Journal of Statistical Planning and Inference 38 (1994) 327–350
North-Holland

Growth curves: a two-stage nonparametric approach

Joël Boularan

Laboratoire de Statistique et Probabilités, URA CNRS 745, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

Louis Ferré

Laboratoire de Mathématiques Appliquées, URA CNRS 1204, Université de Pau et des Pays de l'Adour, Avenue de l'Université, 64000 PAU, France

Philippe Vieu

Laboratoire de Statistique et Probabilités, URA CNRS 745, Université Paul Sabatier, 118 route de Narbonne, 31062 Toulouse Cedex, France

Received 2 December 1991; revised manuscript received 15 February 1993

Abstract

We develop in this paper a two-stage nonparametric method for estimating growth curves. Our two-stage additive model consists in assuming that dependence between height and age is a sum of two components, the former being the same for all the individuals. We estimate both functional components by nonparametric kernel smoothing techniques.

We first give theoretical results concerning L_1, L_2 and L_∞ rates of convergence for our estimates. Then we discuss how to choose the smoothing degree in an optimal way. This method is then applied to a real example.

AMS Subject Classification: Primary 65D10; secondary 62D25

Key words: Additive model; growth curves; kernel smoothing; nonparametric mixed effects model; two-stage procedure

1. Introduction

The treatment of longitudinal data (where J curves are measured on $j=1,\ldots,J$ subjects) has received attention for a long time. Particularly, one wishes to separate

Correspondence to: Prof. P. Vieu, Universite Paul Sabatier, Laboratoire de Statistique et Prob., U.A.C.N.R.S. 745, 118 Route de Narbonne, 31062 Toulouse Cedex, France

SSDI 0378-3758(93)E0019-D

what is common to the whole population from what is specific to each individual. This is the notion of subject-specific model as described, for instance, by Zeger et al. (1988). In the growth curves setting, many parametric models have been investigated with such a goal. Most familiar are those of Potthof and Roy (1964), Grizzle and Allen (1969) and Rao (1965, 1967), where polynomial models are considered and examined through a generalized analysis of variance model. A main limitation of such an approach is that it is available only for balanced data. When the data are unbalanced and when the models are not necessarily assumed to be linear, it is possible to fit individually each curve (with 'classical' growth curve models) and to work afterward on the parameter set (Houllier, 1987; Caussinus and Ferré, 1989) to investigate the relations and differences between the curves. To estimate common and specific effects, Laird and Ware (1982), in the linear case, and Lindstrom and Bates (1990), in the nonlinear one, use mixed effects models in which the common part is the fixed effect while the specific ones are the random effects of the model. Then the maximum likelihood estimators are obtained from the EM algorithm. This requires, of course, a knowledge of what is common and what is specific. In practical situations the main difficulty is to deciding which parametric model to assume.

Alternatively, nonparametric approaches have been developed in order to avoid this difficulty by estimating the relationship between height and age over a large class of smooth functions. In a previous work, Gasser et al. (1984) proposed to study growth curves by considering that there is a specific relationship for each individual between height Y^j and age T. Basically, their model can be written as

$$Y^j = m^j(T) + \text{error}. \tag{1.1}$$

Then these authors constructed J distinct kernel regression estimates of the functions m^j. Of course, this model is more general, but the main drawback of this 'direct' approach is that nonparametric regression estimates may behave quite poorly for small sample sizes and, unfortunately, this is quite often the case in practice. A first natural approach would be to investigate the model

$$Y^j = \bar{Y} + m_2^j(T) + \text{error}, \tag{1.2}$$

where $\bar{Y}$ is the averaged height over the J individuals. This two-stage approach would have the advantage that the mean part can be estimated very precisely by using the data of all the J individuals, but, of course, it would need to assume that all the J individuals have been measured at the same time. In this paper, we propose to investigate the following additive model:

$$Y = m_1(T) + m_2(T) + \text{error}, \tag{1.3}$$

where m_2 is a random function satisfying $E^{T=t}(m_2(T)) = 0$. Actually, this model is a nonparametric mixed effects model, and it combines the advantages of both models (1.1) and (1.2). It allows one to use possible common structure over all the individuals to estimate the curve m_1 by performing the regression from the data measured *for all* *the J individuals*, and so we may expect to have a more precise estimation than the one of the direct method derived from model (1.1). On the other hand, the random part m_2 represents the specific component of each individual in the population. Given the jth individual, we define the function

$$m_2^j(t)=E^{T=t}(Y^j-m_1(T)),$$

which will be estimated by performing the regression based on the data collected for the jth individual. This will be made clear, both theoretically, and with a real example, later on in the paper. The curve m_1 can be seen as a mean curve over all the individuals, however, if one is interested only in having mean curve that represents the average behaviour of growth in the population, it would be interesting to use Gasser and Kneip (1991) mean curve notion. Their approach differs from ours since we intend to estimate suitably the individual curves by emphasizing differences between them. For instance, when the population is not homogeneous and when the general shape of the curve differs between one groups, the meaning of an 'average behaviour' on the whole population is not clear. Nevertheless, it is consistent to consider a common effect in the curves to take account of the fact of their belonging to the same population. So, our method would be particularly relevant to distinguish the different groups. Then it could make sense to investigate the 'average behaviour' inside of each group, for instance by applying method of Gasser and Kneip. This 'average behaviour' could also be estimated by applying the method presented herein for the groups rather than for the individual, i.e. by considering the model:

$$Y=m_1(T)+m^g(T)+\text{error}, \tag{1.4}$$

where $m^g(T)$ would be the effect of the gth group. This could also lead straightforwardly to a three-stage method of estimating individual curves. For simplicity, we deal only with the two-stage approach, but the reader will see that our results could be easily extended to a model like (1.4).

Compared with (1.2), our model has the advantage of being usable also in situations where all the individuals are not measured at the same ages (and, moreover, also when the number of data is not the same for all the individuals). Of course, in case of balanced data the optimal smoothing will give an estimate of m_1 which is very close to the average height, and so our model (1.3) will look like (1.2). The model (1.3) is closely related to those presented by Besse and Ramsay (1986) and Rice and Silverman (1991) in which the estimations are carried out via nonlinear (or smoothed) principal component analysis. However, the problem of choice of optimal estimates (related to the choice of an optimal dimensionality) has not yet been solved. To overcome this difficulty, we propose to estimate mean and individual curves by kernel methods (see Section 2).

In addition to the differences in the model, our problem differs from the ones treated in all the papers mentioned above since we consider that time is random and not fixed. In the context of growth curve, this has been investigated by Mack and Müller (1989b) in a different context. Let us emphasize that this assumption is consistent with the fact that we are possibly dealing with unbalanced measurements. Nevertheless, our model also includes the treatment of the usual growth curve data sets, where heights are given at 'fixed ages', while ages are actually independent realizations (of time) which have been discretized and ordered (note that this ordering does not affect our results since we are dealing with sums). The results obtained in the context of fixed design regression (see, for instance, the discussion on bandwidth choice in Section 3) are no longer available, and we give in Section 3 the theoretical asymptotics about the estimates of the presented model in the situation of random design. Particularly, rates of convergence (in L_1, L_2 and L_∞ settings) are given together with a data-driven way of selecting optimal estimates of m_1 and m_2^j. In Section 4, an illustration is given through a real data set which is treated both by our two-stage method and by a 'direct' kernel method derived from model (1.1). Finally, the proofs of the theoretical results of Section 3 are given in Section 5.

Let us finally mention that our method is presented in this paper in the growth curve setting. However, our results are valid for any repeated measurements problem when T is not necessarily time but may be any random variable.

2. Kernel estimates

Recall that the relation between height Y and time T is defined according to the following two-stage model:

$$Y=m_1(T)+m_2(T)+\varepsilon,$$

where ε is a random variable with zero mean and where, to make the model identifiable, we assume that

$$E^{T=t}(Y)=m_1(T), \quad \text{i.e. } E^{T=t}(m_2(T))=0. \tag{2.1}$$

Assume that we have N measurements of heights and that the number of subjects depends on N, i.e. $J=J_N$. We have measured the height Y_i^j of each of the J_N individuals $j=1,\dots,J_N$ at n_j different times t_i^j, $i=1,\dots,n_j$. For any $j=1,\dots,J_N$, and for any $i=1,\dots,n_j$, we have

$$Y_i^j=m_1(t_i^j)+m_2^j(t_i^j)+\varepsilon_i^j, \tag{2.2}$$

where

$$m_2^j(t)=E^{T=t}(Y^j-m_1(T)),$$

and ε_i^j are independent random variables with zero mean. The function m_1 will be estimated by using all the data as if they had been collected for one single individual. We use as estimates of m_1 the familiar Watson–Nadaraya kernel regression smoothers. These estimates are defined from a kernel function K and from a bandwidth h by

$$\hat{m}_1(t)=\hat{g}_1(t)/\hat{f}_1(t), \tag{2.3}$$

where

$$\hat{f}_1(t)=\frac{1}{Nh}\sum_{j=1}^{J_N}\sum_{i=1}^{n_j} K((t-t_i^j)/h),$$

and

$$\hat{g}_1(t)=\frac{1}{Nh}\sum_{j=1}^{J_N}\sum_{i=1}^{n_j} Y_i^j K((t-t_i^j)/h).$$

Note that we have

$$N=\sum_{j=1}^{J_N} n_j.$$

Surveys on asymptotic properties of these estimates, together with definitions of other kinds of regression estimates, can be found in Collomb (1981) and Härdle (1990). We also wish to mention works by Jennen–Steinmetz and Gasser (1988), Mack and Müller (1989a, b) and by Fan (1992), in which other forms of kernel estimates are defined and compared with the Watson–Nadaraya estimate defined above. Because it is based on a sample of size N, we expect our estimate to have the classical rate of convergence $N^{-4/5}$. This will be stated precisely in Section 3.

Once we have estimated the mean curve m_1, we estimate for each individual the individual component by doing kernel regression on the residuals, i.e.

$$\hat{m}_2^j(t)=\hat{g}_2^j(t)/\hat{f}_2^j(t), \tag{2.4}$$

where

$$\hat{f}_2^j(t)=\frac{1}{n_j h_j}\sum_{i=1}^{n_j} K((t-t_i^j)/h_j)$$

and

$$\hat{g}_2^j(t)=\frac{1}{n_j h_j}\sum_{i=1}^{n_j} (Y_i^j-\hat{m}_1(t_i^j))K((t-t_i^j)/h_j).$$

Here the bandwidth h_j can vary from one individual to the other. This allows different degrees of smoothing according to the structure of the relation between height and time for each individual since it is desirable to take into account special features due to each individual. This will be of particular interest when the number of measurements n_j differs from one individual to the other.

3. Asymptotic properties

In this section, we give rates of convergence of these estimates, and then the task of the choice of the smoothing factors h and h_j will be investigated. We first focus our attention on convergence of the kernel estimates when several measures of errors are considered (L_1, L_2 and L_∞ errors). Let us introduce some assumptions. Our nonparametric model is defined by assuming that

$$\begin{aligned} &m_1,\ m_2^j \text{ and } f \text{ have two continuous derivatives,} \\ \text{and} \\ &\psi_j(\cdot)=E((Y^j)^2 \mid T=.) \text{ exists and is continuous on the real line.} \end{aligned} \tag{3.1}$$

In order to avoid problems with possible null values for the denominator, we have to restrict our attention to some compact set. This is done via some bounded weight function w having a compact support S and such that

$$\inf_{t\in S} f(t)>\xi \quad \text{for some } \xi>0. \tag{3.2}$$

Concerning our estimates, it is assumed that the kernel K is a Hölder-continuous, probability density function being square-integrable, symmetric and having a finite and nonnull second-order moment. To get L_∞ results, the usual assumptions on the bandwidth h are:

$$h\to 0 \quad \text{as } N\to\infty, \tag{3.3}$$

and

$$\frac{Nh}{\log N}\to\infty \quad \text{as } N\to\infty, \tag{3.4}$$

while for any j the individual bandwidths h_j have to satisfy:

$$h_j\to 0 \quad \text{as } N\to\infty, \tag{3.5}$$

and

$$\frac{n_j h_j}{\log n_j}\to\infty \quad \text{as } N\to\infty. \tag{3.6}$$

Relations between h and h_j are controlled by assuming that

$$h=\mathrm{o}(h_j). \tag{3.7}$$

To ensure that we do not have very large differences in the data collected for individuals, we required that for any $j=1,\ldots,J_N$ we have

$$\frac{|n_j-n|}{n}=\mathrm{o}(h^2), \tag{3.8}$$

where n is the average number of measurements for one individual

$$n=N/J_N.$$

About the estimation of the individual curve m_2^j, we have to assume that

$$\frac{\log J_N}{J_N}=\mathrm{O}(\inf\{h;\ h_j^4\}), \tag{3.9}$$

in order to control the way the number of individuals J_N grows up to infinity.

Remark 1. To make clear the relation between the number of subjects and the number of data points, note that in the familiar cases when $h=C_0N^{-1/5}$ and $h_j=C_jn^{-1/5}$, condition (3.9) is satisfied when $J_N=C_1n^{4/5}$ (i.e., $J_N=C_2N^{4/9}$), which is, in practice, quite a reasonable assumption on J_N.

Theorem 1. (i) *Assume that conditions* (2.1), (3.1)–(3.4) *and* (3.8) *hold. Then we have*

$$\sup_t |\hat{m}_1(t)-m_1(t)|w(t)=\mathrm{O}(h^2)+\mathrm{O}\left(\sqrt{\frac{\log N}{Nh}}\right)+\mathrm{O}\left(\sqrt{\frac{\log\log J_N}{J_N}}\right), \textit{ a.s.}$$

(ii) *If in addition* (3.9), (3.5)–(3.7) *hold, then we have*

$$\sup_t |\hat{m}_2^j(t)-m_2^j(t)|w(t)=\mathrm{O}(h_j^2)+\mathrm{O}\left(\sqrt{\frac{\log n_j}{n_jh_j}}\right), \textit{ a.s.}$$

To get L_1 or L_2 consistency, we may weaken the conditions on the bandwidths in the following way:

$$Nh\to\infty \quad \text{as } N\to\infty, \tag{3.10}$$

and

$$n_jh_j\to\infty \quad \text{as } N\to\infty. \tag{3.11}$$

Theorem 2. (i) *Assume that conditions* (2.1), (3.1)–(3.3), (3.5) *and* (3.10) *hold. Then we have*

$$E\int |\hat{m}_1(t)-m_1(t)|w(t)\,\mathrm{d}t=\mathrm{O}(h^2)+\mathrm{O}\left(\sqrt{\frac{1}{Nh}}\right)+\mathrm{O}\left(\sqrt{\frac{\log\log J_N}{J_N}}\right).$$

(ii) *If in addition* (3.5), (3.7), (3.9) *and* (3.11) *hold, then we have*

$$E\int |\hat{m}_2^j(t)-m_2^j(t)|w(t)\,\mathrm{d}t=\mathrm{O}(h_j^2)+\mathrm{O}\left(\sqrt{\frac{1}{n_jh_j}}\right).$$

Theorem 3. (i) *Assume that conditions* (2.1), (3.1)–(3.3), (3.8) and (3.10) *hold. Then we have*

$$E\int (\hat{m}_1(t)-m_1(t))^2w(t)\,\mathrm{d}t=Bh^4+\frac{V}{Nh}+\mathrm{O}\left(\frac{\log\log J_N}{J_N}\right)+\mathrm{o}\left(h^4+\frac{1}{Nh}\right).$$

(ii) *If in addition* (3.5), (3.7), (3.9) and (3.11) *hold, then we have*

$$E\int (\hat{m}_2^j(t)-m_2^j(t))^2w(t)\,\mathrm{d}t=B^jh_j^4+\frac{Vj}{n_jh_j}+\mathrm{o}\left(h_j^4+\frac{1}{n_jh_j}\right).$$

The constants B, V, B^j and V^j are given in the proof, respectively, by (5.13), (5.14), (5.18) and (5.19).

These theorems show that our estimates may reach the optimal global rates of convergence as defined in Stone (1982), by suitable bandwidth choices. Look for instance at the L_1 or L_2 error (same thing applies also for the L_∞ error upto an additional log factor). Taking h of the order of $N^{-1/5}$, h_j of the order of $n^{-1/5}$ and J_N large enough, the estimate of the mean curve m_1 achieves the optimal rate of convergence $N^{-4/5}$, while the estimates of the individual curves only achieve the optimal rate $n^{-1/5}$. If we consider a direct model like (1.1), kernel estimates of the functions m^j converge at the rate $n^{-1/5}$. Of course, the sum of our two estimates also converges at the rate $n^{-1/5}$. However, even if both methods (direct and two-stage) give the same rate of convergence for the individual curves, our two-stage procedure makes one step of the estimation much more precise since the mean curve is estimated with considerably better rate. This is particularly true when the number of individuals is large. Indeed, our procedure uses the idea of mixed effects model and estimates one individual curve by using (through estimation of m_1) all the information given by all other individuals. This is reasonable because all the individuals are coming from the same population.

As it is the case in any functional estimation problem, the smoothing factors h and h_j take a prominent role as illustrated by the theorems above. The aim of the remaining part of this section is to discuss optimal data-driven choices of them. What is usually done is to look for bandwidths which are optimal in terms of L_2 errors. Theorem 3 is not directly useful for this, since the minimizers of the leading terms in expressions i and ii depend on the unknown functions to be estimated. In other settings, several techniques have been developed to select the smoothing factor (see the surveys by Marron (1988) in density and Vieu (1992) in regression). The most popular are cross-validation (Härdle and Marron, 1985), bootstrapping (Härdle and Bowman 1987) and plug-in methods (Gasser and Herrmann, 1991). As pointed out in the above-mentioned surveys, all these techniques are actually well-known when the estimates do not involve some random denominator. However, until now, only the cross-validation procedure can be shown to have optimality properties in cases of random denominator. This is the reason why we focus our attention here on cross-validation. Future research could study the behaviour of other bandwidth selection rules and compare them with cross-validation. According to the conclusions of both of the above-mentioned surveys, the most probable is that none among these three kinds of methods could be shown to be better than all the others.

Estimating m_1 by $\hat{m}_1$ is a classical regression problem and so we choose as bandwidth the minimizer of the following global cross-validation score function:

$$\mathrm{CV}(h)=\frac{1}{N}\sum_{j=1}^{J_N}\sum_{i=1}^{n_j}(Y_i^j-\hat{m}_1^{-j}(t_i^j))^2\,w(t_i^j),$$

where $\hat{m}_1^{-j}$ is the kernel estimate of m_1 constructed when the data collected for the jth individual have been suppressed, i.e.,

$$\hat{m}_1^{-j}(t)=\hat{g}_1^{-j}(t)/\hat{f}_1^{-j}(t),$$

with

$$\hat{f}_1^{-j}(t)=\frac{1}{(N-n_j)h}\sum_{k\neq j}\sum_{i=1}^{n_k} K((t-t_i^k)/h),$$

and

$$\hat{g}_1^{-j}(t)=\frac{1}{(N-n_j)h}\sum_{k\neq j}\sum_{i=1}^{n_k} Y_j^k K((t-t_i^k)/h).$$

In the usual regression setting, the Härdle and Marrons's technique consisted in deleting only one data point, but further works (see e.g. Härdle and Vieu (1991)) gave motivations for leaving out more than one data point. In our growth curves point of view, it is quite clear that all the measures collected for the jth subject may have influence on the measurement at the ith time. So it makes sense to leave out all the data concerning the jth individual in constructing the cross-validation function, rather than only one pair. This '*leave one curve out*' technique agrees with the points of view given in related settings by Rice and Silverman (1991) and by Hart and Wehrly (1992). The next theorem states asymptotic optimality of this bandwidth selection criterion with respect to the following L_2 measure of errors

$$\mathrm{MISE}(h)=E\int(\hat{m}_1(t)-m_1(t))^2 w(t)\,\mathrm{d}t.$$

In order to simplify the computations, we will make the classical assumption (which is reasonable in the light of Theorem 3(i)):

$$h\in H_N \text{ where } H_N=[a_1 N^{-1/5-\delta};\ a_2 N^{-1/5+\delta}] \quad \text{for some } \delta,\ 0<\delta<1/5. \tag{3.12}$$

Theorem 4. *Assume that conditions* (2.1), (3.1), (3.2), (3.8) *and* (3.12) *hold. Then the bandwidth* $\hat{h}=\arg\inf_{h\in H_N}\mathrm{CV}(h)$ *satisfies*

$$\inf_{h\in H_N}\mathrm{MISE}(h)/\mathrm{MISE}(\hat{h})\to 1, \textit{ a.s.}, \quad \textit{as } N\to\infty.$$

Let us denote by $\tilde{m}_1$ the estimate $\hat{m}_1$ for which we use $\hat{h}$ as bandwidth. Denoting by

$$Z_i^j=Y_i^j-\tilde{m}_1(t_i^j),$$

we choose as bandwidth the minimizer of the following individual cross-validation score function

$$\mathrm{CV}_j(h)=\frac{1}{n_j}\sum_{i=1}^{n_j}(Z_i^j-\hat{m}_2^{j,-i}(t_i^j))^2 w(t_i^j).$$

Here $\hat{m}_2^{j,-i}$ is the usual '*leave one point out*' kernel estimate, i.e.,

$$\hat{m}_2^{j,-i}(t)=\hat{g}_2^{j,-i}(t)/\hat{f}_2^{j,-i}(t),$$

with

$$\hat{f}_2^{j,-i}(t)=\frac{1}{(n_j-1)h_j}\sum_{k\neq i} K\left(\frac{t-t_k^j}{h}\right),$$

and

$$\hat{g}_2^{j,-i}(t)=\frac{1}{(n_j-1)h_j}\sum_{k\neq i} Z_k^j K\left(\frac{t-t_k^j}{h}\right).$$

Introducing as a measure of errors the quantity

$$\mathrm{MISE}_j(h)=E\int(\hat{m}_2^j(t)-m_2^j(t))^2 w(t)\,\mathrm{d}t,$$

we have the following asymptotic optimality property over the set of bandwidths

$$H_j=[b_1 n_j^{-1/5-\delta}; b_2 n_j^{-1/5+\delta}]. \tag{3.13}$$

Theorem 5. *Assume that conditions* (2.1), (3.1), (3.2), (3.8), (3.9), (3.12) *and* (3.13) *hold. For each j, the bandwidth*

$$\hat{h}_j=\arg\inf_{h\in H_j}\mathrm{CV}_j(h)$$

satisfies

$$\inf_{h\in H_j}\mathrm{MISE}_j(h)/\mathrm{MISE}_j(\hat{h}_j)\to 1,\ a.s.,\quad as\ N\to\infty.$$

These five theorems will be proved in Section 5. Let us close this section by mentioning possible extensions of our results. These generalizations are easy to state by using classical nonparametric regression techniques, but they are not given in this paper in order to save space and not to give long and tedious notations and calculations. For instance, the rates of convergence in Theorems 1–3 can be improved using higher-order kernels and stronger smoothness assumptions, as done by Gasser and Müller (1979) in regression setting. Also the cross-validation procedure can be improved by allowing the bandwidths h and h_j to be location adaptive (i.e., depending on t) as in Vieu (1991a) for regression. Also the conditions (3.12) and (3.13) could be suppressed and the minimizations can be done over all the set of positive real numbers by using techniques described in Härdle and Kelly (1987). Another possible extension is to state similar results without assuming independence between the data (see e.g. Györfi et al. (1989)). In place of presenting such obvious extensions, we prefer to illustrate our two-stage method through some real data set.

4. An illustration

The data set consists of $N=1107$ measurements of height (in cm). There are exactly 41 measurements for each of $J_N=27$ individuals (14 girls and 13 boys) from birth to

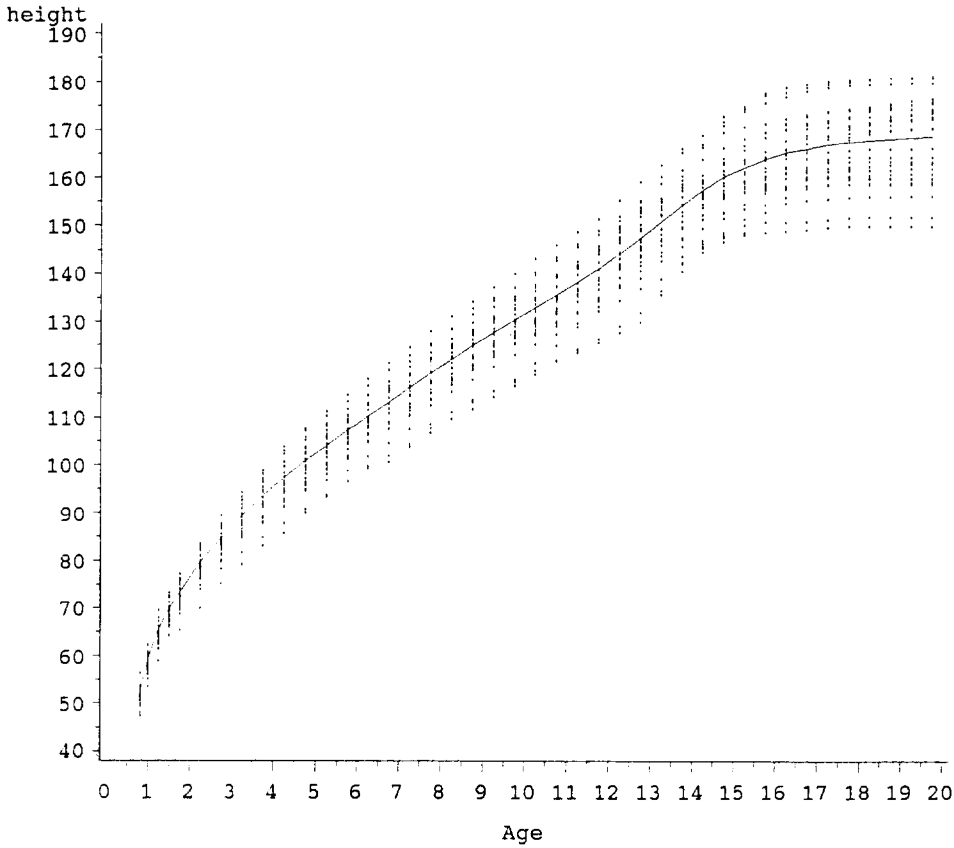

Fig. 1. Height (in cm) against age (in years). Kernel estimate of the mean curve $m_1(t)$.

adult age (near 19 years). The cloud of data points is plotted in Fig. 1 together with kernel estimation of m_1. Our kernel estimate of m_1 is exactly the average height at each time, since the discretization phenomenon (see Section 1) leads to a set of balanced data. The bandwith obtained by cross-validation was the shortest period between two observations (i.e., 0.17). We can see 'more or less' the classical general form of human growth with a spurt at puberty, followed by a flat. This does not appear very clearly here, because spurt depends on the time of puberty, and so it differs between males and females. Females begin their acceleration sooner than males and so, in the mean curve of Fig. 1, it is difficult to locate acceleration. Note that our goal was not to define a mean curve m_1 as interpretable as possible, but it was to introduce it as a first stage of our estimation procedure. This will be made clear later in the discussion of Fig. 3.

Then, we computed individual kernel estimates to see how each individual differs from the mean curve. The curves are presented in Fig. 2. The bandwidths were selected

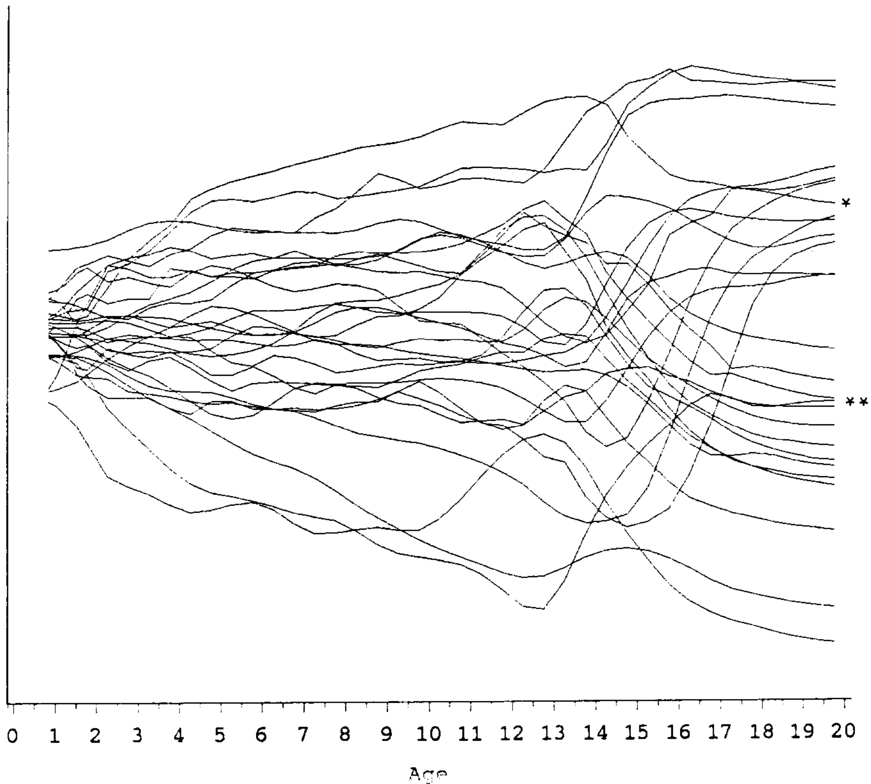

Fig. 2. Kernel estimates of the 27 individual curves $m_2^j(t)$. ($*$) curve of the individual No. 6; ($**$) curve of the individual No. 22.

using the procedure described in Theorem 5 and the 27 selected h_j were all between 0.55 and 1.45. At first glance, we see clearly four groups according to final sizes. Roughly, they are: high males, males, females and small females. A deeper examination of these plots reveals two kinds of curves: (1) these with an increase and then a decrease (females) and (2) with a big spurt (males). This phenomenon appears more clearly in Fig. 3, where these curves are presented in two groups according to the sex.

The observation of Figs. 2 and 3 reveals two individuals with special behaviours. Individual no. 22 is a male but his final height is about the mean adult height of the female group. On the contrary individual no. 6 is a female with final size like a male. The plot shows us that no. 22's curve has the same general form of males group upto a gap all over the period, confirming the fact that he belongs to the male group. Similarly, this plot shows that individual no. 6 has a female growth.

Finally, in order to show the accuracy of our additive two-stage method, we compared our results with those of a direct kernel method derived from model (1.1).

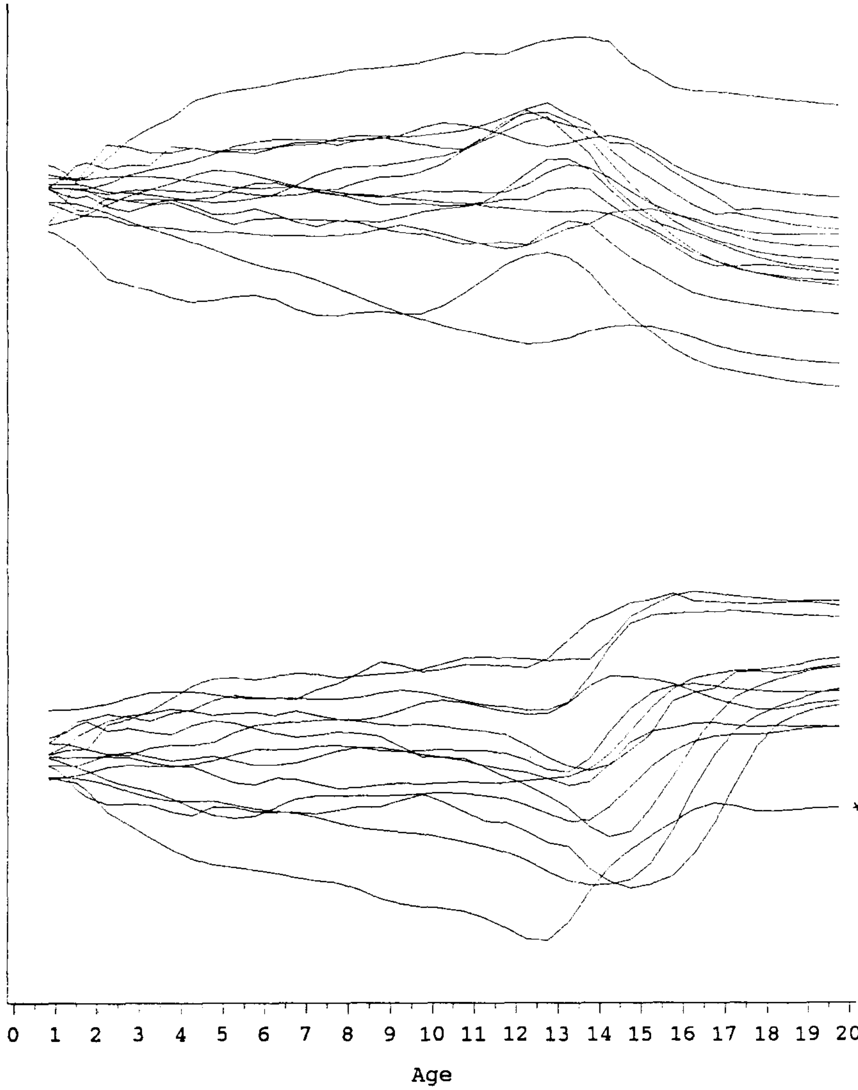

Fig. 3. Kernel estimates of the 27 individual curves $m_2^j(t)$ grouped according to sex. Above: females; below: males. ($*$) curve of the individual No. 6; ($**$) curve of the individual No. 22.

We computed the individual direct kernel estimates of m^j, with bandwidth obtained by ordinary regression cross-validation (Härdle and Marron, 1985). Because of the quite smooth aspect of human growth and because each individual estimation is performed with the same sample size, the selected bandwidth was constant ($=0.55$) over all the 27 individuals. We have drawn the corresponding 27 individual plots, and two phenomena appeared. Firstly, we noted that for both methods (two-stage and

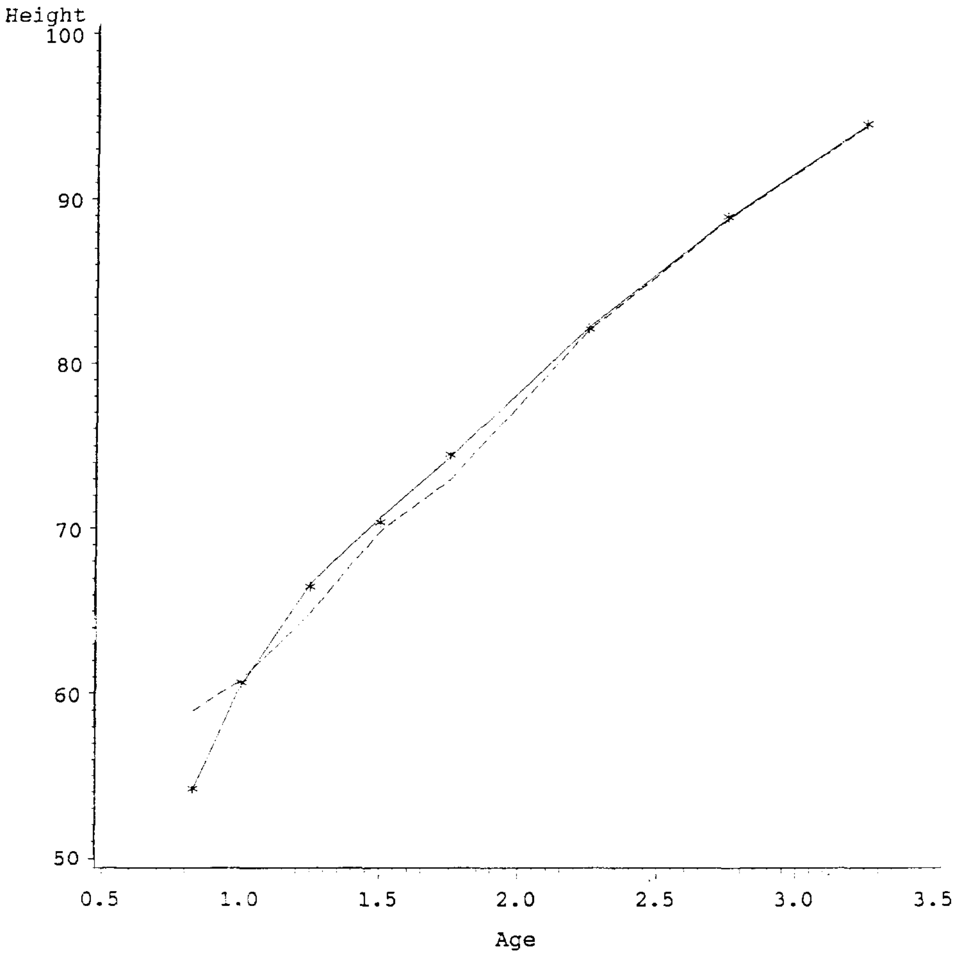

Fig. 4. Direct estimated curve $\hat{m}^j$ (dashed line) and two-stages estimated curve $\hat{m}^1 + \hat{m}_2^j$ (solid line) for the individual number 6.

direct methods) each individual curve looked roughly similarly smoothed. Secondly, we noted that main difference between both curves comes from small ages (i.e., $T \leqslant 5$ years) while both estimates are almost perfectly coinciding for $T \geqslant 10$. To highlight this fact, we present both curves in Fig. 4 (for one individual and for $0 < T < 5$). Same plots for other individuals look similar and are presented (all together) in Fig. 5.

We computed for each individual and for each method the residual sum of squares (RSS) and also the residual absolute sum (RSA). Among the 27 individuals the two-stage method provided smaller RSS (RSA) for 27 (25) individuals. In average over the sample, the RSA was around 0.1 for the two-stage method, while it was 0.2 for the direct method. This improvement comes mainly from early years, i.e. for $0 < T < 5$.

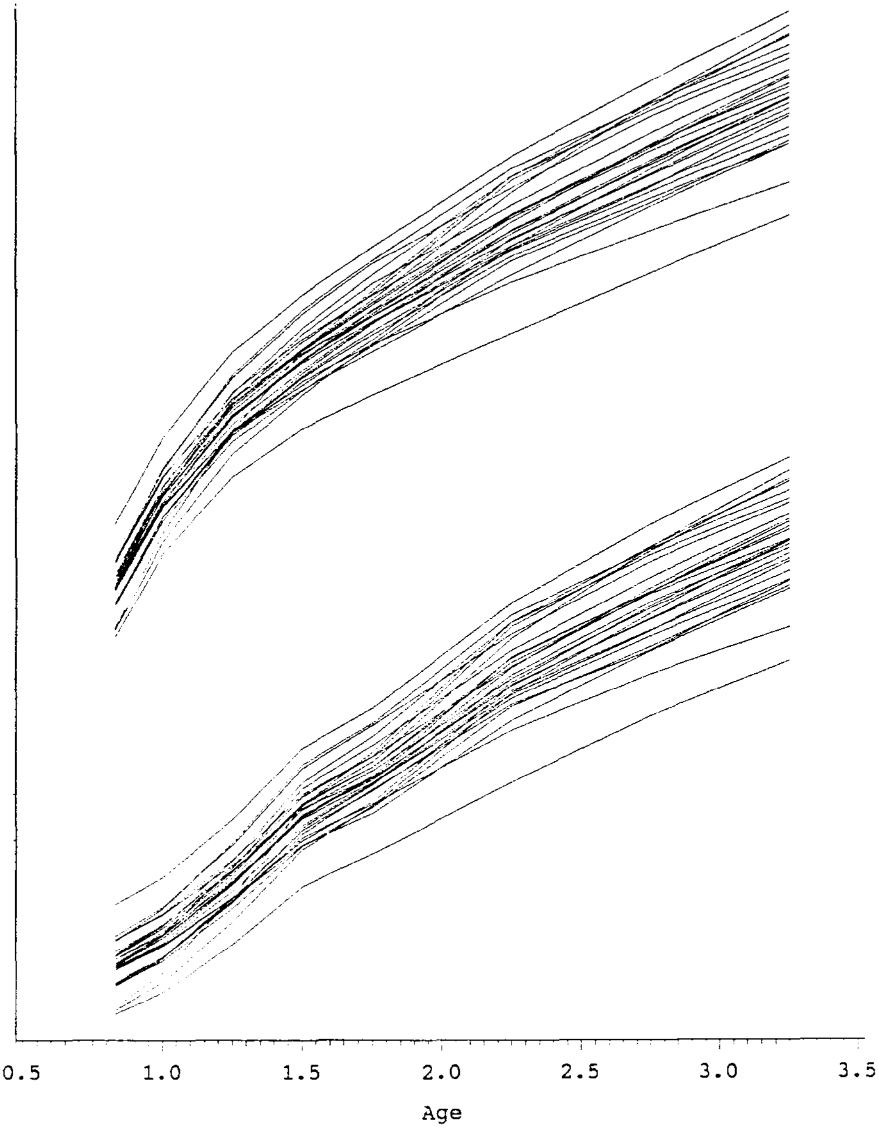

Fig. 5. Direct estimated curve $\hat{m}^j$ (below) and two-stages estimated curve $\hat{m}^1+\hat{m}_2^j$ (above) for all the 27 individuals.

Two-stage estimation gives smaller residuals for early years, without being undersmoothed. So an interesting feature is that, at least for this example, our two-stage procedure tends to reduce the well-known boundary effects. This can be empirically explained as follows. Estimating m_1 provides small boundary errors because a small bandwidth (0.17) is used since we have at hand large data (1107 exactly). Then in estimating the individual component, even if large bandwiths are used, we do not have large boundary error since data points have small amplitude compared with those used in estimating directly m^j. In other words, the main boundary error comes from the fact that the curves are increasing, and the two-stage method estimates this general increasing tendancy with a smaller bandwidth than the direct one does. Of course, such an improvement cannot be observed for old ages since in this area the curves are very flat and, therefore, boundary effects are not a real problem. Our algorithm was implemented on Fortran 77 on a PC. All the plots were produced using SAS software V.04. Further details, about the program as well as about this particular data set are available on request.

5. Proofs of theorems

Proof of Theorem 1(i). Let us denote $g_1 = m_1 f$ and $g_2^j = m_2^j f$. Under the conditions (3.1) and (3.2) we have (see e.g. Collomb (1984))

$$\sup_t |\hat{m}_1(t) - m_1(t)| w(t) = \mathrm{O}\left(\sup_t |\hat{g}_1(t) - g_1(t)| w(t) \right) + \mathrm{O}\left(\sup_t |\hat{f}_1(t) - f_1(t)| w(t) \right), \text{ a.s.} \tag{5.1}$$

Well-known results in density estimation (see e.g. Györfi et al. 1989, Theorem 5.3.1) allow us to write

$$\sup_t |\hat{f}_1(t) - f_1(t)| w(t) = \mathrm{O}(h^2) + \mathrm{O}\left(\sqrt{\frac{\log N}{Nh}} \right) \text{ a.s.} \tag{5.2}$$

On the other hand, formula (3.3.4) of Györfi et al. (1989) gives

$$\sup_t |\hat{g}_1(t) - E\hat{g}_1(t)| w(t) = \mathrm{O}\left(\sqrt{\frac{\log N}{Nh}} \right) \text{ a.s.} \tag{5.3}$$

Let us now study the bias of $\hat{g}_1(t)$, when we consider it as an estimate of $g_1 = m_1 f$. We can write

$$E\hat{g}_1(t) = \mathrm{I} + \mathrm{II}, \tag{5.4}$$

where

$$\mathrm{I} = h^{-1} \int m_1(u) K((t-u)/h) f(u) \, \mathrm{d}u$$

and

$$\mathrm{II} = \frac{1}{Nh} \sum_{i=1}^{J_N} \int m_2^j(u) K((t-u)/h) f(u) \, \mathrm{d}u.$$

Taylor expansion gives

$$\mathrm{I}=g_1(t)+\frac{h^2}{2}g_1''(t)\int v^2K(v)\,\mathrm{d}v+\mathrm{o}(h^2). \tag{5.5}$$

Similarly, we have

$$\begin{aligned}\mathrm{II}&=\frac{1}{N}\sum_{i=1}^{J_N}\sum_{i=1}^{n_j}\left[g_2^j(t)+\frac{h^2}{2}g_2^{j''}(t)\int v^2K(v)\,\mathrm{d}v\right]+\mathrm{o}(h^2)\\&=\frac{1}{N}\sum_{j=1}^{J_N}n_j\left[g_2^j(t)+\frac{h^2}{2}g_2^{j''}(t)\int v^2K(v)\,\mathrm{d}v\right]+\mathrm{o}(h^2)\\&=f(t)\frac{1}{J_N}\sum_{j=1}^{J_N}m_2^j(t)+\sum_{j=1}^{J_N}\frac{n_j-n}{N}g_2^j+\frac{h^2}{2N}\sum_{j=1}^{J_N}n_jg_2^{j''}(t)\int v^2K(v)\,\mathrm{d}v+\mathrm{o}(h^2).\end{aligned}$$

The first term on the right-hand side of this last equality is treated using (2.1) and the strong law of large numbers. The other terms are treated by using (3.1) and (3.8). So we have

$$\mathrm{II}=\frac{h^2}{2N}\sum_{j=1}^{J_N}n_jg_2^{j''}(t)\int v^2K(v)\,\mathrm{d}v+\mathrm{O}\left(\sqrt{\frac{\log\log J_N}{J_N}}\right)+\mathrm{o}(h^2),\ \text{a.s.} \tag{5.6}$$

Finally, we get from (5.4)–(5.6) that

$$\begin{aligned}E\hat{g}_1(t)=g_1(t)&+\frac{h^2}{2}\left[g_1''(t)+\frac{1}{N}\sum_{j=1}^{J_N}n_jg_2^{j''}(t)\right]\int v^2K(v)\,\mathrm{d}v\\&+\mathrm{O}\left(\sqrt{\frac{\log\log J_N}{J_N}}\right)+\mathrm{o}(h^2).\end{aligned} \tag{5.7}$$

Note that (because of (3.1) and (3.2)) the o and the O are uniform over t belonging to the support of w. So, Theorem 1(i) follows directly from (5.1)–(5.3) and (5.7).

Proof of Theorem 2(i). We will just sketch this proof since it follows the same ideas as before. Well-known results in L_1 density estimation [see e.g. Theorems 7.3 and 7.6 in Devroye (1986)] give

$$E\int|\hat{f}(t)-f(t)|w(t)\,\mathrm{d}t=\mathrm{O}(h^2)+\mathrm{O}\left(\sqrt{\frac{1}{Nh}}\right). \tag{5.8}$$

Note that similar results are given with more generality by Holmström and Klemelä (1992). Similarly, usual results in L_1 regression (see e.g. Krzyzak (1985) and Wand (1990)) give:

$$E\int|\hat{g}_1(t)-E\hat{g}_1(t)|w(t)\,\mathrm{d}t=\mathrm{O}\left(\sqrt{\frac{1}{Nh}}\right). \tag{5.9}$$

Now, using (3.2) and (5.7) we get

$$\int |E\hat{g}_1(t)-g_1(t)|w(t)\,\mathrm{d}t=\mathrm{O}(h^2)+\mathrm{O}\left(\sqrt{\frac{\log\log J_N}{J_N}}\right),\ \text{a.s.} \tag{5.10}$$

Finally, the claimed result follows from (5.8)–(5.10).

Proof of Theorem 3(i). Because of (5.2) we have

$$E\int(\hat{m}_1(t)-m_1(t))^2 w(t)\,\mathrm{d}t=E(\hat{m}_1(t)-m_1(t))^2(\hat{f}_1(t)/f(t))^2 w(t)\,\mathrm{d}t$$
$$+\mathrm{o}(E\int(\hat{m}_1(t)-m_1(t))^2 w(t)\,\mathrm{d}t). \tag{5.11}$$

On the other hand, we have

$$E\int(\hat{m}_1(t)-m_1(t))^2(\hat{f}_1(t)/f(t))^2 w(t)\,\mathrm{d}t=\mathrm{III}+\mathrm{IV}, \tag{5.12}$$

where

$$\mathrm{III}=\int E[\hat{g}_1(t)-m_1(t)\hat{f}_1(t)-E(\hat{g}_1(t)-m_1(t)\hat{f}_1(t)]^2 w(t)f(t)^{-2}\,\mathrm{d}t$$

and

$$\mathrm{IV}=\int[E(\hat{g}_1(t)-m_1(t)\hat{f}_1(t))]^2 w(t)f(t)^{-2}\,\mathrm{d}t.$$

Proceeding as to get (5.7), we can show that we have

$$E\hat{f}_1(t)=f(t)+\frac{h^2}{2}f''(t)\int v^2K(v)\,\mathrm{d}v+\mathrm{o}(h^2),$$

and this result together with (5.7) leads to

$$\mathrm{IV}=Bh^4+\mathrm{O}\left(\sqrt{\frac{\log\log J_N}{J_N}}\right)+\mathrm{o}(h^4), \tag{5.13}$$

where

$$B=\frac{1}{4}\left[\int v^2K(v)\,\mathrm{d}v\right]^2\int\left[g_1''(t)+\frac{1}{N}\sum_{j=1}^{J_N}n_j g_2^{j''}(t)-f''(t)m_1(t)\right]^2$$
$$\times w(t)f(t)^{-2}\,\mathrm{d}t.$$

Let us now compute III. Using first the fact that the pairs (Y_i^j, t_i^j) are i.i.d. and then using (5.13), we can write

$$E[\hat{g}_1(t)-m_1(t)\hat{f}_1(t)-E(\hat{g}_1(t)-m_1(t)\hat{f}_1(t))]^2=$$

$$=\frac{1}{N^2h^2}\sum_{j=1}^{J_N} n_j E[Y_1^j-m_1(t)]^2 K^2((t-t_1^j)/h)$$

$$+\frac{1}{Nh^2}[E(Y_1^j-m_1(t))K((t-t_1^j)/h)]^2$$

$$=\frac{1}{N^2h^2}\sum_{j=1}^{J_N} n_j E[Y_1^j-m_1(t)]^2 K^2((t-t_1^j)/h)+\mathrm{o}(h^4).$$

Using now Taylor expansion and proceeding as we did for (5.6) we finally have

$$E[\hat{g}_1(t)-m_1(t)\hat{f}_1(t)-E(\hat{g}_1(t)-m_1(t)\hat{f}_1(t))]^2=$$

$$=\frac{1}{N^2h}\sum_{j=1}^{J_N} n_j E[((Y_1^j)^2|t_i^j=t)-m_1(t)^2-2m_1(t)m_2^j(t)]f(t)\int K^2(v)\,\mathrm{d}v$$

$$+\mathrm{o}\left(\frac{1}{Nh}\right)$$

$$=\frac{1}{Nh}E[((Y_1^1)^2|t_1^1=t)-m_1(t)^2]f(t)\int K^2(v)\,\mathrm{d}v+\mathrm{o}\left(\frac{1}{Nh}\right).$$

So, we have

$$\mathrm{III}=\frac{V}{Nh}+\mathrm{o}\left(\frac{1}{Nh}\right),$$

where

$$V=\int K^2(v)\,\mathrm{d}v\int E[((Y_1^1)^2|t_1^1=t)-m_1(t)^2]\frac{w(t)}{f(t)}\,\mathrm{d}t. \tag{5.14}$$

Finally, the claimed result follows from (5.11)–(5.14).

Proof of Theorem 1(ii). Let us first state a preliminary result that will be useful later. We have for some finite constant C

$$|\hat{m}_2^j(t)-\ddot{m}_2^j(t)|<C\sup_i\left|m_1(t_i^j)-\hat{m}_1(t_i^j)\right|\frac{1}{n_jh_j}\sum_{i=1}^{n_j}\left|K((t-t_i^j)/h_j)\right|,$$

where

$$\ddot{m}_2^j(t)=\sum_{i=1}^{n_j}(Y_i^j-m_1(t_i^j))K((t-t_i^j)/h)\Bigg/\sum_{i=1}^{n_j}K((t-t_i^j)/h).$$

Since the last term on the right-hand side of this inequality is just kernel estimate of f (with new kernel $|K|$) we get from Theorem 1(i)

$$\sup_t |\hat{m}_2^j(t) - \ddot{m}_2^j(t)| w(t) = \mathrm{O}\left(h^2 + \sqrt{\frac{\log N}{Nh}} \right) + \mathrm{O}\left(\sqrt{\frac{\log \log J_N}{J_N}} \right), \text{ a.s.}$$

Finally, we get from (3.7)–(3.9)

$$\sup_t |\hat{m}_2^j(t) - \ddot{m}_2^j(t)| w(t) = \mathrm{o}\left(h_j^2 + \sqrt{\frac{1}{n_j h_j}} \right), \text{ a.s.} \tag{5.15}$$

Let us now consider $\ddot{m}_2^j$. In fact, $\ddot{m}_2^j$ is just a kernel estimate of the regression function $E[(Y_1^j - m_1(t_1^j)) | t_1^j] = m_2^j(t_1^j)$. Therefore, well-known L_∞ results in regression estimation (see e.g. Györfi et al. (1989, Theorem 3.3.2)) lead to

$$\sup_t |\ddot{m}_j^2(t) - m_2^j(t)| w(t) = \mathrm{O}(h_j^2) + \mathrm{O}\left(\sqrt{\frac{\log n_j}{n_j h_j}} \right), \text{ a.s.} \tag{5.16}$$

Finally, (5.15) and (5.16) complete this proof.

Proof of Theorem 2(ii). Because of (5.15) above (similarly) we have

$$E \int |\hat{m}_2^j(t) - m_2^j(t)| w(t)\, \mathrm{d}t = E \int |\ddot{m}_2^j(t) - m_2^j(t)| w(t)\, \mathrm{d}t + \mathrm{o}\left(h_j^2 + \sqrt{\frac{1}{n_j h_j}} \right),$$

and as before, $\ddot{m}_2^j$ being a classical kernel estimate of m_2^j, we have

$$E \int |\ddot{m}_2^j(t) - m_2^j(t)| w(t)\, \mathrm{d}t = \mathrm{O}(h_j^2) + \mathrm{O}\left(\sqrt{\frac{1}{n_j h_j}} \right),$$

and this proof is complete.

Proof of theorem 3(ii). Using the decomposition

$$\begin{aligned} E \int (\hat{m}_2^j(t) - m_2^j(t))^2 w(t)\, \mathrm{d}t &= E \int (\ddot{m}_2^j(t) - m_2^j(t))^2 w(t)\, \mathrm{d}t \\ &+ E \int (\ddot{m}_2^j(t) - \hat{m}_2^j(t))^2 w(t)\, \mathrm{d}t + 2E \int (\ddot{m}_2^j(t) - m_2^j(t))(\ddot{m}_2^j(t) \\ &- \hat{m}_2^j(t)) w(t)\, \mathrm{d}t, \end{aligned}$$

and using again (5.15) we get

$$\begin{aligned} E \int (\hat{m}_2^j(t) - m_2^j(t))^2 w(t)\, \mathrm{d}t &= E \int (\ddot{m}_2^j(t) - m_2^j(t))^2 w(t)\, \mathrm{d}t \\ &+ \mathrm{o}\left(h_j^4 + \frac{1}{n_j h_j} \right). \end{aligned} \tag{5.17}$$

Now, treating the bias part of $\ddot{m}_2^j$ by using results on L_2 regression estimation (see e.g. Vieu (1991b, Section 6)) we have

$$\int E[\ddot{m}_2^j(t) - E(\ddot{m}_2^j(t))]^2 w(t)\,\mathrm{d}t = B^j h_j^4 + o(h_j^4), \tag{5.18}$$

where

$$B^j = \frac{1}{4}\left[\int v^2 K(v)\,\mathrm{d}v\right]^2 \int [g_2^{j''}(t) - f''(t) m_2^j(t)]^2 w(t) f(t)^{-2}\,\mathrm{d}t.$$

The variance component of $\ddot{m}_2^j$ is treated by using formula 4.22 in Collomb (1976), and we get

$$\int E[\ddot{m}_2^j(t) - m_2^j(t)]^2 w(t)\,\mathrm{d}t = \frac{V^j}{n_j h_j} + \mathrm{o}\left(\frac{1}{n_j h_j}\right), \tag{5.19}$$

where

$$V^j = \int K^2(v)\,\mathrm{d}v \int [E[(Y_1^1)^2 | t_1^1 = t] - m_1(t)^2 - m_2^j(t)^2] \frac{w(t)}{f(t)}\,\mathrm{d}t.$$

Our result follows now from (5.17)–(5.19).

Proof of Theorem 4. Let us denote

$$\mathrm{ASE}(h) = \frac{1}{N} \sum_{j=1}^{J_N} \sum_{i=1}^{n_j} (m_1(t_i^j) - \hat{m}_1(t_i^j))^2 w(t_i^j),$$

and

$$\mathrm{ASE}^0(h) = \frac{1}{N} \sum_{j=1}^{J_N} \sum_{i=1}^{n_j} (m_1(t_i^j) - \hat{m}_i^{-j}(t_i^j))^2 w(t_i^j).$$

It follows directly from the definitions of $\hat{m}_1$ and m_1^{-j} (see Härdle and Marron (1985) for details) that ASE and ASE^0 are asymptotically equivalent over $h \in H_n$. On the other hand, because of results in Marron and Härdle (1986), ASE and MISE are also asymptotically equivalent over $h \in H_n$. Therefore, Theorem 4 will be proved as long as we show

$$\mathrm{CV}(h) = \mathrm{ASE}^0(h) + C + \mathrm{CT}(h), \tag{5.20}$$

where C is a finite constant independent of h and where

$$\sup_{h \in H_n} [\mathrm{CT}(h)/\mathrm{MISE}(h)] = \mathrm{o}(1), \text{ a.s.} \tag{5.21}$$

To get the decomposition (5.20) it suffices to define C by

$$C = \frac{1}{N} \sum_{j=1}^{J_N} \sum_{i=1}^{n_j} (Y_i^j - m_1(t_i^j))^2 w(t_i^j),$$

and

$$\mathrm{CT}(h) = \frac{2}{N} \sum_{j=1}^{J_N} \sum_{i=1}^{n_j} (Y_i^j - m_1(t_i^j))(m_1(t_i^j) - \hat{m}_1^{-j}(t_i^j)) w(t_i^j).$$

Note that CT can be written as

$$\mathrm{CT}(h)=\frac{2}{N}\sum_{j=1}^{J_N}\sum_{i=1}^{n_j}\Gamma_i^j(m_1(t_i^j)-\hat{m}_1^{-j}(t_i^j))w(t_i^j)$$

where Γ_i^j is such that

$$E(\Gamma_i^j|t_i^j)=0. \tag{5.22}$$

Note that Lemma 4 in Härdle and Marron (1985) treats the same quantity but with ε_i^j in place of Γ_i^j. Looking at their proof, it appears that the only property of ε_i^j they use is the fact that these random variables satisfy (5.22). Therefore, the proof of their result is also valid in our situation and we get

$$\sup_{h\in H_n}\left[\mathrm{CT}(h)\Big/\left(h^4+\frac{1}{nh}\right)\right]=\mathrm{o}(1),\text{ a.s.} \tag{5.23}$$

Finally, (5.21) follows from (5.23) together with Theorem 3(i), and the proof is complete.

Proof of Theorem 5. Let us denote

$$\mathrm{CV}_j^0(h)=\frac{1}{n_j}\sum_{i=1}^{n_j}(W_i^j-\hat{m}_2^{j,-i}(t_i^j))^2w(t_i^j),$$

where

$$W_i^j=Y_i^j-m_1(t_i^j).$$

Our theorem is valid when CV_j is changed into CV_j^0 because of Theorem 1 in Härdle and Marron (1985). Therefore, all we have to prove is that for some finite quantity D (D is independent of $h\in H_j$) we have

$$\sup_{h\in H_j}|\mathrm{CV}_j(h)-\mathrm{CV}_j^0(h)+D|/\mathrm{MISE}_j(h)=\mathrm{o}(1),\text{ a.s.} \tag{5.24}$$

We have

$$\mathrm{CV}_j(h)-\mathrm{CV}_j^0(h)=T_1(h)+2T_2+2T_3(h),$$

with

$$T_1(h)=\frac{1}{n_j}\sum_{i=1}^{n_j}(m_1(t_i^j)-\tilde{m}_1(t_i^j))^2w(t_i^j),$$

$$T_2=\frac{1}{n_j}\sum_{i=1}^{n_j}\varepsilon_i^j(m_1(t_i^j)-\tilde{m}_1(t_i^j))w(t_i^j),$$

and

$$T_3(h)=\frac{1}{n_j}\sum_{i=1}^{n_j}(\hat{m}_2^j(t_i^j)-m_2^j(t_i^j))(m_1(t_i^j)-\tilde{m}_1(t_i^j))w(t_i^j).$$

From Theorems 3(ii) and 1(i) it follows that

$$\sup_{h \in H_n} |T_1(h)| / \mathrm{MISE}_j(h) = \mathrm{o}(1), \text{ a.s.} \tag{5.25}$$

while Theorems 3(ii), 1(i) and 1(ii) give us

$$\sup_{h \in H_n} |T_3(h)| / \mathrm{MISE}_j(h) = \mathrm{o}(1), \text{ a.s.} \tag{5.26}$$

Choosing $D = -2T_2$ and using (5.25) and (5.26), we finally get (5.24) and so this proof is complete.

Acknowledgements

The data set were kindly provided by Professors J. Pontier, M.O. Pernin and M. Sempé, and we would like to express our sincere gratitude to them. We also wish to thank an anonymous referee and Associate Editor for their advise and comments. Many discussions with Professor H. Caussinus were very helpful all along this work, and we would like to thank him very much.

References

Besse, P. and J.O. Ramsay (1986). Principal component analysis of sampled functions. *Psychometrika* **51**, 285–311.

Caussinus, H. and L. Ferré (1992). Comparing the parameters of a model for several units by mean of principal component analysis. *Comput. Statist. Data Anal.* **13**, 269–280.

Collomb, G. (1976). Estimation non paramétrique de la régression, Thesis. University P. Sabatier, Toulouse, France.

Collomb, G. (1981). Estimation non paramétrique de la régression: revue bibliographique. *Internat. Statist. Rev.* **49**, 75–93.

Collomb, G. (1984). Propriétés de convergence presque complète du prédicteur à noyau, *Z. Wahrsch. und verw. Gebiete* **66**, 441–460.

Devroye, L. (1986). A Course in Density Estimation. Birkhauser.

Fan, J. (1992). Design-adaptive nonparametric regression. *J. Amer. Statist. Assoc.*, **87**, 998–1004.

Gasser, T. and E. Herrmann (1991). Data-adaptive kernel estimation. In: G.G. Roussas, Ed., *Nonparametric Functional Estimation and Related Topics.* Kluwer Academic Publishers, Dordrecht, 67–80.

Gasser, T. and A. Kneip (1991). Analysis of samples of curves. In: G.G. Roussas, Ed., *Nonparametric Functional Estimation and Related Topics.* Kluwer Academic Publishers, Dordrecht, 99–110.

Gasser, T. and H.G. Müller (1979). Kernel estimation of regression functions. In: Gasser and Rosenblatt, Eds., *Smoothing Techniques for Curve Estimation*, Springer, Heidelberg, 23–68.

Gasser, T., H.G. Müller, W. Kölher, L. Molinari and A. Prader (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12**, 210–219.

Grizzle, J.E. and D.M. Allen (1969). Analysis of growth and dose response curves. *Biometrics* **25**, 357–381.

Györfi, L., W. Härdle, P. Sarda and P. Vieu (1989). *Nonparametric Curve Estimation from Time Series.* Lectures Notes in Statistics, Vol. 60, Springer, Berlin.

Härdle, W. (1990). *Applied Nonparametric Regression.* Oxford Univ. Press, Boston.

Härdle, W. and A. Bowman (1987). Bootstrapping in nonparametric regression. *J. Amer. Statist. Assoc.* **83**, 127–141.

Härdle, W. and G. Kelly (1987). Nonparametric kernel regression estimation. Optimal bandwidth choice. *Statistics* **18**, 21–35.

Härdle, W. and J.S. Marron (1985). Optimal bandwidth selection in nonparametric regression funcation estimation. *Ann. Statist.* **13**, 1465–1481.

Härdle, W. and P. Vieu (1991). Kernel regression smoothing of time series. *J. Time Ser. Anal.* **13**, 209–232.

Hart, J.D. and T.E. Wehrly (1992). Consistency of cross-validation when data are curves. *Stochastic Process Appl.*, to appear.

Hölmstrom, L. and J. Klemelä (1992). Asymptotic bounds for the expected L_1 error of a multivariate kernel density estimator. *J. Multivariate Anal.* **42**, 245–266.

Houllier, F. (1987). Comparaison de courbes et de modeles de croissance: choix d'une distance entre individus. *Statistique et Analyse des Donnėes* **12**, 17–36.

Jennen-Steinmetz, C. and H.G. Gasser (1988). A unifying approach to nonparametric regression estimation. *J. Amer. Statist. Assoc.* **83**, 1084–1089.

Krzyzak, A. (1985). The rates of convergence of kernel regression estimates and classification rules. *IEEE Trans. Inform. Theory* **32**, 668–679.

Laird, N.M., J.H. Ware (1982). Random effects models for longitudinal data *Biometrics* **38**, 963–974.

Lindstrom, M.J. and D.M. Bates (1990). Nonlinear mixed effects models for repeated measures data. *Biometrics* **46**, 673–687.

Mack, Y.P. and H.G. Müller (1989a). Convolution type estimators for nonparametric regression. *Statist. Probat. Let.* **7**, 229–239.

Mack, Y.P. and H.G. Müller (1989b). Derivative estimation in nonparametric regression with random predictor variable. *Sankhya Ser A* **51**, 59–72.

Marron, J.S. (1988). Automatic smoothing parameter selection: a survey. *Empirical Economics* **13**, 187–208.

Marron, J.S. and W. Härdle (1986). Random approximations of some measures accuracy in nonparametric curve estimation. *J. Multivariate Anal.* **20**, 91–113.

Potthof, R.R. and S.N. Roy (1964). A generalized multiple analysis of variance model useful especially for growth curve problems. *Biometrika* **51**, 313–326.

Rao, C.R. (1965) M. The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–458.

Rao, C.R. (1967). Least squares theory using an estimated dispersion matrix and its application to measurement of signals. *Proc. 5th Berkeley Symp. Math. Statist. Probability*, **1**, 355–372.

Rice, J. and B. Silverman (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc.* **53**, 223–243.

Stone, C. (1982). Optimal global rates in nonparametric regression. *Ann. Statist.* **10**, 1040–1053.

Vieu, P. (1991a). Nonparametric regression: optimal local bandwidth choice. *J. Roy. Statist. Soc.* **53**, 453–464.

Vieu, P. (1991b). Quadratic errors for nonparametric estimates under dependence. *J. Multivariate Anal.* **39**, 324–347.

Vieu, P. (1992). Bandwidth selection for kernel regression: a survey. "Computer Intensive Methods in Statistics" (N. Härdle and L. Simar eds) in Statistics and Computing, Vol 1. Physica Verlag. Berlin.

Wand, M. (1990). On exact L_1 rates of convergence in nonparametric kernel regression. *Scand. Statist.* **17**, 251–256.

Zeger, S.L. K. Liang and P. Albert (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics* **44**, 1049–1060.