The Annals of Statistics
1988, Vol. 16, No. 1, 82–112

CONVERGENCE AND CONSISTENCY RESULTS FOR SELF-MODELING NONLINEAR REGRESSION[1]

BY ALOIS KNEIP AND THEO GASSER

Universität Heidelberg and Zentralinstitut für Seelische Gesundheit, Mannheim

This paper is concerned with parametric regression models of the form $Y_{ij} = f(t_{ij}, \theta_i) + \text{error}$, $i = 1, \dots, n$, $j = 1, \dots, T_i$, where the continuous function f may depend nonlinearly on the known regressors t_{ij} and the unknown parameter vectors θ_i. The assumption of an a priori known f is dropped and replaced by the requirement that qualitative information about the structure of the model is available or can be generated by a preliminary exploratory data analysis. This framework—allowing both f and the individual parameter vectors to be unknown—necessitates a detailed discussion of identifiability of model and parameters. A method is then proposed for the simultaneous estimation of f and θ_i by making use of the prior information. An iterative algorithm simplifying computation of the estimates is presented, and for $\min\{n, T_1, \dots, T_n\} \to \infty$ conditions for strong uniform consistency of the resulting estimators of f and strong consistency of the estimators of θ_i are established. Some examples illustrating the method are included.

1. Introduction. Many experiments in biomedicine and in the physical sciences are initiated to study a biological (chemical,...) process by a number of independent realizations. Therefore, based on some experimental design, at consecutive times (or ages,...) t_{ij}, observations Y_{ij}, $j = 1, \dots, T_i$, $i = 1, \dots, n$, are obtained for a sample of individuals (or experimental units,...) of size n. In such situations it is often adequate to assume that Y_{ij}, the jth observation on the ith individual, satisfies the regression model

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

Here the (unknown) regression function $f_i \in C(\mathbb{R}, \mathbb{R})$ *may depend nonlinearly on the known regressors* $t_{ij} \in [a, b] =: J$, $a, b \in \mathbb{R}$ *known*; $a < b$. *The unknown error terms* ε_{ij} *are realizations of independent random variables with expectation* 0.

Often the sample of individual regression curves will show a certain homogeneity in structure leading to the assumption that they are generated according to some parametric model

$$f_i(t) = f(t_i, \theta_i), \qquad \text{for all } t \in J, i = 1, \dots, n. \tag{2}$$

The function $f \in C(J \times \mathscr{R}, \mathbb{R})$ *represents the typical functional shape common to all* f_i, *and it is usually unknown. It may depend nonlinearly on the unknown individual parameter vectors* $\theta_i \in \mathscr{R}$, $i = 1, \dots, n$, *representing the individual variation.* $\mathscr{R}$ *denotes some subset of a (finite-dimensional) Euclidean space* $\mathscr{E}$.

Received March 1985; revised June 1987.

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

AMS 1980 *subject classifications*. Primary 62J02; secondary 62G05, 62F11.

Key words and phrases. Nonlinear regression, model selection, method of sieves, consistency.

Classical nonlinear regression analysis is based on the assumption that the model function f is specified a priori. Then the individual parameter vectors θ_i can be estimated by nonlinear least squares, yielding also an estimate of the individual regression functions f_i. The requirement of a prespecified model function is very restrictive, and in many applications the specification of f does not rely on more than a clever guess.

Self-modeling nonlinear regression (SEMOR) offers more flexibility: It suffices to specify an *appropriate class of functions* containing the true model. SEMOR then estimates the "best-fitting" model function and the corresponding parameter vectors simultaneously. In some applications the individual parameters will be of main interest, whereas in other applications the pattern of f will be more relevant. In any case, the choice of a reasonable class of functions containing the true model is a crucial step for SEMOR. It is not possible to construct an estimation method, which, without making use of any a priori knowledge about the structure of the model, automatically grinds out an appropriate model function and interpretable parameters. This is a consequence of the problem of identifiability arising within model (2), to be discussed in Section 2. Some examples will be given, the most important, in view of previous applications, being the following.

Parametric model with a population parameter. Assume

$$f_i(\cdot) = G_\lambda(\cdot, \theta_i) =: f(\cdot, \theta_i), \qquad i = 1, \dots, n,$$

for some unknown but fixed $\lambda \in L \subset \mathbb{R}^l$, $l \in \mathbb{N}$, where $G: L \to C(J \times \mathscr{R}, \mathbb{R})$ is a known continuous function. Day (1966) investigated a particular case of this approach, relevant for modeling growth data, with $\mathscr{R} = \mathbb{R}^2_+ \times \mathbb{R}$, $\mathbb{R}_+ :=]0, \infty[$, and

$$G_\lambda(t; \theta_i) = \theta_i^{(1)}\left(1 + \exp\left(-\theta_i^{(2)}t + \theta_i^{(3)}\right)\right)^{-1/\lambda}, \qquad \lambda \in \mathbb{R}_+;\ \theta_i = \left(\theta_i^{(1)}, \theta_i^{(2)}, \theta_i^{(3)}\right)^{\mathrm{T}}.$$

Shape-invariant modeling (SIM). In this case the lack of knowledge about the functional model is not formalized by some finite-dimensional parameter, but it is embedded in some general structural form.

The simplest SIM assumes $\mathscr{R} = \mathbb{R}^2_+ \times \mathbb{R}^2$ and

$$f_i(t) = \theta_i^{(1)} z\left(\frac{t - \theta_i^{(3)}}{\theta_i^{(2)}}\right) + \theta_i^{(4)} =: f(t, \theta_i),$$

for some unknown but fixed function $z \in C(\mathbb{R}, \mathbb{R})$.

Such a model was first proposed and applied by Lawton, Sylvestre and Maggio (1972), introducing a special version of the estimation procedure defined in Section 3. A similar technique was later used by Stuetzle et al. (1980) for modeling human height growth.

After a discussion of the problems of identifiability in Section 2, the estimation procedure is given in Section 3. In Section 4 an algorithm is proposed, simplifying computation of the estimates. It is a generalization of an algorithm used by Lawton, Sylvestre and Maggio (1972). Within Section 5 consistency properties of the SEMOR estimators are investigated. Further aspects such as numerical considerations and applications will be treated in a forthcoming paper.

2. Model identification. The parametric regression model defined by (1) and (2) is supplemented by the following assumption.

ASSUMPTION 1. The individual parameter vectors $\theta_1, \theta_2, \theta_3, \dots$ are realizations of i.i.d. random variables. The common distribution function F possesses a compact support $S(F) \subseteq \mathscr{R}$.

For simplicity, we will in the following not distinguish notationally between a random variable and its realization. The correct interpretation will become clear from the context.

Note that the basic concepts and results of this paper do not depend on that $\theta_1, \theta_2, \dots$ *are* realizations of a random variable. It is merely required that asymptotically the sequence $(\theta_i)_{i=1,2,\dots}$ *behaves* as if it were generated by a r.v. with distribution function F [i.e., $(1/n)\Sigma_{i=1}^n g(\theta_i) \to \int g(\theta)\, dF(\theta)$ for any bounded continuous function g]. However, Assumption 1 in its present form avoids formal problems and is a natural requirement in the given context.

Model (2) together with Assumption 1 [F and $S(F)$ being possibly unknown] will be referred to as the "SEMOR MODEL." In the following we will assume model (1) and that for some model function f the SEMOR MODEL holds.

We now can formulate the estimation problem to be considered more precisely.

STATEMENT OF THE PROBLEM. *The purpose of this paper is to introduce a general method for estimating the model function f [on $J \times S(F)$] as well as the parameter vectors $\theta_1, \theta_2, \dots$ from* (1) *and* (2), *by using a priori information about the SEMOR MODEL.*

It should be noted that by Assumption 1 only the values of f on $J \times S(F)$ are important for modeling the underlying process.

Given that all possible realizations of the underlying process were known, there still exist infinitely many different functions and parameter vectors satisfying the SEMOR MODEL. If, e.g., $\mathscr{R} = \mathbb{R}$ and $f(t, \theta) = e^{t \cdot \theta}$, then obviously for each $\theta \in \mathscr{R}$, $f(\cdot, \theta) = g(\cdot, \vartheta_\theta)$ with $g(t, \vartheta) = 2^{t \cdot \vartheta}$ and $\vartheta_\theta = \ln 2 \cdot \theta$, and the SEMOR MODEL still holds when replacing there f by g, θ_i by ϑ_{θ_i} and F by the corresponding distribution function of the r.v. ϑ_{θ_1}. Consequently, it is impossible to deduce f and $\theta_1, \theta_2, \dots$ from the observations only. Any estimation procedure has to be based on a general concept formalizing the use of a priori knowledge to obtain identifiability of the SEMOR MODEL. This is the problem of this section.

The problem of identifiability to be considered is two-fold: There is a "global" one, to guarantee that f can be uniquely determined on $J \times S(F)$, and there is an "individual" one, to ensure identifiability of the individual parameter vectors, given f on $J \times S(F)$. Note that only the global problem is peculiar to SEMOR, the individual one is a problem of ordinary nonlinear regression, too. To solve the global problem we have to use a priori information in order to distinguish f from any function g with $g \neq f$ [on $J \times S(F)$] such that for *each* $\theta \in S(F)$, there exists a $\vartheta_\theta \in \mathscr{R}$ satisfying $f(\cdot, \theta) = g(\cdot, \vartheta_\theta)$. With respect to all other functions, at least asymptotically ($n \to \infty$), the sample of individual regression curves [derived from (1)] suffices for identification. To solve the individual problem we have to guarantee that the mapping $f_i \to \theta_i$ is uniquely determined for all i.

The following prior information can be expected within applications:

(a) Knowledge about the underlying process derived from the field of application.
(b) Information about common structural properties of the individual regression curves obtained by a preliminary analysis [e.g., by nonparametric regression and differentiation; see Gasser and Müller, (1984)].

Such prior knowledge will usually lead to at least qualitative assumptions about f, i.e., about the functional dependence between the possible realizations of the given process and an adequate set of parameters. The SIM approach used by Lawton, Sylvestre and Maggio (1972), for example, was motivated by finding a common qualitative structure characterizing the shape of all individual curves of the sample, with individual variation consisting of scale and shift differences.

In the following we will assume that a priori knowledge suffices to specify some structural properties of the model function and an adequate parameter space $\mathscr{R}$. This can be formalized by determining the set M, say, of all functions $g \in C(J \times \mathscr{R}, \mathbb{R})$ satisfying the resulting conditions. In certain cases the requirement $f \in M$ will already solve the problem of identifiability. Note that if according to (a) and (b) it is possible to specify f completely, this leads to $M := \{f\}$.

Sometimes, however, M will determine f only up to an equivalence class of functions being observationally *and* structurally equivalent and differing only with respect to some *homeomorphic parameter transformations*. In this case the selection of an element out of this class will often be a matter of convenience or of ease of interpretation. One way to do this is to introduce an appropriate normalizing condition (e.g., with respect to expectation or variance of the parameters and/or certain values of f). Such a normalizing condition has to be defined on the basis of the transformation law characterizing the given equivalence class, ensuring that it enables identification but imposes no real restriction.

Normalizing conditions are important within the whole range of parametric modeling of data. For example, consider ANOVA. The standard model of one-way classification with random effects involves a population parameter, which is not identifiable according to the model structure. Identifiability then is reached by the normalizing condition that the expectation of the factor is 0.

Such conditions can be formalized by determining the corresponding operator N, which assigns to each element g of the equivalence class, being associated with a distribution function F_g, a unique parameter transformation leading to the normalized element f, being associated with a normalized distribution function F.

A priori information leading to the construction of a set M and a "normalizing" operator N now has to guarantee identifiability of the model. This is formalized by Definition 1. We will use the notation:

(a) $H(\mathscr{R})$ denotes the set of all homeomorphisms from $\mathscr{R}$ onto $\mathscr{R}$.

(b) $\mathscr{F}$ denotes the set of all distributions functions F^* defined on $\mathscr{E}$, possessing a compact support $S(F^*) \subset \mathscr{R}$.

(c) A function $g \in C(J \times \mathscr{R}, \mathbb{R})$ is called "observationally equivalent to f" (abbreviated $g \models f$) if for each $\theta \in S(F)$ there exists a $\vartheta \in \mathscr{R}$ so that $f(\cdot, \theta) = g(\cdot, \vartheta)$.

Furthermore, $g|_{J \times S(F)}$ denotes restriction of g to $J \times S(F)$.

(d) For some $g \in C(J \times \mathscr{R}, \mathbb{R})$ with $g \models f$, an $F^* \in \mathscr{F}$ is called "SEMOR-equivalent to F when associated with g" (abbreviated $F^* \cong_g F$) if it is the distribution function of a random variable ϑ_1 satisfying $f(\cdot, \theta_1) = g(\cdot, \vartheta_1)$.

Definition 1. For some $M \subset C(J \times \mathscr{R}, \mathbb{R})$ and some operator N: $M \times \mathscr{F} \to H(\mathscr{R})$ with $g(\cdot, N(g, F^*)(\cdot)) \in M$ for any $(g, F^*) \in M \times \mathscr{F}$, the SEMOR MODEL is called "completely identifiable by M and N" if the following conditions are satisfied:

(i) $f \in M$.

(ii) For all $g \in M$ with $g|_{J \times S(F)} = f|_{J \times S(F)}$ and each $\theta \in S(F)$,

$$g(\cdot, \theta) \neq g(\cdot, \vartheta), \quad \text{for any } \vartheta \in \mathscr{R} \text{ with } \theta \neq \vartheta.$$

(iii) For all $g \in M$ and $F^* \in \mathscr{F}$ with $g \models f$ and $F^* \cong_g F$,

$$f(\cdot, \theta) = g(\cdot, N(g, F^*)(\theta)), \quad \text{for each } \theta \in S(F). \tag{3}$$

Remark. An important special case of this concept of identifiability is the following: A priori knowledge might be sufficient to specify M such that there *does not* exist a $g \in M$ with $g \neq f$ [on $J \times S(F)$] and $g \models f$. Then the SEMOR MODEL is already completely determined by M, and no normalizing condition is required. In terms of Definition 1, this means that the SEMOR MODEL is completely identifiable by M and the (trivial) operator N defined by $N(g, F^*) =$ id for all $(g, F^*) \in M \times \mathscr{F}$ (abbreviated $N \equiv$ id), where id denotes the identity function on $\mathscr{R}$.

Example 1: Parametric model with a population parameter. Assume that the SEMOR MODEL holds with

$$f(t, \theta) := G_\lambda(t, \theta), \quad \text{for all } \theta \in \mathscr{R},\ t \in J,$$

for some unknown population parameter $\lambda \in L \subset \mathbb{R}^l$, $l \in \mathbb{N}$, where G: $L \to C(J \times \mathscr{R}, \mathbb{R})$ is a known continuous function.

In this case the SEMOR MODEL is completely identifiable by $M := \{G_{\lambda^*} | \lambda^* \in L\}$ (and $N \equiv$ id), given the following conditions are satisfied:

(a) Definition 1 (ii) holds.

(b) For all $\lambda^* \in L$ with $\lambda^* \neq \lambda$, there exists a $\theta \in S(F)$ such that $G_\lambda(\cdot, \theta) \neq G_{\lambda^*}(\cdot, \vartheta)$ for all $\vartheta \in \mathscr{R}$.

These conditions are satisfied for Day's model.

EXAMPLE 2: SIM. Assume $\mathscr{R} = \mathbb{R}_+^2 \times \mathbb{R}^2$ and let the SEMOR MODEL hold with

$$(4) \qquad f(t,\theta) := \theta^{(1)} z\left(\frac{t-\theta^{(3)}}{\theta^{(2)}}\right) + \theta^{(4)}, \quad \text{for all } \theta \in \mathscr{R},\ t \in J,$$

for some unknown function $z \in C(\mathbb{R},\mathbb{R})$.

This model is much larger and identifiability needs more consideration. Let $\mathbf{M}$ denote the set of all functions $g \in C(J \times \mathscr{R}, \mathbb{R})$ such that there exists a $z_g \in C(\mathbb{R},\mathbb{R})$ with $g(t,\vartheta) = \vartheta^{(1)} z_g((t-\vartheta^{(3)})/\vartheta^{(2)}) + \vartheta^{(4)}$ for any $\vartheta = (\vartheta^{(1)}, \vartheta^{(2)}, \vartheta^{(3)}, \vartheta^{(4)})^{\mathrm{T}} \in \mathbb{R}_+^2 \times \mathbb{R}^2$.

For $\alpha, \beta > 0$ and $\gamma, \delta \in \mathbb{R}$ we define the homeomorphism $h_{\alpha,\beta,\gamma,\delta}$ by

$$h_{\alpha,\beta,\gamma,\delta}(\vartheta) := \left(\vartheta^{(1)} \cdot \alpha,\ \vartheta^{(2)} \cdot \beta,\ \vartheta^{(3)} + \vartheta^{(2)} \cdot \gamma,\ \vartheta^{(4)} + \vartheta^{(1)} \cdot \delta\right)^{\mathrm{T}}.$$

$h_{\alpha,\beta,\gamma,\delta}$ has the property that $g \in \mathbf{M}$ implies $g(\cdot, h_{\alpha,\beta,\gamma,\delta}(\cdot)) \in \mathbf{M}$. It is a "structural homeomorphism" preserving the structure of the shape-invariant model.

In real applications [compare Lawton, Sylvestre and Maggio (1972)], the SIM approach will often be motivated by finding a common qualitative structure characterizing the shape of individual curves, interindividual variation consisting only of scale and shift differences. This can be used when looking for conditions guaranteeing identifiability.

DEFINITION 2. For some $p \in \mathbb{N}$ and $q_1, \ldots, q_p \in \{2, 1, -2, -1\}$ a continuously differentiable function v: $J^* \to \mathbb{R}$ (for some interval $J^* \subseteq \mathbb{R}$) possesses a "$(q_1, \ldots, q_p)$-succession of characteristic points" iff there exist p points $x_1 < x_2 < \cdots < x_p$ so that for each $r = 1, \ldots, p$, x_r is the location of a (true) local maximum of v, local minimum of v, local maximum of v' or local minimum of v' as q_r is equal to 2, 1, -2 or -1. v has "exactly one $(q_1, \ldots, q_p)$-succession of characteristic points," if the points $x_1, \ldots, x_p$ with the above property are uniquely determined. Then $x_1, \ldots, x_p$ are called "locations of the characteristic points."

In terms of this definition a sample of curves possesses a common qualitative structure, if for some $p \in \mathbb{N}$ and some $q_1, \ldots, q_p \in \{2, 1, -2, -1\}$, *each* individual regression curve has exactly one $(q_1, \ldots, q_p)$-succession of characteristic points. Having found such a structure within a given sample of curves, it is natural to assume that this is an intrinsic property of the whole population, and that also z possesses such a $(q_1, \ldots, q_p)$-succession of characteristic points. For $p \geq 2$ these requirements already suffice to guarantee Definition 1(ii) and unidentifiability reduces to the structural homeomorphisms $h_{\alpha,\beta,\gamma,\delta}$. To ask for $p \geq 2$ avoids overparametrization, since otherwise the structure might be so simple that the individual parameter vectors cannot be uniquely determined (e.g., an exponential decay). One way of dealing with the structural homeomorphisms is to fix the values of locations and amplitudes of some characteristic points. On the other hand, the structure of SIM implies that each individual curve is a scale–shift transformation of the "basic curve" $z = f(\cdot, (1,1,0,0)^{\mathrm{T}})$. In order to interpret z and the individual parameters, it will sometimes be advantageous to assume that the basic curve quantitatively represents an "average curve" with average locations and amplitudes of the characteristic points representing the common pattern. This is equivalent to the normalizing condition $E\theta_1 = (1, 1, 0, 0)^{\mathrm{T}}$, with E denoting expectation.

These considerations motivate the following proposition providing sufficient conditions for identifiability.

PROPOSITION 1. *Assume that for some* $p \geq 2$ *and some* $q_1, \dots, q_p \in \{2, 1, -2, -1\}$, $f(\cdot, S(F)) \subset Z$, *where* Z, *respectively* $\bar{Z}$, *denote the collection of all functions* v *possessing exactly one, respectively no,* $(q_1, \dots, q_p)$*-succession of characteristic points. Under the assumption that* (4) *holds, the SEMOR MODEL is then identifiable in the following ways*:

(a) *Suppose, additionally, there can be determined constants* $c_1, c_2, c_3, c_4 \in \mathbb{R}$ *with* $c_1 < c_2$ *and* $c_3, c_4 > 0$ $(c_3 \neq c_4)$ *such that for some* $r_1, r_2 \in \{1, \dots, p\}$, *we can assume that* $z \in Z_{(c_1, c_2, c_3, c_4)}$. *Let* $Z_{(c_1, c_2, c_3, c_4)}$ *denote the set of all functions* $v \in Z$ *with the property that* $x_{r_1} = c_1$, $x_{r_2} = c_2$, $v(x_{r_1}) = c_3$ *and* $v(x_{r_2}) = c_4$, *with* $x_1, \dots, x_p$ *being the locations of the characteristic points of* v.

Then the SEMOR MODEL is completely identifiable by

$$M := \left\{ g \in \mathbf{M} | z_g \in Z_{(c_1, c_2, c_3, c_4)} \cup \bar{Z} \right\}$$

and $N \equiv id$.

(b) *Assuming* $E\theta_1 = (1, 1, 0, 0)^{\mathrm{T}}$ *and* $z \in Z$, *the SEMOR MODEL is completely identifiable by*

$$M := \left\{ g \in \mathbf{M} | z_g \in Z \cup \bar{Z} \right\}$$

and

$$N(g, F^*) := h_{e_1, e_2, e_3, e_4}, \quad \text{with } e_r := \int \theta^{(r)} \, dF^*(\theta), \qquad r = 1, 2, 3, 4,$$

$$\text{for all } g \in M \text{ and } F^* \in \mathscr{F}.$$

A proof is contained in the Appendix.

REMARK. Defining M in such a way that it also contains functions z_g possessing no $(q_1, \dots, q_p)$-succession of characteristic points has technical reasons: Estimating f and studying consistency (compare Sections 3 and 5) is simpler if M can easily be closed by defining bounds for z and z'. Note that, instead of restricting $\mathbf{M}$ to the set of all $g \in \mathbf{M}$ with $z_g \in Z_{(c_1, c_2, c_3, c_4)} \cup \bar{Z} \subset Z \cup \bar{Z}$ [instead of $z_g \in Z \cup \bar{Z}$ as in Proposition 1(b)], the normalizing condition used in Proposition 1(a) can also be formalized by defining a corresponding (nontrivial) normalizing operator N.

The next example is illustrative. It is not taken from a real application.

EXAMPLE 3. Let $J := [0,1]$, $\mathscr{R} = \mathbb{R}_+^2 \times J$ and assume

$$f(t,\theta) := \begin{cases} \theta^{(1)} z_1\left(t^{\theta^{(2)}}\right), & \text{for } 0 \le t \le \theta^{(3)}, \\ \theta^{(1)} z_1\left(t^{\theta^{(2)}}\right) z_2\left(\left(t-\theta^{(3)}\right)^{\theta^{(2)}}\right), & \text{for } \theta^{(3)} \le t \le 1, \end{cases}$$

for some unknown functions $z_1, z_2 \in C(J, [0,\infty[)$. Furthermore, suppose the following additional model assumptions:

(a) z_1 is continuously differentiable and strictly monotonically increasing.

(b) The ("switch-off") function z_2 is continuously differentiable and monotonically decreasing with $z_2(0) = 1$, $z_2(1) = 0$ and $|z_2'(0)| \ge C$, where C is a known constant with $C > 0$.

(c) There exists some $a^* \in]0,1[$ such that $a^* \le \theta^{(3)}$ for all $\theta \in S(F)$.

Let M denote the set of all functions g that can be modeled in this way, when replacing f by g and z_1, z_2 by some continuously differentiable functions $z_{g,1}, z_{g,2}$ with $z_{g,1}$ being monotonically increasing and $z_{g,2}$ satisfying (b). Again there are structural homeomorphisms: If for $\alpha, \beta > 0$ we set $h_{\alpha,\beta}(\vartheta) := (\alpha \cdot \vartheta^{(1)}, \beta \cdot \vartheta^{(2)}, \vartheta^{(3)})^{\mathrm{T}}$, then obviously $g \in M$ implies $g(\cdot, h_{\alpha,\beta}(\cdot)) \in M$. To resolve this unidentifiability, we might, e.g., use the normalizing conditions $E\theta_1^{(1)} = 1$ and $E\theta_1^{(2)} = 1$.

Under these assumptions the SEMOR MODEL is completely identifiable by M and

$$N(g, F^*) := h_{e_1,e_2}, \quad \text{with} \quad (e_1, e_2)^{\mathrm{T}} := \int \left(\theta^{(1)}, \theta^{(2)}\right)^{\mathrm{T}} dF^*(\theta),$$

$$\text{for all } g \in M \text{ and } F^* \in \mathscr{F}.$$

The proof is omitted.

The above examples might illustrate the scope of additional model assumptions leading to identifiability according to Definition 1. A general approach covering a large number of possible applications can be described as follows: As a first step one should try to determine (via a priori knowledge) for some subset $\mathscr{R}$ of a Euclidean space, some $k \ge 0$, $J_1 \subset \mathbb{R}^{l_1}, \dots, J_k \subset \mathbb{R}^{l_k}$, $l_1, \dots, l_k \in \mathbb{N}$, and some $L \subset \mathbb{R}^l$, $l \ge 0$, an operator G: $L \times C(J_1, \mathbb{R}) \times \cdots \times C(J_k, \mathbb{R}) \times J \times \mathscr{R} \to \mathbb{R}$ so that the corresponding SEMOR MODEL holds with

$$(5) \qquad f(t,\theta) := G_{\lambda, z_1, \dots, z_k}(t,\theta), \qquad \theta \in \mathscr{R},\ t \in J,$$

where $z_1 \in C(J_1, \mathbb{R}), \dots, z_k \in C(J_k, \mathbb{R})$ are unknown fixed functions and $\lambda \in L$ is an unknown fixed parameter vector [for notational convenience we assume that for $k = 0$ (5) reduces to Example 1, whereas for $l = 0$ there is no population parameter]. Obviously, this covers all our examples, and includes also the two-component SIM introduced by Stuetzle et al. (1980) for modeling human height growth. Sometimes identification is immediate from the structure of G (e.g., Day's model). Otherwise, one might proceed as follows:

Restrict the class of functions to be considered by using a priori information and/or common structural characteristics of the sample of all estimated individual curves (Examples 2 and 3). The resulting subsets $Z_1 \subset C(J_1, \mathbb{R}), \ldots, Z_k \subset C(J_k, \mathbb{R})$ should guarantee that with

$$M := \{g \in C(J \times \mathscr{R}, \mathbb{R}) | g(\cdot, \vartheta) = G_{\lambda_g, z_{g,1}, \ldots, z_{g,k}}(\cdot, \vartheta) \tag{6}$$
$$\text{for some } \lambda_g \in L,\ z_{g,1} \in Z_1, \ldots, z_{g,k} \in Z_k\},$$

possibly by using additional regularity conditions, either

(a) the SEMOR MODEL is completely identifiable by M and $N \equiv \mathrm{id}$ [Examples 1 and 2 under Proposition 1(a)]; or

(b) the remaining unidentifiability is due to structural homeomorphisms (Examples 2 and 3) and can be resolved by defining appropriate normalizing conditions determining a nontrivial N [Examples 3 and 2 under Proposition 1(b)].

When the model function f has been properly identified this does not necessarily hold for the different functional components [e.g., $f(t, \theta) = \theta z_1(t) z_2(t)$]. For $p \in \{1, \ldots, k\}$ identifiability of z_p on some set $I_p \subset J_p$ additionally requires that for any $(\lambda^*, z_1^*, \ldots, z_k^*) \in L \times Z_1, \ldots, Z_k$ with

$$G_{\lambda^*, z_1^*, \ldots, z_k^*}(t, \theta) = f(t, \theta), \quad \text{for all } (t, \theta) \in J \times S(F),$$

it holds that

$$z_p^*(x) = z_p(x), \quad \text{for all } x \in I_p.$$

Under the conditions of Proposition 1, z is identifiable on the interval $I_{\mathrm{SIM}} := [\inf_{\theta \in S(F)} (a - \theta^{(3)})/\theta^{(2)}, \sup_{\theta \in S(F)} (b - \theta^{(3)})/\theta^{(2)}]$. Considering Example 3, let us assume that there exist a $\bar{\theta} \in S(F)$ with $\bar{\theta}^{(3)} = 1$ and a $\check{\theta} \in S(F)$ with $f(t, \check{\theta}) = 0$ for some $t > 0$. It is easy to see that then z_1 and z_2 are identifiable on their whole domain J.

3. The SEMOR approach. In the following we will again assume model (1) and the SEMOR MODEL. Having identified the SEMOR MODEL, a natural way to obtain estimates of f and $\theta_1, \ldots, \theta_n$ seems to be the following approach: Solve the least-squares problem

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}\left(t_{ij}, \tilde{\theta}_i\right)\right)^2 = \min_{g \in M} \min_{\vartheta_1, \ldots, \vartheta_n \in \mathscr{R}} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - g(t_{ij}, \vartheta_i)\right)^2,$$

and then normalize the solutions obtained, using the operator N.

However, M will usually be infinite dimensional. In this case, from a computational, as well as from a statistical point of view, no reasonable estimates can be obtained in the above way. With respect to $\mathscr{R}$ the problem arises that the *existence* of parameter vectors solving a nonlinear least-squares problem can only be guaranteed when minimizing within a compact set. Hence, M and $\mathscr{R}$ have to be replaced by some appropriately restricted subsets.

For defining a reasonable estimation method, we thus need some additional conditions. It will be assumed that we can find an $M \subset C(J \times \mathscr{R}, \mathbb{R})$, an operator N: $M \times \mathscr{F} \to H(\mathscr{R})$ and some $D \subseteq \mathscr{R}$ such that Assumption 2 holds.

ASSUMPTION 2. (a) The SEMOR MODEL is completely identifiable by M and N.

(b) D is compact and $S(F) \subseteq D$.

(c) $N|_{M \times \mathscr{F}_D}$ is measurable, where $\mathscr{F}_D$ denotes the set of all $F^* \in \mathscr{F}$ with $S(F^*) \subseteq D$.

REMARK. Within the rest of the paper (including the Appendix) spaces of continuous functions will be endowed with the topology of compact convergence, whereas $\mathscr{F}_D$ will be considered with respect to the weak topology, induced by convergence in distribution. Measurability of N on $M \times \mathscr{F}_D$ refers to the corresponding Borel sets.

A reasonable way of restricting the function space M is indicated by Proposition 2.

PROPOSITION 2. *Let $\mathscr{M} \subset M$ be a set of functions uniformly bounded and equicontinuous on $J \times D$. Then for each $\varepsilon > 0$ there exists a compact set $\mathscr{S} \subset \mathbb{R}^r$ (for some $r \in \mathbb{N}$) and a continuous mapping a: $\mathscr{S} \to \mathscr{M}$ such that*

$$\min_{a_s \in A} \max_{x \in J \times D} |a_s(x) - g(x)| < \varepsilon, \quad \text{for any } g \in \mathscr{M},$$

where $A := \{a_s | s \in \mathscr{S}\}$.

The proof is straightforward.

We will call a set $A \subset M$ being defined by $A := \{a_s | s \in \mathscr{S}\}$ for some $\mathscr{S} \in \mathbb{R}^r$, $r \in \mathbb{N}$, and some continuous mapping a: $\mathscr{S} \to M$ a "parametric subspace of M" [e.g., any set of polynomials of a given order is a parametric subspace of $C(\mathbb{R}, \mathbb{R})$]. Equicontinuity of a set of functions roughly means that all functions have a similar "degree of smoothness." Thus Proposition 2 says that if there exists some knowledge concerning bounds and the degree of smoothness of f, one can always find a parametric subspace of M containing functions arbitrarily close to f on the "relevant" domain $J \times D$ (note that the proposition only guarantees existence; the problem of selecting an appropriate A will be considered later). The procedure of restricting an infinite-dimensional class of functions to a parametric subspace has analogies in nonparametric regression: As mentioned by Geman and Hwang (1982) most well-known nonparametric estimators of a regression function can be interpreted as least-squares estimators with respect to some parametric family of functions.

These considerations lead to the idea of determining estimates of f and $\theta_1, \ldots, \theta_n$ as outlined above by replacing there $\mathscr{R}$ by D and M by some compact parametric subspace A suitable for approximating f (compactness of A is required in order to guarantee existence of solutions of the least-squares problem).

DEFINITION 3. A method for the simultaneous estimation of f and $\theta_1, \ldots, \theta_n$ is called "self-modeling nonlinear regression" (SEMOR) if for given $M \subset C(J \times \mathscr{R}, \mathbb{R})$, $D \subset \mathscr{R}$ and N: $M \times \mathscr{F} \to H(\mathscr{R})$ satisfying Assumption 2, estimates $\hat{f}$ and $\hat{\theta}_1, \dots, \hat{\theta}_n$ are obtained as follows:

(i) Determine least-squares estimates $\tilde{f}$ and $\tilde{\theta}_1, \dots, \tilde{\theta}_n$ by solving

$$\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{j=1}^{T_i}\left(Y_{ij} - \tilde{f}\left(t_{ij}, \tilde{\theta}_i\right)\right)^2 \tag{7}$$
$$= \min_{a_s \in A}\ \min_{\vartheta_1, \dots, \vartheta_n \in D} \frac{1}{n}\sum_{i=1}^{n}\frac{1}{T_i}\sum_{j=1}^{T_i}\left(Y_{ij} - a_s\left(t_{ij}, \vartheta_i\right)\right)^2,$$

where $A \subset M$ is a prespecified compact parametric subspace of M (A is called "approximation space for M").

(ii) Normalize these estimates. Set

$$\hat{f} := \tilde{f}(\cdot, N(\tilde{f}, \tilde{F})(\cdot))$$

and

$$\hat{\theta}_i := N(\tilde{f}, \tilde{F})^{-1}(\tilde{\theta}_i), \quad \text{for all } i = 1, \dots, n,$$

with $\tilde{F}$ denoting the empirical distribution function of $\tilde{\theta}_1, \dots, \tilde{\theta}_n$. $\hat{f}$ and $\hat{\theta}_1, \dots, \hat{\theta}_n$ are called (D, A, N)-SEMOR estimates of f and $\theta_1, \dots, \theta_n$.

Obviously, (7) is a nonlinear least-squares problem with $n \cdot d + r$ parameters, where d denotes the number of components of the individual parameter vectors. Thus Lemma 2 of Jennrich (1969) together with the measurability of N guarantees the existence of measurable SEMOR estimators.

REMARK. (a) If the SEMOR MODEL is identifiable by M (and thus $N \equiv \text{id}$), step (ii) of the SEMOR procedure is obviously superfluous. In this case $\hat{f} = \tilde{f}$ and $\hat{\theta}_i = \tilde{\theta}_i$.

(b) The regularity condition that there can be found a compact subset $D \subset \mathscr{R}$ containing the true parameter vectors is required in ordinary nonlinear regression, too [compare Jennrich (1969) or Wu (1981)]. It is thus not peculiar to SEMOR. Finding appropriate bounds in order to determine such a D will generally be no problem, since these bounds might be chosen arbitrarily large (respectively, small).

It is easy to see that for a parametric model with a population parameter any compact subset of M (containing λ) can be used as approximation space. If M is defined by (6) for a continuous operator G and some sets $L, Z_1, \dots, Z_k$, we might define an approximation space A by replacing in (6) L by some compact subspace $L^* \subset L$ and Z_p, $p = 1, \dots, k$, by some compact approximation spaces A_p^* of Z_p (e.g., spaces of spline functions). Then together with an estimate $\hat{f} := G_{\lambda_{\hat{f}}, z_{\hat{f},1}, \dots, z_{\hat{f},k}}$ of f, SEMOR also yields estimates $z_{\hat{f},p}$ of the functional components z_p, $p = 1, \dots, k$.

EXAMPLE: SIM. Assume the conditions of Proposition 1(a).

As a first step one might determine constants $\alpha, \alpha_1, \alpha_2, \alpha_3, \alpha_4$ and $\beta_1, \beta_2, \beta_3, \beta_4$ such that for all $\theta \in S(F) \sup_{x \in J}|f(x, \theta)| \le \alpha$, $\alpha_1 \le f(x_{r_1}, \theta) \le \alpha_2$, $\alpha_3 \le$

$|f(x_{r_1}, \theta) - f(x_{r_2}, \theta)| \leq \alpha_4$ and $\beta_1 \leq x_{r_1} \leq \beta_2$, $\beta_3 \leq x_{r_2} - x_{r_1} \leq \beta_4$ [x_{r_1}, x_{r_2}: r_1th, respectively r_2th, characteristic point of $f(\cdot, \theta)$, as specified by the normalizing conditions]. Rough guesses of these constants might be obtained by analyzing the observations (this might be done visually or by nonparametric methods).

Given these constants, with $\bar{\alpha}_3 := \alpha_3/|c_3 - c_4|$, $\bar{\alpha}_4 := \alpha_4/|c_3 - c_4|$, $\bar{\beta}_3 := \beta_3/(c_2 - c_1)$ and $\bar{\beta}_4 := \beta_4/(c_2 - c_1)$, the compact subset

$$D := \{\theta \in \mathscr{R} | \bar{\alpha}_3 \leq \theta^{(1)} \leq \bar{\alpha}_4, \bar{\beta}_3 \leq \theta^{(2)} \leq \bar{\beta}_4, \beta_1 - \theta^{(2)}c_1 \leq \theta^{(3)} \leq \beta_2 - \theta^{(2)}c_1, \\ \alpha_1 - \theta^{(1)}c_3 \leq \theta^{(4)} \leq \alpha_2 - \theta^{(1)}c_3\}$$

of $\mathscr{R} = \mathbb{R}_+^2 \times \mathbb{R}^2$ satisfies Assumption 2(b).

Together with J, D characterizes the "relevant" domain $J^* = [a^*, b^*] := [\inf_{\vartheta \in D}(a - \vartheta^{(3)})/\vartheta^{(2)}, \sup_{\vartheta \in D}(b - \vartheta^{(3)})/\vartheta^{(2)}]$ of z. In order to approximate z on J^*, one might then determine a set A^* of cubic spline functions based on a corresponding knot sequence on J^* [see de Boor (1978)].

This leads to defining A as the set of all functions $g \in M$ with the property that there exists a $z_g^* \in A^*$ with $\sup_{t \in J^*}|z_g^*(t)| \leq \bar{\alpha}_4 \alpha$ such that

$$z_g(t) = \begin{cases} z_g^*(a^*) + z_g^{*\prime}(a^*)(t - a^*), & \text{for } t \leq a^*, \\ z_g^*(t), & \text{for } a^* \leq t \leq b^*, \\ z_g^*(b^*) + z_g^{*\prime}(b^*)(t - b^*), & \text{for } t \geq b^*. \end{cases}$$

It can easily be seen that A is a compact parametric subspace of M.

Under the conditions of Proposition 1(b), appropriate sets D and A can be defined analogously. However, to determine D rough guesses of average amplitudes and locations of some characteristic points are additionally required.

It should be noted that the actual data analytic and computational steps involved are simpler than the formalism might imply. Only reasonable guesses of $\beta_1, \beta_2, \beta_3, \beta_4$ are of interest (in order to determine J^*), $\alpha, \alpha_1, \alpha_2, \alpha_3, \alpha_4$ are merely required for technical reasons. They might be chosen extremely large (respectively small) such that the above conditions are "obviously" satisfied.

REMARK. As mentioned above, (7) is a nonlinear least-squares problem with a usually large number of $n \cdot d + r$ parameters. One might ask, whether so many parameters can be estimated without running into the kind of problems that typically occur, when the proportion of the number of observations to the number of parameters is too small. Let $T := T_1 = T_2 = \cdots = T_n$, so that the number of observations available is $n \cdot T$. A prerequisite for a conventional nonliner regression analysis is that T/d is "large enough." When this is fulfilled, the relative increase by the r parameters characterizing the approximation space will generally be modest, since $r \ll n \cdot d < n \cdot T$. For example, assume a SIM, and let $n = 25$ and $T := T_1, \ldots, T_n = 25$. Then the number of observations per parameter is 6.25 for a fixed model and 5.21 for SEMOR, when using as much as 24 knots for a cubic spline approximation.

4. The computation of SEMOR estimates. Following Definition 3, as the first step to obtain SEMOR estimates $\hat{f}$ and $\hat{\theta}_1, \ldots, \hat{\theta}_n$, one has to solve (7) being, as mentioned above, a nonlinear least-squares problem with $n \cdot d + r$ parameters. Thus one might doubt the practical computability of SEMOR estimates for large n, since then—from numerical reasons—it will be extremely difficult to solve a nonlinear least-squares problem with so many parameters *directly*. However, there exists an iterative procedure decomposing this overall minimization problem into a sequence of low-dimensional least-squares problems with d or r parameters, respectively. This algorithm is motivated by the following argument, which is due to Lawton, Sylvestre and Maggio (1972). Obviously, if $\tilde{f}$ were known the parameter vectors $\tilde{\theta}_1, \dots, \tilde{\theta}_n$ could be determined by solving n independent least-squares problems with d parameters, respectively (which is much simpler than solving an $n \cdot d$ parameter least-squares problem). On the other hand, if we knew $\tilde{\theta}_1, \dots, \tilde{\theta}_n$ we could determine $\tilde{f}$ by just solving a least-squares problem with r parameters. This leads to the obvious idea that one might solve (7) by using the following iterative scheme.

ALGORITHM.

START: *Select a first guess $\tilde{f}^0$ of f. Then determine for each $i = 1, \dots, n$, an initial approximation $\tilde{\theta}_i^0$ of $\tilde{\theta}_i$ by solving*

$$\frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^0\left(t_{ij}, \tilde{\theta}_i^0\right)\right)^2 = \min_{\vartheta \in D} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^0\left(t_{ij}, \vartheta\right)\right)^2.$$

$(h+1)$ST ITERATION STEP ($h \in \mathbb{N}_0 := \mathbb{N} \cup \{0\}$): *Compute an improved approximation $\tilde{f}^{h+1}$ of $\tilde{f}$ by solving*

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^{h+1}\left(t_{ij}, \tilde{\theta}_i^h\right)\right)^2 = \min_{a_s \in A} \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - a_s\left(t_{ij}, \tilde{\theta}_i^h\right)\right)^2,$$

where $\tilde{\theta}_i^h, \dots, \tilde{\theta}_n^h$ denote the parameter vectors obtained in the previous iteration. For each $i = 1, \dots, n$ determine an $(h+1)$st approximation $\tilde{\theta}_i^{h+1}$ of $\tilde{\theta}_i$ by solving

$$\frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^{h+1}\left(t_{ij}, \tilde{\theta}_i^{h+1}\right)\right)^2 = \min_{\vartheta \in D} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^{h+1}\left(t_{ij}, \vartheta\right)\right)^2.$$

STOPPING CONDITION: *Stop iteration if for some prespecified $\delta > 0$,*

$$\frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^h\left(t_{ij}, \tilde{\theta}_i^h\right)\right)^2 - \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - \tilde{f}^{h+1}\left(t_{ij}, \tilde{\theta}_i^{h+1}\right)\right)^2 < \delta.$$

We now have to investigate convergence properties of this algorithm. To simplify notation we hereby use the following abbreviations:

(a) $\vec{\theta}^h := (\tilde{\theta}_1^h, \dots, \tilde{\theta}_n^h)$ for $h \in \mathbb{N}_0$.

(b) For $g \in A$ and $\vec{\vartheta} = (\vartheta_1, \dots, \vartheta_n) \in D^n$,

$$q(g, \vec{\vartheta}) := \frac{1}{n} \sum_{i=1}^{n} \frac{1}{T_i} \sum_{j=1}^{T_i} \left(Y_{ij} - g\left(t_{ij}, \vartheta_i\right)\right)^2.$$

We then obtain

PROPOSITION 3. (i) *There exists a* $c \geq 0$ *so that* $q(\tilde{f}^h, \vec{\theta}^h) \to c$ *as* $h \to \infty$.
(ii) *For each limit point* $(\tilde{f}', \vec{\theta}') \in A \times D^n$ *of the sequence* $(\tilde{f}^h, \vec{\theta}^h)_{h \in \mathbb{N}}$,

$$c = q(\tilde{f}', \vec{\theta}') = \min_{\vec{\vartheta} \in D^n} q(\tilde{f}', \vec{\vartheta}) = \min_{a_s \in A} q(a_s, \vec{\theta}'). \tag{8}$$

A proof of this proposition is contained in the Appendix. Proposition 3 implies the following corollary.

COROLLARY 1. *If for any* $(\tilde{f}', \vec{\theta}') \in A \times D^n$ *satisfying*

$$q(\tilde{f}', \vec{\theta}') = \min_{\vec{\vartheta} \in D^n} q(\tilde{f}', \vec{\vartheta}) = \min_{a_s \in A} q(a_s, \vec{\theta}') > \min_{a_s \in A} \min_{\vec{\vartheta} \in D^n} q(a_s, \vec{\vartheta}),$$

we also have

$$q(\tilde{f}', \vec{\vartheta}') > q(\tilde{f}^0, \vec{\theta}^0),$$

then any limit point $(\tilde{f}, \tilde{\theta}_1, \ldots, \tilde{\theta}_n)$ *of the sequence* $(\tilde{f}^h, \tilde{\theta}_1^h, \ldots, \tilde{\theta}_n^h)$ *is a solution of* (7).

Proposition 3 motivates the stopping rule used above: We stop iteration if the iterates $(\tilde{f}^h, \vec{\theta}^h)$ are "close enough" to some limit point $(\tilde{f}', \vec{\theta}')$ [$\equiv (a_{s'}, \vec{\theta}')$ for some $s' \in \mathscr{S}$] satisfying (8). Moreover, Proposition 3(ii) implies that such a limit point is a stationary point of q if s' is an inner point of $\mathscr{S}$ and if q is differentiable (with respect to s and $\vec{\theta}$) at $(a_s, \vec{\theta}) = (a_{s'}, \vec{\theta}')$. Thus we can infer from Proposition 3 and Corollary 1 that the above iteration can be considered as an algorithm for solving (7). It might end up in another stationary point of q (this might happen to any other algorithm for solving this nonlinear least-squares problem, too), but it follows from Corollary 1 that if the initial guess $(\tilde{f}^0, \vec{\theta}^0)$ is "close" to a solution $(\tilde{f}, \vec{\theta}) = (\tilde{f}, \tilde{\theta}_1, \ldots, \tilde{\theta}_n)$ of (7), then the iteration solves the minimization problem (7).

5. Consistency. Sufficient conditions for strong consistency of SEMOR estimators will be established as $\min\{n, T_1, \ldots, T_n\} \to \infty$.

Some (weak) assumptions regarding design and error term are required.

ASSUMPTION 3. Let $(Y_{ij})_{i, j \in \mathbb{N}}$, $(\varepsilon_{ij})_{i, j \in \mathbb{N}}$ and $(\theta_i)_{i \in \mathbb{N}}$ be sequences of random variables with the following property: There is a sequence $(t_{ij})_{i, j \in \mathbb{N}}$ of design points such that for all $n, T_1, \ldots, T_n \in \mathbb{N}$, model (1) and the SEMOR MODEL hold for a.e. sequence of realizations of Y_{ij}, ε_{ij} and θ_i, $i = 1, \ldots, n$, $j = 1, \ldots, T_i$. Furthermore, let the following hold:

(i) The random variables ε_{ij} are independent with expectation 0, finite variances σ_{ij}^2 and finite fourth moments α_{ij}^4 satisfying: (a) For each $i \in \mathbb{N}$ there exists a $c_i > 0$ so that $\sigma_{ij}^2 \leq c_i$ and $\alpha_{ij}^4 \leq c_i$ for all $j \in \mathbb{N}$. (b) There exists a constant $c < \infty$, so that $(1/n)\sum_{i=1}^n c_i \leq c$ for all $n \in \mathbb{N}$.

(ii) For each subinterval $\mathscr{J} \subset J$ there exists a $\tilde{T} \in \mathbb{N}$ and an $\varepsilon > 0$ so that

$$\frac{1}{T} \sum_{j=1}^{T} \chi_{\mathscr{J}}(t_{ij}) > \varepsilon, \quad \text{for all } T \geq \tilde{T} \text{ and every } i \in \mathbb{N},$$

where $\chi_{\mathscr{J}}$ denotes the indicator function of $\mathscr{J}$.

REMARK. Assumption 3(ii) holds for most of the reasonable designs arising in practice. Given a systematic design with $t_{ij} = t_j$ (for each $i \in \mathbb{N}$), it is merely required that the relative number of all t_j falling into some specified subinterval of J does not converge to 0. Also the case of missing observations is included. It has then to be assumed that, as $\inf_{i \in \mathbb{N}} T_i$ increases, the relative number of missings per individual tends to zero with some uniform rate. The case of controlled "jittered" design is covered as well.

To establish consistency results for SEMOR, one inevitably has to allow the approximation space to "grow" with sample size, i.e., one has to construct an appropriate sieve of approximation spaces [compare Grenander (1981)]. Using that, due to Proposition 2, (on $J \times D$) any compact subset of M can be approximated up to an arbitrarily small error by parametric subspaces, we get the following theorem.

THEOREM. *Let Assumption 3 hold, and suppose that there are some given $D \subset \mathscr{R}$, $M \subset C(J \times \mathscr{R}, \mathbb{R})$ and N: $M \times \mathscr{F} \to H(\mathscr{R})$ satisfying Assumption 2. Assume that there can be determined a compact subset $\mathscr{M}$ of M containing at least one function $g \in \mathscr{M}$ with $g = f$ [on $J \times S(F)$]. Furthermore, suppose that, when restricted to $\mathscr{M} \times \mathscr{F}_D$, $N(\cdot,\cdot)$ and $N(\cdot,\cdot)^{-1}$ are continuous at each $(g, F^*) \in \mathscr{M} \times \mathscr{F}_D$ with $g \models f$ and $F^* \cong_g F$.*

Let $\{A_r\}$, $r \in \mathbb{N}$, denote a family of function spaces with the following properties:

(1) *For each $r \in \mathbb{N}$, A_r is a compact parametric subspace of $\mathscr{M} \subset M$.*
(2) *For each $\varepsilon > 0$ there exists a $r_\varepsilon \in \mathbb{N}$ so that for all $r \geq r_\varepsilon$ and every $g \in \mathscr{M}$,*

$$\inf_{a \in A_r} \sup_{x \in J \times D} |g(x) - a(x)| < \varepsilon.$$

Moreover, assume mappings $m \to n_m$ and $(i, m) \to T_{i,m}$, $m, n_m, i, T_{i,m} \in \mathbb{N}$, so that, as $m \to \infty$, $n_m \to \infty$ and $\inf_{i \in \mathbb{N}} T_{i,m} \to \infty$.

Furthermore, let for $r, m \in \mathbb{N}$, $\hat{f}_{r,m}$ and $\hat{\theta}_{1,r,m}, \ldots, \hat{\theta}_{n_m,r,m}$ denote (D, A_r, N)-SEMOR estimators being determined with respect to the random variables Y_{ij}, $i = 1, \ldots, n_m$, $j = 1, \ldots, T_{i,m}$.

Finally, let S be some compact subset of $\mathscr{R}$ with $S(F) \subseteq S$ such that Definition 1(iii) still holds when replacing $S(F)$ by S in (3). Under these conditions, we obtain for each sequence $(r(m))_{m \in \mathbb{N}}$ with $r(m) \to \infty$ as $m \to \infty$:

(a) *With probability 1,*

$$\lim_{m \to \infty} \hat{f}_{r(m),m}(t, \theta) = f(t, \theta), \quad \textit{uniformly for all } (t, \theta) \in J \times S.$$

(b) *For each $i \in \mathbb{N}$ with probability 1,*

$$\lim_{m \to \infty} \hat{\theta}_{i, r(m), m} = \theta_i.$$

(c) *With probability* 1 *for every* $\varepsilon > 0$,

$$\frac{1}{n_m} \# \left\{ i \in \{1, \ldots, n_m\} \mid \|\theta_i - \hat{\theta}_{i, r(m), m}\|_2 \geq \varepsilon \right\} \to 0, \quad as\ m \to \infty.$$

REMARKS. (a) The assumption of the theorem that one should determine a compact subset $\mathscr{M}$ of M containing some g with $g = f$ [on $J \times S(F)$], imposes some restriction. In view of possible applications, it is a rather weak one. By Ascoli's theorem a subset $\mathscr{M} \subset M$ is compact if it is closed, equicontinuous and if for each $x \in J \times \mathscr{R}$, $\{g(x) | g \in \mathscr{M}\}$ is bounded. Usually it will suffice to have a rough idea about appropriate bounds and about the "degree" of smoothness of the model function. This will generally not be too difficult, since these bounds might be chosen arbitrarily large, whereas the "degree" of smoothness assumed might be arbitrarily small.

(b) Assume that f can be modeled by (5) for some *continuous* operator G, and that for some $p \in \{1, \ldots, k\}$ the pth functional component z_p is identifiable on a compact set $I_p \subset J_p$. If $\mathscr{M}$ is defined by (6) for some compact sets $\mathscr{L} \subset L$, $\mathscr{Z}_1 \subset Z_1, \ldots, \mathscr{Z}_k \subset Z_k$, under the conditions of the theorem we then obtain that a.s. $z_{\hat{f}_{r(m), m}, p} \to z_p$ uniformly on I_p. This is an immediate consequence of assertion (a), since identifiability implies that for any $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $(\lambda^*, z_1^*, \ldots, z_k^*) \in \mathscr{L} \times \mathscr{Z}_1, \ldots, \mathscr{Z}_k$ with $\sup_{x \in I_p} \|z_p^*(x) - z_p(x)\|_2 \geq \varepsilon$, $\sup_{(t, \theta) \in S(F)} \|G_{\lambda^*, z_1^*, \ldots, z_k^*}(t, \theta) - f(t, \theta)\|_2 \geq \delta$.

(c) For a given model let S be the largest set such that identifiability of f on $J \times S(F)$ implies identifiability of f on $J \times S$. For example, for a parametric model with a population parameter one obtains $S = \mathscr{R}$. According to the theorem, SEMOR then consistently estimates f on $J \times S$.
Within many models S will depend on $S(F)$. Given a finite number of observations, the problem then arises how to interpret a SEMOR estimate $\hat{f}$, if $S(F)$ is completely unknown. In any case, in view of assertion (c) of the theorem it is reasonable to look for a set $\overline{D}$, where the sequence $\{\hat{\theta}_i\}_{i=1, \ldots, n}$ of the parameter estimates clusters, and to study $\hat{f}$ on $J \times \overline{D}$. Within most applications, however, it will be possible to deal with this problem by making use of the particular model structure. As an example assume a SIM. Interpreting $\hat{f}$ will generally be based on analyzing $z_{\hat{f}}$, which under the above conditions is a consistent estimator of z on I_{SIM}. Since I_{SIM} depends on $S(F)$, one might study $z_{\hat{f}}$ within an interval I^*, which can be expected to satisfy $I^* \subset I_{\mathrm{SIM}}$. In particular, one might use

$$I^* := \left[\left(a - \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^{(3)} \right) \Big/ \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^{(2)}, \left(b - \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^{(3)} \right) \Big/ \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_i^{(2)} \right].$$

6. Conclusions. A first step in a conventional nonlinear regression analysis consists in choosing a parametric model, which seems appropriate in view of the visual structure of the data and which incorporates a priori knowledge. After estimating parameters, the adequacy of the model is checked by applying goodness-of-fit tests. If the fit is poor, the model function is modified and the same strategy is pursued. This "controlled guessing" depends on the skill and the patience of the statistician and does not constitute a generally applicable and reproducible method.

Self-modeling nonlinear regression is a general concept to determine both an adequate model and the parameters from the data, based on qualitative, structural information only. Further basic conditions are that the parameter space $\mathscr{R}$ (and, in particular, its dimension) can be fixed, and that the model is identifiable, possibly after applying some normalizing operation. The examples given illustrate that this is usually much easier than specifying a priori the quantitative functional dependence on parameters and design. Spaces of spline functions will often be a natural choice when defining the approximation spaces needed for estimating the model. A crucial assumption for SEMOR is the availability of replicate data. In many fields, this is common. In other fields, the new possibilities offered by SEMOR might be an incentive to plan experiments for more than one experimental unit. This should be done at least once, in order to gauge the model within some framework.

In this paper we have presented the concepts of SEMOR and we have given some theoretical justification. After this first step, data-analytic and algorithmic problems need to be discussed at more length (our experience in this respect was so far quite encouraging). Further theoretical problems also need clarification.

APPENDIX

PROOF OF PROPOSITION 1. Let g denote a function with $g \in \mathbf{M}$ and $g \models f$ such that the corresponding function $z_g \in C^1(\mathbb{R}, \mathbb{R})$ is in $Z \cup \bar{Z}$. Then for each $\theta \in S(F)$ there exists a $\vartheta_{g,\theta} \in \mathscr{R}$ so that

$$\theta^{(1)} z\left(\frac{t-\theta^{(3)}}{\theta^{(2)}}\right) + \theta^{(4)} = f(t,\theta) = g(t, \vartheta_{g,\theta}) = \vartheta_{g,\theta}^{(1)} z_g\left(\frac{t-\vartheta_{g,\theta}^{(3)}}{\vartheta_{g,\theta}^{(2)}}\right) + \vartheta_{g,\theta}^{(4)},$$

for all $t \in \mathscr{J}$ $[\theta = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \theta^{(4)})^{\mathrm{T}};\ \vartheta_{g,\theta} = (\vartheta_{g,\theta}^{(1)}, \vartheta_{g,\theta}^{(2)}, \vartheta_{g,\theta}^{(3)}, \vartheta_{g,\theta}^{(4)})^{\mathrm{T}}]$. By assumption, for each $\theta \in S(F)$ the resulting regression curve $f(\cdot, \theta)$ possesses exactly one $(q_1, \dots, q_p)$-succession of characteristic points located at some points $x_1^\theta < \cdots < x_p^\theta$. Thus, for each $r = 1, \dots, p$, depending on the value of q_r, the functions $z((t-\theta^{(3)})/\theta^{(2)})$ and $z_g((t-\vartheta_{g,\theta}^{(3)})/\vartheta_{g,\theta}^{(2)})$, respectively their derivatives, have a (true) local extremum at $t = x_r^\theta$. Consequently, $z, z_g \in Z$, and for each $\theta \in S(F)$ it holds for all $r = 1, \dots, p$,

$$x_r^\theta = \theta^{(2)} x_r + \theta^{(3)} = \vartheta_{g,\theta}^{(2)} x_{g,r} + \vartheta_{g,\theta}^{(3)} \tag{9}$$

and

$$f(x_r^\theta, \theta) = \theta^{(1)} z(x_r) + \theta^{(4)} = \vartheta_{g,\theta}^{(1)} z_g(x_{g,r}) + \vartheta_{g,\theta}^{(4)} = g(x_{g,r}, \vartheta_{g,\theta}), \tag{10}$$

with $x_1, \dots, x_p$ and $x_{g,1}, \dots, x_{g,p}$ denoting the locations of the characteristic points of z and z_g.

To prove assertions (a) and (b) we now have to show that Definition 1(i)–(iii) are satisfied with respect to the definitions of M and N within (a) and (b):

(a) Clearly, (i) holds. To show (ii) and (iii), we only have to consider the case $z, z_g \in Z_{(c_1, c_2, c_3, c_4)} \subset Z$. Since then $x_{r_1} = c_1 = x_{g,r_1}$, $x_{r_2} = c_2 = x_{g,r_2}$, $z(x_{r_1}) = c_3 = z_g(x_{g,r_1})$ *and* $z(x_{r_2}) = c_4 = z_g(x_{g,r_2})$, (9) *and* (10) *hold if and only if*

$\vartheta_{g,\theta} = (\vartheta_{g,\theta}^{(1)}, \vartheta_{g,\theta}^{(2)}, \vartheta_{g,\theta}^{(3)}, \vartheta_{g,\theta}^{(4)})^{\mathrm{T}} = (\theta^{(1)}, \theta^{(2)}, \theta^{(3)}, \theta^{(4)})^{\mathrm{T}} = \theta$ for each $\theta \in S(F)$. Hence, (ii) and (iii) are fulfilled.

(b) Obviously, (i) holds. As has been shown above, for each $g \in M$ with $g \models f$, the corresponding function z_g is in Z and satisfies (9) and (10). By assumption, $p \geq 2$, $x_1 < x_2$ and $x_{g,1} < x_{g,2}$. Moreover, without restriction, we can assume that $z(x_1) \neq z(x_2)$ and $z_g(x_{g,1}) \neq z_g(x_{g,2})$. [Otherwise, since neither z nor z_g can be constant within $]x_1, x_2[$, respectively $]x_{g,1}, x_{g,2}[$, there has to exist an additional local maximum or minimum of z, respectively z_g, at a point $\bar{x} \in]x_1, x_2[$, respectively $\bar{x}_g \in]x_{g,1}, x_{g,2}[$. Then, obviously, $z(x_1) \neq z(\bar{x})$ and $z_g(x_{g,1}) \neq z_g(\bar{x}_g)$ and (10) holds when replacing there x_2 by $\bar{x}$, $x_{g,2}$ by $\bar{x}_g$ and x_2^{θ} by the corresponding characteristic point of $f(\cdot, \theta)$]. We thus obtain for every $\theta \in S(F)$

$$\begin{aligned} \vartheta_{g,\theta}^{(1)} &= \theta^{(1)} \frac{z(x_2) - z(x_1)}{z_g(x_{g,2}) - z_g(x_{g,1})} =: \theta^{(1)} \alpha_g, \\ \vartheta_{g,\theta}^{(2)} &= \theta^{(2)} \frac{x_2 - x_1}{x_{g,2} - x_{g,1}} =: \theta^{(2)} \beta_g, \\ \vartheta_{g,\theta}^{(3)} &= \theta^{(3)} + \theta^{(2)}\left(x_1 - \beta_g x_{g,1}\right) =: \theta^{(3)} + \theta^{(2)} \gamma_g, \\ \vartheta_{g,\theta}^{(4)} &= \theta^{(4)} + \theta^{(1)}\left(z(x_1) - \alpha_g z_g(x_{g,1})\right) =: \theta^{(4)} + \theta^{(1)} \delta_g. \end{aligned} \tag{11}$$

Since for each $r = 1, \ldots, p$, x_r and $x_{g,r}$ are uniquely determined, this implies that for any $\theta \in S(F)$, $g(\cdot, \vartheta) = f(\cdot, \theta)$ holds if and only if $\vartheta = \vartheta_{g,\theta} = h_{\alpha_g, \beta_g, \gamma_g, \delta_g}(\theta)$ with $\alpha_g, \beta_g, \gamma_g, \delta_g$ defined by (11). Hence, (ii) holds. Furthermore, since $E\theta_1 = (1, 1, 0, 0)^{\mathrm{T}}$, we obtain $E\vartheta_{g,\theta} = Eh_{\alpha_g, \beta_g, \gamma_g, \delta_g}(\theta_1) = (\alpha_g, \beta_g, \gamma_g, \delta_g)^{\mathrm{T}}$, and thus (iii) is fulfilled. □

PROOF OF PROPOSITION 3. (i) By construction

$$q\left(\tilde{f}^h, \vec{\theta}^h\right) \geq q\left(\tilde{f}^{h+1}, \vec{\theta}^h\right) \geq q\left(\tilde{f}^{h+1}, \vec{\theta}^{h+1}\right) \geq 0,$$

which proves assertion (i) and shows that

$$c := \inf_{h \in \mathbb{N}} q\left(\tilde{f}^h, \vec{\theta}^h\right) = \inf_{h \in \mathbb{N}} q\left(\tilde{f}^{h+1}, \vec{\theta}^h\right) = q\left(\tilde{f}', \vec{\theta}'\right),$$

for any limit point $(\tilde{f}', \vec{\theta}') \in A \times D^n$ of $(\tilde{f}^h, \vec{\theta}^h)_{h \in \mathbb{N}}$.

(ii) Let $(\tilde{f}', \vec{\theta}')$ be an arbitrary limit point of the sequence $(\tilde{f}^h, \vec{\theta}^h)_{h \in \mathbb{N}}$ and let $(\tilde{f}^{h_r}, \vec{\theta}^{h_r})_{r \in \mathbb{N}}$ be a subsequence converging to $(\tilde{f}', \vec{\theta}')$. Since q is continuous on $A \times D^n$, by Lemma 1 of Jennrich (1969), $\min_{a_s \in A} q(a_s, \vec{\theta}')$ [$\equiv \min_{s \in S} q(a_s, \vec{\theta}')$] and $\min_{\vec{\vartheta} \in D^n} q(\tilde{f}', \vec{\vartheta})$ are also continuous in $\vec{\theta}' \in D^n$ and $\tilde{f}' \in A$. Consequently,

$$\begin{aligned} c = q\left(\tilde{f}', \vec{\theta}'\right) &= \lim_{r \to \infty} q\left(\tilde{f}^{h_r}, \vec{\theta}^{h_r}\right) = \lim_{r \to \infty} q\left(\tilde{f}^{h_r+1}, \vec{\theta}^{h_r}\right) \\ &= \lim_{r \to \infty} \min_{a_s \in A} q\left(a_s, \vec{\theta}^{h_r}\right) = \min_{a_s \in A} q\left(a_s, \vec{\theta}'\right) \end{aligned}$$

and

$$q\left(\tilde{f}', \vec{\theta}'\right) = \lim_{r \to \infty} q\left(\tilde{f}^{h_r}, \vec{\theta}^{h_r}\right) = \lim_{r \to \infty} \min_{\vec{\vartheta} \in D^n} q\left(\tilde{f}^{h_r}, \vec{\vartheta}\right) = \min_{\vec{\vartheta} \in D^n} q\left(\tilde{f}', \vec{\vartheta}\right). \quad \square$$

For the proof of the theorem we need two lemmas. Within these lemmas and within the proof of the theorem we will use the following notation: For $m \in \mathbb{N}$,

$\tilde{f}_m$ and $\tilde{\theta}_{1,m}, \ldots, \tilde{\theta}_{n_m,m}$ denote the least-squares estimators obtained within step (i) of the SEMOR procedure, yielding the $(D, A_{r(m)}, N)$-SEMOR estimators $\hat{f}_{r(m),m} := \tilde{f}_m(\cdot, N(\tilde{f}_m, \tilde{F}_m)(\cdot))$ and $\hat{\theta}_{i,r(m),m} := N(\tilde{f}_m, \tilde{F}_m)^{-1}(\tilde{\theta}_{i,m})$, $i = 1, \ldots, n_m$. $\tilde{F}_m$ denotes the empirical distribution function of $\tilde{\theta}_{1,m}, \ldots, \tilde{\theta}_{n_m,m}$ and $\tilde{h}_m := N(\tilde{f}_m, \tilde{F}_m)$.

LEMMA A. *Under the assumptions of the theorem, we have:*

(a) *For each* $i \in \mathbb{N}$ *with probability* 1 *for* $m \to \infty$,

$$\frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \big(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta) \big) \varepsilon_{ij} \to 0,$$

uniformly for all $\vartheta \in D$ *and each* $g \in \mathscr{M}$.

(b) *With probability* 1 *for* $m \to \infty$,

$$\frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \big(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta_i) \big) \varepsilon_{ij} \to 0,$$

uniformly for all $g \in \mathscr{M}$ *and every sequence* $(\vartheta_i)_{i \in \mathbb{N}} \subset D$.

PROOF. First we will prove assertion (a). To simplify notation, for any $m, i, p \in \mathbb{N}$, we will use the abbreviations

$$\check{c} := \sup_{g, g' \in \mathscr{M}} \sup_{\vartheta, \vartheta' \in D} \sup_{t \in J} |g(t, \vartheta) - g'(t, \vartheta')|,$$

$$t_{r,p} := a + (r-1)\frac{b-a}{p}, \qquad \text{for } r = 1, \ldots, p,$$

$$J_{r,i,m,p} := \big\{ j \in \{1, \ldots, T_{i,m}\} | t_{ij} \in [t_{r,p}, t_{r+1,p}) \big\}, \quad \text{for } r = 1, \ldots, p-1,$$

$$J_{p,i,m,p} := \big\{ j \in \{1, \ldots, T_{i,m}\} | t_{ij} \in [t_{p,p}, b] \big\},$$

$$T_{r,i,m,p} := \# J_{r,i,m,p}, \qquad \text{for } r = 1, \ldots, p,$$

$$\xi_{r,i,m,p} := \left| \frac{1}{T_{i,m}} \sum_{j \in J_{r,i,m,p}} \varepsilon_{ij} \right|, \qquad \text{for } r = 1, \ldots, p.$$

(Since $\mathscr{M}$, D and $J := [a, b]$ are compact, $\check{c} < \infty$.)

Select an arbitrary $i \in \mathbb{N}$ and an arbitrary $\varepsilon > 0$.

The compactness of J and D implies that each $g \in \mathscr{M}$ is uniformly continuous on $J \times D$. Furthermore, by Ascoli's theorem, it follows from the compactness of $\mathscr{M}$ that $\mathscr{M}$ is equicontinuous. Consequently, there exists a $p \in \mathbb{N}$, $p \equiv p_{\varepsilon,i}$, so that for each $g \in \mathscr{M}$ and all $\vartheta \in D$,

$$\sup_{\substack{t, t' \in J \\ |t-t'| \le (b-a)/p}} |g(t, \vartheta) - g(t', \vartheta)| \le \frac{\varepsilon}{4\sqrt{c_i}}.$$

Here c_i denotes a constant satisfying $\sigma_{ij}^2 \le c_i$ for all $j \in \mathbb{N}$. Its existence is guaranteed by Assumption 3(i)(a).

We now obtain for all $\vartheta \in D$, each $g \in \mathscr{M}$ and every $m \in \mathbb{N}$,

$$
\begin{aligned}
&\left| \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta) \right) \varepsilon_{ij} \right| \\
&\quad \le \left| \sum_{r=1}^{p} \left(f(t_{r,p}, \theta_i) - g(t_{r,p}, \vartheta) \right) \frac{1}{T_{i,m}} \sum_{j \in J_{r,i,m,p}} \varepsilon_{ij} \right| \\
&\qquad + \left| \frac{1}{T_{i,m}} \sum_{r=1}^{p} \sum_{j \in J_{r,i,m,p}} \left[\left(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta) \right) \right. \right. \\
&\qquad\qquad \left. \left. - \left(f(t_{r,p}, \theta_i) - g(t_{r,p}, \vartheta) \right) \right] \varepsilon_{ij} \right| \\
&\quad \le \left| \check{c} \sum_{r=1}^{p} \frac{1}{T_{i,m}} \sum_{j \in J_{r,i,m,p}} \varepsilon_{ij} \right| \\
&\qquad + \left\{ \frac{1}{T_{i,m}} \sum_{r=1}^{p} \sum_{j \in J_{r,i,m,p}} \left[\left(f(t_{ij}, \theta_i) - f(t_{r,p}, \theta_i) \right) \right. \right. \\
&\qquad\qquad \left. \left. - \left(g(t_{ij}, \vartheta) - g(t_{r,p}, \vartheta) \right) \right]^2 \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \varepsilon_{ij}^2 \right\}^{1/2} \\
&\quad \le \check{c} \sum_{r=1}^{p} \xi_{r,i,m,p} \\
&\qquad + \left\{ \left[\frac{1}{T_{i,m}} \sum_{r=1}^{p} \sum_{j \in J_{r,i,m,p}} \left(\frac{\varepsilon}{2\sqrt{c_i}} \right)^2 \right] \left[\frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) + c_i \right] \right\}^{1/2} \\
&\quad \le \check{c} \sum_{r=1}^{p} \xi_{r,i,m,p} \\
&\qquad + \left\{ \left[\frac{1}{T_{i,m}} \sum_{r=1}^{p} T_{r,i,m,p} \left(\frac{\varepsilon}{2\sqrt{c_i}} \right)^2 \right] c_i \right. \\
&\qquad\qquad \left. + \left| \left[\frac{1}{T_{i,m}} \sum_{r=1}^{p} T_{r,i,m,p} \left(\frac{\varepsilon}{2\sqrt{c_i}} \right)^2 \right] \left[\frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) \right] \right| \right\}^{1/2} \\
&\quad = \check{c} \sum_{r=1}^{p} \xi_{r,i,m,p} + \left\{ \frac{\varepsilon^2}{4} + \left| \frac{\varepsilon^2}{4c_i} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) \right| \right\}^{1/2} \\
&\quad \le \check{c} \sum_{r=1}^{p} \xi_{r,i,m,p} + \frac{\varepsilon}{2} + \left\{ \left| \frac{\varepsilon^2}{4c_i} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) \right| \right\}^{1/2}.
\end{aligned}
$$

Since $\inf_i T_{i,m} \to \infty$ as $m \to \infty$, it follows from Assumption 3(ii) that $T_{r,i,m,p} = \# J_{r,i,m,p} \to \infty$ as $m \to \infty$ for all $r = 1, \dots, p$. Since, furthermore, $T_{r,i,m,p} \le T_{i,m}$ and since due to Assumption 3(i)(a) $\sigma_{ij}^2 \le c_i < \infty$ for each $j \in \mathbb{N}$, the strong law of large numbers implies that for all $r = 1, \dots, p$ as $m \to \infty$,

$$\xi_{r,i,m,p} = \left| \frac{1}{T_{i,m}} \sum_{j \in J_{r,i,m,p}} \varepsilon_{ij} \right| \to 0 \quad \text{a.s.}$$

Moreover, since by Assumption 3(i)(a) $E(\varepsilon_{ij}^2 - \sigma_{ij}^2)^2 \le \alpha_{ij}^4 \le c_i < \infty$ for every $j \in \mathbb{N}$, the strong law of large numbers also yields

$$\frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) \to 0 \quad \text{a.s. as } m \to \infty$$

and hence

$$\zeta_{i,m} := \left\{ \left| \frac{\varepsilon^2}{4c_i} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) \right| \right\}^{1/2} \to 0 \quad \text{a.s as } m \to \infty.$$

Combining these considerations, we obtain

$$\lim_{\tilde{m} \to \infty} P\left(\sup_{m \ge \tilde{m}} \left[\sup_{\vartheta \in D} \sup_{g \in \mathscr{M}} \left| \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta) \right) \varepsilon_{ij} \right| \right] < \varepsilon \right)$$
$$\ge \lim_{\tilde{m} \to \infty} P\left(\sup_{m \ge \tilde{m}} \left| \check{c} \sum_{r=1}^{p} \xi_{r,i,m,p} + \frac{\varepsilon}{2} + \zeta_{i,m} \right| < \varepsilon \right) = 1.$$

Since ε is arbitrary this proves assertion (a).

It remains to prove assertion (b).

Choose an arbitrary $\varepsilon > 0$. Then select a $p \in \mathbb{N}$, $p \equiv p_\varepsilon$, so that for any $g \in \mathscr{M}$ and all $\vartheta \in D$,

$$\sup_{\substack{t, t' \in J \\ |t - t'| \le (b-a)/p}} |g(t, \vartheta) - g(t', \vartheta)| \le \frac{\varepsilon}{4\sqrt{c}}.$$

Hereby c is the constant given by Assumption 3(i)(b) with the property that $(1/n)\sum_{i=1}^n c_i \le c < \infty$ for all $n \in \mathbb{N}$.

Using the notation given above, the following holds for all $(\vartheta_i)_{i \in \mathbb{N}} \subset D$, each $g \in \mathscr{M}$ and every $m \in \mathbb{N}$,

$$\left| \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta_i) \right) \varepsilon_{ij} \right|$$
$$\le \check{c} \frac{1}{n_m} \sum_{i=1}^{n_m} \sum_{r=1}^{p} \xi_{r,i,m,p}$$
$$+ \left\{ \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{r=1}^{p} \right.$$
$$\times \sum_{j \in J_{r,i,m,p}} \left[\left(f(t_{ij}, \theta_i) - f(t_{r,p}, \theta_i) \right) - \left(g(t_{ij}, \vartheta_i) - g(t_{r,p}, \vartheta_i) \right) \right]^2$$

$$\left. \times \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \varepsilon_{ij}^2 \right\}^{1/2}$$

$$\leq \check{c} \frac{1}{n_m} \sum_{i=1}^{n_m} \sum_{r=1}^{p} \xi_{r,i,m,p} + \left\{ \left[\frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{r=1}^{p} T_{r,i,m,p} \left(\frac{\varepsilon}{2\sqrt{c}} \right)^2 \right] \right.$$

$$\left. \times \left[\frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) + c \right] \right\}^{1/2}$$

$$\leq \check{c} \frac{1}{n_m} \sum_{i=1}^{n_m} \sum_{r=1}^{p} \xi_{r,i,m,p} + \frac{\varepsilon}{2} + \left\{ \left| \frac{\varepsilon^2}{4c} \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2 \right) \right| \right\}^{1/2}.$$

As mentioned above, the strong law of large numbers implies for all $i \in \mathbb{N}$ and each $r = 1, \ldots, p$,

$$\xi_{r,i,m,p} \to 0 \quad \text{a.s. as } m \to \infty.$$

Moreover,

$$0 \leq \lim_{m \to \infty} E\xi_{r,i,m,p} \leq \lim_{m \to \infty} \sqrt{E\xi_{r,i,m,p}^2} \leq \lim_{m \to \infty} \sqrt{\frac{c_i}{T_{i,m}}} = 0.$$

Thus for all $\tilde{m} \in \mathbb{N}$ with probability 1,

$$\lim_{\acute{m} \to \infty} \sum_{r=1}^{p} \frac{1}{n_{\tilde{m}}} \sum_{i=1}^{n_{\tilde{m}}} \left(\xi_{r,i,\acute{m},p} - E\xi_{r,i,\acute{m},p} \right) = 0.$$

It follows that a.s.

$$\lim_{\tilde{m} \to \infty} \lim_{\acute{m} \to \infty} \sum_{r=1}^{p} \frac{1}{n_{\tilde{m}}} \sum_{i=1}^{n_{\tilde{m}}} \left(\xi_{r,i,\acute{m},p} - E\xi_{r,i,\acute{m},p} \right) = 0.$$

On the other hand, by Assumption 3(i)(a), $E\xi_{r,i,m,p}^2 \leq c_i/T_{i,m}$ for all $i, m \in \mathbb{N}$ and each $r = 1, \ldots, p$, whereas from $(1/n_m)\sum_{i=1}^{n_m} c_i \leq c < \infty$ (for all $m \in \mathbb{N}$), it follows $\sum_{i=1}^{\infty} c_i/i^2 < \infty$. Since, additionally, independence of the r.v. ε_{ij}, $i, j \in \mathbb{N}$, implies independence of the r.v. $\xi_{r,i,m,p}$, $i \in \mathbb{N}$, we can deduce from the strong law of large numbers that for all $\acute{m} \in \mathbb{N}$ with probability 1,

$$\lim_{\tilde{m} \to \infty} \sum_{r=1}^{p} \frac{1}{n_{\tilde{m}}} \sum_{i=1}^{n_{\tilde{m}}} \left(\xi_{r,i,\acute{m},p} - E\xi_{r,i,\acute{m},p} \right) = 0.$$

It follows that a.s.

$$\lim_{\acute{m} \to \infty} \lim_{\tilde{m} \to \infty} \sum_{r=1}^{p} \frac{1}{n_{\tilde{m}}} \sum_{i=1}^{n_{\tilde{m}}} \left(\xi_{r,i,\acute{m},p} - E\xi_{r,i,\acute{m},p} \right) = 0.$$

Combining this, we can infer that a.s.

$$\lim_{\tilde{m} \to \infty} \lim_{\acute{m} \to \infty} \sum_{r=1}^{p} \frac{1}{n_{\tilde{m}}} \sum_{i=1}^{n_{\tilde{m}}} \left(\xi_{r,i,\acute{m},p} - E\xi_{r,i,\acute{m},p} \right)$$

$$= \lim_{\acute{m} \to \infty} \lim_{\tilde{m} \to \infty} \sum_{r=1}^{p} \frac{1}{n_{\tilde{m}}} \sum_{i=1}^{n_{\tilde{m}}} \left(\xi_{r,i,\acute{m},p} - E\xi_{r,i,\acute{m},p} \right) = 0$$

and, consequently,

$$\sum_{r=1}^{p} \frac{1}{n_m} \sum_{i=1}^{n_m} \left(\xi_{r,i,m,p} - E\xi_{r,i,m,p}\right) \to 0 \quad \text{a.s. as } m \to \infty.$$

Since by Assumption 3(i)(b), $(1/n_m)\sum_{i=1}^{n_m}\sqrt{c_i} = O(1)_{(m\to\infty)}$, which implies

$$0 \le \lim_{m\to\infty} \sum_{r=1}^{p} \frac{1}{n_m} \sum_{i=1}^{n_m} E\xi_{r,i,m,p} \le \lim_{m\to\infty} \sum_{r=1}^{p} \frac{1}{n_m} \sum_{i=1}^{n_m} \sqrt{\frac{c_i}{T_{i,m}}} = 0,$$

we can thus conclude

$$\frac{1}{n_m} \sum_{i=1}^{n_m} \sum_{r=1}^{p} \xi_{r,i,m,p} \to 0 \quad \text{a.s. as } m \to \infty.$$

On the other hand, it follows from Assumption 3(i) that $E(\varepsilon_{ij}^2 - \sigma_{ij}^2)^2 \le \alpha_{ij}^4 \le c_i$ and $\sum_{i=1}^{\infty} c_i/i^2 < \infty$. Hence we can infer from the strong law of large numbers that as $m \to \infty$,

$$\frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2\right) \to 0 \quad \text{a.s.},$$

and therefore

$$\zeta_m := \left\{ \left| \frac{\varepsilon^2}{4c} \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(\varepsilon_{ij}^2 - \sigma_{ij}^2\right) \right| \right\}^{1/2} \to 0 \quad \text{a.s.}$$

Combining again these considerations, we finally obtain

$$\lim_{\tilde m\to\infty} P\left(\sup_{m\ge \tilde m} \sup_{\vartheta_1,\dots,\vartheta_{n_m}\in D} \sup_{g\in\mathscr{M}} \left| \frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(f(t_{ij},\theta_i) - g(t_{ij},\vartheta_i)\right)\varepsilon_{ij} \right| < \varepsilon \right)$$

$$\ge \lim_{\tilde m\to\infty} P\left(\sup_{m\ge\tilde m} \left| \check{c} \frac{1}{n_m} \sum_{i=1}^{n_m} \sum_{r=1}^{p} \xi_{r,i,m,p} + \frac{\varepsilon}{2} + \zeta_m \right| < \varepsilon \right) = 1.$$

Since ε is arbitrary this proves assertion (b). □

Lemma B. *Under the assumptions of the theorem, with probability* 1

$$\frac{1}{n_m} \sum_{i=1}^{n_m} \sup_{t\in J} \left(f(t,\theta_i) - \tilde f_m(t,\tilde\theta_{i,m})\right)^2 \to 0, \quad \text{as } m \to \infty. \tag{12}$$

Proof. The proof will be given in two steps.

Step 1. With probability 1

$$\frac{1}{n_m} \sum_{i=1}^{n_m} \frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \left(f(t_{ij},\theta_i) - \tilde f_m(t_{ij},\tilde\theta_{i,m})\right)^2 \to 0, \quad \text{as } m \to \infty. \tag{13}$$

Under the assumptions of the theorem there exists an $f^* \in \mathscr{M}$ with $f^*|_{J\times S(F)} = f|_{J\times S(F)}$. Moreover, there exists a sequence $(f_{r(m)})_{m\in\mathbb{N}}$ of functions with $f_{r(m)} \in A_{r(m)}$ for each $m \in \mathbb{N}$ so that for $m \to \infty$, $f_{r(m)}(x) \to f^*(x)$ uniformly for all $x \in J \times D$. By definition, for each $m \in \mathbb{N}$ the least-squares estimators $\tilde{f}_m, \tilde{\theta}_{1,m}, \ldots, \tilde{\theta}_{n_m,m}$ minimize $(1/n_m)\sum_{i=1}^{n_m} 1/T_{i,m}\sum_{j=1}^{T_{i,m}}(Y_{ij} - g(t_{ij}, \vartheta_i))^2$ with respect to all $g \in A_{r(m)}$ and all $\vartheta_1, \ldots, \vartheta_{n_m} \in D$. Therefore for each $m \in \mathbb{N}$,

$$\frac{1}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(Y_{ij} - f_{r(m)}(t_{ij}, \theta_i)\right)^2 \geq \frac{1}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(Y_{ij} - \tilde{f}_m(t_{ij}, \tilde{\theta}_{i,m})\right)^2.$$

Taking into account the model assumption $Y_{ij} = f(t_{ij}, \theta_i) + \varepsilon_{ij}$, this yields for each $m \in \mathbb{N}$,

$$\begin{aligned}
&\frac{1}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(f(t_{ij}, \theta_i) - f_{r(m)}(t_{ij}, \theta_i)\right)^2\\
&\quad+\frac{2}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(f(t_{ij}, \theta_i) - f_{r(m)}(t_{ij}, \theta_i)\right)\varepsilon_{ij}\\
&\quad\geq \frac{1}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(f(t_{ij}, \theta_i) - \tilde{f}_m(t_{ij}, \tilde{\theta}_{i,m})\right)^2\\
&\quad\quad+\frac{2}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(f(t_{ij}, \theta_i) - \tilde{f}_m(t_{ij}, \tilde{\theta}_{i,m})\right)\varepsilon_{ij}.
\end{aligned}$$

Since a.s. $\theta_i \in S(F) \subset D$ for all $i \in \mathbb{N}$ and thus, by construction of the sequence $(f_{r(m)})m \in \mathbb{N}$,

$$\frac{1}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\left(f(t_{ij}, \theta_i) - f_{r(m)}(t_{ij}, \theta_i)\right)^2$$

a.s. converges to 0 as $m \to \infty$, the assertion of Step 1 follows from Lemma A(b).

STEP 2. (13) implies (12).

This will be proved by contradiction. Assume that for some sequence of realizations of the r.v. (13) holds, whereas (12) is wrong. Then there exists an $\varepsilon > 0$ and a sequence $(m_l)_{l\in\mathbb{N}}$ such that for each $l \in \mathbb{N}$,

$$\frac{1}{n_{m_l}}\sum_{i=1}^{n_{m_l}}\sup_{t\in J}\left(f(t, \theta_i) - \tilde{f}_{m_l}(t, \tilde{\theta}_{i,m_l})\right)^2 \geq \varepsilon.$$

Furthermore, due to the equicontinuity of $\mathscr{M}$ and the compactness of $J \times D$, there exists a $p \in \mathbb{N}$ so that for each $g \in \mathscr{M}$ and all $\vartheta, \tilde{\vartheta} \in D$,

$$\sup_{\substack{t, t'\in J\\ |t-t'|\leq(b-a)/p}}\left|\left(f(t, \vartheta) - g(t, \tilde{\vartheta})\right)^2 - \left(f(t', \vartheta) - g(t', \tilde{\vartheta})\right)^2\right| < \frac{\varepsilon}{2}.$$

Using the notation introduced within the proof of Lemma A, this yields

$$\frac{1}{n_m}\sum_{i=1}^{n_m}\frac{1}{T_{i,m}}\sum_{j=1}^{T_{i,m}}\Big(f(t_{ij},\theta_i)-\tilde{f}_m\big(t_{ij},\tilde{\theta}_{i,m}\big)\Big)^2$$
$$\geq \frac{1}{n_m}\sum_{i=1}^{n_m}\sum_{r=1}^{p}\frac{T_{r,i,m,p}}{T_{i,m}}\left[\sup_{t\in[t_{r,p},\,t_{r+1,p}]}\Big(f(t,\theta_i)-\tilde{f}_m\big(t,\tilde{\theta}_{i,m}\big)\Big)^2-\frac{\varepsilon}{2}\right]$$
$$\geq \frac{1}{n_m}\sum_{i=1}^{n_m}\min_{r=1,\dots,p}\frac{T_{r,i,m,p}}{T_{i,m}}\left[\sup_{t\in J}\Big(f(t,\theta_i)-\tilde{f}_m\big(t,\tilde{\theta}_{i,m}\big)\Big)^2-\frac{\varepsilon}{2}\right].$$

On the other hand, since $\inf_{i\in\mathbb{N}}T_{i,m}\to\infty$, Assumption 3(ii) implies the existence of an $\tilde{m}\in\mathbb{N}$ and of a $\gamma>0$ with the property that for all $m\geq\tilde{m}$ and every $i\in\mathbb{N}$,

$$\min_{r=1,\dots,p}\frac{T_{r,i,m,p}}{T_{i,m}}\geq\gamma.$$

Combining these relations, one can conclude that for each $l\in\mathbb{N}$ with $m_l\geq\tilde{m}$,

$$\frac{1}{n_{m_l}}\sum_{i=1}^{n_{m_l}}\frac{1}{T_{i,m_l}}\sum_{j=1}^{T_{i,m_l}}\Big(f(t_{ij},\theta_i)-\tilde{f}_{m_l}\big(t_{ij},\tilde{\theta}_{i,m_l}\big)\Big)^2$$
$$\geq\gamma\frac{1}{n_{m_l}}\sum_{i=1}^{n_{m_l}}\left[\sup_{t\in J}\Big(f(t,\theta_i)-\tilde{f}_{m_l}\big(t,\tilde{\theta}_{i,m_l}\big)\Big)^2-\frac{\varepsilon}{2}\right]\geq\gamma\frac{\varepsilon}{2}.$$

This is a contradiction to (13). □

Proof of the theorem. (a) To prove assertion (a) we have to show that a.s. as $m\to\infty$

$$\hat{f}_{r(m),m}(t,\vartheta)=\tilde{f}_m\big(t,N(\tilde{f}_m,\tilde{F}_m)(\vartheta)\big)\to f(t,\vartheta), \tag{14}$$

uniformly for all $(t,\vartheta)\in J\times S$.

To prove that (14) holds with probability 1, it suffices to show that with probability 1,

$$\tilde{f}\vDash f \text{ and } \tilde{F}\cong_{\tilde{f}} F, \tag{15}$$

for each limit point $(\tilde{f},\tilde{F})\in\mathscr{M}\times\mathscr{F}_D$ of $(\tilde{f}_m,\tilde{F}_m)_{m\in\mathbb{N}}$.

This can easily be seen by the following argument: Assume that for some sequence of realizations of the r.v. (14) does not hold, whereas (15) does. Then there exist an $\varepsilon>0$ and a subsequence $(\tilde{f}_{m_l},\tilde{F}_{m_l})_{l\in\mathbb{N}}$ of $(\tilde{f}_m,\tilde{F}_m)_{m\in\mathbb{N}}$ such that for each $l\in\mathbb{N}$,

$$\sup_{(t,\vartheta)\in J\times S}\Big|\tilde{f}_{m_l}\big(t,N(\tilde{f}_{m_l},\tilde{F}_{m_l})(\vartheta)\big)-f(t,\vartheta)\Big|\geq\varepsilon. \tag{16}$$

Due to the compactness of D, $\mathscr{F}_D$ is sequentially compact. Since by construction $\tilde{F}_{m_l}\in\mathscr{F}_D$ for each $l\in\mathbb{N}$, together with the compactness of $\mathscr{M}$, this implies the existence of a limit point $(\tilde{f},\tilde{F})\in\mathscr{M}\times\mathscr{F}_D$ of $(\tilde{f}_{m_l},\tilde{F}_{m_l})_{l\in\mathbb{N}}\subset\mathscr{M}\times\mathscr{F}_D$ and of a further subsequence $(\tilde{f}_{m_l(p)},\tilde{F}_{m_l(p)})_{p\in\mathbb{N}}$ satisfying $\tilde{f}_{m_l(p)}\to\tilde{f}$ and $\tilde{F}_{m_l(p)}\to_{\mathscr{D}}\tilde{F}$ as $p\to\infty$. From (15) follows $\tilde{f}\vDash f$ and $\tilde{F}\cong_{\tilde{f}}F$. Thus, by assumption, $N|_{\mathscr{M}\times\mathscr{F}_D}$ is continuous at $(\tilde{f},\tilde{F})$ yielding $N(\tilde{f}_{m_l(p)},\tilde{F}_{m_l(p)})\to N(\tilde{f},\tilde{F})$. Using (16), we can conclude

$$\varepsilon \leq \lim_{p\to\infty} \sup_{(t,\vartheta)\in J\times S} \left| \tilde{f}_{m_l(p)}\Big(t, N\big(\tilde{f}_{m_l(p)}, \tilde{F}_{m_l(p)}\big)(\vartheta)\Big) - f(t,\vartheta) \right|$$
$$= \sup_{(t,\vartheta)\in J\times S} \left| \tilde{f}\big(t, N(\tilde{f}, \tilde{F})(\vartheta)\big) - f(t,\vartheta) \right|,$$

establishing a contradiction to the definition of S.

By Lemma B, (12) holds almost surely. Moreover, by the Glivenko–Cantelli theorem and by the definition of $S(F)$, with probability 1

(17) $$F_m \to_{\mathscr{D}} F, \quad \text{as } m \to \infty, \quad \big(F_m\text{: empirical distribution function of } \theta_1, \dots, \theta_{n_m}\big),$$

(18) $$\theta_i \in S(F), \quad \text{for all } i \in \mathbb{N}.$$

Consequently, the *common* probability of (12), (17) and (18) is 1, too. Theorefore to prove assertion (a) of the theorem, it suffices to show that combining (12), (17) and (18) *implies* (15).

Thus let us assume that (12), (17) and (18) hold for some sequence of realizations of the r.v. Under this assumption (15) can be derived as follows.

Let $(\tilde{f}, \tilde{F}) \in \mathscr{M} \times \mathscr{F}_D$ denote an *arbitrary* limit point of $(\tilde{f}_m, \tilde{F}_m)_{m\in\mathbb{N}}$ and let $(\tilde{f}_{m_l}, \tilde{F}_{m_l})_{l\in\mathbb{N}}$ be some subsequence converging to $(\tilde{f}, \tilde{F})$. We have to show that $\tilde{f} \models f$ and $\tilde{F} \cong_{\tilde{f}} F$. This will be done in two steps.

STEP 1. $\tilde{f} \models f$.

By definition, $\tilde{f}_{m_l} \to \tilde{f}$ (in the topology of compact convergence), implying $\tilde{f}_{m_l}(t,\theta) \to \tilde{f}(t,\theta)$ uniformly for all $(t,\theta) \in J \times D$. Since $\tilde{\theta}_{i,m_l} \in D$ for each $l \in \mathbb{N}$, it follows from (12) that as $l \to \infty$,

$$\frac{1}{n_{m_l}} \sum_{i=1}^{n_{m_l}} \sup_{t\in J} \big(f(t,\theta_i) - \tilde{f}(t, \tilde{\theta}_{i,m})\big)^2 \to 0.$$

This implies

$$\frac{1}{n_{m_l}} \sum_{i=1}^{n_{m_l}} \min_{\vartheta\in D} \sup_{t\in J} \big(f(t,\theta_i) - \tilde{f}(t,\vartheta)\big)^2 \to 0,$$

and hence, due to (18),

(19) $$\int_{S(F)} \left[\min_{\vartheta\in D} \sup_{t\in J} \big(f(t,\theta) - \tilde{f}(t,\vartheta)\big)^2\right] dF_{m_l}(\theta) \to 0.$$

Following Lemma 1 of Jennrich (1969),

$$q(\theta) := \min_{\vartheta\in D} \sup_{t\in J} \big(f(t,\theta) - \tilde{f}(t,\vartheta)\big)^2$$

is a continuous function of $\theta \in S(F)$. Thus combining (17) and (19) yields

$$\int_{S(F)} \left[\min_{\vartheta\in D} \sup_{t\in J} \big(f(t,\theta) - \tilde{f}(t,\vartheta)\big)^2\right] dF(\theta) = 0.$$

Due to the continuity of q and due to the definition of $S(F)$, this induces that for each $\theta \in S(F)$, $q(\theta) = 0$ [$S(F)$ is the *smallest* closed set S with $F(S) = 1$].

Thus for each $\theta \in S(F)$ there is a $\vartheta_\theta \in D \subset \mathscr{R}$ satisfying

$$\sup_{t \in J} \left(f(t,\theta) - \tilde{f}(t,\vartheta_\theta) \right)^2 = 0,$$

implying $\tilde{f} \models f$.

STEP 2. $\tilde{F} \cong_{\tilde{f}} F$.

Lemma 2 of Jennrich (1969) implies the existence of a measurable mapping $\bar{h}$: $\mathscr{R} \to D$ with

$$\sup_{t \in J} \left(f(t,\theta) - \tilde{f}(t,\bar{h}(\theta)) \right)^2 = \min_{\vartheta \in D} \sup_{t \in J} \left(f(t,\theta) - \tilde{f}(t,\vartheta) \right)^2,$$

for all $\theta \in \mathscr{R}$. Since we can infer from step 1 that $\sup_{t \in J}(f(t,\theta) - \tilde{f}(t,\bar{h}(\theta)))^2 = 0$ for any $\theta \in S(F)$, it follows that the distribution function F^* of the r.v. $\bar{h}(\theta_1)$ satisfies $F^* \cong_{\tilde{f}} F$. From Definition 1(iii) and from the properties of N, we can now conclude that $h := N(\tilde{f}, F^*)$ satisfies $\bar{h}|_{S(F)} = h|_{S(F)}$ and $\tilde{f}(\cdot, h(\cdot)) \in M$. By Definition 1(ii), this implies that for each $\theta \in S(F)$,

(20) $$f(\cdot,\theta) = \tilde{f}(\cdot,h(\theta)) \neq \tilde{f}(\cdot,\vartheta), \quad \text{for all } \vartheta \in D \text{ with } \vartheta \neq h(\theta).$$

Hence, $\tilde{F} \cong_{\tilde{f}} F$ *obviously holds if and only if* $\tilde{F} = F^*$.

Given $m \in \mathbb{N}$, $\varepsilon, \delta > 0$ and $\vartheta \in D$, we will use the abbreviations

$$I_{m,\varepsilon} := \#\left\{ i \in \{1,\dots,n_m\} \mid \sup_{t \in J} \left(f(t,\theta_i) - \tilde{f}(t,\tilde{\theta}_{i,m}) \right)^2 \geq \varepsilon \right\},$$

$$S_\delta(\vartheta) := \left\{ \bar{\vartheta} \in D \mid \|\vartheta - \bar{\vartheta}\|_2 < \delta \right\},$$

$$R_{m,\varepsilon,\delta} := \#\left\{ i \in \{1,\dots,n_m\} \mid \inf_{\substack{\vartheta \in D \\ \vartheta \notin S_\delta(h(\theta_i))}} \sup_{t \in J} \left(f(t,\theta_i) - \tilde{f}(t,\vartheta) \right)^2 < \varepsilon \right\}.$$

It follows from (12) that for each $\varepsilon > 0$,

$$\frac{1}{n_{m_l}} I_{m_l,\varepsilon} \to 0, \quad \text{as } l \to \infty.$$

Moreover, it can be obtained from (20) and (18) that for every $\delta > 0$,

$$\lim_{\varepsilon \to 0} \lim_{l \to \infty} \frac{1}{n_{m_l}} R_{m_l,\varepsilon,\delta} = 0.$$

Let v denote an arbitrary continuous function from $\mathscr{R}$ into $\mathscr{R}$, and select an arbitrary $\gamma > 0$. The function v is uniformly continuous on the compact set D. Thus there exists a $\delta > 0$ satisfying

$$\|v(\vartheta) - v(\bar{\vartheta})\|_2 < \gamma, \quad \text{for all } \vartheta, \bar{\vartheta} \in D \text{ with } \|\vartheta - \bar{\vartheta}\| < \delta.$$

Now set

$$\text{CON} := \sup_{\vartheta, \bar{\vartheta} \in D} \|v(\vartheta) - v(\bar{\vartheta})\|_2.$$

Since D is compact, we have $\text{CON} < \infty$. Hence there exists an $\varepsilon > 0$ such that

$$\lim_{l \to \infty} \frac{1}{n_{m_l}} R_{m_l,\varepsilon,\delta} < \frac{\gamma}{\text{CON}}.$$

It follows from the definitions of $I_{m_l,\varepsilon}$ and $R_{m_l,\varepsilon,\delta}$ that for each $l \in \mathbb{N}$,

$$\frac{1}{n_{m_l}} \sum_{i=1}^{n_{m_l}} \left\| v(h(\theta_i)) - v\left(\tilde{\theta}_{i,m_l}\right) \right\|_2$$

$$< \mathrm{CON} \frac{1}{n_{m_l}} I_{m_l,\varepsilon} + \mathrm{CON} \frac{1}{n_{m_l}} R_{m_l,\varepsilon,\delta} + \gamma \left[\frac{1}{n_{m_l}} \left(n_{m_l} - I_{m_l,\varepsilon} - R_{m_l,\varepsilon,\delta} \right) \right],$$

so that for l sufficiently large

$$\frac{1}{n_{m_l}} \sum_{i=1}^{n_{m_l}} \left\| v(h(\theta_i)) - v\left(\tilde{\theta}_{i,m_l}\right) \right\|_2 < 2\,\mathrm{CON} \frac{\gamma}{\mathrm{CON}} + \gamma = 3\gamma.$$

Since γ is arbitrary, this implies

$$\frac{1}{n_{m_l}} \sum_{i=1}^{n_{m_l}} \left\| v(h(\theta_i)) - v\left(\tilde{\theta}_{i,m_l}\right) \right\|_2 \to 0, \quad \text{as } l \to \infty, \tag{21}$$

inducing thus

$$\left\| \int v(\theta)\, dF_{m_l}^*(\theta) - \int v(\theta)\, d\tilde{F}_{m_l}(\theta) \right\|_2 \to 0, \quad \text{as } l \to \infty,$$

where for $m \in \mathbb{N}$, F_m^* denotes the empirical distribution function of $h(\theta_1), \dots, h(\theta_{n_m})$. Since it can be deduced from (17) that $F_m^* \to_{\mathscr{D}} F^*$, we can conclude that

$$\left\| \int v(\theta)\, dF^*(\theta) - \int v(\theta)\, d\tilde{F}_{m_l}(\theta) \right\|_2 \to 0, \quad \text{for } l \to \infty.$$

Since v is arbitrary, we thus obtain

$$\tilde{F}_{m_l} \to_{\mathscr{D}} F^*,$$

which proves $\tilde{F} = F^*$ and, consequently, $\tilde{F} \cong_{\tilde{f}} F$.

(b) Choose an arbitrary $i \in \mathbb{N}$. To prove assertion (b) of the theorem we have to show that with probability 1 as $m \to \infty$,

$$\theta_i = \lim_{m\to\infty} \hat{\theta}_{i,r(m),m} = \lim_{m\to\infty} \tilde{h}_m^{-1}\left(\tilde{\theta}_{i,m}\right). \tag{22}$$

To prove this it suffices to show that with probability 1,

$$\begin{aligned} &\theta_i = N\left(\tilde{f}, \tilde{F}\right)^{-1}(\tilde{\theta}), \\ &\qquad \text{for each limit point } \left(\tilde{\theta}, \tilde{f}, \tilde{F}\right) \text{ of } \left(\tilde{\theta}_{i,m}, \tilde{f}_m, \tilde{F}_m\right)_{m\in\mathbb{N}}. \end{aligned} \tag{23}$$

This can be seen by the following argument: Assume that for some sequence of realizations of the r.v. (23) holds, whereas (22) is wrong. Then there exists an open neighborhood U of θ_i and a subsequence $(\hat{\theta}_{i,r(m_l),m_l})_{l\in\mathbb{N}}$ of $(\hat{\theta}_{i,r(m),m})_{m\in\mathbb{N}}$ with the property that for each $l \in \mathbb{N}$, $\hat{\theta}_{i,r(m_l),m_l} = \tilde{h}_{m_l}^{-1}(\tilde{\theta}_{i,m_l}) \notin U$. But, since D and $\mathscr{M}$ are compact whereas $\mathscr{F}_D$ is sequentially compact, there exists a limit point $(\tilde{\theta}, \tilde{f}, \tilde{F}) \in D \times \mathscr{M} \times \mathscr{F}_D$ of $(\tilde{\theta}_{i,m_l}, \tilde{f}_{m_l}, \tilde{F}_{m_l})_{l\in\mathbb{N}}$ and a further subsequence

$(\tilde{\theta}_{i, m_{l(s)}}, \tilde{f}_{m_{l(s)}}, \tilde{F}_{m_{l(s)}})_{s \in \mathbb{N}}$ converging to $(\tilde{\theta}, \tilde{f}, \tilde{F})$. Due to (a), $\tilde{f} \vDash f$ and $\tilde{F} \cong_{\tilde{f}} F$. Since, by assumption, $N(\cdot, \cdot)^{-1}$ (restricted to $\mathscr{M} \times \mathscr{F}_D$) is continuous at $(\tilde{f}, \tilde{F})$, it follows that $\tilde{h}_{m_{l(s)}}^{-1} \to h^{-1} := N(\tilde{f}, \tilde{F})^{-1}$. This implies

$$\tilde{h}_{m_{l(s)}}^{-1}(\vartheta) \to h^{-1}(\vartheta), \quad \text{uniformly for all } \vartheta \in D, \tag{24}$$

and thus $\tilde{h}_{m_{l(s)}}^{-1}(\tilde{\theta}_{i, m_{l(s)}}) \to h^{-1}(\tilde{\theta})$. This establishes a contradiction to (23), since because of $\tilde{h}_{m_{l(s)}}^{-1}(\tilde{\theta}_{i, m_{l(s)}}) \notin U$ for each $s \in \mathbb{N}$ it holds $h^{-1}(\tilde{\theta}) \notin U$ and thus $h^{-1}(\tilde{\theta}) \neq \theta_i$.

Due to Lemma A,

$$\frac{1}{T_{i,m}} \sum_{j=1}^{T_{i,m}} \big(f(t_{ij}, \theta_i) - g(t_{ij}, \vartheta) \big) \varepsilon_{ij} \to 0, \quad \text{as } m \to \infty, \tag{25}$$

$$\text{uniformly for all } \vartheta \in D \text{ and each } g \in \mathscr{M},$$

holds with probability 1. In order to prove assertion (b) of the theorem, it suffices then to show that combining (12), (17), (18) and (25) *implies* (23).

Thus let us assume that (12), (17), (18) and (25) hold for some sequence of realizations of the r.v. Under this assumption (23) can be derived in the following way.

Let $(\tilde{\theta}, \tilde{f}, \tilde{F}) \in D \times \mathscr{M} \times \mathscr{F}_D$ denote an *arbitrary* limit point of the sequence $(\tilde{\theta}_{i,m}, \tilde{f}_m)_{m \in \mathbb{N}}$ and let $(\tilde{\theta}_{i, m_l}, \tilde{f}_{m_l}, \tilde{F}_{m_l})_{l \in \mathbb{N}}$ be a subsequence converging to $(\tilde{\theta}, \tilde{f}, \tilde{F})$. Set $h := N(\tilde{f}, \tilde{F})$.

It follows from (a) that $h(S(F)) \subset D$. Thus, by definition of $\tilde{\theta}_{i, m_l}$, obviously for each $l \in \mathbb{N}$,

$$\frac{1}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(Y_{ij} - \tilde{f}_{m_l}(t_{ij}, h(\theta_i)) \big)^2 \geq \frac{1}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(Y_{ij} - \tilde{f}_{m_l}(t_{ij}, \tilde{\theta}_{i, m_l}) \big)^2.$$

Taking into account the model assumption $Y_{ij} = f(t_{ij}, \theta_i) + \varepsilon_{ij}$, this yields for all $l \in \mathbb{N}$,

$$\begin{aligned} &\frac{1}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(f(t_{ij}, \theta_i) - \tilde{f}_{m_l}(t_{ij}, h(\theta_i)) \big)^2 \\ &\quad + \frac{2}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(f(t_{ij}, \theta_i) - \tilde{f}_{m_l}(t_{ij}, h(\theta_i)) \big) \varepsilon_{ij} \\ &\geq \frac{1}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(f(t_{ij}, \theta_i) - \tilde{f}_{m_l}(t_{ij}, \tilde{\theta}_{i, m_l}) \big)^2 \\ &\quad + \frac{2}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(f(t_{ij}, \theta_i) - \tilde{f}_{m_l}(t_{ij}, \tilde{\theta}_{i, m_l}) \big) \varepsilon_{ij}. \end{aligned}$$

Due to (25), and since (12), (17) and (18) imply (15), it follows that

$$\frac{1}{T_{i,m_l}} \sum_{j=1}^{T_{i,m_l}} \big(f(t_{ij}, \theta_i) - \tilde{f}_{m_l}(t_{ij}, \tilde{\theta}_{i, m_l}) \big)^2 \to 0, \quad \text{as } l \to \infty.$$

Discarding all summations over i within the proof of step 2 of Lemma B and then applying the same arguments, it can be shown that this implies

$$\sup_{t\in J}\left(f(t,\theta_i)-\tilde{f}_{m_l}\left(t,\tilde{\theta}_{i,m_l}\right)\right)^2\to 0,\quad \text{as } l\to\infty.$$

Since $\tilde{f}_{m_l}\to\tilde{f}$ (*uniformly* on $J\times D$) and $\tilde{\theta}_{i,m_l}\to\tilde{\theta}$, this yields

$$0=\sup_{t\in J}\left(f(t,\theta_i)-\tilde{f}(t,\tilde{\theta})\right)^2=\sup_{t\in J}\left(f(t,\theta_i)-\tilde{f}\left(t,h\left(h^{-1}(\tilde{\theta})\right)\right)\right)^2. \tag{26}$$

By assumption, the SEMOR MODEL is completely identifiable by $M\supset\mathscr{M}$ and N, and thus Definition 1(ii) is fulfilled. Since, as shown above, $f=\tilde{f}(\cdot,h(\cdot))$ on $J\times S(F)$, it follows that (26) holds if and only if $\theta_i=h^{-1}(\tilde{\theta})$. This proves assertion (b).

(c) It remains to prove assertion (c) of the theorem. Obviously, it suffices to show that if for some sequence of realizations of the r.v. (12), (17) and (18) are satisfied, for each $\varepsilon>0$,

$$Q_{\varepsilon,m}:=\frac{1}{n_m}\#\left\{i\in\{1,\dots,n_m\}\,|\,\|\theta_i-\hat{\theta}_{i,r(m),m}\|_2\geq\varepsilon\right\}\to 0, \tag{27}$$

$$\text{as } m\to\infty.$$

This will be proved by contradiction.

Let us assume that for some sequence of realizations (12), (17) and (18) hold, whereas (27) is wrong for some $\varepsilon>0$. Then there exists a $\delta>0$ and a subsequence $(n_{m_l})_{l\in\mathbb{N}}$ of $(n_m)_{m\in\mathbb{N}}$ with the property that

$$Q_{\varepsilon,m_l}>\delta,\quad\text{for each } l\in\mathbb{N}. \tag{28}$$

There exists a limit point $(\tilde{f},\tilde{F})\in\mathscr{M}\times\mathscr{F}_D$ of the sequence $(\tilde{f}_{m_l},\tilde{F}_{m_l})_{l\in\mathbb{N}}$ as well as some subsequence $(\tilde{f}_{m_{l(s)}},\tilde{F}_{m_{l(s)}})_{s\in\mathbb{N}}$ converging to $(\tilde{f},\tilde{F})$. As has been shown, $\tilde{f}\models f$, $\tilde{F}\cong_{\tilde{f}}F$ and $N(\tilde{f}_{m_{l(s)}}\tilde{F}_{m_{l(s)}})\to h:=N(\tilde{f},\tilde{F})$ with the homeomorphism $h\in H(\mathscr{R})$ satisfying (20). Furthermore, as has been derived within step 2, (21) holds for *each* continuous function $v\colon\mathscr{R}\to\mathscr{R}$, and thus, in particular, for $v=h^{-1}$. Hence one obtains

$$\frac{1}{n_{m_{l(s)}}}\sum_{i=1}^{n_{m_{l(s)}}}\left\|\theta_i-h^{-1}\left(\tilde{\theta}_{i,m_{l(s)}}\right)\right\|_2\to 0,\quad\text{as } s\to\infty.$$

In the same way as in the proof of assertion (b), we can now derive (24). Since $\hat{\theta}_{i,r(m_{l(s)}),m_{l(s)}}=\tilde{h}^{-1}_{m_{l(s)}}(\tilde{\theta}_{i,m_{l(s)}})$, this yields

$$\frac{1}{n_{m_{l(s)}}}\sum_{i=1}^{n_{m_{l(s)}}}\|\theta_i-\hat{\theta}_{i,r(m_{l(s)}),m_{l(s)}}\|^2\to 0,\quad\text{as } s\to\infty,$$

implying obviously

$$Q_{\varepsilon,m_{l(s)}}\to 0,\quad\text{for } s\to\infty.$$

This contradicts (28). □

REFERENCES

DAY, N. E. (1966). Fitting curves to longitudinal data. *Biometrics* **22** 276–291.

DE BOOR, C. (1978). *A Practical Guide to Splines*. Springer, New York.

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

GEMAN, S. and HWANG, C.-R. (1982). Nonparametric maximum likelihood estimation by the method of sieves. *Ann. Statist.* **10** 401–414.

GRENANDER, U. (1981). *Abstract Inference*. Wiley, New York.

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

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

STUETZLE, W., GASSER, TH., MOLINARI, L., LARGO, R. H., PRADER, A. and HUBER, P. J. (1980). Shape-invariant modelling of human growth. *Ann. Human Biol.* **7** 507–528.

WU, C.-F. (1981). Asymptotic theory of nonlinear least squares estimation. *Ann. Statist.* **9** 501–513.

INSTITUT FÜR ANGEWANDTE MATHEMATIK
UNIVERSITÄT HEIDELBERG
IM NEUENHEIMER FELD 294
6900 HEIDELBERG
FEDERAL REPUBLIC OF GERMANY

ZENTRALINSTITUT FÜR SEELISCHE GESUNDHEIT
P.O.B. 5970
6800 MANNHEIM 1
FEDERAL REPUBLIC OF GERMANY