ASYMPTOTIC AND BOOTSTRAP CONFIDENCE BOUNDS FOR THE STRUCTURAL AVERAGE OF CURVES[1]

BY KONGMING WANG[2] AND THEO GASSER

SUNY Health Science Center and University of Zürich

For analyzing samples of curves, Kneip and Gasser proposed a "structural analysis" for estimating an average curve which gives the average amplitude and dynamics of the sample curves. An important step of their method is the estimation of the warping functions in order to eliminate the differences in dynamics between different curves. It is of interest to compute confidence bounds for the structural average curve. First, we derive the necessary asymptotic results to obtain confidence intervals based on asymptotic normality. Due to the complex form of the asymptotic formulas and due to their asymptotic nature, bootstrap procedures are studied in a second step. A small archive SACI (Structural Averaging and Confidence Intervals), written in Fortran, can be obtained from the Web site "http://www.unizh.ch/biostat" to compute the structural average curve, to construct confidence regions at a structural point and to compute confidence bars at other points.

1. Introduction. Consider a nonparametric regression model for m subjects or experimental units:

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

Here the f_i are unknown curves recorded at times $t_{ij} \in J \equiv [a_0, a_1] \subset \mathbf{R}$. The ε_{ij} are independent error terms with mean $E(\varepsilon_{ij}) = 0$ and variance $\mathrm{var}(\varepsilon_{ij}) = \sigma_i^2 > 0$. Under the assumption that all $\{f_i\}$ are smooth and have a similar pattern, one wants to estimate an average curve which reflects the common structure of the individual curves. Such problems arise in studying some biological, chemical or physical process or any other development. One example is growth of humans or animals, where growth evolves at a different intensity and at a different pace in different individuals. Another example is speech signals, where the same words are spoken with varying loudness and speed.

The variation on the time axis makes the cross-sectional average $(f_1 + \cdots + f_m)/m$ inadequate as an average curve. Assume for (1) a model of the form $f_i(t) = s(g_i(t))$, where s is some "shape function" and $\{g_i\}$ is an identically independently distributed random sample of the developmental rates with $E(g_i(t)) = t$. Then by the law of large numbers, $(f_1 + \cdots + f_m)/m \to E(f_1) \neq s$ whenever s is nonlinear.

In order to cope with the unfamiliar variation on the time axis, Kneip and Gasser (1992) introduced "structural analysis." The central idea of the structural analysis is the estimation of individual warping functions $\{g_i\}$ which align individual curves to a common time scale. Naturally, these warping functions are required to be strictly increasing and smooth. This proceeds in the following steps (further details are given in Section 2).

1. The regression functions f_i and/or their derivatives are estimated nonparametrically.
2. Individual structural points are identified from the estimated functions. Roughly speaking, structural points are features that are common to all or most curves. Extremes and inflection points are examples for structural points.
3. Warping functions are constructed from structural points such that individual structural points are shifted to the respective average location. In-between smooth monotonic interpolation is used.
4. The structural average is then estimated by averaging the aligned curves.

This method proved to be successful for studying human growth for variables inaccessible so far because of their small size and relatively large residual variation [Gasser, Kneip, Binding, Prader and Molinari (1991) and Gasser, Kneip, Zieger, Molinari, Prader and Largo (1994)]. The success can be attributed to the possibility of pooling information over many subjects to derive the development of a "typical" child.

Even pediatricians acquainted with these results asked about the effect of random variability on the estimated structural average curve. The purpose of this paper is to describe methods for constructing confidence bounds for the structural average. In the case of a single curve, one can only assess the variations in amplitude caused by noise. If a sample of similar curves are observed, as discussed here, not only the variability in amplitude but also the variability in locating structural points can be estimated. This is intuitively obvious. Therefore the confidence bounds take two forms. At a structural point, a two-dimensional confidence region can be formed such that it contains the structural point with a specified probability. At a nonstructural point, only confidence bars for variation in amplitude are needed.

Two methods are presented, one based on the asymptotic normal approximation and the other on bootstrapping. One reason for studying asymptotic normality is that large samples are available in some applications so that the asymptotic normal approximation could work. Another reason is that asymptotic normality of a statistic is often a necessary and sufficient condition for bootstrapping [cf. Mammen (1992)]. To prove asymptotic normality we assume that the curves in the sample are fixed unknown curves. If the curves are considered to be an identically independently distributed sample from a function space, then the effects of noise can be ignored (cf. Section 4), and asymptotic normality holds under general conditions. Thus we will not discuss it in detail. It is shown that both the warping functions and the structural average curve are asymptotically normal with a proper scaling. Asymptotic bias and

variances are given in closed forms. In principle, these bias and variances could be estimated from the data by plug-in methods, leading to asymptotic confidence bounds. The complicated form makes this unrealistic and leads us to propose bootstrap methods.

Bootstrap confidence bounds for an unknown function (density or regression function) have been often studied [Härdle and Bowman (1988), Hall (1992a, b), Härdle and Mammen (1993), Härdle and Marron (1991) and others]. A focus of those methods is the correction of bias of nonparametric fitting since naive bootstrap has bootstrap bias zero. For a sample of curves, bias correction is even more important than for a single curve. This is essentially due to the fact that variance decreases with the sample size m but not bias.

When bootstrapping, we should again make a distinction, whether the f_i $(i = 1, \ldots, m)$ are to be considered as fixed unknown curves or as an i.i.d. random sample from some function space. In the first case, variations caused by the $\{\varepsilon_{ij}\}$ are bootstrapped. Thus, resampling properly estimated residuals works fine. In the second case, the leading term comes from random variation of $\{f_i\}$ while the effects of $\{\varepsilon_{ij}\}$ can be ignored to a first-order approximation. Naive bootstrapping of whole functions then works. See Section 4 for details, where a combination of the two approaches is also discussed.

Figure 1 depicts the results of an illustrative simulation based on 50 functions sampled at 100 equispaced points. The noise level is $\sigma = 1$. As to be
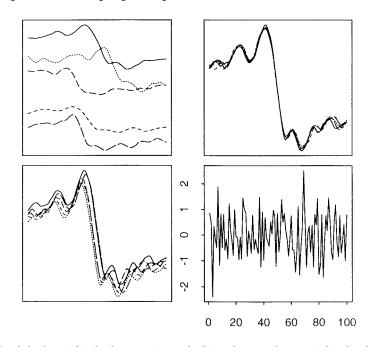

FIG. 1. *Top left*: *five individual curves (smoothed) in the sample*; *top right*: *five bootstrapped average curves by resampling estimated residuals*; *bottom left*: *five bootstrapped average curves by resampling curves with replacement* (*the range of the averages is about* −6 *to* 6); *bottom right*: 100 *estimated residuals from one curve data.*

expected, bootstrapping when assuming a random sample of functions leads to more variability between bootstrap realizations of the average curve. For more information on the simulation, see Section 4.3.

The paper is organized as follows: the structural analysis of Kneip and Gasser (1992) is summarized in Section 2. Results on asymptotic normal approximation are in Section 3. Bootstrap methods are presented in Section 4. Comments on proofs of the theorems are gathered in Section 5.

2. Structural analysis. The procedure outlined in Section 1 for aligning curves which are similar in structure—and usually created by one and the same basic process—has intuitive appeal. Elementary features (or landmarks) common to the curves are aligned to their average location; from this discrete function, a smooth warping function is created by smooth interpolation. Despite the intuitive appeal, a relatively complicated mathematical formalism is required which we will try to keep simple [see Kneip and Gasser (1992) for more details]. First we need some definitions.

DEFINITION 2.1. The function space $C^{(\nu)}(J)$ consists of all ν-times continuously differentiable functions on $J \equiv [a_0, a_1]$.

DEFINITION 2.2. The kernel estimator $\hat{f}_i$ of f_i, with kernel W and bandwidth b, is

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

where $s_{i, j} = (t_{ij+1} + t_{ij})/2$, $j = 1, \ldots, n_i - 1$, $s_{i, 0} = t_{i1} - (s_{i, 1} - t_{i1})$ and $s_{i, n_i} = t_{in_i} + (t_{in_i} - s_{i, n_i-1})$.

For simplicity we will use $\hat{f}_i^{(j)} = d^j \hat{f}_i / dt^j$ as derivative estimator, but a kernel estimator with special kernels could also be used. The bandwidth b depends on the order of the derivative. Also other smoothing methods like local polynomials or splines can be used without changing asymptotic results (note that everything is based on fixed design, where most methods are asymptotically equivalent).

The extraction of structural features can be viewed as a mapping from function space to the real line. This leads to the definition of structural functionals. For simplicity we will often concentrate on extrema of $f_i^{(\gamma)}$ associated with "*e*-functionals."

DEFINITION 2.3. A structural functional T is a mapping $T: \mathscr{D} \subset C^{(\nu)}(J) \mapsto (a_0, a_1)$. An *e*-functional of order $(\gamma + 1)$ locates an extrema of $f_i^{(\gamma)}$ (ordinary extrema are first order).

A further interesting class of features are characterized by "*p*-functionals" which locate a certain percentage of decrease (or increase) in a monotone segment (e.g., half lifetime).

DEFINITION 2.4. Assume that $f_i^{(\gamma)}$ is monotone on $[\alpha_i, \beta_i] \subset (a_0, a_1)$, where α_i and β_i are extremes of $f_i^{(\gamma)}$. For some $p \in (0, 1)$, a p-functional of order γ locates a point $p(f_i) \in (\alpha_i, \beta_i)$ such that

$$f_i^{(\gamma)}(p(f_i)) = pf_i^{(\gamma)}(\alpha_i) + (1 - p)f_i^{(\gamma)}(\beta_i).$$

REMARK 2.1. One can include the case where some extremum is not found in some function by assigning a "missing" site (a point not in $[a_0, a_1]$). This can be handled also in practical terms [Kneip and Gasser (1992)].

A basic and plausible assumption is some structural homogeneity of the class of functions $\mathcal{R} \subset C^{(\nu)}(J)$ considered. For structural functionals $T_r: \mathcal{D}_r \mapsto (a_0, a_1)$ $(r = 1, \ldots, l)$, let us assume the following:

(i) $\{f_i: i = 1, 2, \ldots\} \subset \mathcal{R} \subset \bigcap_{r=1}^l \mathcal{D}_r$;
(ii) T_r is continuous;
(iii) No missings are allowed [Kneip and Gasser (1992)].

For convenience, these functionals are numbered so that $T_r(f_i) < T_s(f_i)$ for $r < s$.

The introduction of warping functions needs some further notation and definitions. The application of structural functionals $T_1, \ldots, T_l$ to a regression function f_i leads to an ordered l-tuple $(x_{i1}, \ldots, x_{il})$ of real numbers. These locations of structural features form the basis for constructing warping functions, which are obtained by mapping these discrete values to a continuous function g_i via a warping operator G.

DEFINITION 2.5. Let $\mathcal{H}_l \equiv \{(x_1, \ldots, x_l) \in J^l: x_r < x_s \text{ for all } r < s\}$. A *warping operator* is a mapping $G: \mathcal{H}_l^2 \mapsto C^1(\mathbf{R})$ such that we have the following:

(i) For all $(u, x) \in \mathcal{H}_l^2$, $G_{(u, x)}(\cdot)$ is a strictly, monotonically increasing real function;
(ii) $G_{(u, x)}(x_r) = u_r$ for $r = 1, \ldots, l$.

Denote $\tau_{ir} = T_r(f_i)$ and $\tau = (\tau_{\cdot 1}, \ldots, \tau_{\cdot l})$ where $\tau_{\cdot r} = (\tau_{1r} + \cdots + \tau_{mr})/m$. Then the warping function for aligning f_i to the average time scale is defined by

$$g_i(\cdot) = G_{(\tau^i, \tau)}(\cdot),$$

where $\tau^i = (\tau_{i1}, \ldots, \tau_{il})$. The definition of G implies some freedom in constructing warping functions between consecutive structural points. While linear interpolation is in practice often good enough, interpolating splines obeying to monotonicity are the method of choice [Kelly and Rice (1990)].

REMARK 2.2. For periodic data, one would estimate a translation $g_i(t) = t + \delta_i$ (Silverman, 1995). Linear interpolation gives this translation since $(\tau_{\cdot(r+1)} - \tau_{\cdot r})/(\tau_{i(r+1)} - \tau_{ir}) = 1$ for all r in the translation case.

Estimating warping functions is then rather straightforward; estimates of structural points are obtained from kernel estimators using some bandwidth

b: $\hat{\tau}_{ir} = T_r(\hat{f}_i(\cdot; b))$. Note that the delicate step is to determine the number and meaning of structural functionals, which is something like "qualitative vague modeling." For details see Kneip and Gasser (1992) and Gasser and Kneip (1995). In the latter paper a method is described for determining consistent features directly from the data, instead of using just a priori knowledge.

The structural average curve

$$f(t) = \frac{1}{m} \sum_{i=1}^{m} f_i(g_i(t))$$

is then estimated by plugging in the estimated $\hat{g}_i$ followed by kernel estimation of f_i with some bandwidth c,

$$\hat{f}(t) = \frac{1}{m} \sum_{i=1}^{m} \hat{f}_i(\hat{g}_i(t); c).$$

Note that c has to be smaller than bandwidth b used for extracting features. This is because averaging regression functions reduces variance and b is involved in the estimation of the derivatives.

Consistency and convergence rates of the estimator $\hat{f}$ have been established by Kneip and Gasser (1992). Let μ_r be the order of the functional T_r, $\bar{\mu} = \max\{\mu_1, \ldots, \mu_l\}$, and $\underline{\mu} = \min\{\mu_1, \ldots, \mu_l\}$. Recall that for an e-functional, $\mu_r = 1$ means an extrema, $\mu_r = 2$ an inflection point, and so on. We assume the following five conditions [cf. Kneip and Gasser (1992)].

(A1) (Equally spaced design, equal number of observations per subject). Assume that $n_1 = \cdots = n_m = n$ and $t_{ij} = t_j = a_0 + j(a_1 - a_0)/n$ for $i = 1, \ldots, m$; $j = 1, \ldots, n$.

(A2) All moments of ε exist.

(A3) (Smooth kernel of order k). For some even integer $k \geq 2$, we require that:
   A3.1. W is symmetric, and $W(u) = 0$ for $|u| > 1$;
   A3.2. W is of order k.
   A3.3. W is $\bar{\mu} + 1$ times continuously differentiable on $\mathbf{R}$, and it is $\bar{\mu} + 2$ times continuously differentiable on $[-1, 1]$.

(A4) (Smoothness). Model (1) holds with $f_i \in C^{(\nu)}(J)$ for some $\nu \geq \bar{\mu} + k$, and $G_{(\cdot, \cdot)}(t)$ is twice continuously differentiable at each $(u, x) \in \mathscr{H}_l^2$.

(A5) (Conditions on bandwidth). It is assumed that $b \to 0$, $n^{1/3}c/\log(n) \to \infty$, $n^{1/(2\bar{\mu}+3)}b/\log(n) \to \infty$, $b^{2\underline{\mu}+1}/c \to e \in (0, \infty)$ and $nb^{2\underline{\mu}+2k+1} \to \pi^2 \in [0, \infty)$.

REMARK 2.3. (a) Assumption (A1) is mainly for simplification. For the number of observations per subject, it is certainly enough if the limits $\lim n_i/n_j = c_{ij} \in (0, \infty)$ exist for $i, j \in \{1, \ldots, m\}$. In this case, it is easy to modify the results presented in this paper. (b) The conditions $n^{1/3}c/\log(n) \to \infty$ and $n^{1/(2\bar{\mu}+3)}b/\log(n) \to \infty$ are needed to bound the supreme norm of the derivatives of $\hat{f}_i(\cdot; c)$ and $\hat{f}_i(\cdot; b)$, respectively. The condition $b^{2\underline{\mu}+1}/c \to e \in (0, \infty)$ is used to balance the variances of $\hat{f}_i(\cdot; c)$ and

$\hat{f}_i^{(\underline{\mu})}(\cdot;b)$. For the case $\underline{\mu} > 0$ (e.g., considering only extremes as structural points), one could take $e = 0$. Finally, $nb^{2\underline{\mu}+2k+1} \to \pi^2$ is used to balance the variance and bias^2 of $\hat{f}_i^{(\underline{\mu})}(\cdot;b)$. (c) The existence of such bandwidths b and c can be seen as follows. If we take $c = O(n^{-\gamma})$ and $b = O(n^{-\delta})$ for some positive constants γ and δ, then assumption (A5) requires that

$$\gamma = \delta(2\underline{\mu} + 1) < 1/3,$$

$$1/(2\underline{\mu} + 2k + 1) \le \delta < 1/(2\bar{\mu} + 3).$$

It is easy to see that this system has solutions. These include the most useful cases $\bar{\mu} - 1 = \underline{\mu} = 0$, $k \ge 3$, $\bar{\mu} = \underline{\mu} = 1$, $k \ge 4$ and $\bar{\mu} = \underline{\mu} = 2$, $k \ge 6$.

3. Asymptotic normality. First, the asymptotics of $\hat{f}$ at the structural points $\tau_{\cdot r}$ is given. This is easier than the general case since no warping function is involved. We restrict ourselves to e-functionals. To give the asymptotic results in concise form, we need some notation.

Let Δ be the delta-function $\Delta(x, y) = 1$ for $x = y$, and $\Delta(x, y) = 0$ for $x \ne y$. Define for $i = 1, \ldots, m$ that

$$B_{ir}(t) = \frac{(-1)^{\mu_r+1}\pi f_i^{(\mu_r+k)}(t)}{(k + \mu_r)! f_i^{(\mu_r+1)}(t)} \int_{-1}^{1} W^{(\mu_r)}(u) u^{(k+\mu_r)} \, du,$$

$$H_{ir}(t) = \frac{\pi\Delta(\underline{\mu}, 0)}{e^k} \frac{f_i^{(k)}(t)}{k!} \int_{-1}^{1} W(u) u^k \, du.$$

These functions are the familiar bias functions for kernel estimators, and they describe the asymptotic bias for structural points and for regression functions

$$B_r^\tau \equiv \frac{\Delta(\mu_r, \underline{\mu})}{m} \sum_{i=1}^{m} B_{ir}(\tau_{ir}), \quad B_r^f = \frac{\Delta(\mu_r, \underline{\mu})}{m} \sum_{i=1}^{m} \left[H_{ir}(\tau_{ir}) + f_i^{(1)}(\tau_{ir})B_{ir}(\tau_{ir})\right].$$

To summarize the structure of the asymptotic variances we define

$$Q_{ir}(t) = \left(\frac{\sigma_i}{f_i^{(\mu_r+1)}(t)}\right)^2 \int_{-1}^{1} \left[W^{(\mu_r)}(u)\right]^2 \, du,$$

$$V_{ir}(t) = \sigma_i^2 e\Delta(\mu_r, \underline{\mu}) \int_{-1}^{1} W^2(u) \, du + \left(f_i^{(1)}(t)\right)^2 Q_{ir}(t).$$

The asymptotic variances and covariance can now be defined by

$$V_r^\tau = \frac{1}{m^2} \sum_{i=1}^{m} Q_{ir}(\tau_{ir}), \qquad V_r^f = \frac{1}{m^2} \sum_{i=1}^{m} V_{ir}(\tau_{ir}),$$

$$\text{Cov}_r = \frac{\Delta(\mu_r, \underline{\mu})}{m^2} \sum_{i=1}^{m} f_i^{(1)}(\tau_{ir}) Q_{ir}(\tau_{ir}).$$

We give the asymptotic normal distribution in the following theorem.

THEOREM 3.1. *Assume conditions* (A1)–(A5). *If T_r is an e-functional, then*

$$\left(nb^{2\mu_r+1}\right)^{1/2}(\hat{\tau}_{.r} - \tau_{.r}) \Rightarrow N\left(B_r^\tau, V_r^\tau\right),$$

$$\left(nb^{2\mu_r+1}\right)^{1/2}\left(\hat{f}(\hat{\tau}_{.r}) - f(\tau_{.r})\right) \Rightarrow N\left(B_r^f, V_r^f\right)$$

and

$$nb^{2\mu_r+1}\,\mathrm{Cov}\left(\hat{\tau}_{.r} - \tau_{.r}, \hat{f}(\hat{\tau}_{.r}) - f(\tau_{.r})\right) \to \mathrm{Cov}_r.$$

In principle, Theorem 3.1 could be used to construct asymptotic confidence regions for $(\tau_{.r}, f(\tau_{.r}))$ as follows. Let S_α be a $100\alpha\%$ confidence region, centered at $(0,0)$, for a two-dimensional standard normal variable. Let $\hat{B}^r$ be the plug-in estimator of the asymptotic bias (B_r^τ, B_r^f), and $\hat{\Sigma}_r$ the plug-in estimator of the asymptotic covariance matrix of $(nb^{2\mu_r+1})^{1/2}(\hat{\tau}_{.r}, \hat{f}(\hat{\tau}_{.r}))$. Then a $100\alpha\%$ confidence region for $(\tau_{.r}, f(\tau_{.r}))$ is given approximately by

$$(2) \qquad \begin{pmatrix} \hat{\tau}_{.r} \\ \hat{f}(\hat{\tau}_{.r}) \end{pmatrix} - \{(nb^{2\mu_r+1})^{-1/2}(x + \hat{B}^r) \colon x \in \hat{\Sigma}_r^{1/2} S_\alpha\}.$$

The plug-in estimators $\hat{B}^r$ and $\hat{\Sigma}_r$ are defined by replacing unknown $f_i^{(s)}$ in the asymptotic bias and covariances given in Theorem 3.1 with estimated values. If some or all σ_i^2, $i = 1, \ldots, m$, are unknown, then use the difference estimator [cf. Gasser, Sroka and Jennen-Steinmetz (1986)], which is strongly consistent:

$$(3) \qquad \hat{\sigma}_i^2 = \frac{2}{3(n-2)} \sum_{j=2}^{n-1} \left(\frac{1}{2}y_{ij-1} + \frac{1}{2}y_{ij+1} - y_{ij}\right)^2, \qquad i = 1, \ldots, m.$$

The following proposition is a direct consequence of Theorem 3.1 and its proof. Simultaneous confidence bounds for f at structural points can be constructed based on the following proposition.

PROPOSITION 3.1. *Assume conditions* (A1)–(A5) *and that all structural functionals are e-functionals. Then*

$$\left(\left(nb^{2\mu_1+1}\right)^{1/2}\left(\hat{f}(\hat{\tau}_{.1}) - f(\tau_{.1})\right), \ldots, \left(nb^{2\mu_l+1}\right)^{1/2}\left(\hat{f}(\hat{\tau}_{.l}) - f(\tau_{.l})\right)\right) \Rightarrow N(B, V).$$

Here $B = (B_1^f, \ldots, B_l^f)$ and $V = \mathrm{diag}(V_1^f, \ldots, V_l^f)$.

Next we investigate confidence bars for $f(t)$ at a nonstructural point t. In this case, all structural functionals are involved in the definition of the warping functions. Both bias and variance can be written in closed forms, but they are more complicated than the formulas in Theorem 3.1 and are of little practical use. Therefore, they are omitted. Interested readers can obtain an earlier version of this paper from the Web site "http://www.unizh.ch/biostat" to get the formulas. The asymptotic distribution of $\hat{f}$ is given by the following theorem.

THEOREM 3.2. *Assume conditions* (A1)–(A5) *and that all structural functionals are e-functionals. Then the random variable* $(nb^{2\bar{\mu}+1})^{1/2}(\hat{f}(t) - f(t))$ *is asymptotically normal for* $t \in (\sup_i \{g_i^{-1}(a_0)\}, \inf_i \{g_i^{-1}(a_1)\})$.

REMARK 3.1. Theorem 3.2 gives the limiting distribution of $\hat{f}(\tau_{.r}) - f(\tau_{.r})$ when $t = \tau_{.r}$, while Theorem 3.1 is for the asymptotics of $\hat{f}(\hat{\tau}_{.r}) - f(\tau_{.r})$. For $\sup_i \{g_i^{-1}(a_0)\} < t_1 < \cdots < t_a < \inf_i \{g_i^{-1}(a_1)\}$, it is easily seen that

$$\left(nb^{2\bar{\mu}+1}\right)^{1/2}\left(\hat{f}(t_1) - f(t_1), \ldots, \hat{f}(t_a) - f(t_a)\right)$$

converges to a multinormal random vector. This fact should be kept in mind when one bootstraps the distribution of $(\hat{f}(t_1), \ldots, \hat{f}(t_a))$.

Comments on the proofs of Theorems 3.1 and 3.2 are given in Section 5.

4. Bootstrap confidence bounds. In previous sections we assumed that $\{f_1, \ldots, f_m\}$ are fixed unknown functions. In other words, we treated the conditional problem given $\{f_1, \ldots, f_m\}$. The unconditional problem is to consider the curves $\{f_1, \ldots, f_m\}$ as a random sample according to some probability law.

For the conditional problem, variations are caused by observation errors $\{\varepsilon_{ij}\}$. Thus, the bootstrapping method described in Section 4.1 resamples estimated residuals. This is referred to as bootstrap within subject. For the general problem, the random variation of $\{f_i\}$ dominates the variations caused by observation errors. For a first-order approximation, the effects on the final confidence region by the observation errors can be ignored. This is justified in Section 4.2 and is referred to as bootstrap across subjects.

Bias correction has been an important issue in bootstrapping confidence bars for a single regression curve. Several approaches have been proposed. One method is to form bootstrapped data by adding estimated residuals to an oversmoothed curve [Härdle and Marron (1991)]. The amount of oversmoothing depends on the order of the kernel function. For a second-order kernel, the bias of kernel estimator depends on the second-order derivative of the regression function. Therefore, the second-order derivative of the oversmoothed curve should converge to the second-order derivative of the regression function. Then the bootstrap estimate of bias is consistent. Härdle and Marron (1991) show that this bootstrap works. Another approach is a two-step procedure [Hall (1992b)]. First, construct confidence bars for the expectation of the kernel estimator. Then estimate bias by using another estimator. An example is the plug-in estimator which depends on estimating the derivatives of the regression curve. The final confidence bar is obtained by combining the two estimates. We will apply the first approach which takes care of the bias problem automatically.

For the average curve f of a sample of curves $\{f_i\}$, confidence bounds take two forms. At a structural point $(\tau_{.r}, f(\tau_{.r}))$, both the location $\tau_{.r}$ and the amplitude $f(\tau_{.r})$ are parameters. Therefore the bounds for $(\tau_{.r}, f(\tau_{.r}))$ consist

of a bounded region in two-dimensional space. At a fixed point t, there is only one parameter $f(t)$. The confidence bounds become confidence bars.

Let $(\hat{\tau}_{.r}, \hat{f}(\hat{\tau}_{.r}))$ be an estimate of $(\tau_{.r}, f(\tau_{.r}))$ based on the data (t_{ij}, y_{ij}). Define $Z = (\hat{\tau}_{.r}, \hat{f}(\hat{\tau}_{.r})) - (\tau_{.r}, f(\tau_{.r}))$. Let $V(Z)$ be the covariance matrix of Z and assume that $V(Z)$ is nonsingular. Denote a ball of radius d by $B(0, d)$. For $\alpha \in (0, 1)$, define γ_α by

$$\gamma_\alpha = \inf\{\gamma: P(V(Z)^{-1/2} Z \in B(0, \gamma)) \geq \alpha\}.$$

Then a $100\alpha\%$ confidence region for $(\tau_{.r}, f(\tau_{.r}))$ is given by

(4) $$\Omega_1 = \{(\hat{\tau}_{.r}, \hat{f}(\hat{\tau}_{.r})) - x: x \in V(Z)^{1/2} B(0, \gamma_\alpha)\}.$$

The bootstrap methods described below estimate $V(Z)$ and γ_α.

At a fixed point $t \in [0, 1]$, z_α is defined by

$$z_\alpha = \inf\{z: P(|\hat{f}(t) - f(t)| \leq V(\hat{f}(t))^{1/2} z) \geq \alpha\}.$$

Here $V(\hat{f}(t))$ stands for the variance of $\hat{f}(t)$. Then a $100\alpha\%$ confidence interval for $f(t)$ is given by

$$(\hat{f}(t) - V(\hat{f}(t))^{1/2} z_\alpha, \hat{f}(t) + V(\hat{f}(t))^{1/2} z_\alpha).$$

If one takes $t = \tau_{.r}$, then this is a confidence interval for the amplitude of f at a structural point. Estimation of $V(\hat{f}(t))$ and z_α by bootstrap methods can be done in the same way as the estimation of $V(Z)$ and γ_α. Therefore it will be omitted.

4.1. *Bootstrapping under the conditional model.*   The conditional problem is treated in this section. Let $Z_i \equiv (\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir})) - (\tau_{ir}, f_i(\tau_{ir}))$, $i = 1, 2, \ldots, m$. Then

$$Z \equiv (\hat{\tau}_{.r}, \hat{f}(\hat{\tau}_{.r})) - (\tau_{.r}, f(\tau_{.r})) = \frac{1}{m} \sum_{i=1}^{m} Z_i.$$

Since Z_i depends only on the ε_{ij}, $Z_1, \ldots, Z_m$ are independent random vectors. Therefore, in order to bootstrap the distribution of Z it is enough to bootstrap the distributions of Z_i, $i = 1, \ldots, m$, separately. This is what we mean by bootstrap within subject.

Let $\check{f}_i(t) \equiv \hat{f}_i(t; h)$ be the kernel estimate of f_i with kernel W and bandwidth h. In order to correct the bootstrap bias automatically, oversmoothing is necessary in estimating $\check{f}_i$. Such a bandwidth h can be selected according to Lemma 4.1. Denote the corresponding structural points of $\check{f}_i$ by $\check{\tau}_{ir} = T_r(\check{f}_i)$. Let $\{\hat{\varepsilon}_{ij}: j = 1, \ldots, n\}$ be the centered estimated residuals computed by

$$\hat{\varepsilon}_{ij} = y_{ij} - \hat{f}_i(t_j; b_0) - \frac{1}{n} \sum_{q=1}^{n} [y_{iq} - \hat{f}_i(t_q; b_0)], \qquad j = 1, \ldots, n,$$

where $b_0 = O(n^{-1/(2k+1)})$ is an optimal bandwidth.

Draw bootstrap residuals $\{\varepsilon_{ij}^*\colon j = 1, \ldots, n\}$ with replacement from $\{\hat{\varepsilon}_{ij}\colon j = 1, \ldots, n\}$ and define a bootstrapped sample

$$y_{ij}^* = \check{f}_i(t_{ij}) + \varepsilon_{ij}^*, \qquad j = 1, \ldots, n.$$

Smooth the data $\{(t_{ij}, y_{ij}^*)\colon j = 1, \ldots, n\}$ with kernel W and bandwidth b and find the corresponding structural point τ_{ir}^*. Then the bootstrapped estimate of Z_i is given by

$$Z_i^* = \big(\tau_{ir}^*, \hat{f}_i^*(\tau_{ir}^*)\big) - \big(\check{\tau}_{ir}, \check{f}_i(\check{\tau}_{ir})\big).$$

Here $\hat{f}_i^*$ is the kernel estimator of f_i from the bootstrapped data $\{(t_{ij}, y_{ij}^*)\colon j = 1, \ldots, n\}$ with kernel W and bandwidth c.

Define $Z^* = (Z_1^* + \cdots + Z_m^*)/m$ and define $\hat{\gamma}_\alpha$ by

$$\hat{\gamma}_\alpha = \inf\big\{\gamma\colon P^*\big(V(Z^*)^{-1/2} Z^* \in B(0, \gamma)\big) \geq \alpha\big\},$$

where P^* is the conditional distribution of $\{\varepsilon_{ij}^*\}$ given $\{\varepsilon_{ij}\}$, and $V(Z^*)$ is the covariance matrix of Z^* under the distribution P^*. The following lemma implies that $V(Z^*)$ is a consistent estimator of $V(Z)$ and that $\hat{\gamma}_\alpha$ is a consistent estimator of γ_α. Replacing $V(Z)$ and γ_α in (4) by $V(Z^*)$ and $\hat{\gamma}_\alpha$, respectively, one obtains a bootstrapped confidence region

$$\Omega_1^* = \big\{\big(\hat{\tau}_{\cdot r}, \hat{f}(\hat{\tau}_{\cdot r})\big) - x\colon x \in V(Z^*)^{1/2} B(0, \hat{\gamma}_\alpha)\big\}$$

for the structural point $(\tau_{\cdot r}, f(\tau_{\cdot r}))$. Comments on the proof of this lemma are in Section 5.

LEMMA 4.1. *Assume that the conditions of Theorem* 3.1 *hold and, in addition,* $h \to 0$ *and* $nh^{2(\mu_r + k) + 3} \to \infty$ *as* $n \to \infty$. *Then the following statements hold*:

(i) *For* $0 \leq q \leq \mu_r + k + 1$, $\check{f}_i^{(q)}(\check{\tau}_{ir})$ *converges to* $f_i^{(q)}(\tau_{ir})$ *in probability.*

(ii) *The variable* $(nb^{2\mu_r + 1})^{1/2}(\tau_{ir}^* - \check{\tau}_{ir}, \hat{f}_i^*(\tau_{ir}^*) - \check{f}_i(\check{\tau}_{ir}))$ *under the bootstrap distribution and the variable* $(nb^{2\mu_r + 1})^{1/2}(\hat{\tau}_{ir} - \tau_{ir}, \hat{f}_i(\hat{\tau}_{ir}) - f_i(\tau_{ir}))$ *under the distribution of* $\{\varepsilon_{ij}\}$ *converge to the same normal variable in distribution as* $n \to \infty$.

(iii) *The probability* $P(Z \in \Omega_1^*)$ *converges to* α *as* $n \to \infty$.

REMARK 4.1. Confidence bars for f at multiple points $a_0 < t_1 < \cdots < t_a < a_1$ can be obtained similarly. The distribution of $(nb^{2\bar{\mu}+1})^{1/2}(\hat{f}(t_1) - f(t_1), \ldots, \hat{f}(t_a) - f(t_a))$ can be approximated by $(nb^{2\bar{\mu}+1})^{1/2}(\hat{f}^*(t_1) - \check{f}(t_1), \ldots, \hat{f}^*(t_a) - \check{f}(t_a))$. This can be justified by a multidimensional version of the proof of

Lemma 4.1, based on Theorem 3.2 instead of Theorem 3.1. Then confidence bounds for $(f(t_1), \ldots, f(t_a))$ can be constructed accordingly.

4.2. *Bootstrapping under the random model.* Assume that $\{f_i: i = 1, \ldots, m\}$ is a random sample from $\mathscr{R} \subset C^{(\nu)}(J)$ according to some probability law $\mathbf{P}$. We will consider the simple case in which f_i, $i = 1, \ldots, m$, are identically independently distributed. Let f_0 be the mean of $\mathbf{P}$. Then $\tau_{\cdot r} = T_r(f_0)$. We will assume that $\{f_i\}$ is independent of $\{\varepsilon_{ij}\}$. Define the notation $E^\varepsilon(\cdot) = E(\cdot | f_i, \ i = 1, \ldots, m)$, $E^{\mathbf{P}}$ for expectation with respect to $\mathbf{P}$, and $\mathbf{E}$ for expectation with respect to the joint distribution of $\{\varepsilon_{ij}, f_i\}$.

Let $Z_0 \equiv (\hat{\tau}_{\cdot r}, \hat{f}(\hat{\tau}_{\cdot r})) - (\tau_{\cdot r}, f_0(\tau_{\cdot r}))$. Then

$$
\begin{aligned}
Z_0 &= \frac{1}{m} \sum_{i=1}^{m} \left[(\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir})) - (\tau_{\cdot r}, f_0(\tau_{\cdot r})) \right] \\
&= \frac{1}{m} \sum_{i=1}^{m} \left[(\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir})) - (\tau_{ir}, f_i(\tau_{ir})) \right] \\
&\quad + \frac{1}{m} \sum_{i=1}^{m} \left[(\tau_{ir}, f_i(\tau_{ir})) - (\tau_{\cdot r}, f_0(\tau_{\cdot r})) \right].
\end{aligned}
\tag{5}
$$

In the last display, the first term is mainly the effect of $\{\varepsilon_{ij}\}$. The second term is purely an effect of $\mathbf{P}$ since it vanishes if $\mathbf{P}$ is a point measure $\mathbf{P} = \delta_{f_0}$. First we summarize some useful facts.

FACT 1.   It follows from the i.i.d. assumption on $\{f_i\}$ that $\{(\tau_{ir}, f_i(\tau_{ir})): i = 1, \ldots, m\}$ is an i.i.d. random sample with mean $(\tau_{\cdot r}, f_0(\tau_{\cdot r}))$.

FACT 2.   Asymptotically we have

$$
Z_0 = \left\{ \frac{1}{m} \sum_{i=1}^{m} \left[(\tau_{ir}, f_i(\tau_{ir})) - (\tau_{\cdot r}, f_0(\tau_{\cdot r})) \right] \right\} (1 + o_p(1)).
$$

This is clear from Theorem 3.1, which states that

$$
(\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir})) - (\tau_{ir}, f_i(\tau_{ir})) = O_p\big((nb^{2\mu_r+1})^{-1/2}\big).
$$

FACT 3.   Let $H(f_1, \ldots, f_m)$ be a functional of $f_1, \ldots, f_m$. Then $\Xi \equiv (\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir})) - E^\varepsilon(\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir}))$ and $H(f_1, \ldots, f_m)$ are uncorrelated. Since $\mathbf{E}(\cdot) = E^{\mathbf{P}} E^\varepsilon(\cdot)$, $\mathbf{E}(\Xi) = 0$. Using the independence of $\{f_i\}$ and $\{\varepsilon_{ij}\}$, we have

$$
\begin{aligned}
&\mathbf{E}\big\{ \Xi \big[H(f_1, \ldots, f_m) - E^{\mathbf{P}}(H(f_1, \ldots, f_m)) \big] \big\} \\
&\quad = E^{\mathbf{P}} \big\{ E^\varepsilon(\Xi) \big[H(f_1, \ldots, f_m) - E^{\mathbf{P}}(H(f_1, \ldots, f_m)) \big] \big\} = 0.
\end{aligned}
$$

FACT 4.   If we choose the bandwidth b small such that $nb^{2\mu+2k+1} \to \pi^2 = 0$ as $n \to \infty$, then

$$
(\hat{\tau}_{ir}, \hat{f}_i(\hat{\tau}_{ir})) - (\tau_{ir}, f_i(\tau_{ir})) = \Xi(1 + o_p(1)).
$$

See assumptions (A3) and (A5) in Section 2 for definitions of k and μ. This is a basic property of kernel estimator—trade-off between bias and variance.

Fact 2 says that the variation induced by **P** dominates the variation caused by $\{\varepsilon_{ij}\}$. For bootstrapping the distribution of the leading term of Z_0, we only need to bootstrap the distribution of **P**. This is what bootstrap across subjects means. From Fact 1, the problem of bootstrapping confidence regions at the structural point $(\tau_{.r}, f_0(\tau_{.r}))$ is equivalent to bootstrap confidence regions for the mean of a two-dimensional distribution.

Let $\{(\tau_{ir}^*, f_i^*(\tau_{ir}^*))\colon i = 1, \ldots, m\}$ be a random sample drawn from $\{(\tau_{ir}, f_i(\tau_{ir}))\colon i = 1, \ldots, m\}$ with replacement. Define

$$Z_0^* = \frac{1}{m} \sum_{i=1}^{m} \big[(\tau_{ir}^*, f_i^*(\tau_{ir}^*)) - \bar{Z}_0\big],$$

where $\bar{Z}_0 = [((\tau_{1r}, f_1(\tau_{1r})) + \cdots + (\tau_{mr}, f_m(\tau_{mr}))]/m$. For $\alpha \in (0, 1)$, let $\hat{\beta}_\alpha$ be the solution of the equation

$$\mathbf{P}^*\big(V(Z_0^*)^{-1/2} Z_0^* \in B(0, \hat{\beta}_\alpha)\big) = \alpha.$$

Here $\mathbf{P}^*(\cdot) = \mathbf{P}(\cdot|\{f_i\})$. Then a $100\alpha\%$ confidence region for $(\tau_{.r}, f_0(\tau_{.r}))$ is given by

$$\Omega_2^* = \big\{\bar{Z}_0 - x\colon x \in V(Z_0^*)^{1/2} B(0, \hat{\beta}_\alpha)\big\}.$$

THEOREM 4.1. *Assume conditions* (A1)–(A5) *and that* $(\tau_{1r}, f_1(\tau_{1r}))$ *has finite second moment. Then the probability of* $(\tau_{.r}, f_0(\tau_{.r}))$ *being in* Ω_2^* *converges to* α *as* $n \to \infty, m \to \infty$.

PROOF. By Fact 2 above, we can ignore the first term of (5) when we approximate the distribution of Z_0. Since $(\tau_{1r}, f_1(\tau_{1r}))$ has finite second moment, the central limit theorem holds for the sequence $\{(\tau_{ir}, f_i(\tau_{ir}))\colon i \geq 1\}$. It follows from Theorem 1 of Mammen (1992) that the Kolmogorov distance between the law of $\bar{Z}_0 - (\tau_{.r}, f_0(\tau_{.r}))$ and the bootstrap law of $Z_0^* - \bar{Z}_0$ converges to zero in probability, as $m \to \infty$. This completes the proof of the theorem. □

In practice, it might be better to take into account both the effects of **P** and the law of $\{\varepsilon_{ij}\}$ if m and n are small. By Facts 3 and 4 above, the variations due to **P** and the variations due to $\{\varepsilon_{ij}\}$ can be bootstrapped separately, because they are asymptotically uncorrelated if we choose the bandwidth b properly.

Let Z^* and $V(Z^*)$ be the bootstrapped quantities of Section 4.1. Define $\hat{\delta}_\alpha$ by the solution of the equation

$$\mathscr{P}\big[(V(Z^*) + V(Z_0^*))^{-1/2}(Z^* + Z_0^*) \in B(0, \hat{\delta}_\alpha)\big] = \alpha,$$

where $\mathscr{P}(\cdot) = P(\cdot|\{f_i\}, \{\varepsilon_{ij}\})$. Then a $100\alpha\%$ confidence region for $(\tau_{\cdot r}, f_0(\tau_{\cdot r}))$ is given by

$$\Omega_3^* = \left\{ (\hat{\tau}_{\cdot r}, \hat{f}_0(\hat{\tau}_{\cdot r})) - x \colon x \in \left(V(Z^*) + V(Z_0^*) \right)^{1/2} B(0, \hat{\delta}_\alpha) \right\}.$$

We refer to Ω_3^* as a confidence region obtained by the "full bootstrap."

4.3. *A simulation study.* We know from Section 4.2 that the variation induced by the randomness of the sample curves dominates the variation induced by $\{\varepsilon_{ij}\}$ (see also Figure 1). In the simulation here, the bootstrapped confidence regions are constructed by resampling curves with replacement. The noise is processed only in the smoothing step. If interested, one can obtain a small archive SACI (Structural Averaging and Confidence Intervals), written in Fortran, from the Web site "http://www.unizh.ch/biostat" to compute the structural average curve, to construct confidence regions at a structural point and to compute confidence bars at multiple points.

The data are generated from the model

$$y_i(t) = a_i s\left(\frac{t - b_i}{c_i} \right) + d_i + e_i(t), \qquad t \in [0, 1]$$

with a base function s given by

$$
\text{(6)} \qquad s(t) = \begin{cases} s_1(t) + 4t \sin(31.4(0.45 - t)), & t \leq 0.45, \\ s_1(t) + 10(s_2(t) - s_1(t))(0.45 - t), & 0.45 \leq t \leq 0.55, \\ s_2(t) + 4(1 - t) \sin(31.4(t - 0.55)), & t \geq 0.55, \end{cases}
$$

$$s_1(t) = 0.25(t - 5)^2 - 8(0.45 - t),$$

$$s_2(t) = -0.25(t + 4)^2 + 8(t - 0.55).$$

The e_i are i.i.d. residuals. To simplify the computation of the exact confidence regions and multiple confidence bars, the parameters are chosen to be independently and normally distributed as follows:

$$
\begin{aligned}
a_i &\sim N(1, \alpha^2), & b_i &\sim N(0, \beta^2), \\
c_i &\sim N(1, \gamma^2), & d_i &\sim N(0, \delta^2), \\
e_i(t) &\sim N(0, \sigma^2), & \text{Cov}(e_i(t), e_i(u)) &= 0 \quad \text{for } t \neq u.
\end{aligned}
$$

Due to the choice of parameters, the structural average curve is s.

Let t_0 be a structural point on s (say a maximum point). The corresponding structural point on y_i is at

$$g_i = c_i t_0 + b_i,$$

with amplitude

$$y_i(g_i) = a_i s(t_0) + d_i + e_i(g_i).$$

Thus, the structural point is estimated by

$$\bar{t} = \frac{1}{m} \sum_{i=1}^{m} g_i,$$

with amplitude estimated by

$$\bar{f} = \frac{1}{m} \sum_{i=1}^{m} y_i(g_i).$$

Note that

$$E(\bar{t}) = t_0, \qquad E(\bar{f}) = s(t_0).$$

Therefore, $(\bar{t}, \bar{f})$ is an unbiased estimate of $(t_0, s(t_0))$.

Since a_i, b_i, c_i, d_i, $e_i(t)$ are independent variables, we can compute the moment generating function. For any real number u,

$$E\big[\exp\big(ue_i(g_i)\big)\big] = E\big\{E\big[\exp(ue_i(g_i))\big|g_i\big]\big\} = E\big\{\exp(u^2\sigma^2/2)\big\} = \exp(u^2\sigma^2/2).$$

This and the fact that $E[g_i e_i(g_i)] = 0$ imply that $\bar{t}$ and $\bar{f}$ are independently and normally distributed. Simple computations give

$$\bar{t} - t_0 \sim N(0, (t_0^2\gamma^2 + \beta^2)/m),$$

$$\bar{f} - s(t_0) \sim N\big(0, (s^2(t_0)\alpha^2 + \delta^2 + \sigma^2)/m\big).$$

Let T and F be the central 95% CIs for t_0 and $s(t_0)$, respectively. Then $C_0 = T \times F$ is a central 90% confidence region for $(t_0, s(t_0))$.

The exact confidence region C_0 is a rectangle with edges parallel to axes. The bootstrapped confidence region C_b is an ellipse. Let C_r be the smallest rectangle which contains C_b and has edges parallel to axes. Then the distance d_b between the centers of C_0 and C_b and the sizes of C_0, C_b, and C_r indicate the performance of the bootstrap method.

The simulation proceeds as follows. Fix the number of curves m and the sample size n per curve. Run the proposed estimation procedure for $K = 500$ Monte Carlo replications. For each Monte Carlo replication, we compute a bootstrapped confidence region for $(t_0, s(t_0))$ with $B = 200$ bootstrap replications. This produces K estimated confidence regions $C_b(k)$, and the associated distances $d_b(k)$, rectangles $C_r(k)$, and area differences

$$A_b(k) = \text{area of } C_b - \text{area of } C_0,$$

$$A_r(k) = \text{area of } C_r - \text{area of } C_0.$$

The following quantities will be computed:

$D = \text{mean of } \big\{d_b(k)\colon k = 1, \ldots, K\big\};$

$V = \text{variance of the sample mean } D;$

$P_b = \text{proportion of Monte Carlo replications such that } \big(t_0, s(t_0)\big) \in C_b;$

P_r = proportion of Monte Carlo replications such that $(t_0, s(t_0)) \in C_r$;

$A1_b$ = mean of $\{A_b(k): k = 1, \ldots, K\}$;

$V1_b$ = variance of the sample mean $A1_b$;

$A1_r$ = mean of $\{A_r(k): k = 1, \ldots, K\}$;

$V1_r$ = variance of the sample mean $A1_r$.

Table 1 gives the simulation results for different combinations of m and n. It follows from $A1_b \approx 0$ that the bootstrapped confidence region is about the same size as that of the exact confidence region. The fact that $A1_r$ is also small implies that the bootstrapped confidence region is close to a rectangle which has edges parallel to axes. Therefore, the bootstrap method performs well in constructing confidence regions for structural average curves. Of course these are limited simulation results. The measure D gives the error of estimating $(t_0, s(t_0))$, particularly the error of estimating $s(t_0)$. Its value is relatively large because we have only 20 or 50 curves and because the amplitudes of the sample curves depend on the parameters a_i (standard deviation 0.4), d_i (standard deviation 2) and $\sigma_i(t)$ (standard deviation 0.5).

A typical run with $m = 50$, $n = 100$ is plotted in Figure 2.

Simulation with another set of parameters is performed to check the performance of the bootstrap method for data with larger shifts. Again the simulation is performed with 500 Monte Carlo replications and 200 bootstrap replications. The results are grouped in Table 2. It can be seen that the result for this parameter set is not quite as good as the result for the previous parameter set. This is probably due to heavier shift variation.

Generally speaking, the bootstrap method for constructing confidence regions works well.

TABLE 1
*Simulation results of bootstrapping confidence regions at a structural point**

(m, n)	(20, 100)	(50, 100)	(100, 100)	(20, 200)	(50, 200)	(100, 200)
P_b	0.878	0.898	0.902	0.884	0.860	0.672
P_r	0.884	0.904	0.928	0.886	0.874	0.686
D	0.6367	0.4092	0.3353	0.6169	0.3880	0.5455
V	0.2149	0.0948	0.0564	0.2097	0.082	0.1053
$A1_b$	−0.008	−0.001	0.0026	−0.0074	−0.0039	−0.0012
$V1_b$	0.0008	0.0001	0.00002	0.001	0.0001	0.00001
$A1_r$	0.0303	0.0133	0.0107	0.0251	0.01	0.0059
$V1_r$	0.0014	0.0001	0.00003	0.0016	0.0001	0.00001

*The first line (m, n) gives the (number of curves, sample points per curve). The parameters for generating sample curves are $\alpha = 0.4$, $\beta = 0.05$, $\gamma = 0.01$, $\delta = 2$ and $\sigma = 0.5$. The confidence level is 90%.
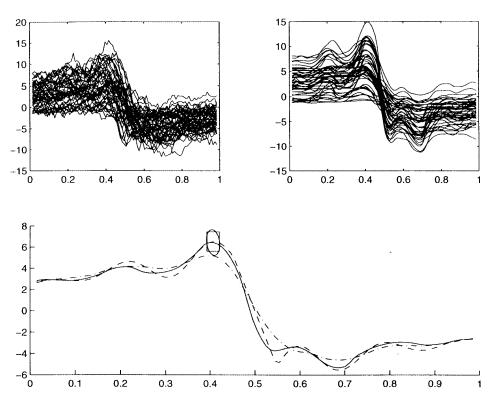
FIG. 2.   *Raw data of* 50 *curves* (*top left*); *shifted smooth curves* (*top right*); *confidence regions* (*bottom*): *base curve* (*dashed*), *estimated structural average* (*solid*), *cross-sectional average* (*dashdot*), 90% *confidence region by bootstrap* (*ellips*) *and* 90% *exact confidence region* (*rectangle*).

It is more involved to compute the exact uniform confidence band. We know that the structural average curve is the base function s, which is estimated by

$$\bar{f}(t) = \frac{1}{m}\sum_{i=1}^{m}[a_i s(t) + d_i + e_i(g_i(t))], \qquad g_i(t) = c_i t + b_i.$$

For $0 < u_1 < \cdots < u_q < 1$, let

$$\bar{f} = \begin{pmatrix} \bar{f}(u_1) \\ \vdots \\ \bar{f}(u_q) \end{pmatrix}, \qquad S = \begin{pmatrix} s(u_1) \\ \vdots \\ s(u_q) \end{pmatrix}, \qquad \ell = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}.$$

Then it can be shown that $\bar{f} - S \sim N(0, V)$ with

$$V = \frac{\sigma^2}{m}I + \frac{\delta^2}{m}\ell\ell' + \frac{\alpha^2}{m}SS'.$$

TABLE 2
*Simulation results with parameters $\alpha = 0.8$, $\beta = 0.1$, $\gamma = 0.05$,
$\delta = 1.5$ and $\sigma = 1$*[*]

(*m, n*)	(20, 200)	(50, 200)	(100, 200)
P_b	0.754	0.756	0.374
D	0.9653	0.5601	1.2448
V	0.5314	0.1804	0.3374
$A1_b$	0.0611	0.0216	0.0219
$V1_b$	0.0508	0.006	0.001

[*]The confidence level is 90%.

For any $\eta \in (0, 1)$, let C_η be defined by

$$\int_{|x| \leq C_\eta} h(x)\, dx = \eta,$$

where h is the desity function of $\bar{f} - S$. Then $\{x: |\bar{f} - x| \leq C_\eta\}$ is a $100\eta\%$ confidence region for S (uniform confidence bars on s). Note that the exact uniform confidence bars are independent of the shift parameters b_i and c_i, because we assume the optimal alignment. However, the shift parameters are involved in the bootstrapped confidence bars because the optimal alignment cannot be obtained from noisy data.

In the following example, we simply compute the bootstrapped 90% confidence bars at each u_i. This of course leads to uniform confidence bars of less than 90% confidence level. Base function and parameters are the same as that used for Table 1, with 50 sample curves and 100 sample points per curve. The confidence bars are plotted in Figure 3.

5. Comments on the proofs. Theorems 3.1 and 3.2 and Lemma 4.1 are proved based on the ideas of Kneip and Gasser (1992). Following the
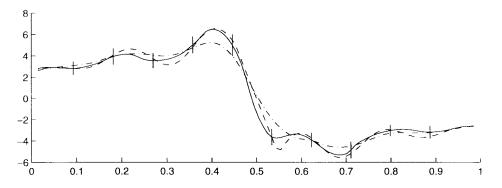

FIG. 3. *Bootstrapped* 90% *confidence bars at* 10 *equally spaced locations*: *base curve* (*dashed*), *estimated structural average* (*solid*) *and cross-sectional average* (*dashdot*).

suggestion of a referee, we cut out the detailed treatment, which can be obtained from an earlier version of this paper from the Web site "http://www.unizh.ch/biostat."

The main step is to expand (Taylor expansion) $\hat{\tau}_{\cdot r} - \tau_{\cdot r}$ and $\hat{f}(\hat{\tau}_{\cdot r}) - f(\tau_{\cdot r})$ as linear functions of the kernel estimators $\hat{f}_i(\tau_{ir}; c) - f_i(\tau_{ir})$ and $\hat{f}_i^{(s)}(\tau_{ir}; b) - f_i^{(s)}(\tau_{ir})$, $i = 1, \ldots, m$; $s = 0, \ldots, \bar{\mu}$, plus a smaller error term. By Theorem 3 of Gasser and Müller (1984), each of these linear terms has a limiting normal distribution if it is properly scaled (Lindeberg condition). The scaling factors are given in Kneip and Gasser (1992).

Recall that the e-functional of order μ_r locates an extrema τ_{ir} of $f_i^{(\mu_r-1)}$. Assume that $f_i^{(\mu_r+1)}(\tau_{ir}) \neq 0$. Then

$$\hat{\tau}_{ir} - \tau_{ir} = -\frac{1}{f_i^{(\mu_r+1)}(\tau_{ir})} \big[\hat{f}_i^{(\mu_r)}(\tau_{ir}; b) - f_i^{(\mu_r)}(\tau_{ir}) + H_{ir}\big],$$

where H_{ir} is a higher order term.

For a p-functional of order μ_r, let α_{qir}, $q = 0, 1$, be the two associated extremes of $f_i^{(\mu_r)}$ such that $f_i^{(\mu_r)}$ is monotone in $(\alpha_{0ir}, \alpha_{1ir})$ and that

$$f_i^{(\mu_r)}(\tau_{ir}) = p_r f_i^{(\mu_r)}(\alpha_{0ir}) + (1 - p_r) f_i^{(\mu_r)}(\alpha_{1ir})$$

for some $p_r \in (0, 1)$. Then

$$\begin{aligned}
\hat{\tau}_{ir} - \tau_{ir} = -\frac{1}{f_i^{(\mu_r+1)}(\tau_{ir})} \big[& \hat{f}_i^{(\mu_r)}(\tau_{ir}; b) - f_i^{(\mu_r)}(\tau_{ir}) + U_{ir} \\
& - p_r\big(\hat{f}_i^{(\mu_r)}(\alpha_{0ir}; b) - f_i^{(\mu_r)}(\alpha_{0ir})\big) \\
& - (1 - p_r)\big(\hat{f}_i^{(\mu_r)}(\alpha_{1ir}; b) - f_i^{(\mu_r)}(\alpha_{1ir})\big) \big].
\end{aligned}$$

Here U_{ir} is a higher order term.

Use the Taylor expansion and the definition of structural average to get that

$$\hat{f}(\hat{\tau}_{\cdot r}) - f(\tau_{\cdot r}) = \frac{1}{m} \sum_{i=1}^{m} \big[\hat{f}_i(\tau_{ir}; c) - f_i(\tau_{ir}) + f_i^{(1)}(\tau_{ir})(\hat{\tau}_{ir} - \tau_{ir}) + R_{ir}\big].$$

Again R_{ir} is a higher order term. Substitute the expansions of $\hat{\tau}_{ir} - \tau_{ir}$ into this formula to get the expansion for $\hat{f}(\hat{\tau}_{\cdot r}) - f(\tau_{\cdot r})$.

Now we need only to find out the formulas for the bias and variance of each term in the expansions, the covariance of any two such terms, and the order of the supremum norm of each term. Most of these formulas are well known for kernel estimators, or they are proved in the lemma of Kneip and Gasser (1992). By carefully combining these facts, one can prove Theorem 3.1.

The proof of Theorem 3.2 is based on the same idea, but it is much more complicated. First, all the structural functionals are involved. Secondly, an extrema might be both a structural point (i.e., an e-functional is specified to

locate this point) and an end point of a monotone segment on which a p-functional is applied.

The proof of Lemma 4.1 is similar to the proof of Theorem 3.1. We know from Theorem 3.1 that $(nb^{2\mu_r+1})^{1/2}(\hat{\tau}_{ir} - \tau_{ir}, \hat{f}_i(\hat{\tau}_{ir}) - f_i(\tau_{ir}))$ is asymptotically normal. An expansion of $(nb^{2\mu_r+1})^{1/2}(\tau_{ir}^* - \check{\tau}_{ir}, \hat{f}_i^*(\tau_{ir}^*) - \check{f}_i(\check{\tau}_{ir}))$ as in the proof of Theorem 3.1 shows that this variable is also asymptotically normal. In order to say that these two variables converge to the same normal variable in distribution, we only need to compare their covariance matrices, which are obtained from the results for kernel estimates of regression functions.

REFERENCES

GASSER, T. and KNEIP, A. (1995). Searching for structure in curve samples. *J. Amer. Statist. Assoc.* **90** 1179–1188.

GASSER, T., KNEIP, A., BINDING, A., PRADER, A. and MOLINARI, L. (1991). The dynamics of linear growth in distance, velocity and acceleration. *Annals of Human Biology* **18** 187–205.

GASSER, T., KNEIP, A., ZIEGER, P., MOLINARI, L., PRADER, A. and LARGO, R. (1994). Development and outcome of indices of obesity in normal children. *Annals of Human Biology* **21** 275–286.

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

GASSER, T., SROKA, L. and JENNEN-STEINMETZ, C. (1986). Residual variance and residual pattern in nonlinear regression. *Biometrika* **73** 625–633.

HALL, P. (1992a). Effect of bias estimation on coverage accuracy of bootstrap confidence intervals for a probability. *Ann. Statist.* **20** 675–694.

HALL, P. (1992b). On bootstrap confidence intervals in nonparametric regression. *Ann. Statist.* **20** 695–711.

HÄRDLE, W. and BOWMAN, A. (1988). Bootstrapping in nonparametric regression: local adaptive smoothing and confidence bands. *J. Amer. Statist. Soc.* **83** 102–110.

HÄRDLE, W. and MAMMEN, E. (1993). Comparing nonparametric versus parametric regression fits. *Ann. Statist.* **21** 1926–1947.

HÄRDLE, W. and MARRON, J. (1991). Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.* **19** 778–796.

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

KELLY, C. and RICE, J. (1990). Monotone smoothing with application to dose-response curves and the assessment of synergism. *Biometrics* **46** 1071–1085.

MAMMEN, E. (1992). *When Does Bootstrap Work? Lecture Notes in Statist.* **77**. Springer, Berlin.

SILVERMAN, B. (1995). Incorporating parametric effects into functional principal components analysis. *J. Roy. Statist. Soc. Ser. B* **57** 673–689.

DEPARTMENT OF PSYCHIATRY
SUNY HEALTH CENTER
BOX 1203
450 CLARKSON AVENUE
BROOKLYN, NEW YORK 11203
E-MAIL: kmw@mendel.neurodyn.hscbklyn.edu

DEPARTMENT OF BIOSTATISTICS, ISPM
UNIVERSITY OF ZÜRICH
SUMATRASTRASSE 30
CH-8006 ZÜRICH
SWITZERLAND