The Annals of Statistics
1992, Vol. 20, No. 3, 1266–1305

STATISTICAL TOOLS TO ANALYZE DATA REPRESENTING A SAMPLE OF CURVES[1]

BY ALOIS KNEIP AND THEO GASSER

Universität Bonn and Zentralinstitut für Seelische Gesundheit

The paper is concerned with data representing a sample of smooth curves which can be considered as independent realizations of an underlying biological (chemical, ...) process. Such samples of curves often possess the following features: There is a typical structural pattern common to all curves of the sample. On the other hand, individual realizations of the typical shape show different dynamics and intensity. In particular, typical peaks are shifted from individual to individual. Differences in dynamics complicate the analysis of samples of curves. For example, the cross-sectional average usually does not reflect an average pattern. Due to shifts, structure is smeared or might even disappear. Our approach consists in synchronizing the individual curves before determining the average or any further statistics. Pointwise averaging of the synchronized curves then leads to an average curve which represents the common structure with average dynamics and average intensity. The method requires the introduction of new statistical objects. They are defined mathematically, their properties are discussed, and possible estimators are proposed. The asymptotic bias and variance of the estimators are derived. An application to visually evoked brain potentials illustrates the approach.

1. Introduction. Many experiments in biomedicine and in the physical sciences are initiated to study a biological (chemical, ...) process by a number of different realizations. At consecutive times (or ages, ...) $t_{ij} \in J := [a_0, a_1] \subset \mathbb{R}$ observations Y_{ij}, $j = 1, \ldots, n_i$; $i = 1, \ldots, m$, are collected for a sample of individuals (or experimental units) of size m. The assumption that each realization of the underlying process generates a smooth curve then leads to a nonparametric regression model

$$Y_{ij} = f_i(t_{ij}) + \epsilon_{ij}, \qquad i = 1, \ldots, m; j = 1, \ldots, n_i, \tag{1.1}$$

where ϵ_{ij} denotes an unknown zero-mean error term, and where the f_i are unknown smooth functions.

Such data are frequent in statistical practice (e.g., growth curves, brain potentials). If a parametric model were available a priori, subsequent analysis would be simplified. However, at the beginning of a data analysis there does usually not exist enough knowledge to build an appropriate parametric model. Existing models may be seriously deficient in classical fields of application, as

Received September 1989; revised September 1991.

[1]This work was performed as part of the research program of the Sonderforschungsbereich 123 (project B1) at the University of Heidelberg, and was made possible by financial support from the Deutsche Forschungsgemeinschaft.

AMS 1980 *subject classification*. 62G07.

Key words and phrases. Nonparametric curve estimation, samples of curves, asymptotic theory.

could be shown, for example, for modeling human height growth across age [Gasser, Müller, Köhler, Molinari and Prader (1984)].

In such situations a nonparametric analysis of the data is called upon. So far, nonparametric techniques for analyzing single curves have been extensively studied [compare, e.g., Eubank (1988), Müller (1988) and Härdle (1990)]. This paper provides some basic concepts and estimators for analyzing samples of curves. In particular, a meaningful definition and an estimator of some average curve is sought. This curve should both summarize the data and generate hypotheses about the mechanisms generating $f_1, \ldots, f_m$. The cross-sectional average, which is frequently used in practice, is generally not adequate in quantitative or qualitative terms. Due to individual shifts, structure is smeared and might even disappear.

Figure 1 illustrates the problem and approach for five (true) curves from a simulated sample of size $m = 40$ (for details see Section 3.2). The five curves

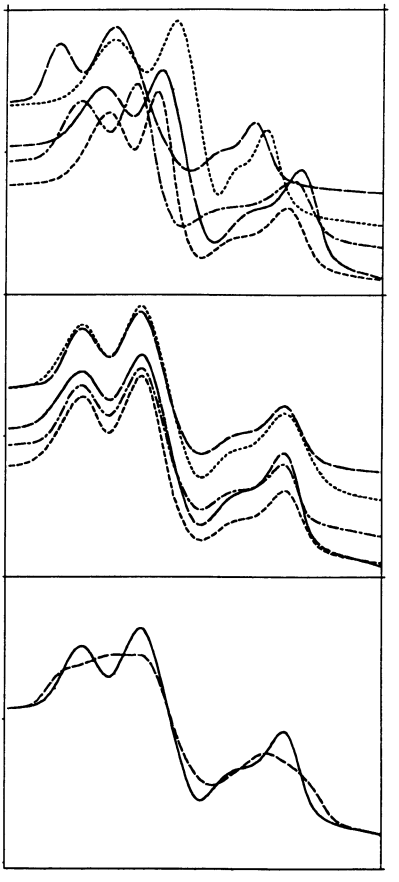

FIG. 1. *Top: five curves from a synthetic process which differ in dynamics and amplitude. Middle: transformed curves obtained after applying time transformations. Bottom: cross-sectional average (dashed curve) versus structural average (solid curve).*

$f_1, \ldots, f_5$ in the upper part reflect some basic features of many real samples of curves:

(i) There is a typical structural pattern common to all curves of the sample.
(ii) Individual realizations of the typical shape show different dynamics and intensity.

Admittedly, the example is simplified. In practice, we have to estimate the curves from noisy data, using nonparametric regression methods. Furthermore, the common shape of real samples of curves may not be as regular.

The basic idea is to synchronize the individual curves before determining the average or any further statistics. Herëby, we rely on structural characteristics like extrema or inflection points occurring in a manner which is qualitatively *equivalent* or *corresponding* in different curves (this vague idea will be partially formalized in Section 2). For the curves in the upper part of Figure 1, this holds for each of the five common extrema. In a first step we determine a set of locations of corresponding structural characteristics. In our example:

1. The individual locations $\tau_{1i}, \ldots, \tau_{5i}$, $i = 1, \ldots, m$, of the five extrema present in all curves are determined.

In practice, such locations have to be estimated from the data, using estimated curves. In the example, the assignment of corresponding structural characteristics is given rather straightforwardly by the succession of extrema. In general, this can be a more delicate problem. However, the decision on the equivalence of features will often be clear from the specific application or by common sense. For example, well-known common characteristics of the human growth process are the so-called mid-growth spurt and the pubertal growth spurt. Biomedical knowledge permits the assignment of—estimated—extrema in the growth velocity curves to these spurts.

Corresponding structural characteristics often indicate equivalent states of the underlying process. Different locations then quantify individual (local) shift differences. Thus, we can eliminate differences in dynamics by aligning individual locations of corresponding structural characteristics to their average locations. Since we want to synchronize *curves*, this can be combined with an interpolation step resulting in smooth, strictly monotone time transformations for individual functions. (the notion "time" should not preclude any other interpretation). In our example the procedure is as follows:

2. Smooth, strictly monotone time transformations g_i are defined, assigning $\tau_{1i}, \ldots, \tau_{5i}$ to their average locations (for details see Section 3).
3. Synchronized curves $f_i^* := f_i(g_i(\cdot))$ are determined (middle part of Figure 1).
4. Pointwise averaging of the synchronized curves yields the *structural average* (lower part of Figure 1).

Evidently, the structural average represents the common pattern better and is more informative than the cross-sectional average.

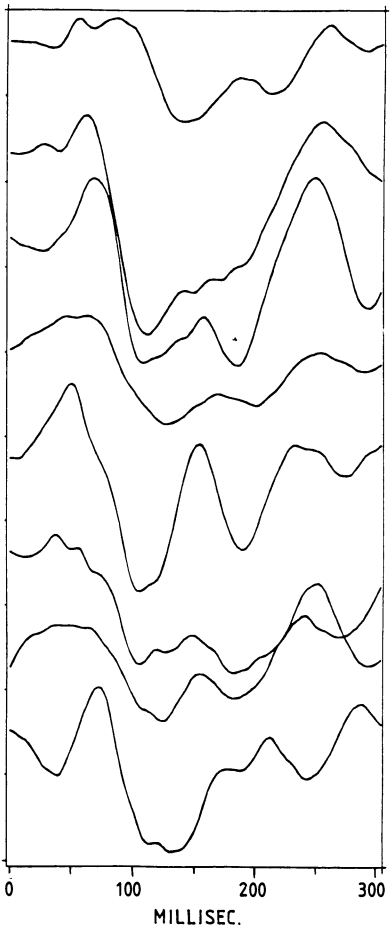

FIG. 2. *Evoked potentials to flashes of light (slightly smoothed) for eight healthy children (10–12 years old) recorded over the visual sensory area.*

The usefulness of the new method is well illustrated for real data, consisting of brain potentials evoked by external stimuli [Brillinger (1988) and Gasser, Pietz, Schellberg and Köhler (1988); the data to be presented stem from the latter paper]. The curves given in Figure 2 represent slightly smoothed average evoked potentials to flashes of light, recorded over the occipital cortex. They belong to 8 healthy children, aged 10–12 years, who are a subsample of $m = 42$ children. We recognize a common structural pattern. At the same time, we note a substantial variation in amplitude but also in the dynamics (called "latency" in neurophysiology) of individual processing of stimuli by the brain.

Traditionally, the cross-sectional average of individual potentials has been used as the average waveform of a group. Subsequently, different groups, or

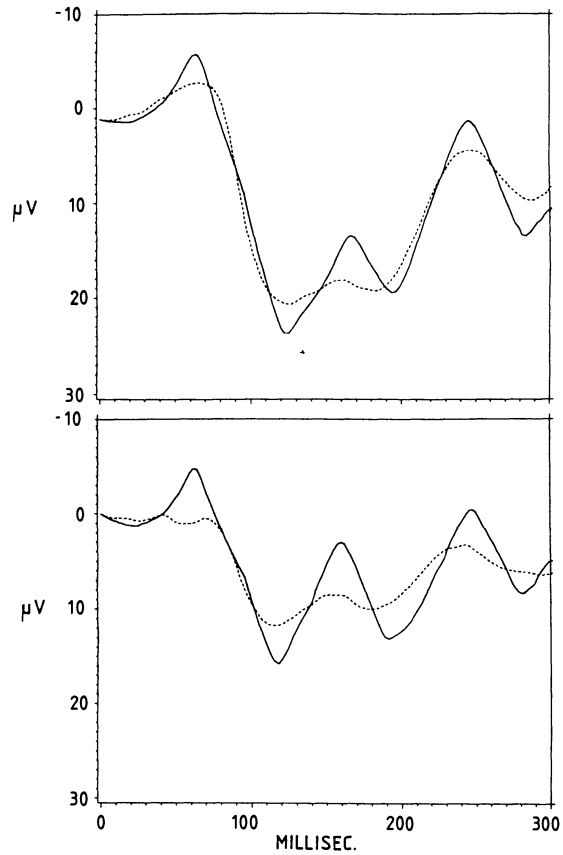

FIG. 3. *Top: cross-sectional average (dotted curve) versus structural average (solid curve) of evoked potentials of* $m = 42$ *healthy children. Bottom: same for* $m = 25$ *mildly retarded children.*

different experimental conditions, have been compared graphically to depict neurophysiological functioning. The estimated cross-sectional average is compared with the estimated structural average in Figure 3 for a sample of healthy children ($m = 42$) and a sample of mildly mentally retarded children ($m = 25$), both aged 10–12 years. The structure of the average potentials conforms to expectation taken from the literature, with maxima at approximately 60, 165 and 245 msec and minima at 130 and 190 msec. The component at 60 msec is considered to be related to primary sensory processing, whereas the latter ones are influenced by sensory and psychological factors.

For the group of healthy children, the cross-sectional average and the structural average show much the same structure, but it is more accentuated for the latter method. Strange enough, the cross-sectional average does not show the primary component at 60 msec for the retarded children (their individual acuity had been tested). This paradox can be solved by considering the structural average, which shows this component in about the same size as

for healthy children. Its disappearance in the cross-sectional average can be attributed to the greater variability in latencies of evoked potentials in retarded children [Gasser, Pietz, Schellberg and Köhler (1988)].

This method has also been applied to longitudinal growth data. It led to new insight into the mechanisms of normal growth [Gasser, Kneip, Binding, Prader and Molinari (1991)] and in the pattern of stunted growth due to chronic renal failure [Schäfer, Seidel, Binding, Gasser, Largo, Prader and Schärer (1990)].

The new approach requires the introduction of mathematical entities at a conceptual level. Therefore, work to be done is threefold: Appropriate statistical objects have to be defined mathematically, their properties have to be discussed and possible estimators have to be considered. In Section 2 we start with the definition of structural functionals which enable, for example, the localization of corresponding extrema. The subsequent estimators rely on nonparametric curve estimation. Section 3 is first devoted to the construction of time transformations which enable the synchronization of individual curves. The definition of a structural average is then straightforward. Section 4 deals with connections to semiparametric model building. An asymptotic theory for the estimators proposed is treated in Section 5. It gives a solid basis for our intuitive idea that an average sample curve can be determined more reliably than an individual curve. The proofs are deferred to the Appendix.

2. Structural functionals.

2.1. *Concepts.* First, we formalize the identification of locations of typical features in individual curves. An example are the locations of extrema in the synthetic data of Figure 1. In general, such a location is obtained by applying a functional T which maps elements of the space of regression functions to time.

We start with some notation: Let $H \subseteq \mathbb{R}$. For any $\nu \in \mathbb{N} \cup \{0\}$, $C^\nu(H)$ will denote the set of all ν times continuously differentiable functions from H into $\mathbb{R}$. As far as this is necessary, for compact H, $C^\nu(H)$ will be endowed with the norm $\|v\|_H^{(\nu)} := \sum_{s=0}^{\nu} \sup_{t \in H} |v^{(s)}(t)|$, $v \in C^\nu(H)$. Note that for any $\mu \leq \nu$, $C^\nu(H) \subset C^\mu(H)$, and $\|v - w\|_H^{(\nu)} < \gamma$ implies $\|v - w\|_H^{(\mu)} < \gamma$, $v, w \in C^{(\nu)}(H)$; $\gamma > 0$.

Recall that $t_{ij} \in J := [a_0, a_1] \subset \mathbb{R}$. In the following discussion it will be assumed that the true regression curves $f_1, f_2, \ldots$ of our sample are elements of $C^{(\nu)}(J)$ for some $\nu \geq 2$.

The shape of a curve is to some extent quantified by the locations and amplitudes of its extrema. Functionals T localizing *corresponding* extrema in different curves necessitate some identification procedure. This might be based on knowledge from the field of application: The traditional approach for evoked potentials uses the absolute maximum (or minimum) in some prespecified time window according to prior experience.

Generally, the following principles can be stated: First, if no unique identification is possible for some function v, we set $T(v) := \alpha$ for some prespecified $\alpha \notin J$ ("missing"). Second, continuity may be used in the sense that for curves

v, w close to each other $T(v)$ and $T(w)$ should also be close, if neither is missing. This is supported by the following proposition.

PROPOSITION 1. *Let $v \in C^{\nu}(J)$ for some $\nu \geq 2$, and let x denote the location of an extremum of v with $\operatorname{sgn}(v''(x)) \neq 0$. Then there exist an $\varepsilon > 0$ and some neighborhood $U \subset C^2(J)$ (referring to $\|\cdot\|_J^{(2)}$) of v such that for any $w \in U$ there is exactly one $x_w \in [x - \varepsilon, x + \varepsilon]$ with $w'(x_w) = 0$ and $\operatorname{sgn}(w''(x_w)) = \operatorname{sgn}(v''(x))$.*

PROOF. The assumptions on v and x imply the existence of some $\varepsilon, \gamma > 0$ such that $|v'(x - \varepsilon)| \geq \gamma$, $|v'(x + \varepsilon)| \geq \gamma$, $\operatorname{sgn}(v'(x - \varepsilon)) = -\operatorname{sgn}(v'(x + \varepsilon))$ and $|v''(t)| \geq \gamma$ for all $t \in [x - \varepsilon, x + \varepsilon]$. The assertion then follows with $U := \{w \in C^2(J) \mid \|v - w\|_J^{(2)} < \gamma\}$. □

We will use $\mathscr{D}_T$ to denote the space of all functions v in the domain of T with $T(v) \neq \alpha$, that is, not missing. The above arguments motivate the following definition.

DEFINITION 1. A functional $T\colon C^2(J) \to]a_0, a_1[\cup \{\alpha\}$ (for some $\alpha \notin J$) is called a *first-order structural e-functional* if the following conditions are satisfied:

(i) T is continuous on $\mathscr{D}_T$.

(ii) For all $v, w \in \mathscr{D}_T$, $v(T(v))$ and $w(T(w))$ are either both local maxima or both local minima of v and w, and $\operatorname{sgn}(v''(T(v))) = \operatorname{sgn}(w''(T(w))) \neq 0$.

To give an example, let $H := [c, d] \subseteq J$. Then

$$T(v) := \begin{cases} \arg\sup_{t \in H} v(t), & \text{if } v \text{ possesses a unique supremum in }]c, d[, \\ \alpha, & \text{else,} \end{cases} \tag{2.1}$$

is a first-order structural e-functional. Another e-functional may be specified by replacing "supremum" by "infimum." This establishes the type of structural functional that might be used for evoked potentials.

Additionally, one might quantify the rate of increase in *corresponding* monotone segments of curves. Points where a specified percentage of total increase or decrease is reached might be useful, and such points will be called *p-points* (p for percentage).

DEFINITION 2. For some $p \in]0, 1[$ a functional $T\colon C^2(J) \to]a_0, a_1[\cup \{\alpha\}$ (for some $\alpha \notin J$) is called a *first-order structural p-functional* if the following conditions are satisfied:

(i) T is continuous on $\mathscr{D}_T$.

(ii) For all $v \in \mathscr{D}_T$ it holds that

$$v(T(v)) = pv(\psi_{0,v}) + (1 - p)v(\psi_{1,v}),$$

where $\psi_{0,v}, \psi_{1,v}$ are successive local extrema of v, and v is strictly monotone in $[\psi_{0,v}, \psi_{1,v}]$. Moreover, either $\psi_{0,v}$ is a local maximum for all v [with $T(v) \neq \alpha$] or it is a local minimum for all v [with $T(v) \neq \alpha$].

In the following we will just say "structural functionals" to denote both structural e- and p-functionals. A further possibility for quantifying individual locations of equivalent structural features is to use, for example, *corresponding* inflection points. This leads to structural functionals for the derivatives.

DEFINITION 3. For some $u \in \mathbb{N}$ a functional T: $C^{u+2}(J) \to]a_0, a_1[\cup \{\alpha\}$ is called a structural functional of order $u + 1$ if there exists a first-order structural functional T_u: $C^2(J) \to]a_0, a_1[\cup \{\alpha\}$ with $T(v) = T_u(v^{(u)})$ for all $v \in C^{u+2}(J)$.

For a structural functional T and for any v, $T(v)$ will be called a *structural point* of v based on T.

2.2. *Practical aspects and estimation.* Following the definitions in Section 2.1, the first step of our approach is to specify a number of structural functionals $T_1, \ldots, T_l$, $l \in \mathbb{N}$.

Evidently, there are many alternative specifications. However, the problem is not to define just any structural functional, but rather to specify functionals which make sense for the given sample of curves and the given problem. They have to quantify individually differing locations of *corresponding* extrema or p-points. Obviously, this necessitates defining structural functionals in such a way that no, or at least very few, missing structural points are to be expected for the true curves. Apart from prior knowledge on the underlying mechanisms, or some other kind of substantive knowledge, the data themselves may suggest reasonable specifications of structural functionals. A simple possibility is visual inspection of nonparametric estimates $\hat{f}_1, \ldots, \hat{f}_m$.

A heuristic automatic procedure has also been developed and successfully applied [Gasser and Kneip (1991)]. It has been tailored to the case, which frequently arises in practice, that interindividual differences between the locations of corresponding structural characteristics are relatively small compared to intraindividual differences between the locations of different structural characteristics (compare the evoked potential data of Figure 2). Without loss of generality let us consider the problem of defining first-order structural functionals identifying common maxima:

(i) First, we try to retrieve all local maxima in each individual curve. This is done by computing the appropriate zeros of the first derivative by kernel estimation.

(ii) All zeros of all curves are sorted into one array and submitted to kernel density estimation.

Common maxima will then give rise to peaks in the resulting density. This leads in a straightforward way to a specification of structural functionals

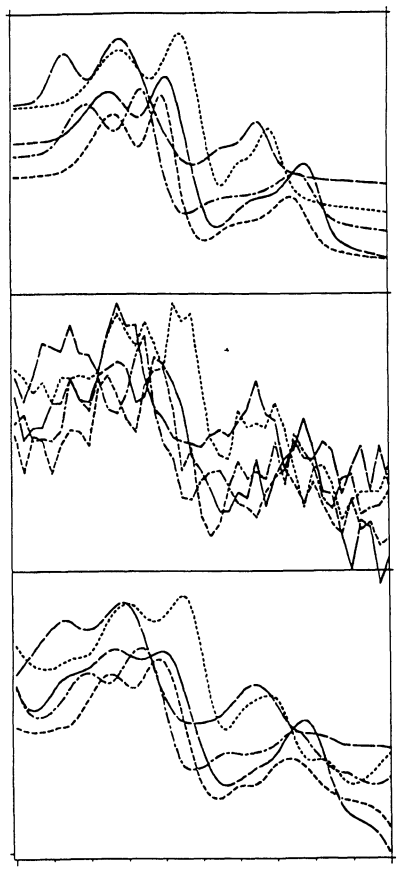

FIG. 4. *Top: five curves from a synthetic process (same as Figure 1). Middle: Simulated data, generated by adding normal residuals to 40 equally spaced functional values of the above curves, followed by linear interpolation. Bottom: Curves obtained from "data" by kernel estimation.*

following (2.1). For the evoked potential example, this method brought out the typical components, that is, extrema, known in neurophysiology. In this example the method allows a more rational selection of sharp time windows, that is, intervals on the time axis where specific extrema are sought.

The exact definition of structural functionals needs some care. This is due to the possible occurrence of random peaks and the possible omission of true peaks in the estimated curves. For example, the simple rule of using successive extrema—as for the simulated true curves in Section 1—will usually not work. The middle part of Figure 4 shows simulated noisy data generated from the synthetic curves of Figure 1. Data are linearly interpolated. The curves estimated from these data reflect the structure of the true curves quite well, but there are some additional small wiggles (lower part of Figure 4; for details see Section 3). Therefore, more robust and more complicated definitions of structural functionals were used. For $v \in C^2(J)$ let E_v denote the set of all

$t \in]a_0, a_1[$ with $v'(t) = 0$. The following specifications are in line with the assignments of *corresponding* structural features one might intuitively use, based on estimated curves.

It is graphically evident that all curves have a unique and corresponding minimum approximately in the middle of the abscissa just after the steep decline. For the true curves it is always the fourth local extremum. For true and *estimated* curves it is a global minimum when excluding the last third of the abscissa:

$$T_4(v) := \begin{cases} \arg \inf_{t \in]a_0, a_1^*[} v(t), & \text{if } v \text{ has a unique infimum in }]a_0, a_1^*[, \\ \alpha, & \text{else.} \end{cases}$$

Here, $a_1^* := a_0 + 2(a_1 - a_0)/3$. $T_3(v)$ is the location of the one local maximum just preceding $T_4(v)$ (always the third local extremum and global maximum in the true curves):

$$T_3(v) := \begin{cases} t_v = \max\{t \in E_v | t < T_4(v)\}, & \text{if } T_4(v) \neq \alpha,\ t_v \text{ exists},\ v''(t_v) < 0, \\ \alpha, & \text{else.} \end{cases}$$

Similarly, structural functionals T_1, T_2, T_5 are defined based on the largest local maximum preceding $T_3(v)$, the global minimum in between $T_1(v)$ and $T_3(v)$, and the global maximum in $[T_3(v), a_1]$. For the true curves this leads to the same structural points as in Section 1. By construction, $T_1(v) < T_2(v) < \cdots < T_5(v)$ holds if all structural points are not missing.

Estimates of the structural points $\tau_{ri} := T_r(f_i)$ are determined from the estimated curves (compare Figures 2 and 4). The procedure can in general terms be described as follows: For any $r \in \{1, \ldots, l\}$ determine nonparametric estimates $\tilde{f}_1, \ldots, \tilde{f}_m$ of $f_1, \ldots, f_m$, using, for example, smoothing splines or kernel estimators. Estimate the structural points $\tau_{r1}, \ldots, \tau_{rm}$ by $\hat{\tau}_{r1} := T_r(\tilde{f}_1), \ldots, \hat{\tau}_{rm} := T_r(\tilde{f}_m)$.

Identifying structural points will, in practice, often require estimating derivatives. Even when determining typical extrema of the curve itself, it will usually be advisable to look for the corresponding zeros of the first derivatives. The respective smoothing method should thus be chosen in a way to guarantee the best possible estimates of f_i', for example. For kernel estimators this might be realized by using specific kernels, tailored to estimating derivatives [compare Gasser, Müller, Köhler and Mammitzsch (1985)].

REMARK. When applying kernel estimators, an appropriate choice of bandwidths is important. Following the theoretical results of Müller (1985, 1989), optimal bandwidths for estimating individual locations of extrema and for estimating the derivative of an individual curve are approximately of the same order of magnitude. The latter bandwidth might be estimated as in Gasser, Kneip and Köhler (1991). It has to be emphasized, however, that in order to obtain a best estimate of a structural average mildly undersmoothing bandwidths (i.e., smaller than the individually optimal ones) have to be used. This is justified theoretically in Section 5.

2.3. *Theoretical considerations.* This section deals first with a formalization of the mechanisms generating the regression functions f_i. Then a uniqueness result is given which provides some insight into the problem of specifying reasonable structural functionals.

As outlined in Section 1, we assume the sample $\{f_i\}_{i=1,\dots,m}$ of true regression functions to be generated by common underlying mechanisms. At the same time, individual realizations will be governed by unknown factors. Together with the assumption of smoothness, this leads to the following formalization:

(2.2) There is some set Ω and a mapping $A\colon \Omega \to C^\nu(J)$, $\nu \geq 2$, such that for any $i \in \{1,\dots,m\}$; $f_i = A(\omega_i)$ holds for some $\omega_i \in \Omega$.

Throughout this paper it is assumed that (2.2) holds. If (2.2) is supplemented by a probability measure based on appropriate σ-algebras, this is equivalent to assuming that $f_1, f_2, \dots$ are independent realizations of a function-valued random variable. This assumption is reasonable when dealing, for example, with the visually evoked potentials of Figure 2.

In most applications it can be assumed that the physical processes are such that all functions in $\mathscr{R} := A(\Omega)$ are uniformly bounded and of a comparable degree of smoothness. Furthermore, one will suppose a continuum of possible "factors" ω, as well as some homogeneity of the population studied, as is the case in most statistical approaches. If there is evidence of distinct subgroups, such as boys and girls when studying human growth, a separate analysis should be performed. Formally, the following additional assumption will thus not impose a severe restriction:

(2.3) $\mathscr{R} := A(\Omega)$ is a compact, connected subspace of $C^\nu(J)$.

Necessary and sufficient for compactness of $\mathscr{R}$ is that

$$\sup_{t\in J} \sup_{\mu=0,1,\dots,\nu} \sup_{v\in\mathscr{R}} |v^{(\mu)}(t)| < \infty,$$

and that $\mathscr{R}^{(\nu)} := \{v^{(\nu)} | v \in \mathscr{R}\}$ is equicontinuous. For compact $\mathscr{R}$, connectedness means that there do not exist two sets $\mathscr{R}_1, \mathscr{R}_2 \neq \emptyset$ with $\mathscr{R} = \mathscr{R}_1 \cup \mathscr{R}_2$ such that $\inf_{v\in\mathscr{R}_1} \inf_{w\in\mathscr{R}_2} \sup_{t\in J} |w(t) - v(t)| > 0$.

Assumption (2.3) refers to the topology induced by $\|\cdot\|_J^{(\nu)}$. However, for any $\mu \leq \nu$, compactness and connectedness of $\mathscr{R}$ with respect to $\|\cdot\|_J^{(\nu)}$ carries over to compactness and connectedness of $\mathscr{R}$, when referring to $\|\cdot\|_J^{(\mu)}$.

Now, assume (2.2) and (2.3), and let us return to the problem of specifying appropriate structural functionals. Such specifications can be done in many different ways. The question arises whether this leads to arbitrariness. This is not the case for functionals T localizing structural features which occur consistently in each possible realization of the underlying process. They have to satisfy

(2.4) $$T(v) \neq \alpha \quad \text{for all } v \in \mathscr{R}.$$

We will expect that (2.4) holds for each of the structural functionals specified

in our real and synthetic examples. Proposition 2 shows that two functionals of the same order satisfying (2.4) are either identical or completely different on $\mathscr{R}$. In other words, given $T(v)$ for *one* curve $v \in \mathscr{R}$, $\mathscr{R}$ itself uniquely determines *all* further values $T(w)$, $w \in \mathscr{R}$.

PROPOSITION 2. *Assume* (2.2) *and* (2.3). *For some* $u \leq \nu - 2$ *let* T: $C^{u+2}(J) \to]a_0, a_1[\cup \{\alpha\}$ *and* T^*: $C^{u+2}(J) \to]a_0, a_1[\cup \{\alpha\}$ *be structural functionals of order* $u + 1$, *and assume that* $T(v) \neq \alpha$, $T^*(v) \neq \alpha$ *for all* $v \in \mathscr{R}$. *Then one of the following relations is fulfilled*:

(a) $T(v) < T^*(v)$ *for all* $v \in \mathscr{R}$.
(b) $T(v) = T^*(v)$ *for all* $v \in \mathscr{R}$.
(c) $T(v) > T^*(v)$ *for all* $v \in \mathscr{R}$.

PROOF. Note that, as indicated above, $\mathscr{R}$ is also compact and connected when referring to $\|\cdot\|_J^{(u+2)}$. By the definition of a structural functional, it follows that T and T^* are continuous on $\mathscr{R}$ (referring to $\|\cdot\|_J^{(u+2)}$).

Now, let $\mathscr{R}_1 := \{w \in \mathscr{R} | T(w) < T^*(w)\}$, $\mathscr{R}_2 := \{w \in \mathscr{R} | T(w) = T^*(w)\}$ and $\mathscr{R}_3 := \{w \in \mathscr{R} | T(w) > T^*(w)\}$. Continuity of T and T^* on $\mathscr{R}$ implies that we obtain for $s = 1, 3$:

(2.5) For any $v \in \mathscr{R}_s$ there exists an open neighborhood $U_s(v) \subset C^{u+2}(J)$ such that $U_s(v) \cap \mathscr{R} = U_s(v) \cap \mathscr{R}_s$.

Consider $\mathscr{R}_2$. The definition of structural functionals implies that $\mathscr{R}_2$ is nonempty only if either both T and T^* localize local maxima or both localize local minima or both localize p-points (for the same $p \in]0, 1[$). Note that if T, T^* are structural e-functionals, $v^{(u+2)}(T(v)) \neq 0$ and $v^{(u+2)}(T^*(v)) \neq 0$ hold for all $v \in \mathscr{R}$. Continuity on $\mathscr{R}$ and the results of Proposition 1 then imply that for any $v \in \mathscr{R}_2$ we have $x_w = T(w) = T^*(w)$ for all $w \in \mathscr{R}$ in a sufficiently small neighborhood of v. It can easily be verified that this generalizes to the case that both T and T^* are structural p-functionals for the same $p \in]0, 1[$. This shows that (2.5) also holds for $s = 2$. We can now conclude from (2.5) that there exist open subspaces $D_1 \supseteq \mathscr{R}_1$, $D_2 \supseteq \mathscr{R}_2$ and $D_3 \supseteq \mathscr{R}_3$ of $C^{u+2}(J)$ such that $D_1 \cap D_2 \cap \mathscr{R} = \emptyset$, $D_1 \cap D_3 \cap \mathscr{R} = \emptyset$, $D_2 \cap D_3 \cap \mathscr{R} = \emptyset$ and $\mathscr{R} \subset D_1 \cup D_2 \cup D_3$. Only one of these open spaces D_1, D_2, D_3 is nonempty, as follows from the connectedness of $\mathscr{R}$. □

The assertion of the proposition is not necessarily true if T and T^* are of different order. Note that (2.4) is not postulated for the estimated curves.

3. Synchronizing regression curves. Assume that l structural functionals $T_1, \ldots, T_l$, $l \in \mathbb{N}$, have been specified, the orders not necessarily being equal. We require that these functionals be defined such that $T_r(v) < T_s(v)$ holds for $r < s$, when $T_r(v) \neq \alpha$ and $T_s(v) \neq \alpha$. Some notation will be used: Set $\mathbf{T}(v) := (T_1(v), \ldots, T_l(v))'$. Furthermore, let $\mathscr{H}_l$ denote the set of all $x = (x_1, \ldots, x_l)' \in (J \cup \{\alpha\})^l$ with $x_r < x_s$ for all $r < s$ with $x_r \neq \alpha$ and $x_s \neq \alpha$.

3.1. *Concepts.* Let us first study the synchronization of the true curves $f_1, \ldots, f_m$. By definition, structural functionals are identifying individual locations of *corresponding* structural characteristics. They may thus be used to quantify local shift differences. Our approach consists in assigning to each f_i a (local) shift function g_i which transforms this curve to a common time scale. The mathematical formalization by an operator G, assigning a shift function g_i to each individual curve f_i, is based on the following ideas:

1. For all $i, j \in \{1, \ldots, m\}$ differences between g_i and g_j may only depend on differences between the structural points $\tau_{ri} := T_r(f_i)$ and $\tau_{rj} := T_r(f_j)$, $r = 1, \ldots, l$, since there is no further information about shift differences. Shift functions have to be strictly monotonically increasing.
2. Structural points are aligned to fixed $x_1, \ldots, x_l \in \mathbb{R}$ such that $g_i(x_r) = T_r(f_i)$, $r = 1, \ldots, l$, for all i [then $f_i(g_i(x_r)) = f_i(T_r(f_i))$].
3. It is most natural to align structural points to their average locations, that is, $x_1 = \bar{\tau}_1, \ldots, x_l = \bar{\tau}_l$. Here, for $r \in \{1, \ldots, l\}$, $\bar{\tau}_l$ denotes the average of all τ_{ri} with $\tau_{ri} \neq \alpha$ (i.e., not missing).
4. In line with the smoothness of individual curves, one will ask for some smoothness of g_i.

This leads to the following definition of shift operators and shift functions.

DEFINITION 4.

(I) An operator G: $\mathscr{H}_l^2 \to C^1(\mathbb{R})$ is called a *shift operator* if the following conditions are satisfied:

(i) For all $(\tau, x) \in \mathscr{H}_l^2$, $G_{(\tau, x)}(\cdot)$ is a strictly monotonically increasing real function. Furthermore, for all $t \in \mathbb{R}$, $G_{(\cdot, \cdot)}(t)$ is continuous.

(ii) For all $\tau = (\tau_1, \ldots, \tau_l)' \in \mathscr{H}_l$ and all $x = (x_1, \ldots, x_l)' \in \mathscr{H}_l$,

$$G_{(\tau, x)}(x_r) = \tau_r$$

for all $r = 1, \ldots, l$ with $\tau_r \neq \alpha$, $x_r \neq \alpha$.

(II) For some shift operator G,

$$g_i(\cdot) := G_{(\mathbf{T}(f_i), \bar{\tau})}(\cdot)$$

is called a *shift function* for f_i based on G. Hereby, $\bar{\tau} := (\bar{\tau}_1, \ldots, \bar{\tau}_l)'$.

As an example $G_{(\tau, x)}$ might be a function defined by smooth, strictly monotonical interpolation of the points $(\tau_1, x_1), \ldots, (\tau_l, x_l)$.

It immediately follows that $(1/m)\sum_{i=1}^m g_i(\bar{\tau}_r) = \bar{\tau}_r$, $r = 1, \ldots, l$, if there are no missing structural points for the true curves. Ideally, one might wish that the average of the shift functions corresponds to the identity. This can be realized by determining normalized shift functions $g_i^* := g_i \circ ((1/m)\sum_{k=1}^m g_k)^{-1}$. In practice, problems might arise when there is a substantial proportion of missing structural points.

By applying shift functions, a synchronized sample is obtained as $\{f_i^*\}_{i \in \{1, \ldots, m\}} = \{f_i(g_i(\cdot)\}_{i \in \{1, \ldots, m\}}$. For these curves differences in individ-

ual dynamics are partially eliminated, and remaining variation will be attributed mainly to intensity.

DEFINITION 5. Let G be a shift operator, and let g_i denote the respective shift functions based on G. Then

$$f(t) := \frac{1}{m}\sum_{i=1}^{m} f_i(g_i(t)), \qquad t \in J^* := \left[\sup_i g_i^{-1}(a_0), \inf_i g_i^{-1}(a_1)\right],$$

is called the *structural average* of $\{f_i\}_{i\in\{1,\ldots,m\}}$ based on $G, \mathbf{T}$.

Often we cannot define f on the whole interval J since the domain of $f_i(g_i(\cdot))$ is $[g_i^{-1}(a_0), g_i^{-1}(a_1)]$ which may not be equal to J. However, J^* necessarily contains the interval $[\bar{\tau}_1, \bar{\tau}_l]$. This problem is associated with the experimental design. For many processes there will exist a known starting point $a_S \le a_0$ and a known end point $a_E \ge a_1$. For the evoked potential example, the starting point is naturally defined to be the stimulus ($a_S = a_0 = 0$). Neurophysiological experience suggests an end point at about 800 msec. When studying human growth processes across age, one can assume that any such process starts at conception and ends before 24 years. It is reasonable to postulate that there are no individual shift differences at such starting or end points. Hence, if these points are known we might use them when defining G, setting $g_i(a_S) = a_S$ and $g_i(a_E) = a_E$. In particular, if we can assume that our measurements cover the whole range of the underlying process, that is, $a_0 = a_S$ and $a_1 = a_E$, we might set $g_i(a_0) = a_0$ and $g_i(a_1) = a_1$. In this case we obtain $J^* = J$. This assumption was made for the simulated data.

As discussed in Section 2.3, the true curves $f_1, f_2, \ldots$ should, in many applications, be formalized as independent realizations of a function-valued random variable F, say. Then a structural average f can be considered as an estimate of the *structural mean* $\mathbf{f} := EF(G_{(\mathbf{T}(F),\, E\mathbf{T}(F))}(\cdot))$, provided the moments exist (E denotes expectation).

3.2. *Practical aspects.* Obviously, there is some arbitrariness in the definition of a shift operator G and of the resulting shift functions. Only the values of g_i at the points $\bar{\tau}_1, \ldots, \bar{\tau}_l$ are fixed. In between these points it is merely required that g_i be a smooth, strictly monotonically increasing function. In some cases (compare Section 4) there will be a priori knowledge determining g_i in a unique way. Otherwise, one might use the approach discussed below. Increasing l by including a further functional T_{l+1} with, for example, $T_r(f_i) < T_{l+1}(f_i) < T_{r+1}(f_i)$ for some $r < l$, results in fixing the values of g at a further point $\bar{\tau}_{l+1} \in]\bar{\tau}_r, \bar{\tau}_{r+1}[$. It is easy to see that, at least if l is reasonably large, the defining properties of G and g_i are quite restrictive and one possible shift function *cannot* be very different from any other. Increasing the number of structural functionals in order to fix the values of g_i at additional points will result merely in marginal changes of the transformed curves.

If no knowledge in between structural points is available, ad hoc procedures might be used to define shift functions. The simplest one is piecewise linear interpolation of the individual structural points at their mean locations. This violates, however, the smoothness requirement. It is not trivial to find a flexible class of smooth interpolants which guarantee strict monotonicity. This is, for instance, not the case for interpolating splines.

Treating known starting and end points a_S and a_E like additional structural points, the following general procedure might be applied [recall that $\tau_{ri} := T_r(f_i)$]:

1. For $l = 0$ let $g_i(t) := t$.
2. For $l = 1$ let $g_i(t) := t + \tau_{1i} - \bar{\tau}_1$.
3. For $l = 2$ let g_i be the straight line with $g_i(\bar{\tau}_1) = \tau_{1i}$ and $g_i(\bar{\tau}_2) = \tau_{2i}$.
4. For $l > 2$ set $g_i(\bar{\tau}_1) = \tau_{1i}, \ldots, g_i(\bar{\tau}_l) = \tau_{li}$. Keeping up with monotonicity, use monotone piecewise cubic interpolants as proposed by Brodie (1980) and investigated by Fritsch and Butland (1984) to determine $g_i(t)$ for $t \in [\bar{\tau}_1, \bar{\tau}_l]$. Outside this interval extrapolate linearly.
5. If some structural points are missing, apply the above rules to the remaining ones.

These rules implicitly specify a shift operator G. Replacing $\tau_{ri}, \bar{\tau}_r$ by x_r, x_r^*, this procedure might be applied to any $(x, x^*) \in \mathscr{H}_l^2$, and it is easy to see that it satisfies the conditions of Definition 4. Resulting shift functions are once continuously differentiable, but not more. A method to obtain twice differentiable functions has recently been published by Kelly and Rice (1990).

3.3. *Estimation*. In practice, shift functions and structural averages have to be estimated. Based on a shift operator G, estimates $\hat{g}_i$ of the shift functions g_i are obtained by

$$\hat{g}_i(\cdot) := G_{(\hat{\tau}_i, \bar{\hat{\tau}})}(\cdot). \tag{3.1}$$

Hereby, $\hat{\tau}_i = (\hat{\tau}_{1i}, \ldots, \hat{\tau}_{li})'$ and $\bar{\hat{\tau}} = (\bar{\hat{\tau}}_1, \ldots, \bar{\hat{\tau}}_l)'$, where $\hat{\tau}_{ri}$ denote estimates of the structural points $\tau_{ri} = T_r(f_i)$ (compare Section 2.2), and $\bar{\hat{\tau}}_r$ are the resulting averages. Again averages are determined by omitting all missing structural points.

Estimated shift functions $\hat{g}_i$ lead to estimates of the synchronized curve $\hat{f}_i(\hat{g}_i(\cdot))$, and straightforwardly to an *estimated structural average*,

$$\hat{f}(t) := \frac{1}{m} \sum_{i=1}^{m} \hat{f}_i(\hat{g}_i(t)), \tag{3.2}$$

which estimates the structural average f for every

$$t \in J^* \cap \left[\sup_i \hat{g}_i^{-1}(a_0), \inf_i \hat{g}_i^{-1}(a_1) \right].$$

Here, $\hat{f}_i$ is a nonparametric estimate of f_i, for example a kernel estimate, which involves the choice of a bandwidth. Theoretical considerations and practical experience show that a mildly undersmoothing bandwidth is advis-

able in this step. Undersmoothing means to choose a smaller bandwidth than the optimal one for estimating an individual curve. The following argument gives some insight (for details see Section 6): Under suitable conditions, it can be shown that the expansion

$$\hat{f}(t) = f(t) + \frac{1}{m}\sum_{i=1}^{m}\left(\hat{f}_i(g_i(t)) - f_i(g_i(t)) + \frac{\partial}{\partial x} f_i\big(G_{(x,\bar{\tau})}(t)\big)\Big|_{x=\tau_i} \cdot (\hat{\tau}_i - \tau_i)\right)$$

$$+ \text{terms of smaller order}$$

holds for large m and n. By undersmoothing individual curves, we reduce the bias, while averaging curves by itself diminishes the variance (first term in the sum). A similar argument applies to the estimators of the structural points, and the choice of their bandwidths, as can be seen from the second term in the sum. Asymptotic theory demonstrates that in this way for large m much better rates can be obtained for the estimated structural average than for an individual nonparametric estimate.

This approach is exemplified by our previous simulated example. Based on a prespecified basic curve ϕ, the true individual curves were generated according to a model $f_i(t) := S_{\theta_i}(\phi(s_{\psi_i}(t)))$, where $S_{\theta_i}(t)$ and $s_{\psi_i}(t)$ were specified parametric functions, monotone in t. For simulation, parameters were selected randomly to generate 40 different individual curves. For each of these 40 curves, Gaussian noise was added to 40 equally spaced functional values in order to simulate observations Y_{ij} satisfying model (1.1) (to get an idea of the variance used, see Figure 4). These observations then served as input to the procedure described above, using kernel estimators. Optimal bandwidths were estimated from the individual data following Gasser, Kneip and Köhler (1991), and averaged. A mildly undersmoothing bandwidth equal to 0.75 times the average optimal one then was used in all steps. Computation of an estimated structural average was based on the five structural functionals defined in Section 2.2. The shift operator G was chosen according to Section 3.2.

This simulation procedure was used to generate and analyze 101 different samples of 40 curves. The resulting 101 estimated structural averages were compared to the true ones. Figure 5 shows the realization with median integrated squared error out of the 101 replicates. The estimated structural average is surprisingly close to the true one. Both represent the common structure well, which is evidently not true for the cross-sectional average. The individual nonparametric estimates (Figure 4) do not show the small shoulder occurring in each of the five curves just before the last maximum. However, this subtle structural feature again appears in the estimated structural average.

The usefulness of the new approach cannot only be seen in synthetic examples, but also in the real evoked potential data. We used (3.1) and (3.2) to determine the estimated structural averages given in Figure 3. A small bandwidth was selected according to prior experience. Quantifying shift differences was based on five structural e-functionals of the type (2.1), three of them localizing maxima (in appropriate time windows including 60, 165 and 245

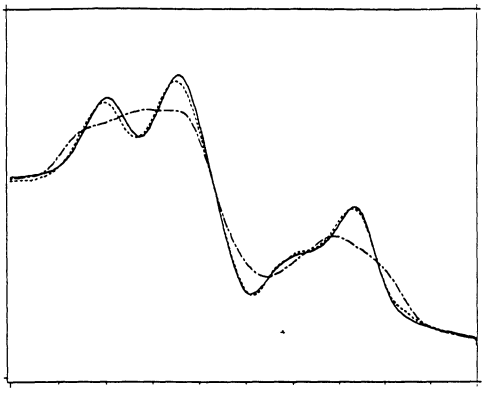

FIG. 5. *Structural average (solid curve) versus estimated structural average (dotted curve) and cross-sectional average (dashed curve).*

msec), and two of them localizing minima (in appropriate time windows including 130 and 190 msec). Synchronization then followed the procedure described in Section 3.2.

3.4. *Some conceptual remarks.* This discussion is based on the true curves f_i. Transforming individual curves using some shift operator G splits the sample: Instead of the original curves $\{f_i\}_{i \in \{1, \dots, m\}}$, we have the synchronized sample $\{f_i(g_i(\cdot))\}_{i \in \{1, \dots, m\}}$ together with $\{g_i\}_{i \in \{1, \dots, m\}}$. This is no loss of information, since knowing $f_i(g_i(\cdot))$ and g_i leads to $f_i = f_i(g_i(g_i^{-1}(\cdot)))$. For any i, g_i is a parametric function with the structural points as individual parameters, leading thus to a partial parametrization of the curves. We obtain the following: The number and amplitudes of the local extrema of f_i and $f_i(g_i(\cdot))$ are identical. Furthermore, if f_i possesses an extremum at $t^* \in J$, then $f_i(g_i(\cdot))$ has an extremum at $g_i^{-1}(t^*)$, and $g_i^{-1}(t^*) = \bar{\tau}_s$ if $t^* = \tau_{ri}$ for some $r \in \{1, \dots, l\}$. Similar relations hold for p-points.

If there are no missing structural points, the following properties hold for the structural average f:

1. Let T_r be a first-order structural e-functional. Then the structural average f has an extremum at $\bar{\tau}_r$, and $f(\bar{\tau}_r) = (1/m)\sum_{i=1}^m f_i(\tau_{ri})$.
2. Let T_r be a first-order structural p-functional, and assume that for some $r_1, r_2 \in \{1, \dots, l\}$, T_{r_1} and T_{r_2} are structural e-functionals such that for all $v \in \mathscr{R}$ we have $v(T_r(v)) = pv(T_{r_1}(v)) + (1 - p)v(T_{r_2}(v))$. Then $[\bar{\tau}_{r_1}, \bar{\tau}_{r_2}]$ is a strictly monotone segment of f, and $f(\bar{\tau}_r) = (1/m)\sum_{i=1}^m f_i(\tau_{ri}) = pf(\bar{\tau}_{r_1}) + (1 - p)f(\bar{\tau}_{r_2})$.

These properties show that the concept of first-order structural functionals is consistent with interpreting different locations of corresponding structural

characteristics as results of shift variation of the individual curves f_i. We see that all curves $f_i(g_i(\cdot))$ of the synchronized sample assume the same values as the original curves f_i, but part of the shift variation has been eliminated.

The above properties do not, in general, carry over to higher-order structural functionals. The derivative of $f_i(g_i(\cdot))$ is $f_i'(g_i(\cdot))g_i'(\cdot)$. Thus, the shift curve introduces an additional amplitude variation, unless g_i is a straight line. If structural homogeneity is mainly expected for derivatives $f_i^{(u)}$, it may be preferable to analyze the sample of derivatives.

Given a sample $\{f_i\}_{i \in \{1, \cdots, m\}}$, we have proposed one possible approach to analyze differences between individual curves, which can be expected to work in many applications. Evidently, two extreme principles are to explain individual differences fully as differences in individual intensity (or amplitude), or as completely as possible by differences in dynamics (or location). Without a priori knowledge we cannot decide in which proportion interindividual differences should be assigned to one of these sources of variability.

The first extreme is represented by principal component analysis, a popular method to analyze samples of curves [Rao (1958)]. It does not account for interindividual shifts since this approach explains interindividual differences completely by amplitude variation. As a consequence, PCA may create additional components which are derivatives of those components showing substantial shifts from subject to subject [Möcks (1986)].

The other extreme, explaining variation mainly by shifts, can be described, for example, as follows: Given an individual curve f_i and a reference curve v, one might try to determine an alignment by minimizing pseudometrics like $\int |f_i(g_i(t)) - v(t)|\, dt$ or, more reasonable, $\int |f_i(g_i(t)) - v(t)|\, dt + \int |f_i'(g_i(t)) - v'(t)|\, dt$ with respect to g_i. However, such an approach does not guarantee that structural characteristics, which intuitively should be aligned, really are aligned. Furthermore, too much variation is explained by shifts, producing shift functions having jumps or being constant within certain intervals. If, for example, f_i and v differ only by an additive constant, methods based on such pseudometrics will generally tend to explain part of the differences by shifts. A further disadvantage is the need for a reference curve v.

The approach presented is more flexible: It allows us to explain by (local) shift differences those phenomena thought to be associated with an individually differing dynamic. Furthermore, the concept of—a spontaneous or stimulated—process with an individually differing dynamic and intensity seems to be fruitful for many disciplines. When collecting data across time, age, and so on, which often starts at some intervention or stimulus, such a concept is frequently in our mind and might facilitate the statistical analysis.

4. Analyzing semiparametric regression models. As might be expected, the above concepts are useful in the context of semiparametric model building. The most frequently used models in nonlinear (parametric) regression share a common structural form [compare Ratkowsky (1983)]. It is assumed that for some $\phi \in C^\nu(\mathbb{R})$, $\nu \geq 2$, and some $[t_0, t_1[\subset \mathbb{R}$ the following

holds for all $i \in \{1, \ldots, m\}$:

There are unique parameters $\vartheta_{1i}, \theta_{1i}, \theta_{2i}, \vartheta_{2i} \in \mathbb{R}$, $\vartheta_{1i}, \theta_{1i} > 0$, such that for some interval $J_i \subseteq J$ with $[\theta_{1i}t_0 + \theta_{2i}, \theta_{1i}t_1 + \theta_{2i}] \subseteq J_i$,

$$f_i(t) = \vartheta_{1i}\phi\left(\frac{t - \theta_{2i}}{\theta_{1i}}\right) + \vartheta_{2i} \quad \text{for } t \in J_i. \tag{4.1}$$

According to (4.1), $f_i|_{J_i}$ arises from linear scale and time transformations of a basic function ϕ. Usually, the analysis is based on a prespecified ϕ (examples are the logistic and the Gompertz function). The importance of this model class led Lawton, Sylvestre and Maggio (1972), Kneip and Gasser (1988) and Härdle and Marron (1990) to consider the semiparametric model where ϕ is unknown. In the first two papers a rather complicated iterative scheme is proposed to estimate ϕ and the individual parameters simultaneously, while the third paper deals with somewhat different questions.

It will now be sketched that the tools developed in this paper lead to a simple procedure to estimate an unknown model function ϕ. Let $x \in [t_0, t_1]$ denote the location of an extremum of $\phi^{(u)}$, $u \leq \nu - 2$. Evidently, each $f_i^{(u)}$ possesses a corresponding extremum at $T(f_i) = \theta_{1i}x + \theta_{2i}$. Structural functionals can localize these extrema, and thus provide information about θ_{1i} and θ_{2i}, even if ϕ is unknown.

Based on Section 2.3, we can give this idea a theoretical basis. Assume that (2.2) and (2.3) hold. We will then assume that (4.1) extends to all functions $v \in \mathscr{R}$ for some $J_v \subseteq J$ and parameters $\vartheta_{1v}, \theta_{1v}, \theta_{2v}, \vartheta_{2v}$. Then the following proposition holds.

PROPOSITION 3. *Under the above assumptions let T be a structural e-functional of order $u + 1$ ($u \leq \nu - 2$) with $T(v) \neq \alpha$ for all $v \in \mathscr{R}$, and with $T(f_i) \in [\theta_{1i}t_0 + \theta_{2i}, \theta_{1i}t_1 + \theta_{2i}]$ for some $i \in \{1, \ldots, m\}$. Then there is an $x \in [t_0, t_1]$ such that:*

(a) *$\phi^{(u)}$ has an extremum at x.*
(b) *$T(v) = \theta_{1v}x + \theta_{2v}$ for all $v \in \mathscr{R}$.*

Using Proposition 2, the proof is straightforward and thus omitted. Under some additional conditions the proposition generalizes to structural p-functionals.

Suppose that the sample of curves yields enough information to specify two structural e-functionals T_1 and T_2 with $T_1(v) < T_2(v)$, $v \in \mathscr{R}$, which satisfy the conditions of Proposition 3. Then there are some $x_1, x_2 \in [t_0, t_1]$, $x_1 < x_2$, such that for all $i \in \{1, \ldots, m\}$,

$$T_1(f_i) = \theta_{1i}x_1 + \theta_{2i}, \qquad T_2(f_i) = \theta_{1i}x_2 + \theta_{2i}. \tag{4.2}$$

Now note that for any ϕ^* with $\phi^*(t) = a^*\phi(b^*t + c^*) + d^*$, $a^*, b^* > 0$, c^*, $d^* \in \mathbb{R}$, a trivial reparametrization can be done such that the model remains valid when replacing ϕ by ϕ^* (and $[t_0, t_1]$ by some $[t_0^*, t_1^*]$). This unidentifiability may be eliminated by appropriate normalizing conditions. A possible nor-

malization is to impose that the parameters $(\vartheta_{1i}, \theta_{1i}, \theta_{2i}, \vartheta_{2i})$ average to $(1, 1, 0, 0)$. This implies that $\phi(t) = (1/m)\sum_{i=1}^{m} f_i(\theta_{1i}t + \theta_{2i})$. Furthermore, considering (4.2), we obtain $x_1 = \bar{\tau}_1 = (1/m)\sum_{i=1}^{m} T_1(f_i)$, $x_2 = \bar{\tau}_2 = (1/m)\sum_{i=1}^{m} T_2(f_i)$. We now can define a shift operator G by

$$G_{(\tau, z)}(t) := \tau_1 + \frac{\tau_2 - \tau_1}{z_2 - z_1}(t - z_1),$$

$\tau = (\tau_1, \tau_2)' \in \mathscr{H}_2$, $z = (z_1, z_2)' \in \mathscr{H}_2$. This leads to

$$g_i(t) := G_{(\mathbf{T}(f_i), \bar{\tau})}(t) = \theta_{1i}t + \theta_{2i}, \qquad i = 1, \ldots, m,$$

$$\phi(t) = \frac{1}{m}\sum_{i=1}^{m} f_i(g_i(t)), \qquad t \in J^* := \left[\sup_i g_i^{-1}(u_{0,i}), \inf_i g_i^{-1}(u_{1,i})\right],$$

where $[u_{0,i}, u_{1,i}] = J_i$. The function ϕ thus can be represented as a structural average, which can be estimated as described in this paper.

The basic ideas of the above approach generalize to some more complicated semiparametric models.

5. Asymptotic theory. Rates of convergence will be derived for the estimators just presented. To simplify the presentation, an equally spaced design with an equal number of observations per subject is assumed. The theorems can be generalized to more complex situations, relaxing Assumption 1.

ASSUMPTION 1.

(a) $n := n_1 = n_2 = \cdots = n_m$, and for all m, n and each $j \in \{1, \ldots, m\}$ we have $t_j := t_{1j} = \cdots = t_{mj}$. Moreover, $t_{j+1} - t_j = (a_1 - a_0)/n$ for all $j \in \{1, \ldots, n-1\}$.

(b) For all i, j the random variables ϵ_{ij} are independent, and for any i, $\epsilon_{i1}, \epsilon_{i2}, \ldots$ are $i.i.d.$ random variables with expectation 0 and variance $\sigma_i^2 < \infty$. Furthermore, for each $\rho \in \mathbb{N}$ there exists a $C_\rho < \infty$ such that $E\epsilon_{ij}^\rho < C_\rho$ for all i (E denotes expectation).

We assume that the analysis is based on l, $l \in \mathbb{N}$, structural functionals $T_1, \ldots, T_l$. We use μ_r to denote the order of the structural functional T_r, $r = 1, \ldots, l$. Recall that $\mu_r := u_r + 1$.

Asymptotic theory relies on $n \to \infty$, while m is held fixed or increases, too. Asymptotics for $m \to \infty$ is based on model (2.2) (compare Section 2.3). We need further conditions on structural functionals and on the shift operator G.

ASSUMPTION 2.

(a) Model (2.2) holds for some $\nu \geq \mu + 2$, where $\mu := \max_{r \in \{1, \ldots, l\}} \mu_r$. $\mathscr{R}$ is a compact subspace of $C^\nu(J)$ (referring to $\|\cdot\|_J^{(\nu)}$).

(b) For any $r \in \{1, \ldots, l\}$ there exists an open subspace $D_r \subset C^{\mu_r + 1}$ (referring to $\|\cdot\|_J^{(\mu_r + 1)}$) such that $\mathscr{R} \subset D_r \subseteq \mathscr{D}_{T_r}$.

(c) $G_{(\cdot, \cdot)}(t)$ is twice continuously differentiable at each $(x, x^*) \in \mathscr{H}_l^2$.

The shift operator given in Section 3.2 satisfies condition (c). Condition (b) is crucial for asymptotic theory. It imposes a real restriction by requiring that $T_r(v) \neq \alpha$ for any $v \in \mathscr{R}$, that is, that the *true* regression curves have no missings in the above sense. This is kind of a homogeneity requirement. If, for example, T_r is defined by (2.1) this condition is fulfilled if any $v \in \mathscr{R}$ possesses a unique supremum in $]c, d[$ [note that then $\mathscr{D}_{T_r}$ is an open subspace of $C^2(J)$].

We use kernel estimators $\tilde{f}_i$ and $\hat{f}_i$ for determining structural points $\hat{\tau}_{ri} = T_r(\tilde{f}_i)$ and for the synchronization step. Other nonparametric techniques could be applied instead.

Given a kernel function W and a bandwidth $b > 0$ a (W, b)-kernel estimate $\hat{f}_{W,b;i}$ of f_i will be defined as follows:

$$\hat{f}_{W,b;i}(x) := \frac{1}{b} \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} W\left(\frac{x-u}{b}\right) du \cdot Y_{ij},$$

where $s_j := (t_{j+1} + t_j)/2$, $j = 1, \ldots, n-1$, $s_0 = t_1 - (s_1 - t_1)$, and $s_n := t_n + (t_n - s_{n-1})$. This is a convolution-type kernel estimator as considered by Gasser and Müller (1984). It is similar to the one proposed by Priestley and Chao (1972).

We consider the following situation with bandwidths $b, c > 0$:

1. (W, b)-kernel estimates are used to determine $\tilde{f}_i$, $i = 1, \ldots, m$.
2. $\hat{f}_1, \ldots, \hat{f}_m$ are obtained as (W, c)-kernel estimates of $f_1, \ldots, f_m$.

Here, W denotes a prespecified kernel function. We assume that:

3. W is a kernel of order k for some even $k \in \mathbb{N}$ and $\nu - \mu \geq k$.

Some additional technical requirements on W are deferred to the Appendix. Together with Assumptions 1 and 2, they are expected to hold throughout.

It should be noted that using the same kernel function within each step of the analysis is in no way necessary. This is done to simplify the notation. Moreover, different bandwidths might be used to estimate different structural points τ_{ri}. In both cases the basic results of Theorems 1–4 remain unchanged.

Before stating the asymptotic bias and variance of the estimators of the structural points, we have to introduce some notation:

Let T_r be a μ_rth-order structural e-functional. We will say that $\mathscr{R}$ is "symmetric at T_r" if there exists an $\varepsilon > 0$ such that for any $v \in \mathscr{R}$ we have $v^{(\mu_r - 1)}(T_r(v) - x) = v^{(\mu_r - 1)}(T_r(v) + x)$ for all $x < \varepsilon$. It should be noted that an assumption of symmetry has to be based on prior knowledge. For example, let $S\colon C^\nu(\mathbb{R}) \to C^\nu(\mathbb{R})$ be an operator such that S_v is a monotone function. Assume that for some $z \in C^\nu(\mathbb{R})$, $v(t) = S_v(z((t - a_v)/b_v))$ holds for all $v \in \mathscr{R}$ and all t in an ε-neighborhood of $T_r(v)$, where $(a_v, b_v) \in \mathbb{R} \times \mathbb{R}_+$ are appropriate parameters. Then $\mathscr{R}$ is symmetric at T_r if $z^{(\mu_r)}$ has a symmetric extremum at some $x \in \mathbb{R}$, and if $x = (T_r(v) - a_v)/b_v$ for all v.

For any m, E, var, cov, resp P will denote expectation, variance, covariance, resp. probability. The following theorem establishes the asymptotic bias

and variance of $\hat{\tau}_{ri}$ when T_r is a structural e-functional. Asymptotic normality can be inferred from Müller (1985, 1989).

THEOREM 1. *In addition to the above assumptions, suppose* $b \to 0$ *as* $n \to \infty$ *such that* $n^{1/(2\mu_r+3)}b/\log n \to \infty$, $r = 1, \dots, l$. *If* T_r *is a structural e-functional, then for all* i,

$$\text{(I)} \qquad E(\hat{\tau}_{ri} - \tau_{ri}) = \beta_{ri} + O(1/(nb^{2\mu_r+1})),$$

where

$$\beta_{ri} = O\left(b^{k-1}/\sqrt{nb^{2\mu_r+1}}\right)$$

if $\mathscr{R}$ *is symmetric at* T_r, *and else*

$$\beta_{ri} := b^k \int_{-1}^{1} W^{(\mu_r)}(x) x^{\mu_r+k}\, dx \frac{(-1)^{\mu_r+1} f_i^{(\mu_r+k)}(\tau_{ri})}{(k+\mu_r)!\, f_i^{(\mu_r+1)}(\tau_{ri})} + o(b^k).$$

$$\text{(II)} \qquad \operatorname{var}(\hat{\tau}_{ri}) = v_{ri} + o(\beta_{ri}^2) + o(1/(nb^{2\mu_r+1})),$$

where

$$v_{ri} := \frac{\sigma_i^2}{nb^{2\mu_r+1}} \int_{-1}^{1} W^{(\mu_r)}(x)^2\, dx / f_i^{(\mu_r+1)}(\tau_{ri})^2.$$

A proof is contained in the Appendix. Within Theorem 1 and the subsequent theorems, convergence established by O and o terms holds uniformly for all $i \in \{1, \dots, m\}$ and $m \in \mathbb{N}$. In each of the theorems this can easily be verified when using the compactness of $\mathscr{R} \supset \{f_i\}_{i \in \{1, \dots, m\}}$ and the uniform boundedness of all moments of ϵ_{ij} (with respect to i). In the following it will not be explicitly mentioned.

The theorem shows that the bias is practically negligible if $\mathscr{R}$ is symmetric at T_r. We then might let $b \to 0$ extremely slow, to obtain $E|\hat{\tau}_{ri} - \tau_{ri}| \approx O(n^{-1/2})$. In the general case, both bias and variance depend on $|1/f_i^{(\mu_r+1)}(\tau_{ri})|$. This term is small if $f_i^{(\mu_r-1)}$ is rapidly increasing/decreasing near its extremum at τ_{ri}. It is large if $f_i^{(\mu_r-1)}$ is very flat in a neighborhood of τ_{ri}. The bias additionally depends on $|f_i^{(\mu_r+k)}(\tau_{ri})|$. Evidently, if $f_i^{(\mu_r-1)}$ is symmetric at τ_{ri}, $f_i^{(\mu_r+\gamma)}(\tau_{ri}) = 0$ for all $\gamma \in \{0, 2, 4, \dots\}$. Since k is even, $|f_i^{(\mu_r+k)}(\tau_{ri})|$ provides some measure of symmetry of the respective extremum. As regards the constants in the bias and variance terms above, we thus can conclude:

(a) The constants are small if the extremum of $f_i^{(\mu_r-1)}$ at τ_{ri} possesses a high degree of symmetry, and/or if $f_i^{(\mu_r-1)}$ is rapidly increasing/decreasing close to τ_{ri}.

(b) The constants are large if the extremum is rather asymmetric, and if $f_i^{(\mu_r-1)}$ is flat in a neighborhood of τ_{ri}.

The best individual rates of convergence are achieved when choosing b to be of

the order $n^{-1/(2k+2\mu_r+1)}$. Then

$$E(\hat{\tau}_{ri} - \tau_{ri})^2 = O(n^{-2k/(2k+2\mu_r+1)}).$$

If $k > 2$, setting $b = C \cdot n^{-1/(k+2\mu_r+1)}$ for some appropriate C minimizes the bias. We then obtain

$$|E\hat{\tau}_{ri} - \tau_{ri}|^2 = O(n^{-2k/(k+2\mu_r+1)}).$$

For $k = 2$ the corresponding rate $O(n^{-4/(2\mu_r+3)})$ cannot quite be reached, since the theorem requires $n^{1/(2\mu_r+3)}b/\log n \to \infty$. Apart from some particular situations, for example, symmetry, it is not possible to reduce the order of the bias any further. This result is not surprising. Let T_r be a first-order structural e-functional identifying corresponding maxima. Suppose we do not smooth at all, and let $\hat{\tau}_{ri}$ denote the location of the maximal observation Y_{ij} within some neighborhood of τ_{ri}. If the peak is asymmetric, $\hat{\tau}_{ri}$ is a biased estimator of τ_{ri}, and this bias does not converge to 0 as $n \to \infty$. Thus, smoothing is necessary to reduce *both* the bias and the variance. The minimum bias choice of b will play an important role when dealing with averages over samples of curves, since averaging in itself diminishes variance.

Theorem 2 gives the asymptotic bias and variance, if T_r is a structural p-functional. We additionally impose the minor restriction that for any $v \in D_r$ the strictly monotonical segment containing $T_r(v)$ is given by $[\psi_0(v), \psi_1(v)]$, where ψ_0, ψ_1 are structural e-functionals ($D_r \subseteq \mathscr{D}_{\psi_0}$, $D_r \subseteq \mathscr{D}_{\psi_1}$).

THEOREM 2. *Let T_r be a structural p-functional. Under the above assumptions we obtain*

(I) $$E(\hat{\tau}_{ri} - \tau_{ri}) = \beta_{ri} + O(1/(nb^{2\mu_r+1})),$$

where

$$\beta_{ri} := b^k \int_{-1}^{1} W^{(\mu_r-1)}(x) x^{k+\mu_r-1}\, dx \times (-1)^{\mu_r} \frac{f_i^{(k+\mu_r-1)}(\tau_{ri}) - p f_i^{(k+\mu_r-1)}(\psi_0(f_i)) - (1-p) f_i^{(k+\mu_r-1)}(\psi_1(f_i))}{(k+\mu_r-1)! f_i^{(\mu_r)}(\tau_{ri})} + o(b^k).$$

(II) $$\operatorname{var}(\hat{\tau}_{ri}) = v_{ri} + o(\max\{b^{2k}, 1/(nb^{2\mu_r-1})\}),$$

where

$$v_{ri} := \frac{\sigma_i^2(1 + p^2 + (1-p)^2)}{nb^{2\mu_r-1}} \int_{-1}^{1} W^{(\mu_r-1)}(x)^2\, dx / f_i^{(\mu_r)}(\tau_{ri})^2.$$

A proof is contained in the Appendix. We see again that $b = C^* n^{-1/(k+2\mu_r+1)}$ (for some appropriate C^*) minimizes the bias if $k > 2$.

The best individual rates of convergence are obtained when choosing $b = C^{**} n^{-1/(2k+2\mu_r+1)}$. Then

$$E(\hat{\tau}_{ri} - \tau_{ri})^2 = O(n^{-2k/(2k+2\mu_r-1)}).$$

Thus, in general, the optimal rates of convergence for μ_rth-order structural p-functionals are better than those of μ_rth-order structural e-functionals.

Based on Theorems 1 and 2, it is possible to obtain the asymptotic bias, variance and covariance of the estimated transformations $\hat{g}_i = G_{(\hat{\tau}_i, \bar{\hat{\tau}})}$, of the transformed curves $\hat{f}_i \circ \hat{g}_i$, and of the estimated structural average. In the following G^* and G^{**} will denote partial derivatives of G, $G^*_{(x,x^*)}(t) := (\partial/\partial\vartheta)G_{(\vartheta,x^*)}(t)|_{\vartheta=x}$ and $G^{**}_{(x,x^*)}(t) := (\partial/\partial\vartheta)G_{(x,\vartheta)}(t)|_{\vartheta=x^*}$ for $(x, x^*, t) \in \mathscr{H}_l^2 \times \mathbb{R}$. For $r = 1, \ldots, l$, let $G^*_{(x,x^*;r)}(t)$ resp. $G^{**}_{(x,x^*;r)}(t)$ denote the rth elements of the vectors $G^*_{(x,x^*)}(t)$ resp. $G^{**}_{(x,x^*)}(t)$.

THEOREM 3. *Under the above assumptions let $b \to 0$ as $n \to \infty$ such that $n^{1/(2\mu_r+3)}b/\log n \to \infty$. Suppose that m ($\equiv m_n$) is a nondecreasing function of n. For $r = 1, \ldots, l$, let $o_r := \mu_r$ if T_r is a structural e-functional, and $o_r := \mu_r - 1$ if T_r is a structural p-functional. Furthermore, let $o := \max\{o_1, \ldots, o_l\}$, $\mu = \max\{\mu_1, \ldots, \mu_l\}$. Then, as $n \to \infty$, we obtain for all $t \in J$,*

$$\text{(I)} \quad E\big(\hat{g}_i(t) - g_i(t)\big) = B_i(t) + O\big(1/(nb^{2\mu+1})\big) = O\big(b^k + 1/(nb^{2\mu+1})\big),$$

where

$$B_i(t) := \sum_{r=1}^{l} \left(G^*_{(\tau_i,\bar{\tau};r)}(t)\beta_{ri} + G^{**}_{(\tau_i,\bar{\tau};r)}(t)\frac{1}{m}\sum_{j=1}^{m}\beta_{rj} \right).$$

$$\text{(II)} \quad \operatorname{var}\big(\hat{g}_i(t)\big) = V_i(t) + o\big(b^{2k} + 1/(nb^{2o+1})\big) = O\big(b^{2k} + 1/(nb^{2o+1})\big),$$

where

$$V_i(t) := \sum_{r=1}^{l} \left(v_{ri}\left(G^*_{\tau_i,\bar{\tau};r)}(t) + \frac{1}{m}G^{**}_{(\tau_i,\bar{\tau};r)}(t) \right)^2 + \frac{1}{m^2}\sum_{k\neq i} v_{rk}G^{**}_{(\tau_k,\bar{\tau};r)}(t)^2 \right).$$

$$\text{(III)} \quad \operatorname{cov}\big(\hat{g}_i(t), \hat{g}_k(t)\big) = \frac{1}{m}\big(C_{ik}(t) + o\big(b^{2k} + 1/(nb^{2o+1})\big)\big) + o(n^{-2})$$
$$= O\big((1/m)\big(b^{2k} + 1/(nb^{2o+1})\big)\big) + o(n^{-2}),$$

where

$$C_{ik}(t) := \sum_{r=1}^{l} \left(v_{ri}G^*_{(\tau_i,\bar{\tau};r)}(t)G^{**}_{(\tau_k,\bar{\tau};r)}(t) + v_{rk}G^*_{(\tau_k,\bar{\tau};r)}(t)G^{**}_{(\tau_i,\bar{\tau};r)}(t) + G^{**}_{(\tau_i,\bar{\tau};r)}(t)G^{**}_{(\tau_k,\bar{\tau};r)}(t)\frac{1}{m}\sum_{s=1}^{m} v_{rs} \right).$$

A proof is contained in the Appendix. In the following, for any $\varepsilon > 0$ and all m, we will set $J_\varepsilon := [\sup_{i\in\{1,\ldots,m\}} g_i^{-1}(a_0) + \varepsilon, \inf_{i\in\{1,\ldots,m\}} g_i^{-1}(a_1) - \varepsilon]$. It should be noted that $[(1/m)\sum_{i=1}^m \tau_{1i}, (1/m)\sum_{i=1}^m \tau_{li}] \subset J_\varepsilon$, if ε is sufficiently small.

THEOREM 4. *Let $\varepsilon > 0$. Under the assumptions of Theorem 3 let $b \log n \to 0$. Furthermore, let $c = O(b)$ such that $nb^{2o+1} = O(nc)$ and $nc^3b \to \infty$ as*

$n \to \infty$. *We then obtain as* $n \to \infty$,

$$\text{(I)} \quad \int_{J_\varepsilon} \left(E\hat{f}_i(\hat{g}_i(t)) - f_i(g_i(t))\right)^2 dt$$

$$= \int_{J_\varepsilon} \left\{ c^k \frac{\int_{-1}^{1} W(x)x^k\,dx}{k!} f_i^{(k)}(g_i(t)) + f_i'(g_i(t))B_i(t) \right\}^2 dt$$

$$+ o(b^{2k}) + O(1/(n^2 b^{4\mu+2}))$$

$$= O(b^{2k} + 1/(n^2 b^{4\mu+2})).$$

$$\text{(II)} \quad \int_{J_\varepsilon} \operatorname{var}\left(\hat{f}_i(\hat{g}_i(t))\right) dt$$

$$= \int_{J_\varepsilon} \left(\frac{\sigma_i^2}{nc} \int_{-1}^{1} W(x)^2\,dx + f_i'(g_i(t))^2 V_i(t) \right) dt$$

$$+ o(b^{2k} + 1/(nb^{2o+1}))$$

$$= O(b^{2k} + 1/(nb^{2o+1})).$$

$$\text{(III)} \quad \int_{J_\varepsilon} \operatorname{cov}\left(\hat{f}_i(\hat{g}_i(t)), \hat{f}_k(\hat{g}_k(t))\right)$$

$$= \frac{1}{m} \left(\int_{J_\varepsilon} f_i'(g_i(t))\, f_k'(g_k(t)) C_{ik}(t)\,dt \right.$$

$$\left. + o(b^{2k} + 1/(nb^{2o+1})) \right) + o(n^{-2})$$

$$= O((1/m)(b^{2k} + 1/(nb^{2o+1}))) + o(n^{-2}).$$

A proof is contained in the Appendix. By the above conditions on c, the error terms depending on c are of smaller or equal magnitude compared to those stemming from the error in estimating the structural points. These conditions can always be fulfilled. The theorem implies the following corollary for the estimated structural average $\hat{f}$.

COROLLARY 1. *Assume the conditions of Theorem* 4. *For any* m $(\equiv m_n)$, *let* $f := (1/m)\sum_{i=1}^{m} f_i(g_i(\cdot))$, $f \equiv f_m$, *denote the respective structural average. Then*

$$E\int_{J_\varepsilon} \left(\hat{f}(t) - f(t)\right)^2 dt = O(b^{2k} + 1/(n^2 b^{4\mu+2}) + 1/(mnb^{2o+1})).$$

Evidently, there is a difference in whether one wants a best estimate of an individual transformed curve $f_i \circ g_i$ or a best estimate of the structural average f. In the second case one has to use smaller bandwidths throughout. This is not surprising since averaging in itself diminishes variance.

Under the conditions of the corollary, we have to choose b to be of the order of magnitude $\max\{n^{-1/(k+2\mu+1)}, (mn)^{-1/(2k+2\mu+1)}\}$ to obtain the best possible rate of convergence for $\hat{f}$. Then

$$E\int_{J_\epsilon}\left(\hat{f}(t) - f(t)\right)^2 dt = O\left(n^{-2k/(k+2\mu+1)} + (mn)^{-2k/(2k+2\mu+1)}\right).$$

If m ($\equiv m_n$) increases fast enough with n, and if k (and ν) become sufficiently large, this approaches n^{-2}. The lower bound n^{-2} is a consequence of the fact that the bias of $\hat{f}$ cannot be made arbitrarily small by increasing m. Existence and magnitude of this lower bound are plausible, since it also arises in the parametric case. Assume a parametric model for the regression curves f_i. When estimating the parameters by nonlinear least squares, it is well known that the resulting estimators have a squared bias of the order n^{-2}. This bias leads to a corresponding bias for shift curves determined from these parameters. A respective—parametric—structural average then shows a best obtainable rate of convergence $O(n^{-2})$.

As an example let $k = 4$ and assume that $T_1, \ldots, T_l$ are all first-order structural functionals. Furthermore, suppose that $\min\{m, n\} \to \infty$ such that $n^{4/7} = O(m)$ as $n \to \infty$. Then we might choose c, b to be of the order $n^{-1/7}$. This yields

$$E\int_{J_\epsilon}\left(\hat{f}(t) - f(t)\right)^2 dt = O(n^{-8/7}).$$

APPENDIX

The kernel function W (of order k) is assumed to satisfy the following conditions:

1. W is symmetric, and $W(u) = 0$ for $|u| > 1$.
2. $$\int W(u)u^j\,du = \begin{cases} 1, & \text{for } j = 0, \\ 0, & \text{for } j = 1, \ldots, k-1, \\ \neq 0, & \text{for } j = k. \end{cases}$$
3. W is $\mu + 1$ times continuously differentiable on $\mathbb{R}$, and it is $\mu + 2$ times continuously differentiable on $[-1, 1]$.

Throughout the following lemma and the proofs of the theorems, convergence established by $o(\cdot)$ [and $O(\cdot)$] terms is to be interpreted to hold uniformly for $i \in \{1, \ldots, m\}$, $m \equiv m_n \in \mathbb{N}$. In either case this will follow from the arguments used to derive the respective results.

LEMMA. *Let $\gamma \in \{0, \ldots, \mu + 2\}$. Under the assumptions of Theorem 1, let $\bar{f}_i$ denote (W, λ)-kernel estimators of f_i with $\lambda \to 0$ and $n\lambda^3 \to \infty$ as $n \to \infty$. For $\mu^* \in \{0, \ldots, \mu + 2\}$, define $W_{\mu^*}\colon \mathbb{R} \to \mathbb{R}$ by $W_{\mu^*}(x) := W^{(\mu^*)}(x)$ for $x \in [-1, 1]$ and $W_{\mu^*}(x) := 0$ for $x \notin [-1, 1]$.*

Furthermore, let $\mathcal{J} \neq J$ *denote some compact subinterval of* J, *and let* $\{\eta_n\}_{n \in \mathbb{N}}$ *be an arbitrary sequence of constants with* $\eta_n \to \infty$ *as* $n \to \infty$. *Then, as* $n \to \infty$, *it holds with* $\kappa := \min\{\nu - \gamma, k\}$:

(a) *There exists a* $C > 0$ (*independent of* i, m) *such that for each* $t \in \mathcal{J}$ $|E(\tilde{f}_i^{(\gamma)}(t) - f_i^{(\gamma)}(t))| \leq \lambda^{\kappa} C \sup_{t \in \mathcal{J}} |f_i^{(\kappa)}(t)|$, *and if* $k + \gamma \leq \nu$,

$$E\left(\tilde{f}_i^{(\gamma)}(t) - f_i^{(\gamma)}(t)\right) = \lambda^k \int_{-1}^{1} W_\gamma(x) x^{k+\gamma}\, dx \frac{(-1)^\gamma f_i^{(k+\gamma)}(t)}{(k+\gamma)!} + o(\lambda^k),$$

where the remainder term is uniform in $t \in \mathcal{J}$.

(b) *For any* $t \in J$,

$$\operatorname{var}\left(\tilde{f}_i^{(\gamma)}(t)\right) = \frac{\sigma_i^2}{n\lambda^{2\gamma+1}} \int_{-1}^{1} W_\gamma(x)^2\, dx + o\left(1/(n\lambda^{2\gamma+1})\right),$$

where the remainder term is uniform in $t \in J$.

(c) *For any* $\rho \in \mathbb{N}$,

$$E\left(\sup_{t \in \mathbb{R}} \left|\tilde{f}_i^{(\gamma)}(t) - E\tilde{f}_i^{(\gamma)}(t)\right|^\rho\right) = o(q_n^\rho).$$

Hereby, $q(n) := \eta_n \log n / \sqrt{n\lambda^{2\gamma+1}}$.

(d) *If* $\gamma \leq \mu + 1$ *and* $n^{1/2\gamma+1}\lambda / \log n \to \infty$, *for each* $\varepsilon > 0$ *and all* $\beta \in \mathbb{N}$,

$$P\left(\sup_{t \in \mathcal{J}} \left|f_i^{(\gamma)}(t) - \tilde{f}_i^{(\gamma)}(t)\right| > \varepsilon\right) = o(1/n^\beta).$$

PROOF. Assertions (a) and (b) follow from Lemma 1 and Theorem 4 of Gasser and Müller (1984) in a straightforward way. The proof of assertion (c) is based on the proof of Theorem 3 of Cheng and Lin (1981).

Evidently, $\tilde{f}_i(x) = 0$ for $x \notin J_\lambda := [a_0 - \lambda, a_1 + \lambda]$. We thus only have to show that

$$E \frac{1}{q(n)^\rho} \sup_{t \in J_\lambda} \left|\tilde{f}_i^{(\gamma)}(t) - E\tilde{f}_i^{(\gamma)}(t)\right|^\rho = o(1). \tag{A1}$$

Define the truncated random variables

$$\epsilon_{ij}^* := \epsilon_{ij} I\left(|\epsilon_{ij}| \leq (n\lambda)^{1/4}\right) \tag{A2}$$

and set

$$\tilde{f}_i^{*(\gamma)}(x) := \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} \frac{1}{\lambda^{\gamma+1}} W_\gamma\left(\frac{x-u}{\lambda}\right) du\, \epsilon_{ij}^*.$$

By using now a moment inequality of the exponential form [compare Lamperti (1966), pages 43 and 44] and assertion (b), analogously to Cheng and Lin (1981), we obtain for every $\varepsilon > 0$,

$$P\left(\frac{1}{q(n)} \left|\tilde{f}_i^{*(\gamma)}(t) - E\tilde{f}_i^{*(\gamma)}(t)\right| > \varepsilon\right) \leq C \cdot n^{-\varepsilon \eta_n^{1/2}}$$

for all $t \in \mathcal{J}$ and all i [$0 < C < \infty$, note that due to Assumption 1(b) C can be chosen independent of i, m].

We now claim that for all $\rho \in \mathbb{N}$,

$$\text{(A3)} \qquad E\frac{1}{q(n)^{\rho}} \sup_{t \in J_\lambda} \left| \bar{f}_i^{*(\gamma)}(t) - E\bar{f}_i^{*(\gamma)}(t) \right|^{\rho} = o(1).$$

This can be seen in the following way:

For some $\delta \in \mathbb{N}$ define $z_{1,n} := a_0 - \lambda$ and $z_{r+1,n} := z_{r,n} + (a_1 - a_0 + 2\lambda)/n^{\delta}$ for each $r \in \{2, \ldots, n^{\delta}\}$. Let $E_n := \{z_{r,n}\}_{r \in \{1, \ldots, n^{\delta}\}}$, and for each $x \in J_\lambda$ let $Z(x)$ denote an element of E_n with the property that $|x - Z(x)| = \min_{x^* \in E_n} |x - x^*|$. Then there exists a generic constant A_ρ such that

$$\text{(A4)} \qquad \frac{1}{q(n)^{\rho}} \sup_{t \in J_\lambda} \left| \bar{f}_i^{*(\gamma)}(t) - E\bar{f}_i^{*(\gamma)}(t) \right| < A_\rho (S_{1i}^{\rho} + S_{2i}^{\rho} + S_{3i}^{\rho}),$$

with

$$S_{1i} := \frac{1}{q(n)} \sup_{t \in J_\lambda} \left| \bar{f}_i^{*(\gamma)}(t) - \bar{f}_i^{*(\gamma)}(Z(t)) \right|,$$

$$S_{2i} := \frac{1}{q(n)} \sup_{t \in J_\lambda} \left| \bar{f}_i^{*(\gamma)}(Z(t)) - E\bar{f}_i^{*(\gamma)}(Z(t)) \right|,$$

$$S_{3i} := \frac{1}{q(n)} \sup_{t \in J_\lambda} \left| E\bar{f}_i^{*(\gamma)}(t) - E\bar{f}_i^{*(\gamma)}(Z(t)) \right|.$$

Since by construction $\eta_n \to \infty$,

$$\text{(A5)} \qquad P_\epsilon(S_{2i} > \varepsilon) \le n^{\delta} C n^{-\varepsilon \eta_n^{1/2}}$$

implies $ES_{2i}^{\rho} \to 0$ as $n \to \infty$.

Let $L := \sup_{x \in \mathbb{R}} |W_\gamma(x)|$. Noting that for any $t \in J$, $\int_{s_{j-1}}^{s_j} W_\gamma((t - x)/\lambda)\, dx = \lambda(W_{\gamma-1}((t - s_j)/\lambda) - W_{\gamma-1}((t - s_{j-1})/\lambda))$, we furthermore obtain that with probability 1,

$$\text{(A6)} \qquad S_{1i} \le \frac{1}{q(n)} (n\lambda)^{1/4} \frac{L(a_1 - a_0)}{n^{\delta} \lambda^{\gamma+1}} n.$$

Thus, choosing δ large enough, $ES_{1i}^{\rho} \to 0$ and $S_{3i}^{\rho} \to 0$ as $n \to \infty$. This proves (A3).

We are now in a position to show (A1): Note that for all t,

$$\text{(A7)} \qquad \bar{f}_i^{(\gamma)}(t) = \bar{f}_i^{*(\gamma)}(t) + \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} \frac{1}{\lambda^{\gamma+1}} W_\gamma\left(\frac{t - u}{\lambda}\right) du(\epsilon_{ij} - \epsilon_{ij}^*).$$

For any $\rho > 0$, all i and each $\beta \in \mathbb{N}$, we obtain

$$\text{(A8)} \qquad \begin{aligned} &\frac{1}{q(n)^{\rho}} E \sup_{t \in J_\lambda} \left| \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} \frac{1}{\lambda^{\gamma+1}} W_\gamma\left(\frac{t-u}{\lambda}\right) du(\epsilon_{ij} - \epsilon_{ij}^*) \right|^{\rho} \\ &\quad < \frac{1}{q(n)^{\rho}} \frac{L^{\rho}}{\lambda^{\rho(\gamma+1)}} \frac{1}{(n\lambda)^{(\rho^*-\rho)/4}} E|\epsilon_{ij}|^{\rho^*} = o(n^{-\beta}) \end{aligned}$$

by choosing $\rho^* > \rho$ sufficiently large, as follows from (A2) and the boundedness of all moments of ϵ_{ij}. When combining this with (A7), then (A1) is a consequence of (A3).

It remains to show assertion (d). Note that under the postulated additional assumption on λ, we might choose the sequence η_n such that $1/q(n) \geq 1$ for all n large enough. By assertion (a) and by the compactness of $\mathscr{R}$, it is easily seen that assertion (d) holds iff for any $\varepsilon > 0$ and all $\beta \in \mathbb{N}$,

$$P\left(\frac{1}{q(n)} \sup_{t \in \mathscr{J}} \left| \hat{f}_i^{(\gamma)}(t) - E\hat{f}_i^{(\gamma)}(t) \right| > \varepsilon \right) = o(n^{-\beta}).$$

This immediately follows from (A7), (A4), (A5), (A6) and (A8). □

PROOF OF THEOREM 1. Let $\gamma := \mu_r + 1$. By assumption, $\mathscr{R}$ is a compact subspace of $C^{\nu}(J)$, $\nu \geq \gamma + 1$, with $C^{\nu}(J)$ being endowed with the norm $\|\cdot\|_J^{(\nu)}$. Together with the continuity of T_r on $\mathscr{R}$, and with $T_r(\mathscr{R}) \subset]a_0, a_1[$, this implies that there is a compact subinterval $\mathscr{J} \subset J$ with $T_r(\mathscr{R}) \subseteq \mathscr{J}$.

In the following, for all $\varepsilon > 0$, $v \in C^{\gamma}(\mathscr{J})$, $t \in \mathscr{J}$, we will use $U_\varepsilon(v)$ and $U_\varepsilon(t)$ to denote ε-neighborhoods of v and t (with respect to $\|\cdot\|_{\mathscr{J}}^{(\gamma)}$ and $|\cdot|$). Furthermore, let $\Lambda := \sup_{x,y \in J \cup \{\alpha\}} |x - y|$.

Note that $\mathscr{R}$ is also a compact subset of $C^{\nu}(\mathscr{J})$ and $C^{\gamma}(\mathscr{J})$ (referring to $\|\cdot\|_{\mathscr{J}}^{(\nu)}$ and $\|\cdot\|_{\mathscr{J}}^{(\gamma)}$), and that $\sup_{v \in \mathscr{R}} \sup_{t \in J} |v^{(q)}(t)| < \infty$ for all $q = 0, \ldots, \gamma + 1$. By definition of a structural e-functional, we have $|v^{(\gamma)}(T_r(v))| > 0$ and $\operatorname{sign}(v^{(\gamma)}(T_r(v))) = \operatorname{sign}(w^{(\gamma)}(T_r(w)))$ for all $v, w \in \mathscr{R}$. Moreover, $T_r(v)$ and, hence, $v^{(\gamma)}(T_r(v))$ are continuous functionals of $v \in \mathscr{R}$ (with respect to both $\|\cdot\|_{\mathscr{J}}^{(\nu)}$ and $\|\cdot\|_{\mathscr{J}}^{(\gamma)}$). Based on the properties of $\mathscr{R}$, we thus can infer that there exist some $d, \delta > 0$ such that for all $v \in \mathscr{R}$,

$$\inf_{t \in U_\delta(T_r(v))} |v^{(\gamma)}(t)| > 2d.$$

Furthermore, there exists an $e > 0$ such that for each $v \in \mathscr{R}$, $U_e(v) \subset D_r$ and

$$\sup_{w \in U_e(v)} |T_r(v) - T_r(w)| < \delta/4.$$

With $\varepsilon := \min\{e, d\}$ we thus obtain for each $v \in \mathscr{R}$,

$$\text{(A9)} \qquad \begin{aligned} \inf_{w \in U_\varepsilon(v)} \inf_{t \in U_{\delta/2}(T_r(w))} |w^{(\gamma)}(t)| &\geq \inf_{w \in U_e(v)} \inf_{t \in U_{\delta/2}(T_r(w))} |v^{(\gamma)}(t)| - d \\ &\geq \inf_{t \in U_\delta(T_r(v))} |v^{(\gamma)}(t)| - d \geq d. \end{aligned}$$

The above considerations now allow us to derive upper bounds for the rates of convergence of $E|\hat{\tau}_{ri} - \tau_{ri}|^\rho$:

If $\tilde{f}_i \in U_\varepsilon(f_i)$ we obviously have $|\tilde{f}_i^{(\gamma)}(t)| \geq d$ for all t in the segment joining τ_{ri} and $\hat{\tau}_{ri}$. Moreover, $\tilde{f}_i^{(\mu_r)}(\hat{\tau}_{ri}) = f_i^{(\mu_r)}(\tau_{ri}) = 0$. A Taylor expansion of $\tilde{f}_i^{(\mu_r)}(\hat{\tau}_{ri})$ thus shows that if $\tilde{f}_i \in U_\varepsilon(f_i)$,

$$|\hat{\tau}_{ri} - \tau_{ri}| \leq \frac{1}{d}\left| f_i^{(\mu_r)}(\tau_{ri}) - \tilde{f}_i^{(\mu_r)}(\tau_{ri})\right|.$$

It follows that for any $\rho \in \mathbb{N}$ and any i,

$$\begin{aligned} E|\hat{\tau}_{ri} - \tau_{ri}|^\rho < \Lambda^\rho \sum_{s=0}^{\gamma} P\left(\sup_{t \in \mathscr{I}} \left| f_i^{(s)}(t) - \tilde{f}_i^{(s)}(t)\right| > \varepsilon/(\gamma+1)\right) \\ + \left(\frac{2^\rho}{d^\rho}\right)\left(\left| f_i^{(\mu_r)}(\tau_{ri}) - E\tilde{f}_i^{(\mu_r)}(\tau_{ri})\right|^\rho \right. \\ \left. + E\left| \tilde{f}_i^{(\mu_r)}(\tau_{ri}) - E\tilde{f}_i^{(\mu_r)}(\tau_{ri})\right|^\rho\right). \end{aligned}$$

By Lemma 1(d), the first term on the right-hand side is $o(1/n^2)$. By Lemma 1(b), $E|\tilde{f}_i^{(\mu_r)}(\tau_{ri}) - E\tilde{f}_i^{(\mu_r)}(\tau_{ri})|^\rho$ can be bounded by $O(1/(\sqrt{nb^{2\mu_r+1}})^\rho)$ for $\rho = 1, 2$. This generalizes to $\rho > 2$, as can easily be seen by using Whittle's (1960) inequality. Lemma 1(a) yields $O(b^{\rho k})$ as a general bound for $|f_i^{(\mu_r)}(\tau_{ri}) - E\tilde{f}_i^{(\mu_r)}(\tau_{ri})|^\rho$. Now consider the special case that $\mathscr{R}$ is symmetric at T_r. Note that then $f_i^{(\mu_r)}(\tau_{ri} - x) = -f_i^{(\mu_r)}(\tau_{ri} + x)$ for all x sufficiently small. The following relations now follow from the arguments of Appendix (1) of Gasser and Müller (1984), partial integration and the symmetry of W:

$$\begin{aligned} f_i^{(\mu_r)}(\tau_{ri}) - E\tilde{f}_i^{(\mu_r)}(\tau_{ri}) &= -E\tilde{f}_i^{(\mu_r)}(\tau_{ri}) \\ &= (-1)^{\mu_r+1}\int_{-1}^{1} W(x)\, f_i^{(\mu_r)}(\tau_{ri} - bx)\, dx \\ &\quad + O(1/(nb^{\mu_r})) = O(1/(nb^{\mu_r})). \end{aligned}$$

Let $\kappa_r := 1/(nb^{2\mu_r+1})$ if $\mathscr{R}$ is symmetric at T_r, and set $\kappa_r := \max\{b^{2k}, 1/(nb^{2\mu_r+1})\}$ else. When combining the above results, we obtain

$$E|\hat{\tau}_{ri} - \tau_{ri}|^\rho = O\left(\kappa_r^{\rho/2}\right). \tag{A10}$$

Based on (A10), it is now possible to derive a more exact approximation to the conditional asymptotic bias and variance. Let $T^*(v) = T_r(v)$ for all $v \in U_\varepsilon(f_i)$, and set $T^*(v) = \tau_{ri}$ for any $v \notin U_\varepsilon(f_i)$. Obviously, (A10) still holds when replacing there τ_{ri} by $\tau_{ri}^* = T^*(f_i)$ and $\hat{\tau}_{ri}$ by $\hat{\tau}_{ri}^* = T^*(\tilde{f}_i)$. A more precise analysis of $\hat{\tau}_{ri} - \tau_{ri}$ leads to

$$\hat{\tau}_{ri} - \tau_{ri} = A(\tilde{f}_i, f_i) + \sum_{s=1}^{3} R_s(\tilde{f}_i, f_i), \tag{A11}$$

where for all $v, w \in C^{\gamma}(J)$,

$$A(v,w) := \frac{1}{a(w)}\left(w^{(\mu_r)}(T^*(w)) - v^{(\mu_r)}(T^*(w))\right),$$

$$R_1(v,w) := \frac{-1}{2a(w)} w^{(\gamma+1)}(\xi_{v,w})(T^*(v) - T^*(w))^2$$

for some measurable mean-value function $\xi_{v,w}$,

$$R_2(v,w) := \frac{1}{a(w)} \int_{T^*(w)}^{T^*(v)} \left(w^{(\gamma)}(\tau) - v^{(\gamma)}(\tau)\right) d\tau,$$

$$R_3(v,w) := \left(T_r(v) - T_r(w) - A(v,w) - \sum_{s=1}^{2} R_s(v,w)\right) \times I(v \notin U_{\varepsilon}(f_i) \text{ or } w \notin U_{\varepsilon}(f_i)).$$

Hereby, $a(w) := \max\{d, w^{(\gamma)}(T^*(w))\}$, and I denotes the indicator function.

The above arguments show that $E|A(\tilde{f}_i, f_i)|^{\rho} = O(\kappa_r^{\rho/2})$. Now combining (A10) with the results of Lemma 1, repeated use of the Cauchy–Schwarz inequality yields $E|A(\tilde{f}_i, f_i)\Sigma_{s=1}^{3} R_s(\tilde{f}_i, f_i)| = o(\kappa_r)$ and $E(\Sigma_{s=1}^{3} R_s(\tilde{f}_i, f_i))^2 = o(\kappa_r)$. Hence,

$$E(\hat{\tau}_{ri} - \tau_{ri})^2 = EA(\tilde{f}_i, f_i)^2 + o(\kappa_r) = O(\kappa_r). \tag{A12}$$

Relation (A12) and $\nu \geq \gamma + 1$ imply $E|R_1(\tilde{f}_i, f_i)| = O(\kappa_r)$. By Lemmas 1(c) and 1(d), we furthermore have $E|R_3(\tilde{f}_i, f_i)| = o(\kappa_r)$. We thus obtain

$$E(\hat{\tau}_{ri} - \tau_{ri}) = EA(\tilde{f}_i, f_i) + ER_2(\tilde{f}_i, f_i) + O(\kappa_r). \tag{A13}$$

By (A12) and (A13) the assertions of the theorem follow from Lemmas 1(a) and 1(b), given that

$$ER_2(\tilde{f}_i, f_i) = O\left(b^q\sqrt{\kappa_r} + \kappa_r\right), \tag{A14}$$

where $q := \min\{\nu - \gamma, k\}$ [when considering the assumptions of the theorem, note that necessarily $q \geq k - 1$ and $b^{k-1}\sqrt{\kappa_r} = o(b^k)$].

It thus only remains to prove (A14). Lemma 1(a) and (A10) imply that

$$R_2(\tilde{f}_i, f_i) = \frac{1}{a(f_i)} \frac{1}{b^{\gamma+1}} \sum_{j=1}^{n} \int_{s_{j-1}}^{s_j} \int_{\tau_{ri}^*}^{\hat{\tau}_{ri}^*} W^{(\gamma)}\left(\frac{\tau - u}{b}\right) d\tau\, du\, \epsilon_{ij} + O\left(b^q\sqrt{\kappa_r}\right)$$

$$=: R_2^*(\tilde{f}_i, f_i) + O\left(b^q\sqrt{\kappa_r}\right).$$

The only difficulty in proving that $ER_2^*(\tilde{f}_i, f_i) = O(1/nb^{2\mu_r+1})$ obviously stems from the fact that $\hat{\tau}_{ri}^*$ and ϵ_{ij} are not independent.

Now, for any $j \in \{1, \ldots, n\}$ and all $t \in J$, let

$$\tilde{f}_{i,j}(t) := \tilde{f}_i(t) - \int_{s_{j-1}}^{s_j} \frac{1}{b} W\left(\frac{t-u}{b}\right) du\, \epsilon_{ij}, \tag{A15}$$

$$\hat{\tau}_{ri,j}^* := T^*(\tilde{f}_{i,j}). \tag{A16}$$

The arguments leading to upper bounds for the rate of convergence of $|\hat{\tau}_{ri} - \tau_{ri}|^\rho$ might also be applied to $|\hat{\tau}^*_{ri,j} - \hat{\tau}^*_{ri}|^\rho$. For any $j \in \{1, \ldots, n\}$, this leads to

$$E\left|\hat{\tau}^*_{ri,j} - \hat{\tau}^*_{ri}\right|^\rho = \Lambda^\rho EI\left(\tilde{f}_i \notin U_{\varepsilon/2}(f_i)\right) + \Lambda^\rho EI\left(\tilde{f}_{i,j} \notin U_{\varepsilon/2}(f_i)\right)$$

$$+ \frac{1}{d^\rho} E \sup_{t \in J} \left| \int_{s_{j-1}}^{s_j} \frac{1}{b^\gamma} W^{(\mu_r)}\left(\frac{t-u}{b}\right) du\, \epsilon_{ij} \right|^\rho = O\left(\frac{1}{n^\rho b^{\rho\gamma}}\right). \tag{A17}$$

(It is evident that the results of Lemma 1 still apply when replacing $\tilde{f}_i$ by $\tilde{f}_{i,j}$.)

Since for any j, $\hat{\tau}^*_{ri,j}$ is independent of ϵ_{ij}, we obtain

$$ER_2^*\left(\tilde{f}_i, f_i\right) = \frac{1}{a(f_i)} \frac{1}{b^{\gamma+1}} \sum_{j=1}^n E \int_{s_{j-1}}^{s_j} \left\{ \int_{\tau^*_{ri}}^{\hat{\tau}^*_{ri}} W^{(\gamma)}\left(\frac{\tau - u}{b}\right) d\tau \right.$$

$$\left. - \int_{\tau^*_{ri}}^{\hat{\tau}^*_{ri,j}} W^{(\gamma)}\left(\frac{\tau - u}{b}\right) d\tau \right\} du\, \epsilon_{ij}.$$

Taylor expansion leads to

$$ER_2^*\left(\tilde{f}_i, f_i\right) = \frac{1}{a(f_i) b^{\gamma+1}} \sum_{j=1}^n E\left\{ \int_{s_{j-1}}^{s_j} W^{(\gamma)}\left(\frac{\hat{\tau}^*_{ri,j} - u}{b}\right)\left(\hat{\tau}^*_{ri} - \hat{\tau}^*_{ri,j}\right) du\, \epsilon_{ij} \right\}$$

$$+ o\left(\frac{1}{nb^{2\mu_r+1}}\right). \tag{A18}$$

The remainder term can be bounded by $\sum_{j=1}^n |L/n \cdot 1/b^{\gamma+2} \cdot (\hat{\tau}^*_{ri,j} - \hat{\tau}^*_{ri})^2 \cdot \epsilon_{ij}|$ which is $o(1/(nb^{2\mu_r+1}))$, as follows from (A17) and an application of the Cauchy–Schwarz inequality.

Similar to (A11) a more detailed analysis of $\hat{\tau}^*_{ri,j} - \hat{\tau}^*_{ri}$ yields

$$\hat{\tau}^*_{ri} - \hat{\tau}^*_{ri,j} = A\left(\tilde{f}_i, \tilde{f}_{i,j}\right) + \sum_{s=1}^3 R_s\left(\tilde{f}_i, \tilde{f}_{i,j}\right). \tag{A19}$$

Using the Cauchy–Schwarz inequality, A(17) and Lemma 1(d), it can easily be checked that for $s = 2, 3$,

$$\frac{1}{a(f_i)} \frac{1}{b^{\gamma+1}} \sum_{j=1}^n E\left\{ \int_{s_{j-1}}^{s_j} W^{(\gamma)}\left(\frac{\hat{\tau}^*_{ri,j} - u}{b}\right) du\, R_s\left(\tilde{f}_i, \tilde{f}_{i,j}\right)\epsilon_{ij} \right\} = o\left(\frac{1}{nb^{2\mu_r+1}}\right).$$

The same relation can be deduced for $s = 1$, when additionally using that by Lemma 1(a) $\sup_{t \in \mathcal{J}} |E\tilde{f}^{(\gamma+1)}_{i,j}(t)| = O(1)$, and that by Lemma 1(c)

$$E \sup_{t \in \mathcal{J}} \left| \tilde{f}^{(\gamma+1)}_{i,j}(t) - E\tilde{f}^{(\gamma+1)}_{i,j}(t) \right|^\rho = o\left((\log n)^{2\rho}/(nb^{2\gamma+3})^{\rho/2}\right) = o(1/b^\rho).$$

Recalling the independence of $\hat{\tau}^*_{ri,j}$ and ϵ_{ij}, (A18) and (A19) thus imply

$$ER_2^*(\tilde{f}_i, f_i) = \frac{\sigma_i^2}{a(f_i)} \frac{-1}{b^{2\gamma+1}} E\left(\sum_{j=1}^{n} \frac{1}{a(\tilde{f}_{i,j})} \int_{s_{j-1}}^{s_j} W^{(\gamma)}\left(\frac{\hat{\tau}^*_{ri,j} - u}{b} \right) du \right.$$

$$\left. \times \int_{s_{j-1}}^{s_j} W^{(\mu_r)}\left(\frac{\hat{\tau}^*_{ri,j} - u}{b} \right) du \right) + o\left(\frac{1}{nb^{2\mu_r+1}} \right).$$

Based on (A17) and on the arguments of Appendix (A1) of Gasser and Müller (1984), it is easily seen that

$$\frac{1}{b^{2\gamma+1}} E \sum_{j=1}^{n} \left\{ \frac{1}{a(\tilde{f}_{i,j})} \int_{s_{j-1}}^{s_j} W^{(\gamma)}\left(\frac{\hat{\tau}^*_{ri,j} - u}{b} \right) du \int_{s_{j-1}}^{s_j} W^{(\mu_r)}\left(\frac{\hat{\tau}^*_{ri,j} - u}{b} \right) du \right\}$$

$$= \frac{1}{nb^{2\gamma+1}} E\left\{ \frac{1}{a(\tilde{f}_i)} \int_{-\infty}^{\infty} W^{(\gamma)}\left(\frac{\hat{\tau}^*_{ri} - u}{b} \right) W^{(\mu_r)}\left(\frac{\hat{\tau}^*_{ri} - u}{b} \right) du \right\}$$

$$+ o\left(\frac{1}{nb^{2\mu_r+1}} \right).$$

The last relation can be inferred by using the arguments of Appendix (A1) of Gasser and Müller (1984). But according to the properties of W, W^q is symmetric at 0 if q is even and it is antisymmetric if q is odd. Hence,

$$\int_{-\infty}^{\infty} W^{(\gamma)}\left(\frac{\hat{\tau}^*_{ri} - u}{b} \right) W^{(\mu_r)}\left(\frac{\hat{\tau}^*_{ri} - u}{b} \right) du = 0$$

if $\hat{\tau}^*_{ri} \in [a_0 + b, a_1 - b]$. Since $b \to 0$ and $\hat{\tau}^*_{ri} \in U_{\delta/2}(\tau_{ri})$, we consequently obtain

$$ER_2^*(\tilde{f}_i, f_i) = o\left(\frac{1}{nb^{2\mu_r+1}} \right),$$

which completes the proof of Theorem 1. □

PROOF OF THEOREM 2. Similar as above, compactness of $\mathscr{R}$ implies the existence of a compact subinterval $\mathscr{J} \neq J$ with $T_r(\mathscr{R}) \subseteq \mathscr{J}$. Let $o_r := \mu_r - 1$. We will use some of the notation introduced in the proof of Theorem 1. In particular, let $\Lambda, \gamma := \mu_r + 1, U_c(v), U_c(t)$ (for some $c > 0$) be defined as before. Furthermore, let $\varepsilon_1 > 0$ and $\varepsilon_2 > 0$ be defined in such a way that for some appropriate $\delta_1, d_1 > 0$ and $\delta_2, d_2 > 0$, (A9) holds, when replacing there T_r by the structural e-functionals ψ_0 and ψ_1. Let $\psi_{0i} := \psi_0(f_i)$ and $\psi_{1i} := \psi_1(f_i)$. For all $v, w \in \mathscr{R}$ we have $|v^{\mu_r}(T_r(v))| > 0$ and $\operatorname{sign}(v^{\mu_r}(T_r(v))) = \operatorname{sign}(w^{\mu_r}(T_r(w)))$. This is evident, since if for some $v \in \mathscr{R}$ it holds $v^{(\mu_r)}(T_r(v)) = 0$, for any small $e > 0$ there has to exist a $w \in U_e(v) \subset D_r$ such that the segment $[\psi_0(w), \psi_1(w)]$ is not monotone. Arguments similar to that used at the beginning of the proof of Theorem 1 now show that there are some $\varepsilon_3, d > 0$ such that for all $v \in \mathscr{R}$

and each $w \in U_{\varepsilon_3}(v)$,

$$|w^{(\mu_r)}(t)| > d \quad \text{for all } t \text{ in the segment joining } T_r(v) \text{ and } T_r(w).$$

Set $\varepsilon := \min\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}$. Let $\psi_0^*, \psi_1^*, T_r^*$ be equivalent to ψ_0, ψ_1, T_r on $U_\varepsilon(f_i)$, and set $T_r^*(v) = \tau_{ri}$, $\psi_0^*(v) = \psi_0^*(f_i)$, $\psi_1^*(v) = \psi_1^*(f_i)$ for all $v \notin U_\varepsilon(f_i)$. We will abbreviate $\psi_0^*(f_i)$, $\psi_1^*(f_i)$, $T_r^*(f_i)$, $\psi_0^*(\tilde{f}_i)$, $\psi_1^*(\tilde{f}_i)$, $T_r^*(\tilde{f}_i)$ by $\psi_{0i}^*, \psi_{1i}^*, \tau_{ri}^*, \hat{\psi}_{0i}^*, \hat{\psi}_{1i}^*, \hat{\tau}_{ri}^*$. By construction, if $\tilde{f}_i \in U_\varepsilon(f_i)$, it holds

$$|\hat{\tau}_{ri} - \tau_{ri}| \le \frac{1}{d}\Big| f_i^{(o_r)}(\tau_{ri}^*) - \tilde{f}_i^{(o_r)}(\tau_{ri}^*) - p\Big(f_i^{(o_r)}(\psi_{0i}^*) - \tilde{f}_i^{(o_r)}\big(\hat{\psi}_{0i}^*\big)\Big) - (1-p)\Big(f_i^{(o_r)}(\psi_{1i}^*) - \tilde{f}_i^{(o_r)}\big(\hat{\psi}_{1i}^*\big)\Big)\Big|.$$

Furthermore, $f_i^{(o_r)}(\psi_{si}^*) = \tilde{f}_i^{(o_r)}(\hat{\psi}_{si}^*) + \tilde{f}_i^{(\gamma)}(\xi_{si})(\psi_{si}^* - \hat{\psi}_{si}^*)^2/2$, $s = 0, 1$, for some suitable mean values ξ_{si}. The compactness of $\mathscr{R}$ implies

$$\sup_{v \in \mathscr{R}} \sup_{w \in U_\varepsilon(v)} \sup_{t \in \mathscr{I}} |w^{(\gamma)}| < \infty.$$

For any ρ there thus exist generic constants H_ρ and H_ρ^* such that

$$E|\hat{\tau}_{ri} - \tau_{ri}|^\rho < \Lambda^\rho \sum_{s=0}^{\gamma} P\Big(\sup_{t \in \mathscr{I}} |f_i^{(s)}(t) - \tilde{f}_i^{(s)}(t)| > \varepsilon/(\gamma+1)\Big) + \frac{H_\rho}{d^\rho} E\Bigg(\Big|f_i^{(o_r)}(\tau_{ri}^*) - \tilde{f}_i^{(o_r)}(\tau_{ri}^*)\Big|^\rho + \sum_{s=0}^{1}\Big[\Big|f_i^{(o_r)}(\psi_{si}^*) - \tilde{f}_i^{(o_r)}(\psi_{si}^*)\Big|^\rho + H_\rho^*\big|\psi_{si}^* - \hat{\psi}_{si}^*\big|^{2\rho}\Big]\Bigg).$$

Combining (A10) with the results of Lemma 1, we thus obtain

$$E|\hat{\tau}_{ri} - \tau_{ri}|^\rho = O\Big(\max\Big\{b^k, 1/\sqrt{nb^{2o_r+1}}\Big\}^\rho\Big). \tag{A20}$$

Based on (A20), it is possible to derive the exact rates of convergence. With $z_i := 1/f_i^{(\mu_r)}(\tau_{ri}^*)$ a closer inspection of $\hat{\tau}_{ri} - \tau_{ri}$ leads to

$$\hat{\tau}_{ri} - \tau_{ri} = A_i + \sum_{q=1}^{3} R_{qi}, \tag{A21}$$

where with $p_0 = p$, $p_1 = (1-p)$,

$$A_i := z_i \cdot \Bigg(f_i^{(o_r)}(\tau_{ri}^*) - \tilde{f}_i^{(o_r)}(\tau_{ri}^*) - \sum_{s=0}^{1} p_s\Big(f_i^{(o_r)}(\psi_{si}^*) - \tilde{f}_i^{(o_r)}(\psi_{si}^*)\Big)\Bigg),$$

$$R_{1i} := z_i \sum_{s=0}^{1} p_s\Big[\Big(\tilde{f}_i^{(\mu_r)}(\psi_{si}^*) - f_i^{(\mu_r)}(\psi_{si}^*)\Big)\Big(\hat{\psi}_{si}^* - \psi_{si}^*\Big) + \tilde{f}_i^{(\gamma)}(\xi_{si})\Big(\hat{\psi}_{si}^* - \psi_{si}^*\Big)^2/2\Big]$$

for some suitable mean values ξ_{si},

$$R_{2i} := z_i\Big(f_i^{(\mu_r)}(\tau_{ri}^*) - \tilde{f}_i^{(\mu_r)}(\tau_{ri}^*)\Big)\big(\hat{\tau}_{ri}^* - \tau_{ri}^*\big) - z_i\tilde{f}_i^{(\gamma)}(\xi_{ri}^*)\big(\tau_{ri}^* - \hat{\tau}_{ri}^*\big)^2/2$$

for some suitable mean values ξ_{ri}^*,

$$R_{4i} := (\hat{\tau}_{ri} - \tau_{ri} - A_i)\, I\big(\tilde{f}_i \notin U_\varepsilon(f_i)\big).$$

Using (A20), Lemma 1 and Theorem 1, an analysis of the above terms immediately leads to the assertion of Theorem 2. □

PROOF OF THEOREM 3. Let $\kappa := b^{2k} + 1/(nb^{2o+1})$. Relations (A10) and (A20) imply that $P(\hat{\tau}_{ri} = \alpha) = o(n^{-2})$ for all $r \in \{1, \ldots, l\}$ and $i \in \{1, \ldots, m\}$. Furthermore, by (A11) and (A21) we obtain $E(\hat{\tau}_{ri} - E\hat{\tau}_{ri})(\hat{\tau}_{si} - E\hat{\tau}_{si}) = o(1/(nb^{2o+1}))$ for $r \neq s$, since $\tilde{f}_i^{(o_r)}(\tau_{ri}) - E\tilde{f}_i^{(o_r)}(\tau_{ri})$ and $\tilde{f}_i^{(o_s)}(\tau_{si}) - E\tilde{f}_i^{(o_s)}(\tau_{si})$ are independent for all b small enough. Using Assumption 2 and the results of Theorems 1 and 2, some easy computations now immediately lead to assertions (I) and (II) of the theorem.

To prove assertion (III), note that by Theorems 1 and 2, $E\hat{\tau}_{ri} \in J$ for all i and r, if n is sufficiently large. For given i, k, $i < k$, let $\bar{\tilde{\tau}}_{i,k;r} := (E\hat{\tau}_{ri} + E\hat{\tau}_{rk} + \sum_{j \neq i,k} \hat{\tau}_{rj} \cdot I(\hat{\tau}_{rj} \neq \alpha))/(2 + \sum_{j \neq i,k} \hat{\tau}_{rj} \cdot I(\hat{\tau}_{rj} \neq \alpha))$. Set $\bar{\tilde{\tau}}_{i,k} := (\bar{\tilde{\tau}}_{i,k;1}, \ldots, \bar{\tilde{\tau}}_{i,k;l})'$ and $\bar{\tau}^* := ((1/m)\sum_{i=1}^m E\hat{\tau}_{1i}, \ldots, (1/m)\sum_{i=1}^m E\hat{\tau}_{li})'$. Using Assumption 2, Theorems 1 and 2 [together with relations (A10) and (A20)] and the independence of $\hat{\tau}_{i_1}$ from $\hat{\tau}_{i_2}$ $(i_1 \neq i_2)$, we obtain

$$
\begin{aligned}
\operatorname{cov}(\hat{g}_i(t), \hat{g}_k(t)) &= \operatorname{cov}\left(G_{(\hat{\tau}_i, \bar{\tilde{\tau}}_{i,k})}(t), G_{(\hat{\tau}_k, \bar{\tilde{\tau}}_{i,k})}(t)\right) \\
&\quad + \frac{1}{m}\sum_{r=1}^{l}\Bigg(v_{ri}G^*_{(\tau_i, \bar{\tau}; r)}(t)G^{**}_{(\tau_k, \bar{\tau}; r)}(t) \\
&\qquad + v_{rk}G^*_{(\tau_k, \bar{\tau}; r)}(t)G^{**}_{(\tau_i, \bar{\tau}; r)}(t) \\
&\qquad + G^{**}_{(\tau_i, \bar{\tau}; r)}(t)G^{**}_{(\tau_k, \bar{\tau}; r)}(t)\frac{1}{m}(v_{ri} + v_{rk})\Bigg) + o\left(\frac{\kappa}{m} + n^{-2}\right),
\end{aligned} \tag{A22}
$$

$$
\begin{aligned}
&\operatorname{cov}\left(G_{(\hat{\tau}_i, \bar{\tilde{\tau}}_{i,k})}(t), G_{(\hat{\tau}_k, \bar{\tilde{\tau}}_{i,k})}(t)\right) \\
&\quad = E\left(G_{(\hat{\tau}_i, \bar{\tilde{\tau}}_{i,k})}(t) - EG_{(\hat{\tau}_i, \bar{\tilde{\tau}}_{i,k})}(t)\right)\left(G_{(\hat{\tau}_k, \bar{\tilde{\tau}}_{i,k})}(t) - G_{(\hat{\tau}_k, \bar{\tau}^*)}(t)\right) \\
&\quad = \sum_{r=1}^{l}\left(G^{**}_{(\tau_i, \bar{\tau}; r)}(t)G^{**}_{(\tau_k, \bar{\tau}; r)}(t)\frac{1}{m^2}\sum_{s \neq i,k} v_{rs}\right) + o\left(\frac{\kappa}{m} + n^{-2}\right).
\end{aligned} \tag{A23}
$$

Combining (A22) and (A23) proves the desired result. □

PROOF OF THEOREM 4. By the compactness of $\mathscr{R}$ we have $\text{con}_1 := \sup_{v=0,1,2}\sup_{v \in \mathscr{R}}\sup_{t \in J}|v^{(\nu)}(t)| < \infty$. Now define an arbitrary continuous operator asserting to each $v \in \mathscr{R}$ a function $\check{v}\colon \mathbb{R} \to \mathbb{R}$ such that $\check{v}|_J = v$ and such that $\check{v}$ and its first two derivatives can be bounded by 2con_1. By definition of our kernel estimators, there is a $\text{con}_2 < \infty$ such that $|\check{f}_i^{(\nu)}(t) - E\hat{f}_i^{(\nu)}(t)| \leq \text{con}_2$ for all $t \in \mathbb{R}$, every $\nu = 0, 1, 2$, and all i.

For $t \in J_\varepsilon$ we obviously have

$$
\begin{aligned}
&\hat{f}_i(\hat{g}_i(t)) - f_i(g_i(t)) \\
&\quad = \hat{f}_i(\hat{g}_i(t)) - \check{f}_i(g_i(t)) = A_{1i}(t) + A_{2i}(t) + R_i(t),
\end{aligned} \tag{A24}
$$

where

$$A_{1i}(t) := \hat{f}_i(g_i(t)) - \check{f}_i(g_i(t)),$$

$$A_{2i}(t) := \check{f}_i(\hat{g}_i(t)) - \check{f}_i(g_i(t)),$$

$$R_i(t) := \int_{g_i(t)}^{\hat{g}_i(t)} \left(\hat{f}_i'(s) - \check{f}_i'(s)\right) ds.$$

It is immediately verified that there exist some $\gamma_1 > a_0$, $\delta_1 < a_1$ such that $g_i(J_\varepsilon) \subset [\gamma_1, \delta_1] \subset J$ for all i, m. Lemma 1(a) thus yields an approximation to the asymptotic bias of $A_{1i}(t)$, $t \in J_\varepsilon$. Analyzing $EA_{2i}(t)$ can be based on a Taylor expansion of $\check{f}_i(\hat{g}_i(t)) - \check{f}_i(g_i(t))$ and on Theorem 3. It is then immediately seen that assertion (I) of the theorem holds if uniformly for all $t \in J_\varepsilon$,

$$ER_i(t) = O\left(1/(nb^{2\mu+1})\right) + o(b^k). \tag{A25}$$

Choose some γ_2, δ_2 with $\gamma_1 > \gamma_2 > a_0$, $\delta_1 < \delta_2 < a_1$, and let $\mathcal{J} := [\gamma_2, \delta_2]$. We can infer from (A10), (A20) and Assumption 2 that $P(\hat{g}_i(J_\varepsilon) \not\subset \mathcal{J}) = o(n^{-2})$. Together with Lemma 1 and Theorems 1–3, it follows that in order to prove (A25) it suffices to show that for all $r \in \{1, \ldots, l\}$,

$$E \sum_{j=1}^{n} \int_{g_i(t)}^{g_i(t)+H_{ir}(t)(\hat{\tau}_{ri}-\tau_{ri})} \frac{1}{c^2} \int_{s_{j-1}}^{s_j} W'\left(\frac{x-u}{c}\right) du\, dx\, \epsilon_{ij} \tag{A26}$$

$$= O\left(1/(nb^{2\mu+1})\right),$$

uniformly for all $t \in J_\varepsilon$, where $H_{ir}(t) := G^*_{(\tau_i, \bar{\tau}; r)}(t) + (1/m) G^{**}_{(\tau_i, \bar{\tau}; r)}(t)$.

Recall the definitions of $\hat{\tau}^*_{ri}$ at the beginning of the proofs of Theorems 1 and 2. Note that the error in replacing $\hat{\tau}_{ri}$ by $\hat{\tau}^*_{ri}$ on the left-hand side of (A26) is of the order $o(1/(nb^{2\mu+1}))$.

Now, define $\tilde{f}_{ri,j}$ and $\hat{\tau}^*_{ri,j}$ by (A15) and (A16). Since $\hat{\tau}^*_{ri,j}$ is independent of ϵ_{ij}, we obtain that (A26) holds if and only if

$$ER^*_{r,i}(t) := \sum_{j=1}^{n} E\left(\int_{g_i(t)}^{g_i(t)+H_{ir}(t)(\hat{\tau}^*_{ri}-\tau_{ri})} \frac{1}{c^2} \int_{s_{j-1}}^{s_j} W'\left(\frac{x-u}{c}\right) du\, dx\, \epsilon_{ij} \right.$$

$$\left. - \int_{g_i(t)}^{g_i(t)+H_{ir}(t)(\hat{\tau}^*_{ri,j}-\tau_{ri})} \frac{1}{c^2} \int_{s_{j-1}}^{s_j} W'\left(\frac{x-u}{c}\right) du\, dx\, \epsilon_{ij} \right)$$

$$= O\left(1/(nb^{2\mu+1})\right),$$

uniformly for all $t \in J_\varepsilon$.

Let T_r be a structural e-functional. Then $ER^*_{r,i}$ might be analyzed in the same way as $ER^*_2(\tilde{f}_i, f_i)$ in the proof of Theorem 1 [note that

$1/(n^2c^3b^{2\mu_r+2}) = o(1/(nb^{2\mu_r+1}))]$. Since $\int W'(x)\,dx = 0$, this leads to

$$
\begin{aligned}
|ER^*_{r,i}(t)| &= \left| E\frac{H_{ir}(t)\sigma_i^2}{nc^2b^{\mu_r+1}a(\tilde{f}_i)} \right. \\
&\qquad \left. \times \int_{-\infty}^{\infty} W'\left(\frac{g_i(t) + H_{ir}(t)(\hat{\tau}^*_{ri} - \tau_{ri}) - u}{c}\right) W^{(\mu_r)}\left(\frac{\hat{\tau}^*_{ri} - u}{b}\right) du \right| \\
&\qquad + o\left(\frac{1}{nb^{2\mu_r+1}}\right) \\
&= \left| E\frac{H_{ir}(t)\sigma_i^2}{ncb^{\mu_r+1}a(\tilde{f}_i)} \right. \\
&\qquad \left. \times \int_{-\infty}^{\infty} W'(x) W^{(\mu_r)}\left(\frac{\hat{\tau}^*_{ri} - g_i(t) - H_{ir}(t)(\hat{\tau}^*_{ri} - \tau_{ri})}{b} + \frac{c}{b}x\right) dx \right| \\
&\qquad + o\left(\frac{1}{nb^{2\mu_r+1}}\right) \\
&= O\left(\frac{1}{nb^{2\mu_r+1}}\right).
\end{aligned}
$$

Alternatively, if T_r is a structural p-functional (A26) can be shown in a very similar manner. Using (A17), the arguments used to prove (A20) yield $E(\hat{\tau}^*_{ri} - \hat{\tau}^*_{ri,j})^\rho = O(1/(n^\rho b^{\rho(o+1)}))$. A further analysis similar to that sketched above [apply expansion (A21) instead of (A11)] then leads to the desired result.

For the proof of assertion (II) note that $R_i(t) = (\hat{f}_i'(\xi_i(t)) - \check{f}_i'(\xi_i(t)))(\hat{g}_i(t) - g_i(t))$ for some suitable mean value function $\xi_i(t)$. It holds $P(\xi_i(t) \notin \mathcal{J}) = o(n^{-2})$. Additionally, using Lemmas 1(a) and 1(c), our assumptions on c imply that $E|\check{f}'(\xi_i(t)) - \hat{f}_i'(\xi_i(t))|^\rho = o(1)$ for any ρ. By Lemma 1(b) and Theorem 3, appropriate Taylor expansions and the Cauchy–Schwarz inequality thus lead to

$$
\operatorname{cov}(R_i(t), A_{1i} + A_{2i}(t) + R_i(t)) = o\left(b^{2k} + \frac{1}{nb^{2o+1}}\right),
$$

uniformly for all $t \in J_\varepsilon$. Using Lemma 1 and Theorem 3 to analyze $\operatorname{var}(A_{1i}(t))$ and $\operatorname{var}(A_{2i}(t))$, it is easily seen that assertion (b) holds, given that

$$
\int_{J_\varepsilon} \operatorname{cov}(A_{1i}(t), A_{2i}(t))\,dt = o\left(\frac{1}{nb^{2o+1}}\right).
$$

To prove this, it suffices to derive that for all $r \in \{1, \ldots, l\}$,

$$
\begin{aligned}
&E\sum_{j=1}^{n} \int_{J_\varepsilon} \left(\frac{1}{c} \int_{s_{j-1}}^{s_j} W\left(\frac{g_i(t) - u}{c}\right) du\, \epsilon_{ij} \right) \check{f}_i'(g_i(t)) H_{ir}(t)\,dt(\hat{\tau}_{ri} - E\hat{\tau}_{ri}) \\
&\qquad = o\left(\frac{1}{nb^{2o+1}}\right).
\end{aligned} \tag{A27}
$$

Using (A11) and (A21), some easy computations show that (A27) is fulfilled if

$$\text{(A28)}\quad \int_{J_\varepsilon}\left|\frac{1}{nb^{o_r+1}}\int_{-1}^{1}W(x)W^{(o_r)}\left(\frac{\xi_{ri}-g_i(t)}{b}+\frac{c}{b}x\right)dx\right|dt=o\left(\frac{1}{nb^{2o+1}}\right),$$

with $\xi_{ri}=\tau_{ri}$ for structural e-functionals, and with either $\xi_{ri}=\tau_{ri}$, $\xi_{ri}=\psi_{0i}$, $\xi_{ri}=\psi_{1i}$ for structural p-functionals. Relation (A28) is evident for $o_r>0$. If $o_r=0$, it can easily be verified when noting that $c/b=O(1)$ and $W^{(o_r)}((\hat{\tau}_{ri}^*-x)/b+(c/b)x)=0$ if $|\hat{\tau}_{ri}^*-x|>b(1+c/b)$.

It remains to prove assertion (III). Recall our assumptions on G. For $t\in J_\varepsilon$, $\tau,\tau^*\in\mathscr{H}_l$ and $x,x^*\in\mathscr{H}_l$,

$$\begin{aligned}\text{(A29)}\quad \hat{f}_i\big(G_{(\tau^*,x^*)}(t)\big)=\hat{f}_i\big(G_{\tau,x}(t)\big)+\check{f}_i'\big(G_{(\tau,x)}(t)\big)\sum_{r=1}^{l}\big(&G_{(\tau,x;r)}^{*}(t)(\tau_r^*-\tau_r)\\&+G_{(\tau,x;r)}^{**}(t)(x_r^*-x_r)\big)+V_{\hat{f}_i,\tau,\tau^*,x,x^*}(t),\end{aligned}$$

where for some $Q_1,Q_2,Q_3<\infty$ the remainder term can be bounded by

$$\begin{aligned}V_{\hat{f}_i,\tau,\tau^*,x,x^*}<{}&Q_1\Big(\sup_{t\in\mathscr{J}}\big|f_i(t)-E\hat{f}_i(t)\big|+\sup_{t\in\mathbb{R}}\big|\hat{f}_i'(t)-E\hat{f}_i'(t)\big|\Big)\\&\times\big(\|\tau-\tau^*\|_2+\|x-x^*\|_2\big)+Q_2\big(\|\tau-\tau^*\|_2^2+\|x-x^*\|_2^2\big)\\&+Q_3\Big(I\big(G_{(\tau^*,x^*)}(J_\varepsilon)\not\subset\mathscr{J}\big)+I\big(G_{(\tau,x)}(J_\varepsilon)\not\subset\mathscr{J}\big)\Big).\end{aligned}$$

Let $i\neq k$. Recall Lemmas 1(a) and 1(c) and the independence of $\hat{f}_i,\hat{\tau}_i$ from $\hat{f}_k,\hat{\tau}_k$. Bounding differences by (A29),

$$\operatorname{cov}\big(\hat{f}_i(\hat{g}_i(t)),\hat{f}_k(\hat{g}_k(t))\big)=\operatorname{cov}\Big(\hat{f}_i\big(G_{(\hat{\tau}_i,\bar{\tau})}(t)\big),\hat{f}_k\big(G_{(\hat{\tau}_k,\bar{\tau})}(t)\big)\Big)$$

might now be analyzed in a way similar to (A22) and (A23). This yields

$$\begin{aligned}&\int_{J_\varepsilon}\operatorname{cov}\big(\hat{f}_i(\hat{g}_i(t)),\hat{f}_k(\hat{g}_k(t))\big)\\&\quad=E\sum_{r=1}^{l}\int_{J_\varepsilon}\big(\hat{f}_i(g_i(t))-E\hat{f}_i(g_i(t))\big)\frac{1}{m}f_k'(g_k(t))G_{(\hat{\tau}_k,\bar{\tau})}^{**}(t)\,dt\big(\hat{\tau}_{ri}-E\hat{\tau}_{ri}\big)\\&\qquad+E\sum_{r=1}^{l}\int_{J_\varepsilon}\big(\hat{f}_k(g_k(t))\\&\qquad\qquad-E\hat{f}_k(g_k(t))\big)\frac{1}{m}f_i'(g_i(t))G_{(\hat{\tau}_i,\bar{\tau})}^{**}(t)\,dt\big(\hat{\tau}_{rk}-E\hat{\tau}_{rk}\big)\\&\qquad+\frac{1}{m}\int_{J_\varepsilon}f_i'(g_i(t))f_k'(g_k(t))C_{ik}(t)\,dt\\&\qquad+o\big(\big(b^{2k}+1/(nb^{2o+1})\big)/m\big)+o(n^{-2}).\end{aligned}$$

Compared to the expressions obtained in the proof of Theorem 3, the first two terms on the right-hand side are additionally induced by the randomness of $\hat{f}_i, \hat{f}_k$ (with respect to ϵ_{ij}). Arguments analogous to those used to deduce (A27) show that they can be bounded by $o(1/(mnb^{2o+1}))$. This completes the proof of the theorem. □

REFERENCES

BRILLINGER, D. R., (1988). Some statistical methods for random process data from seismology and neurophysiology. *Ann. Statist.* **16** 1–54.

BRODLIE, K. W. (1980). A review of methods for curve and function drawing. In *Mathematical Methods for Computer Graphics and Design* (K. W. Brodlie, ed.) 1–77. Academic, London.

CHENG, K. F. and LIN, P. E. (1981). Nonparametric estimation of a regression function. *Z. Wahrsch. Verw. Gebiete* **57** 223–233.

EUBANK, R. (1988). *Spline Smoothing and Nonparametric Regression*. Dekker, New York.

FRITSCH, F. N. and BUTLAND, J. (1984). A method for constructing local monotone piecewise cubic interpolants. *SIAM J. Sci. Statist. Comput.* **5** 300–304.

GASSER, TH. and KNEIP, A. (1991). Searching for structure in samples of curves. Technical report.

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

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

GASSER, TH., MÜLLER, H. G., KÖHLER, W. and MAMMITZSCH, V. (1985). Kernels for nonparametric curve estimation. *J. Roy. Statist. Soc. Ser. B* **47** 238–252.

GASSER, TH., MÜLLER, H. G., KÖHLER, W., MOLINARI, L. and PRADER, A. (1984). Nonparametric regression analysis of growth curves. *Ann. Statist.* **12** 210–229.

GASSER, TH., PIETZ, J., SCHELLBERG, D. and KÖHLER, W. (1988). Visual evoked potentials of mildly mentally retarded and control children. *Developmental Medicine Child Neurology* **30** 638–645.

HÄRDLE, W. (1990). *Applied Nonparametric Regression*. Econometric Society Monograph Series. Cambridge Univ. Press.

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

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

KNEIP, A. and GASSER, TH. (1988). Convergence and consistency properties for self-modeling nonlinear regression. *Ann. Statist.* **16** 82–113.

LAMPERTI, J. (1966). *Probability*. Benjamin, New York.

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

MÖCKS, J. (1986). The influence of latency variations in principal component analysis of event-related potentials. *Psychophysiology* **23** 480–484.

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

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

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

PRIESTLEY, M. B. and CHAO, M. T. (1972). Non-parametric curve fitting. *J. Roy. Statist. Soc. Ser. B* **34** 385–392.

RAO, C. R. (1958). Some statistical methods for the comparison of growth curves. *Biometrics* **14** 1–17.

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

SCHÄFER, F., SEIDEL, C., BINDING, A., GASSER, TH., LARGO, PRADER and SCHÄRER, K. (1990). Pubertal growth in chronic renal failure. *Pediatric Research* **28** 5–10.

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

WIRTSCHAFTSTHEORIE II
UNIVERSITÄT BONN
5300 BONN 1
GERMANY

ZENTRALINSTITUT FÜR SEELISCHE GESUNDHEIT
6800 MANNHEIM 1
GERMANY