$$\bar{f}_i(t) = E\hat{f}_i(t), \qquad \tilde{f}_i(t) = \hat{f}_i(t) - \bar{f}_i(t),$$

$$\hat{G}_i^h(t, \vartheta) = \hat{f}_i(\vartheta_2 t + \vartheta_3) - \vartheta_1 \hat{\phi}^{h-1}(t) - \vartheta_4,$$

$$G_i^h(t, \vartheta) = f_i(\vartheta_2 t + \vartheta_3) - \vartheta_1 \frac{1}{N} \sum_{l=1}^{N} f_l\left(\hat{\theta}_{l2}^{h-1} t + \hat{\theta}_{l3}^{h-1}\right) - \vartheta_4,$$

$$\overline{G}_i^h(t, \vartheta) = \bar{f}_i(\vartheta_2 t + \vartheta_3) - \vartheta_1 \frac{1}{N} \sum_{l=1}^{N} \bar{f}_l\left(\theta_{l2}^{h-1} t + \hat{\theta}_{l3}^{h-1}\right) - \vartheta_4,$$

$$\overline{G}_i(t, \vartheta) = \bar{f}_i(\vartheta_2 t + \vartheta_3) - \vartheta_1 \frac{1}{N} \sum_{l=1}^{N} \bar{f}_l(\theta_{l2} t + \theta_{l3}) - \vartheta_4,$$

$$\tilde{G}_i^h(t, \vartheta) = \hat{G}_i^h(t) - \overline{G}_i^h(t), \qquad \tilde{G}_i(t, \vartheta) = \hat{G}_i(t) - G_i(t).$$

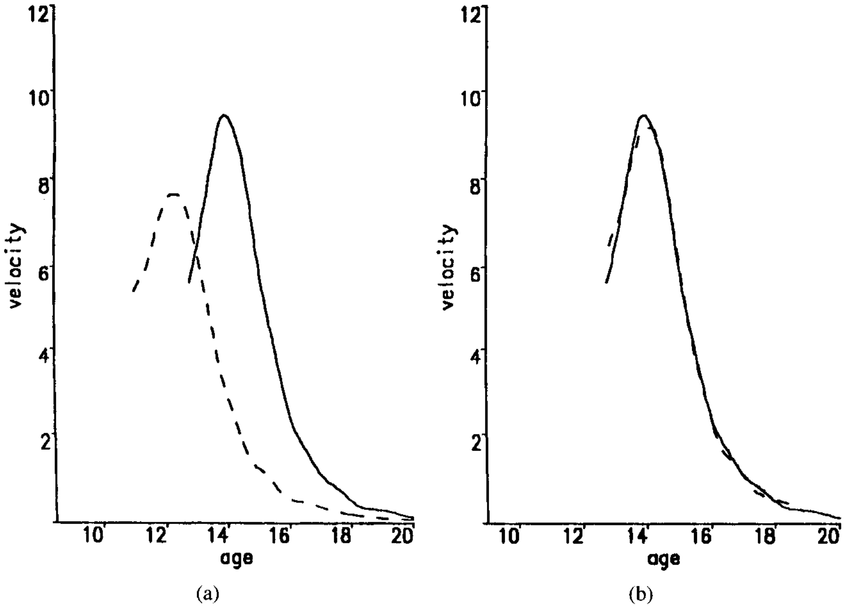

FIG. 3. (a) *Final estimate of model function for human growth between puberty entrance and age 20. Solid line: model for boys, dotted line: model for girls.* (b) *Model for boys (solid line) and model for girls (dashed line) fitted to boys model.*

Moreover, by $\hat{G}_{i,\vartheta}, \tilde{G}_{i,\vartheta}, \overline{G}_{i,\vartheta}$ we denote the derivative of the above terms with respect to ϑ. The identifiability of the model implies the existence of Γ_{ji}. The following lemma provides basic results for $\bar{f}_i$ and $\tilde{f}_i$.

LEMMA 1. *Under the above assumptions, we obtain, as* $n \to \infty$:

(i) *Let* $I := \{\inf_{\vartheta \in D}(\vartheta_2 \bar{a}_0 + \vartheta_3), \sup_{\vartheta \in D}(\vartheta \bar{a}_1 + \vartheta_2)] \subset J$. *Then*

$$\sup_i \sup_{t \in I} |\bar{f}_i(t) - f_i(t)| = O(b^k),$$

$$\sup_i \sup_{t \in I} |\bar{f}_i'(t) - f_i'(t)| = O(b^k),$$

$$\sup_i \sup_{t \in I} |\bar{f}_i''(t) - f_i''(t)| = O(b^{k-1}).$$

(ii) $\sup_{\vartheta \in D} \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2 \, dt = O_p([nb^{2\nu+1}]^{-1})$ *for* $i = 1, \dots, N$, $\nu \in \{0, 1\}$.

(iii) *For a compact set* $L \subset \mathbb{R}^\gamma$, $\gamma \in \mathbb{N}$, *let* $u: J \times L \to \mathbb{R}$ *denote a Lipschitz-continuous function. Then for any* $\eta > 0$,

$$\sup_{\vartheta \in D} \sup_{z \in L} \int_J \tilde{f}_i(\vartheta_2 t + \vartheta_3) u(t, z) \, dt = o_p\Big(\sup_{t \in J} \sup_{z \in L} |u(t, z)| n^{-1/2 + \eta} \Big).$$

(iv) *Furthermore, for any fixed* $\vartheta \in D$, $z \in L$, *it holds that*

$$\int_J \tilde{f}_i(\vartheta_2 t + \vartheta_3) u(t, z)\, dt = O_p\left(n^{-1/2} \left\{ \frac{1}{n\vartheta_2^2} \sum_{t_{ij} \in J_{\theta_i}} u\left(\frac{t_{ij} - \vartheta_3}{\vartheta_2}, z \right)^2 \right\}^{1/2} \right),$$

where $\bar{a}_{0,\vartheta} = \vartheta_2 \bar{a}_0 + \vartheta_3$, $\bar{a}_{1,\vartheta} = \vartheta_2 \bar{a}_1 + \vartheta_3$ *and*

$$J_{\theta_i} := \left[(\bar{a}_0 - \theta_{i3})/\theta_{i2}, (\bar{a}_1 - \theta_{i3})/\theta_{i2} \right].$$

(v) *Asymptotically, it holds that*

$$n^{1/2} \int_J \tilde{f}_i(\vartheta_2 t + \vartheta_3) u(t, z)\, dt \text{ is } AN\left(0, \frac{1}{\vartheta_2} \int_J u(t, z)^2\, dt \right).$$

PROOF. Statement (i) of the lemma follows directly from results of Gasser and Müller (1984). Note that by assumption the f_i are μ-times continuously differentiable, where $\mu \geq k + 1$. Consider assertion (ii). By our assumptions on W and on the ε_{ij} we obtain $\sup_{\vartheta \in D} \mathrm{E} \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt = O([nb^{2\nu+1}]^{-1})$. Furthermore, for any $\alpha \in \mathbb{N}$ there exists a $B_\alpha < \infty$ such that for all $\vartheta \in D$,

$$\text{(A.1)} \qquad E\left\{ \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt - E \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt \right\}^{2\alpha} \leq B_\alpha n^{-2\alpha} b^{-\alpha(4\nu+1)}.$$

This follows from Whittle's (1960) bounds for the moments of quadratic forms [compare the proof of Lemma A2 of Gasser, Kneip and Köhler (1991)]. Moreover, we conclude from our assumptions that there exists $0 < \rho < \infty$ such that

$$\sup_{\substack{\vartheta, \vartheta^* \in D \\ \|\vartheta - \vartheta^*\|_2 \leq n^{-\rho}}} \left| \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt - \int_J \tilde{f}_i^{(\nu)}(\vartheta_2^* t + \vartheta_3^*)^2\, dt \right| = o_p\left([nb^{2\nu+1/2}]^{-1} \right).$$

On the other hand, (A.1) implies that for any set Ω_n containing at most $(\bar{a}_1 - \bar{a}_0) n^\rho$ elements and for all $\varepsilon, \eta > 0$,

$$\begin{aligned} P&\left(\sup_{\vartheta \in \Omega_n} \left| n^{1-\eta} b^{2\nu+1/2} \left\{ \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt - E \int_J \tilde{f}_i^{(\nu)}(\vartheta t + \vartheta_3)^2\, dt \right\} \right| > \varepsilon \right) \\ &\leq \#\Omega_n \varepsilon^{-2\alpha} \\ &\quad \times E\left(\left| n^{1-\eta} b^{2\nu+1/2} \left\{ \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt - E \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2\, dt \right\} \right| \right)^{2\alpha} \\ &= O(n^\rho n^{-2\alpha\eta}). \end{aligned}$$

If α is taken sufficiently large, $n^{\rho}n^{-2\alpha\eta} = o(1)$ holds. When combining these arguments, we conclude that for any sufficiently small $\eta > 0$,

$$\sup_{\vartheta \in D} \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2 \, dt = \sup_{\vartheta \in D} E \int_J \tilde{f}_i^{(\nu)}(\vartheta_2 t + \vartheta_3)^2 \, dt + o_p\left(\left[n^{1-\eta} b^{2\nu+1/2} \right]^{-1}\right).$$

This proves assertion (ii). Using Whittle's (1960) bounds for the moments of linear forms, some straightforward computations show that for any $\alpha \in \mathbb{N}$ there exists a $B_\alpha^* < \infty$ such that $E\{\int_J \tilde{f}_i(\vartheta_2 t + \vartheta_3) u(t, z)\, dt\}^{2\alpha} \le B_\alpha^* n^{-\alpha}$ for all $\vartheta \in D$, $z \in L$. A partitioning argument similar to the one used above now leads to assertion (iii). Statement (iv) is straightforward. Assertion (v) follows from (iv) and the standard central limit theorem. □

For the proof of the two theorems, the following lemma, which describes the iteration path, is crucial. Depending on the rate of convergence achieved for $\hat{\theta}_i^{h-1}$, the lemma yields rates of convergence and asymptotic expansions for $\hat{\theta}_i^h$.

LEMMA 2. *For $h \in \mathbb{N}$ assume that the parameter estimates $\hat{\theta}_1^{h-1}, \dots, \hat{\theta}_N^{h-1}$ obtained in the $(h-1)$th iteration step satisfy $\|\hat{\theta}_i^{h-1} - \theta_i\|_2 = O_p(\alpha_n)$, $i = 1, \dots, N$, where α_n is a sequence of constants with $\alpha_n \to 0$ as $n \to \infty$. We then obtain for any small $\eta > 0$,*

$$\|\theta_i - \hat{\theta}_i^h\|_2 = O_p\left(\alpha_n^2\right) + o_p\left(n^{-1/2+\eta}\right), \qquad i = 1, \dots, N.$$

PROOF. Using $\hat{G}_i^h(t, \vartheta) = G_i^h(t, \vartheta) + \{\overline{G}_i^h(t, \vartheta) - G_i^h(t, \vartheta)\} + \tilde{G}_i^h(t, \vartheta)$, we conclude from Lemma 1 and the consistency of the $\hat{\theta}_i^{h-1}$ that

$$\sup_{\vartheta \in D} \left| \int_J \hat{G}_i^h(t, \vartheta)^2 \, dt - \int_J G_i(t, \vartheta)^2 \, dt \right| = o_p(1). \tag{A.2}$$

Note that $G_i(t, \theta_i) = 0$ for the true parameter θ_i. Since by definition of $\tilde{\theta}_i^h$, $\int_J \hat{G}_i^h(t, \theta_i)^2 \, dt \ge \int_J \hat{G}_i^h(t, \tilde{\theta}_i^h)^2 \, dt$, (A.2) implies that $\int_J G_i(t, \tilde{\theta}_i^h)^2 \, dt = o_p(1)$. By the identifiability of the parameters this leads to

$$\|\tilde{\theta}_i^h - \theta_i\|_2 = o_p(1). \tag{A.3}$$

The probability that $\hat{\theta}_i^h$ is not an interior point of D tends to zero and we obtain

$$\begin{aligned} o_p(n^{-2}) &= \frac{d}{d\vartheta} \int_J \hat{G}_i^h(t, \vartheta)^2 \, dt|_{\vartheta = \tilde{\theta}_i^h} = 2 \int_J \hat{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right) \hat{G}_i^h\left(t, \tilde{\theta}_i^h\right) dt \\ &= 2 \int_J \left\{ \overline{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right) + \tilde{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right) \right\} \left\{ \overline{G}_i^h\left(t, \tilde{\theta}_i^h\right) + \tilde{G}_i\left(t, \tilde{\theta}_i^h\right) \right\} dt. \end{aligned} \tag{A.4}$$

We now analyze the terms on the right-hand side of (A.4) starting with $\overline{G}_i^h(t, \tilde{\theta}_i^h)$. Set $\overline{\phi}(t) := \sum_{i=1}^n 1/b \int_{s_{i(j-1)}}^{s_{ij}} W((t-v)/b)\, dv\, \phi(t_i)$, $t \in \mathbb{R}$, and note that for all i,

$$(\text{A.5}) \qquad \bar{f}_i(\theta_{i2}t + \theta_{i3}) = \theta_{i1}\bar{\phi}(t) + \theta_{i4} \quad \text{for all } t \in \bar{J}.$$

Our assumptions on $\hat{\theta}_i^{h-1}$ imply that δ^h and δ^{*h}, defined as

$$\delta^h = \frac{1}{N}\sum_{l=1}^{N}\frac{\theta_{l1}}{\theta_{l2}}\left(\hat{\theta}_{l2}^{h-1} - \theta_{l2}\right), \qquad \delta^{*h} = \frac{1}{N}\sum_{l=1}^{N}\frac{\theta_{l1}}{\theta_{l2}}\left(\hat{\theta}_{l3}^{h-1} - \theta_{l3}\right),$$

are $O_p(\alpha_n)$. However, if δ^h, δ^{*h} are sufficiently small, then

$$(\text{A.6}) \qquad \bar{\phi}\left((1+\delta^h)t + \delta^{*h}\right) = \bar{\phi}(t) + \bar{\phi}'(t)(\delta^h t + \delta^{*h}) + O_p\left(\alpha_n^2\right)$$

holds for $t \in \bar{J}$. On the other hand,

$$(\text{A.7}) \qquad \frac{1}{N}\sum_{i=1}^{N}\bar{f}_i\left(\hat{\theta}_{i2}^{h-1}t + \hat{\theta}_{i3}^{h-1}\right) = \bar{\phi}(t) + \bar{\phi}'(t)(\delta^h t + \delta^{*h}) + O_p\left(\alpha_n^2\right).$$

Let $\theta_i^* = (\theta_{i1}, \theta_{i2}(1+\delta^h), \theta_{i3} + \theta_{i2}\delta^{*h}, \theta_{i4})$. By assumption (2.2), there exists an $\varepsilon_i > 0$ such that (A.5) generalizes to all $t \in [\bar{a}_0 - \varepsilon_i, \bar{a}_1 + \varepsilon_i]$. Hence, if δ^h, δ^{*h} are small enough, it follows that $\bar{f}_i(\theta_{i2}^* t + \theta_{i3}^*) = \theta_{i1}^*\bar{\phi}((\delta^h + 1)t + \delta^{*h}) + \theta_{i4}^*$ for all $t \in \bar{J}$. Since $\delta^h = O_p(\alpha_n) = o_p(1)$, $\delta^{*h} = O_p(\alpha_n) = o_p(1)$, we conclude from (A.6) and (A.7) that $\sup_{t\in J}|\bar{G}_i^h(t, \theta_i^*)| = o_p(\alpha_n^2)$. Based on a Taylor expansion of $\bar{G}_i^h(t, \tilde{\theta}_i^h)$, (A.4) now can be written as

$$(\text{A.8}) \qquad \begin{aligned} &\int\left\{\bar{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right) + \tilde{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right)\right\} \\ &\quad \times\left\{\bar{G}_{i,\vartheta}^h(t, \xi_i)^T\left(\tilde{\theta}_i^h - \theta_i^*\right) + \tilde{G}_i^h(t, \theta_i^*) + \tilde{G}_i^h\left(t, \tilde{\theta}_i^h\right)\right\} dt \\ &= o_p(n^{-2}). \end{aligned}$$

Recalling the definitions of $\tilde{G}_{i,\vartheta}^h$ and $\tilde{G}_i^h$, Lemma 1(ii) implies that

$$\int_{\bar{J}}\|\tilde{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right)\|_2^2\, dt = O_p\left([nb^3]^{-1}\right), \qquad \int_{\bar{J}}|\tilde{G}_i^h\left(t, \tilde{\theta}_i^h\right)|^2\, dt = O_p\left([nb]^{-1}\right).$$

Applying the Cauchy–Schwarz inequality, we obtain

$$\int_{\bar{J}}\tilde{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right)\bar{G}_i^h(t, \theta_i^*)\, dt = O_p\left(\frac{\alpha_n^2}{n^{1/2}b^{3/2}}\right),$$

$$\int_{\bar{J}}\tilde{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right)\tilde{G}_i^h\left(t, \tilde{\theta}_i^h\right) dt = O_p\left([nb^2]^{-1}\right).$$

Because of our assumptions on b it holds that $(n^{1/2}b^{3/2})^{-1} = o_p(1)$, and since $\int_{\bar{J}}\bar{G}_{i,\vartheta}^h(t, \tilde{\theta}_i^h)\bar{G}_i^h(t, \theta_i^*)\, dt = o_p(\alpha_n^2)$, (A.8) reduces to

$$(\text{A.9}) \qquad \begin{aligned} &\int_{\bar{J}}\left\{\bar{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right) + \tilde{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right)\right\}\bar{G}_{i,\vartheta}^h(t, \xi_i)^T\, dt\left(\theta_i^* - \tilde{\theta}_i^h\right) \\ &\quad = \int_{\bar{J}}\bar{G}_{i,\vartheta}^h\left(t, \tilde{\theta}_i^h\right)\tilde{G}_i^h\left(t, \tilde{\theta}_i^h\right) dt + O_p\left(\alpha_n^2 + [nb^2]^{-1}\right). \end{aligned}$$

Since the elements of $\overline{G}^h_{i,\vartheta}$ are continuously differentiable functions of t and ϑ, invoking Lemma 1 (iii) in order to bound the first term on the right-hand side leads to

$$\text{(A.10)} \qquad \int_{\mathcal{J}} \overline{G}^h_{i,\vartheta}(t, \tilde{\theta}^h_i) \tilde{G}^h_i(t, \tilde{\theta}^h_i)\, dt = o_p(n^{-1/2+\eta})$$

for any small $\eta > 0$. Lemma 1, (A.3) and $\|\theta^*_i - \theta_i\|_2 = o_p(\alpha_n)$ imply that

$$\left| \int_{\mathcal{J}} \{\overline{G}^h_{i,\vartheta}(t, \tilde{\theta}^h_i) + \tilde{G}^h_{i,\vartheta}(t, \tilde{\theta}^h_i)\} \overline{G}^h_{i,\vartheta}(t, \xi_i)^T\, dt - \int_{\mathcal{J}} G_{i,\vartheta}(t, \theta_i) G_{i,\vartheta}(t, \theta_i)^T\, dt \right| = o_p(1).$$

By assumption, the matrix $\int_{\mathcal{J}} G_{i,\vartheta}(t, \theta_i) G_{i,\vartheta}(t, \theta_i)^T\, dt$ is regular. This allows us to derive from (A.9) and (A.10) that for any small $\eta > 0$,

$$\text{(A.11)} \qquad \|\theta^*_i - \tilde{\theta}^h_i\|_2 = O_p(\alpha_n^2) + o_p(n^{-1/2+\eta}).$$

Note that by assumption $[nb^2]^{-1} = o(n^{-1/2})$. Clearly,

$$\text{(A.12)} \qquad \|\theta_i - \tilde{\theta}^h_i\|_2 = O_p(\alpha_n) + o_p(n^{-1/2+\eta})$$

is an immediate consequence. Relations (A.11) and (A.12) allow us to derive asymptotic expansions for $\theta_i - \hat{\theta}^h_i$. By definition, $\hat{\theta}^h_i$ is determined by normalizing $\tilde{\theta}^h_i$. Recall that $\theta^*_{i1} = \theta_{i1}$, $\theta^*_{i2} = \theta_{i2}(1 + \delta^h)$, $\theta^*_{i3} = \theta_{i3} + \theta_{i2}\delta^{*h}$, $\theta^*_{i4} = \theta_{i4}$, and $\delta^h, \delta^{*h} = O_p(\alpha_n)$. Using (A.11) our normalization procedure leads to the asserted bounds on the parameter estimates. □

PROOF OF THEOREM 1. By assumption there exists $\beta > 0$ such that $\|\hat{\theta}^0_i - \theta_i\|_2 = O_p(n^{-\beta})$ for all i. Invoking Lemma 2 we obtain $\|\hat{\theta}^h_i - \theta_i\|_2 = O_p(n^{-\beta 2^h}) + o_p(n^{-1/2+\eta})$ for all $h \in \mathbb{N}$ and any small $\eta > 0$. Since by assumption $h^* \geq -\log(\beta)/\log(2)$, this implies that for all small $\eta > 0$ and all i,

$$\text{(A.13)} \qquad \|\hat{\theta}^{h^*-1}_i - \theta_i\|_2 = o_p(n^{-1/2+\eta}).$$

Recall relations (A.9), (A.11) and (A.12) in the proof of Lemma 2. We obtain $\alpha_n = n^{-1/2+\eta}$ for $h := h^*$, and $\|\theta^*_i - \theta_i\|_2 = o_p(n^{-1/2+\eta})$,

$$\text{(A.14)} \qquad \|\theta^*_i - \tilde{\theta}^{h^*}_i\|_2 = o_p(n^{-1/2+\eta}), \qquad \|\theta_i - \tilde{\theta}^{h^*}_i\|_2 = o_p(n^{-1/2+\eta}).$$

We consider the first term on the right side of (A.9). When using (A.14) and Lemma 1 a straightforward expansion allows us to derive that

$$\text{(A.15)} \qquad \int_{\mathcal{J}} \overline{G}^{h^*}_{i,\vartheta}(t, \tilde{\theta}^{h^*}_i) \tilde{G}^{h^*}_i(t, \tilde{\theta}^{h^*}_i)\, dt = \int_{\mathcal{J}} G_{i,\vartheta}(t, \theta_i) \tilde{G}_i(t, \theta_i)\, dt + O_p\left([nb^2]^{-1} + b^k n^{-1/2}\right).$$

Since by Lemma 1(iv), $\int_{\mathcal{J}} G_{i,\vartheta}(t, \theta_i) \tilde{G}_i(t, \theta_i)\, dt = O_p(n^{-1/2})$, we immediately obtain

$$\text{(A.16)} \qquad \|\theta^*_i - \tilde{\theta}^{h^*}_i\|_2 = O_p(n^{-1/2}).$$

When using (A.16), (A.13) and Lemma 1 to analyze the left-hand side of (A.9) we immediately obtain

$$\int_{\mathscr{J}} \overline{G}_{i,\vartheta}^{h*}(t, \tilde{\theta}_i^{h*}) \overline{G}_{i,\vartheta}^{h*}(t, \xi_i)^T \, dt \big(\theta_i^* - \tilde{\theta}_i^{h*}\big)$$

$$= \int_{\mathscr{J}} G_{i,\vartheta}(t, \theta_i) G_{i,\vartheta}(t, \theta_i)^T \, dt \big(\theta_i^* - \tilde{\theta}_i^{h*}\big) + O_p(b^k n^{-1/2}) + O_p(n^{-1+2\eta}).$$

We thus arrive at

$$\text{(A.17)} \quad \begin{aligned} &\big(\theta_i^* - \tilde{\theta}_i^{h*}\big) \int_{\mathscr{J}} G_{i,\vartheta}(t, \theta_i) G_{i,\vartheta}(t, \theta_i)^T \, dt \\ &\quad = \int_{\mathscr{J}} G_{i,\vartheta}(t, \theta_i) \tilde{G}_{i,\vartheta}(t, \theta_i) \, dt + O_p\Big(\big[nb^2\big]^{-1} + b^k n^{-1/2}\Big). \end{aligned}$$

After applying our normalization procedure, (A.17) leads to

$$\text{(A.18)} \quad \begin{aligned} \theta_i - \hat{\theta}_i &= \left\{ \int_{\mathscr{J}} G_{i,\vartheta}(t, \theta_i) G_{i,\vartheta}(t, \theta_i)^T \, dt \right\}^{-1} \int_{\mathscr{J}} G_{i,\vartheta}(t, \theta_i) \tilde{G}_i(t, \theta_i) \, dt \\ &\quad - Q_i \frac{1}{N} \sum_{l=1}^{N} \left\{ \int_{\mathscr{J}} G_{l,\vartheta}(t, \theta_l) G_{l,\vartheta}(t, \theta_l)^T \, dt \right\}^{-1} \\ &\quad \times \int_{\mathscr{J}} G_{l,\vartheta}(t, \theta_l) \tilde{G}_l(t, \theta_l) \, dt + O_p\Big(\big[nb^2\big]^{-1} + b^k n^{-1/2}\Big) \\ &= O_p\Big(n^{-1/2} + \big[nb^2\big]^{-1} + b^k n^{-1/2}\Big). \end{aligned}$$

We recall that ε_{ij} and ε_{kl}, and hence $\tilde{f}_j$ and $\tilde{f}_l$ are independent for $j \neq l$. Based on Lemma 1(v) and the Cramér–Wold device, straightforward computations now lead to the assertions of the theorem. □

Proof of Theorem 2. We write again $\hat{f}_i^* = \bar{f}_i^* + \tilde{f}_i$, where $\hat{f}_i^*$ is the function estimate obtained by applying a kernel estimator with bandwidth b^*. It holds that

$$\text{(A.19)} \quad \begin{aligned} &\frac{1}{N} \sum_{i=1}^{N} \hat{f}_i^*\big(\hat{\theta}_{i2} t + \hat{\theta}_{i3}\big) \\ &\quad = \frac{1}{N} \sum_{i=1}^{N} \hat{f}_i^*(\theta_{i2} t + \theta_{i3}) \\ &\qquad + \frac{1}{N} \sum_{i=1}^{N} \bar{f}_i^{*\prime}(\theta_{i2} t + \theta_{i3}) \Big\{ \big(\hat{\theta}_{i2} - \theta_{i2}\big) t + \big(\hat{\theta}_{i3} - \theta_{i3}\big) \Big\} \\ &\qquad + \frac{1}{2N} \sum_{i=1}^{N} \bar{f}_i^{*\prime\prime}(\xi_{i,t}) \Big\{ \big(\hat{\theta}_{i2} - \theta_{i2}\big) t + \big(\hat{\theta}_{i3} - \theta_{i3}\big) \Big\}^2 \end{aligned}$$

$$+\frac{1}{N}\sum_{i=1}^{N}\tilde{f}_i^{*\prime}(\theta_{i2}t+\theta_{i3})\Big\{\big(\hat{\theta}_{i2}-\theta_{i2}\big)t+\big(\hat{\theta}_{i3}-\theta_{i3}\big)\Big\}$$
$$+\frac{1}{2N}\sum_{i=1}^{N}\tilde{f}_i^{*\prime\prime}(\xi_{i,t})\Big\{\big(\hat{\theta}_{i2}-\theta_{i2}\big)t+\big(\hat{\theta}_{i3}-\theta_{i3}\big)\Big\}^2$$

for all $t \in \bar{J}$, where $\xi_{i,t}$ denote suitable mean values. Let us now analyze the terms on the right side of (A.19). Recalling the independence of $\hat{f}_j$ and $\hat{f}_i$ for $j \neq i$, (A.18) implies that the second term on the right-hand side is bounded by $O_p([Nn]^{-1/2} + [nb^2]^{-1} + b^k n^{-1/2})$. Considering the third term, (A.18) leads to the bound $O_p(n^{-1})$. We conclude from the results of Gasser and Müller (1984) that for all i and any $t \in J$, $|\tilde{f}_i^{*\prime}(\theta_{i2}t + \theta_{i3})| = O_p([nb^3]^{-1/2})$. Together with (A.18) it follows that the fourth term on the right is bounded by $O_p([nb^{3/2}]^{-1})$. To analyze the fifth term, first note that for any small $\eta > 0$ it holds that

$$\sup_{t\in J}|\tilde{f}_i^{*\prime\prime}(t)| = o_p\Big(n^{\eta}\log n/[nb^{*5}]^{1/2}\Big).$$

This follows from Lemma 1 in Kneip and Gasser (1992), which in the version required in the present context is an easy generalization of a result by Cheng and Lin (1981). Together with (A.18) it follows that

$$\frac{1}{2N}\sum_{i=1}^{N}\tilde{f}_i^{*\prime\prime}(\xi_{i,t})\Big\{\big(\hat{\theta}_{i2}-\theta_{i2}\big)t+\hat{\theta}_{i3}-\theta_{i3}\Big\}^2 = O_p\left(\frac{n^{\eta}\log n}{n^{3/2}b^{*2.5}}\right),$$

which implies the assertion of the theorem. □

REFERENCES

BOCK, R. D., WAINER, H., PETERSEN, A., THISSEN, D., MURRAY, J. and ROCHE, A. (1973). A parametrization for individual human growth curves. *Human Biology* **45** 63–80.

CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **57** 223–234.

DEMING, J. (1957). Application of the Gompertz curve to the observed pattern of growth in length of 48 individual boys and girls during the adolescent cycle of growth. *Human Biology* **29** 83–122.

GASSER, T., KNEIP, A. and KÖHLER, W. (1991). A flexible and fast method for automatic smoothing. *J. Amer. Statist. Assoc.* **86** 643–652.

GASSER, T., KÖHLER, W. MÜLLER, H. G., LARGO, R., MOLINARI, L. and PRADER, A. (1985). Human height growth: correlation and multivariate structure of velocity and acceleration. *Annals of Human Biology* **12** 501–515.

GASSER, T. and MÜLLER, H. G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* **11** 171–185.

GASSER, T., MÜLLER, H. G., KÖHLER, W., PRADER, A., LARGO, R. and MOLINARI, L. (1985). An analysis of the mid-growth and adolescent spurts of height based on acceleration. *Annals of Human Biology* **12** 129–148.

HÄRDLE, W. and MARRON, S. (1990). Semiparametric comparison of regression curves. *Ann. Statist.* **18** 63–90.

JENNRICH, R. I. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40** 633–643.

KNEIP, A. and GASSER, T. (1988). Convergence and consistency results for self-modeling nonlinear regression. *Ann. Statist.* **16** 82–112.

KNEIP, A. and GASSER, T. (1992). Statistical tools to analyze samples of curves. *Ann. Statist.* **20** 1266–1305.

LAWTON, W. H., SYLVESTRE, E. A. and MAGGIO, M. S. (1972). Self-modeling regression. *Technometrics* **14** 513–532.

MARUBINI, E., RESELE, L. F. and BARGHINI, G. (1971). A comparative fitting of the Gompertz and logistic functions to longitudinal height data during adolescence in girls. *Human Biology* **43** 237–252.

MÜLLER, H. G. (1985). Kernel estimators of zeros and location and size of extrema of regression functions. *Scand. J. Statist.* **12** 221–232.

MÜLLER, H. G. (1988). *Nonparametric Regressions Analysis of Longitudinal Data.* Springer, Berlin.

MÜLLER, H. G. (1989). Adaptive nonparametric peak estimation. *Ann. Statist.* **17** 1053–1070.

PREECE, M. A. and BAINES, M. J. (1978). A new family of mathematical models describing the human growth curve. *Annals of Human Biology* **5** 1–24.

RATKOWSKY, D. A. (1983). *Nonlinear Regression Modeling.* Dekker, New York.

STÜTZLE, W., GASSER, T., MOLINARI, L., LARGO, R. H., PRADER, A. and HUBER, P. J. (1980). Shape-invariant modeling of human growth. *Annals of Human Biology* **7** 507–528.

WHITTLE, P. (1960). Bound on the moments of linear and quadratic forms in independent variables. *Theory Probab. Appl.* **5** 302–305.

INSTITUTE DE STATISTIQUE
UNIVERSITÉ CATHOLIQUE DE LOUVAIN
34 VOIE DU ROMAN PAYS
B-1348 LOUVAINE LA NEUVE
BELGIUM

INSTITUT FÜR WIRTSCHAFTSTHEORIE II
UNIVERSITÄT BONN
ADENAUERALLEE 24-26
D-53113 BONN
GERMANY