The Annals of Statistics
1987, Vol. 15, No. 2, 568–586

STRONG CONSISTENCY OF LEAST-SQUARES ESTIMATORS IN THE MONOTONE REGRESSION MODEL WITH STOCHASTIC REGRESSORS[1]

BY N. CHRISTOPEIT AND G. TOSSTORFF

University of Bonn

In this paper it is shown that in the monotone regression model the unknown regression function can be consistently estimated by the least-squares method.

1. Introduction. A situation frequently faced by a researcher who is analyzing real data is the following. A sequence y_t, $t = 1, 2, \dots,$ of real-valued observable outputs is to be explained by observable inputs x_t. This is usually done by approximating the y-values by functions $f(x, \theta)$, where the choice of the function class $f(\cdot, \theta)$ is based on theoretical a priori considerations and the parameter θ is to be chosen so as to minimize a certain criterion—usually the least-squares distance. If a sound theoretical basis is available, it will often be possible to reduce the problem to one in which θ is a finite-dimensional parameter. In cases where no theory at all is available that could suggest an appropriate function type, one will often resort to the use of plots in connection with ad hoc assumptions to specify such a class (cf. [13]).

A minimal and often theoretically well-motivated assumption about the relationship between two variables x and y is that of monotonicity, e.g., in many situations, rising inputs will tend to increase the output, without any specific functional relationship being more plausible than others.

For scalar valued x_t Kruskal [8] gives an algorithm that computes, for given observations $(x_1, \dots, x_T, y_1, \dots, y_T)$, values $\hat\theta_{T,1}, \dots, \hat\theta_{T,T}$ that preserve the order of the x_t in the sense that $x_i \le x_j \Rightarrow \hat\theta_{T,i} \le \hat\theta_{T,j}$ and that approximate the y_t-values by minimizing the least-squares distance

$$\frac{1}{T}\sum_{t=1}^{T} |y_t - \theta_t|^2, \tag{1.1}$$

in the class of all admissible (i.e., order preserving) θ_t (cf. [12]). This algorithm is equivalent to formula (2) in [6].

Going beyond description, the underlying statistical model would assume that the y_t are monotone transformations of the x_t, disturbed by random shocks

$$y_t = \theta_0(x_t) + \varepsilon_t, \tag{1.2}$$

where θ_0 is the "true" monotone function. Minimizing (1.1) under the restriction

Received February 1985; revised October 1986.

[1] This work was supported by Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 303.

AMS 1980 *subject classifications.* Primary 62J02; secondary 62F12.

Key words and phrases. Consistency, least-squares estimators, nonlinear regression, monotone regression.

of order preservation then gives approximate values $\hat{\theta}_{T,1},\dots,\hat{\theta}_{T,T}$ of θ_0 at the observed x_t-values. A desirable property of this procedure would be that, asymptotically, we can get complete knowledge of the function θ_0. To show that this is indeed true is the objective of this paper.

For deterministic x_t the problem may be looked upon as one of determining the unknown mean $\theta(x)$ of a family F_x, $x \in \mathscr{X}$, of distribution functions by drawing random samples $y_t \sim F_{x_t}$. In this case the variable x is merely treated as a parameter. Using an "explicit" formula for the least-squares estimate, consistency results for the determination of the location parameter $\theta(x)$ have been obtained in [6], [9] and [14]. The model (1.2) is, however, in a certain way more comprehensive, admitting predetermined stochastic regressors and hence dependence among the y_t. $\theta_0(x_t)$ is then not merely a location parameter but may contribute essential features of the stochastic properties of the output y_t. Knowledge of the regression or "response" function θ_0 may therefore be of vital importance for the choice of input level or dosage to obtain a certain desired output.

The approach taken in this paper will be along the lines of nonlinear regression theory as exposed, e.g., in [7]. To this end, we may think of the model (1.2) as a special case of the general nonlinear regression model

$$y_t = f(x_t, \theta_0) + \varepsilon_t, \tag{1.3}$$

in which $f(x,\theta)$ is a known function and the unknown parameter θ_0 is to be estimated by the least-squares method, i.e., by minimizing

$$Q_T(\theta; x_1,\dots,x_T, y_1,\dots,y_T) = \frac{1}{T}\sum_{t=1}^{T} |y_t - f(x_t,\theta)|^2, \tag{1.4}$$

over all values θ in some (topological) parameter space. Any θ minimizing (1.4) and depending in a measurable way on the observations $(x_1,\dots,x_T, y_1,\dots,y_T)$ will be called a least-squares estimator. For the special case of the monotone regression model, where $f(x,\theta) = \theta(x)$ and Θ is an appropriate function space, we shall be interested in the question whether any such sequence $\hat{\theta}_T$, $T = 1,2,\dots,$ of least-squares estimators is strongly consistent, i.e., if $\hat{\theta}_T \to \theta_0$ a.e. in the topology of the parameter space.

The paper is organized as follows. In Section 2, after formulation and a brief discussion of the assumptions, the problem is imbedded in a suitable setting with infinite-dimensional parameter space and the main result is stated. In order to obtain compactness of any sequence $\hat{\theta}_T$ of monotone least-squares estimators—a property that is crucially exploited in the proof of the main theorem—it is shown that the monotone least-squares estimators are uniformly bounded (in T) on sets J_N exhausting the support of any asymptotic distribution of the regressors. This is done in Lemma 4 and the discussion following it, and leads us to consider modified estimators $\hat{\theta}_T^N$ that are uniformly bounded (in T) on the whole regressor sample space. Lemmas 1–3 are prerequisites to establish this result (Proposition 1). In Section 3, the proof of the main theorem is given; besides on Proposition 1, it relies mainly on Proposition 2, which gives the asymptotics for terms appearing in the binomial expansion of the sum of squared errors (1.4). In both sections, some care needs to be taken concerning certain null sets, leading to the various remarks. In Section 4, a truncation procedure is suggested to cope with the effect of over- or underestimation at the endpoints of a compact interval.

2. Assumptions and preliminary results. On Euclidean d-space, we have the canonical (i.e., coordinatewise) partial ordering, which will be denoted by "$\leq$." Later on we shall also use the symbol "$<$" to mean: $x < y \Leftrightarrow x^i < y^i$ for all coordinates. A function θ: $\mathbb{R}^d \to \mathbb{R}$ will be called monotone increasing if $x \leq y$ implies $\theta(x) \leq \theta(y)$. The notations $[x, y]$ and (x, y) will be used for the multidimensional intervals $\{z\colon x \leq z \leq y\}$ and $\{z\colon x < z < y\}$, respectively. We shall assume that the x_t take finite values in some finite or infinite interval $\mathscr{X}$ of Euclidean d-space.

Without further mentioning, it will be assumed henceforth that θ_0 is a measurable monotone increasing function.

It will be convenient if we work on the canonical probability space $\Omega = \mathscr{X}^{\mathbb{N}} \times \mathbb{R}^{\mathbb{N}}$, whose elements are infinite sequences $\omega = (x_1, x_2, \dots, y_1, y_2, \dots)$, endowed with the product Borel σ-field and with the distribution measure μ. Besides, we shall use the space of finite histories $\Omega_T = \mathscr{X}^T \times \mathbb{R}^T$ with elements $\omega_T = (x_1, \dots, x_T, y_1, \dots, y_T)$, endowed with the Borel σ-field and the projection μ_T of μ. The random variables x_t, y_t and ε_t are then well defined on Ω as the projections or by (1.2).

We shall make the following assumptions:

Let $F_T(x, \omega) = (1/T)\Sigma_{t=1}^T 1_{[x_t \leq x]}$ denote the empirical distribution function of the x_t (1_A = indicator of A).

(A.1) The ε_t, $t = 1, 2, \dots$, form a martingale difference sequence with respect to some increasing family $\mathscr{F}_t$, $t = 0, 1, \dots$, of σ-algebras, i.e.,

$$E(\varepsilon_t | \mathscr{F}_{t-1}) = 0 \quad \text{a.e., } t = 1, 2, \dots.$$

For all t, x_t is measurable with respect to $\mathscr{F}_{t-1}$.

(A.2) $\sup_t E(|\varepsilon_t|^{2\alpha} | \mathscr{F}_{t-1}) < \infty$ a.e. for some $\alpha > 1$.

(A.3) For every nondegenerate interval J in $\mathscr{X}$

$$\liminf_{T \to \infty} \int_J dF_T > 0 \quad \text{a.e.}$$

Let $\Delta(\omega) = \{x \in \mathscr{X}\colon \limsup_{T \to \infty} \int_{\{x\}} dF_T(\omega) > 0\}$ (a countable set), and define distribution functions F_T^c, F_T^d by

$$\int_A dF_T^c = \int_{A \setminus \Delta} dF_T, \qquad \int_A dF_T^d = \int_{A \cap \Delta} dF_T,$$

for every Borel set A. Note that $F_T = F_T^c + F_T^d$.

(A.4) (i) $\Delta(\omega)$ is a.e. equal to some fixed countable set Δ.

(ii) If, for each ω, $(T'(\omega))$ is a subsequence of (T) such $F_{T'(\omega)}^c(\cdot, \omega)$ converges weakly to some (not necessarily probability) distribution function $F^c(\cdot, \omega)$ as

$T'(\omega) \to \infty$, then with probability one, F^c does not charge any monotone graph (cf. [3] for a definition).

REMARK 1. If we put $\sigma_t^2 = E(\varepsilon_t^2 | \mathscr{F}_{t-1})$, then (A.2) implies that

$$\sup_t \sigma_t^2 = \bar{\sigma}^2 < \infty \quad \text{a.e.}$$

Moreover, by Chow's theorem (cf. the proof of Lemma 2), $(1/T)\Sigma_{t=1}^T(\varepsilon_t^2 - \sigma_t^2) \to 0$ a.e. and hence

$$\limsup_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \le \bar{\sigma}^2 \quad \text{a.e.}$$

A more precise statement will be made in Lemma 2 below.

REMARK 2. We shall assume that we are given fixed versions of the conditional expectations in (A.2). Then the set (of μ-measure 1) of those ω for which the a.e.-statements in (A.2)–(A.4) do actually hold will depend on the chosen interval J [in (A.3)] and the collection $(T'(\omega))$ of subsequences [in (A.4)]. Note, however, that it suffices to require (A.3) for all intervals J with rational endpoints. Hence, let $\tilde{\Omega}_0$ denote a set of μ-measure one on which (A.2) (together with its two consequences in Remark 1), (A.3) and (A.4)(i) hold, and let Ω_0' denote a subset of $\tilde{\Omega}_0$ of full measure on which (A.4)(ii) is true for a given collection $(T'(\omega))$. Note that, by Helly's theorem, every subsequence $(F_{T'(\omega)}(\cdot, \omega))$ contains a further subsequence $(F_{T''(\omega)}(\cdot, \omega))$ that converges weakly to some (right-continuous, not necessarily probability) distribution function. Hence the chief requirement of (A.4)(ii) is that no new point charges are created in $\mathscr{X} \setminus \Delta$ asymptotically. Moreover, as a consequence of (A.3), every limit distribution function $F(\cdot, \omega)$ assigns positive measure to every nondegenerate interval in $\mathscr{X}$.

A field of application for which (A.1)–(A.4) provide a natural setting is the determination of the regression function of a bivariate distribution, i.e., let (x_t, y_t), $t = 1, 2, \ldots$, be independent observations distributed as (X, Y) and suppose it is known that the regression function $\theta_0(x) = E(Y|X = x)$ is monotone increasing in x. Then the experiment may be written in the form (1.2) with $\varepsilon_t = y_t - E(y_t|x_t)$, and the ε_t form a martingale difference sequence with respect to the σ-algebras $\mathscr{F}_t = \sigma\{x_{t+1}, x_t, y_t, \ldots, x_1, y_1\}$. The other assumptions in (A.1)–(A.4) amount to certain regularity conditions to be imposed on the distribution of (X, Y) (see also Remark 3 below).

REMARK 3. (A.3) and (A.4) are automatically satisfied for i.i.d. x_t whose common distribution charges every nondegenerate interval and has a continuous part not charging any monotone graph [(A.4)(i) is then a consequence of the Glivenko–Cantelli theorem]. More generally, let x_t be *stationary ergodic* with common distribution function $F = F^c + F^d$ having an absolutely continuous part F^c. Then the Glivenko–Cantelli theorem applies to yield Δ = support of F^d. Moreover, by ergodicity, $F_T^c \to F^c$.

In order to put the monotone regression model (1.2) into the framework of the general nonlinear regression model (1.3), let

$$f(x, \theta) = \theta(x),$$

where θ ranges in the class $\mathscr{M}$ of all measurable monotone increasing functions on $\mathscr{X}$. The kind of consistency we shall be interested in is of the following type: A sequence $\hat{\theta}_T$, $T = 1, 2, \ldots,$ of least-squares estimators of θ_0 will be called (strongly) *consistent* on $A \subset \mathscr{X}$ if, with probability one,

(2.1) $\quad \hat{\theta}_T(x) \to \theta_0(x) \quad$ as $T \to \infty$ for all continuity points $x \in A$ of θ_0.

The corresponding topology to be chosen on $\mathscr{M}$ will be the topology of pointwise convergence on a countable dense subset of $\mathscr{X}$ (cf. Section 3). Let $\hat{\theta}_{T,t}(\omega_T)$, $t = 1, \ldots, T$, denote the unique order preserving values minimizing (1.1) corresponding to the observations $\omega_T = (x_1, \ldots, x_T, y_1, \ldots, y_T)$ (cf. Section 1). A monotone least-squares estimator (based on the first T observations) is then any measurable (with respect to the Borel σ-field on Θ) function $\omega_T \to \hat{\theta}_T(\cdot, \omega_T) \in \mathscr{M}$ such that $\hat{\theta}_T(x_t, \omega_T) = \hat{\theta}_{T,t}(\omega_T)$, $t = 1, \ldots, T$. A natural candidate is the following. Define

$$\text{(2.2)} \quad \hat{\theta}_T(x, \omega_T) = \begin{cases} \max\{\hat{\theta}_{T,t}\colon x_t \le x\}, & \text{if } x_t \le x \text{ for some } t = 1, \ldots, T, \\ \min\{\hat{\theta}_{T,t}, t = 1, \ldots, T\}, & \text{else.} \end{cases}$$

A different canonical choice is given in [6]. Also, for one-dimensional x_t, one might prefer to interpolate the $\hat{\theta}_{T,t}$-values linearly.

The vector $(\hat{\theta}_{T,1}, \ldots, \hat{\theta}_{T,T})$ is the projection of $(y_1, \ldots, y_T)$ on a convex cone that changes with the $(x_1, \ldots, x_T)$ in such a way that the mapping $\omega_T \to \hat{\theta}_{T,t}(\omega_T)$ becomes measurable. This property carries over to the mapping $\omega_T \to \hat{\theta}_T(\omega_T)$ if we endow $\mathscr{M}$ with the Borel σ-field corresponding to the topology of pointwise convergence on a countable dense subset, so that $\hat{\theta}_T$ as defined by (2.2) is indeed a least-squares estimator as defined in the Introduction.

Our purpose is to prove the following result:

THEOREM 1. *Under* (A.1)–(A.4), *every sequence* $\hat{\theta}_T$, $T = 1, 2, \ldots,$ *of monotone least-squares estimators is strongly consistent on* $\operatorname{int}(\mathscr{X})$.

The proof will be given in Section 3.

As pointed out in the Introduction, our approach is along the lines of asymptotic theory for the general nonlinear regression model. A salient feature of this approach is the need for a certain compactness property of the parameter space (cf. the proof of Theorem 1). The space $\mathscr{M}$ does not have this property. What we need is a uniform bound for the true function θ_0 *and* the $\hat{\theta}_T$. We shall therefore first show that the problem may be reduced to one in which all monotone functions involved are uniformly bounded.

To this end denote

$$\mathscr{G}_T(\omega_T) = \Big\{\hat{\theta} \in \mathscr{M}\colon Q_T(\hat{\theta}; \omega_T) = \inf_{\theta \in \mathscr{M}} Q_T(\theta; \omega_T),$$
$$\underline{\hat{\theta}}_T(\omega_T) \le \hat{\theta}(x) \le \bar{\hat{\theta}}_T(\omega_T) \text{ for all } x \in \mathscr{X}\Big\},$$

where we have put $\underset{\smile}{\hat{\theta}}_T = \min\{\hat{\theta}_{T,t},\ t = 1,\dots,T\}$ and $\bar{\hat{\theta}}_T = \max\{\hat{\theta}_{T,t},\ t = 1,\dots,T\}$. Obviously, $\hat{\theta}_T(\omega_T)$ as defined by (2.2) is in $\mathscr{G}_T(\omega_T)$. Let Δ_N be finite subsets of Δ such that $\Delta_N \nearrow \Delta$, and let $\tilde{J}_N$ be finite open intervals (with rational endpoints) such that $\mathrm{cl}(\tilde{J}_N) \subset \mathrm{int}(\mathscr{X})$, $\tilde{J}_N \nearrow \mathrm{int}(\mathscr{X})$. Put $J_N = \tilde{J}_N \cup \Delta_N$.

LEMMA 1. *Under* (A.4) (*letting* $\bar{A}$ *denote complement*),

$$\lim_{N\to\infty} \liminf_{T\to\infty} \int_{\bar{J}_N} dF_T = 0 \quad a.e.$$

PROOF. For fixed $\omega \in \tilde{\Omega}_0$, let $(T'(\omega))$ be a subsequence such that (omitting ω in the sequel) $F^c_{T'} \to F^c$, $\int_{\{x_i\}} dF^d_{T'} \searrow a_i$ or $\nearrow a_i$ for all $x_i \in \Delta$. Then it is easily verified that $\Sigma a_i \le 1$. Anticipating some results explained in more detail below (cf. Sections 3(a) and (b) and the following discussion), and defining F^d as in (3.6), it can be shown that on Ω'_0

$$\int_{\bar{J}_N} dF^c_{T'} \to \int_{\bar{J}_N} dF^c \quad \text{and} \quad \int_{\bar{J}_N} dF^d_{T'} \to \int_{\bar{J}_N} dF^d,$$

as $T' \to \infty$ [cf. (3.2) and (3.7)]. But $\int_{\bar{J}_N} dF^c \searrow 0$ on Ω'_0, and since $\bar{J}_N \cap \Delta \subset \Delta \setminus \Delta_N$, $\int_{\bar{J}_N} dF^d \searrow 0$ as $N \to \infty$. $\square$

If $\mathscr{X}$ is open, we might have simply taken $J_N = \tilde{J}_N$ (since then $\bar{\tilde{J}}_N \cap \Delta \searrow \varnothing$).

LEMMA 2. *Suppose that, for every* ω, $(T'(\omega))$ *is a subsequence such that* $(1/T'(\omega))\Sigma_{t=1}^{T'(\omega)}\sigma_t^2(\omega) \to \sigma^2(\omega) < \infty$ *a.e. Then, under* (A.1)–(A.2) *and* (A.4),

$$\lim_{T'\to\infty} \frac{1}{T'} \sum_{t=1}^{T'} \varepsilon_t^2 = \sigma^2 \quad a.e.$$

and

$$\lim_{N\to\infty} \limsup_{T\to\infty} \frac{1}{T'} \sum_{x_t\in J_N} \varepsilon_t^2 = \sigma^2 \quad a.e.$$

(*Here we use* $(1/T')\Sigma_{x_t\in J}$ *as a shorthand notation for* $(1/T')\Sigma_{t=1,\ x_t\in J}^{T'}$.)

PROOF. Note first that, by Remark 1, σ^2 is a.e. finite. Then, for each interval J, $\xi_t = 1_{[x_t\in J]}(\varepsilon_t^2 - \sigma_t^2)$ defines a martingale difference sequence such that $\Sigma_{t=1}^{\infty} E(|\xi_t|^{\alpha}|\mathscr{F}_{t-1})/t^{\alpha} < \infty$ a.e. [with $\alpha > 1$ from (A.2)]. Hence, by Chow's theorem (cf. [11], Theorem 3.3.1)

$$\frac{1}{T}\sum_{t=1}^{T} 1_{[x_t\in J]}\left(\varepsilon_t^2 - \sigma_t^2\right) \to 0 \quad \text{a.e.} \tag{2.3}$$

This shows the first assertion and that

$$s_N = \limsup_{T'\to\infty} \frac{1}{T'} \sum_{x_t\in J_N} \sigma_t^2 = \limsup_{T'\to\infty} \frac{1}{T'} \sum_{x_t\in J_N} \varepsilon_t^2 \quad \text{a.e.}$$

But

$$\liminf_{T'\to\infty}\left(\frac{1}{T'}\sum_{t=1}^{T'}\sigma_t^2-\frac{1}{T'}\sum_{x_t\in J_N}\sigma_t^2\right)=\liminf_{T'\to\infty}\frac{1}{T'}\sum_{x_t\notin J_N}\sigma_t^2$$

$$\leq\bar{\sigma}^2\liminf_{T'\to\infty}\int_{\bar{J}_N}dF_T=:\bar{\sigma}^2\alpha_N,$$

whence

$$\sigma^2-s_N\leq\bar{\sigma}^2\alpha_N.$$

Since $s_N\leq\sigma^2$ and $\alpha_N\searrow 0$ (by Lemma 1), this implies

$$\lim_{N\to\infty}s_N=\sigma^2\quad\text{a.e.}\qquad\square$$

REMARK 4. Again, the exceptional ω-set will depend on the collection $(T'(\omega))$. Let Ω_1' denote an ω-set of full measure on which the assertions of Lemma 2 hold for a given collection $(T'(\omega))$.

LEMMA 3. *Let* $G_T(x,e,\omega)=(1/T)\sum_{t=1}^T 1_{[x_t(\omega)\leq x,\ \varepsilon_t(\omega)\leq e]}$ *denote the joint empirical distribution function of the* (x_t,ε_t). *Then, under* (A.1)–(A.3), *with probability one,*

$$\lim_{N\to\infty}\limsup_{T\to\infty}\int_{[x,y]\times(-\infty,N]}dG_T>0$$

for all $x<y$ *with rational endpoints. Moreover, the above conclusion is valid for all* $x\in\Delta$ *with* $[x,y]$ *replaced by* $\{x\}$.

PROOF. For fixed N, let $\alpha_t=E(1_{[\varepsilon_t\leq N]}|\mathscr{F}_{t-1})$, $\sigma_t^2=E(\varepsilon_t^2|\mathscr{F}_{t-1})$. Then

$$\int_{[x,y]\times(-\infty,N]}dG_T=\frac{1}{T}\sum_{t=1}^T 1_{[x_t\in[x,y]]}\left(1_{[\varepsilon_t\leq N]}-\alpha_t\right)+\frac{1}{T}\sum_{t=1}^T 1_{[x_t\in[x,y]]}$$

$$-\frac{1}{T}\sum_{t=1}^T 1_{[x_t\in[x,y]]}E\left(1_{[\varepsilon_t>N]}|\mathscr{F}_{t-1}\right).$$

Since $\xi_t=1_{[x_t\in[x,y]]}(1_{[\varepsilon_t\leq N]}-\alpha_t)$ defines a martingale difference sequence [with respect to $(\mathscr{F}_t)$] such that $\sum_{t=1}^\infty E(\xi_t^2|\mathscr{F}_{t-1})/t^2<\infty$ a.e., it follows from Chow's theorem that

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^T\xi_t=0\quad\text{a.e.}$$

The third term on the right-hand side may be estimated (in absolute value) by

$$\frac{1}{T}\sum_{t=1}^T E\left(1_{[\varepsilon_t>N]}|\mathscr{F}_{t-1}\right)\leq\frac{1}{N^2}\frac{1}{T}\sum_{t=1}^T\sigma_t^2\leq\frac{\bar{\sigma}^2}{N^2},$$

by virtue of the (conditional) Chebyshev inequality. Consequently,

$$\limsup_{T\to\infty}\int_{[x,y]\times(-\infty,N]} dG_T \ge \limsup_{T\to\infty}\int_{[x,y]} dF_T - \frac{\bar{\sigma}^2}{N^2},$$

from which the assertion follows in view of assumption (A.3) and Remark 1. □

Note that, for $x < y$, the assertion holds even with lim sup replaced by lim inf.

REMARK 5. The ω-set on which Lemma 3 is valid is some fixed subset of $\tilde{\Omega}_0$ (cf. Remark 2) and is determined by the choice of certain conditional expectations. Call it Ω_0 in the sequel.

Let now $\operatorname{int}(\mathscr{X}) = (a, b)$, the cases $a = -\infty$ and $b = \infty$ being admitted [as a shorthand notation for $\mathsf{X}_1^d(-\infty, b_i)$, etc.]. For $x < y$, let

$$I_T(\omega, x, y) = \{t\colon 1 \le t \le T,\ x_t \in (x, y)\}$$

($x = -\infty$ and $y = \infty$ admitted). Define

$$l_T^+(\omega, x) = \begin{cases} \min\limits_{t\in I_T(\omega,x,b)} \hat{\theta}_{T,t}(\omega_T), & \text{if } I_T(\omega,x,b) \ne \varnothing,\\ \bar{\hat{\theta}}_T(\omega_T), & \text{if } I_T(\omega,x,b) = \varnothing \end{cases}$$

and

$$l_T^-(\omega, x) = \begin{cases} \max\limits_{t\in I_T(\omega,a,x)} \hat{\theta}_{T,t}(\omega_T), & \text{if } I_T(\omega,a,x) \ne \varnothing,\\ \underline{\hat{\theta}}_T(\omega_T), & \text{if } I_T(\omega,a,x) = \varnothing. \end{cases}$$

LEMMA 4. *Under* (A.1)–(A.3), *for all* $\omega \in \Omega_0$,

(2.4)(i) $$L^+(\omega, x) := \sup_T l_T^+(\omega, x) < \infty$$

and

(2.4)(ii) $$L^-(\omega, x) := \inf_T l_T^-(\omega, x) > -\infty$$

for all rational $x \in \operatorname{int}(\mathscr{X})$.

PROOF. Note first that, by virtue of (A.3), $I_T(\omega, a, x) \ne \varnothing$ and $I_T(\omega, x, b) \ne \varnothing$ for T large enough (depending on ω). Suppose that for fixed $\omega \in \Omega_0$ there is a subsequence (T') such that

$$\min_{t\in I_{T'}(\omega,x,b)} \hat{\theta}_{T',t}(\omega_{T'}) \nearrow \infty.$$

For notational simplicity, let us assume that $(T') = (T)$ and omit ω in the sequel. Choose $x < y$, $y \in \operatorname{int}(\mathscr{X})$ and rational, N so large that $\sup_{x\le z\le y}\theta_0(z) \le N$ and T_N so large that $I_T(x, y) \ne \varnothing$ and

$$\hat{\theta}_{T,t} \ge 3N, \quad \text{for all } T \ge T_N,\ t \in I_T(x, y).$$

Then, for every $\theta \in \mathscr{M}$,

$$
\begin{aligned}
Q_T(\theta) &\geq \frac{1}{T} \sum_{t \in I_T(x,y)} |y_t - \hat{\theta}_{T,t}|^2 \\
&\geq \frac{1}{T} \sum_{\substack{t \in I_T(x,y) \\ y_t \leq 2N}} N^2 \geq \frac{1}{T} \sum_{\substack{t \in I_T(x,y) \\ \varepsilon_t \leq N}} N^2 \\
&= N^2 \int_{(x,y) \times (-\infty, N]} dG_T,
\end{aligned}
$$

for all $T \geq T_N$. Hence, by virtue of Lemma 3,

$$\limsup_{T \to \infty} Q_T(\theta) = \infty.$$

In particular, if we choose $\theta = \theta_0$, then

$$\limsup_{T \to \infty} Q_T(\theta_0) = \limsup_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \varepsilon_t^2 \leq \bar{\sigma}^2 < \infty.$$

Hence, by contradiction, (2.4)(i) must hold on Ω_0. (2.4)(ii) is proved in a similar way. □

Note that the same argument shows that

$$\text{(2.5)(i)} \qquad \lambda^+(x) = \sup_T \min_{x_t = x} \hat{\theta}_{T,t} = \sup_T \hat{\theta}_{T,t}(x) < \infty$$

and

$$\text{(2.5)(ii)} \qquad \lambda^-(x) = \inf_T \max_{x_t = x} \hat{\theta}_{T,t} = \inf_T \hat{\theta}_{T,t}(x) > -\infty,$$

on Ω_0 for all $x \in \Delta$.

The following inequalities follow easily from the definitions. For $\underline{y} < x < \bar{y}$,

$$
\text{(2.6)} \qquad
\begin{aligned}
L^-(\underline{y}) \leq L^-(x) \leq \inf_{T, \hat{\theta} \in \mathscr{G}_T} \hat{\theta}(x) \\
\leq \sup_{T, \hat{\theta} \in \mathscr{G}_T} \hat{\theta}(x) \leq L^+(x) \leq L^+(\bar{y}).
\end{aligned}
$$

Let us remind ourselves that our aim is to show (2.1) for every sequence $\hat{\theta}_T$, $T = 1, 2, \ldots$, of least-squares estimators in $\mathscr{M}$. But corresponding to any such sequence there exists a sequence $\tilde{\theta}_T \in \mathscr{G}_T$, $T = 1, 2, \ldots$, which is asymptotically equivalent in the sense that, with probability one,

$$\left| \hat{\theta}_T(x) - \tilde{\theta}_T(x) \right| \to 0, \quad \text{for all } x$$

(simply put $\tilde{\theta}_T(x) = \underline{\hat{\theta}}_T \vee \hat{\theta}_T(x) \wedge \bar{\hat{\theta}}_T$). Hence, in order to show consistency of any sequence of the least-squares estimators, *we may confine ourselves to sequences* $\hat{\theta}_T \in \mathscr{G}_T$, $T = 1, 2, \ldots$.

Let $\tilde{J}_N = (\underline{y}_N, \bar{y}_N)$, Δ_N and J_N be defined as above, and put

$$L_N^- = L^-(\underline{y}_N) \wedge \min_{x \in \Delta_N} \lambda^-(x), \qquad L_N^+ = L^+(\bar{y}_N) \vee \max_{x \in \Delta_N} \lambda^+(x).$$

Then, by virtue of Lemma 4 and (2.5), (2.6) for every T, $\hat{\theta}_T \in \mathscr{G}_T$, and all $x \in J_N$

$$-\infty < L_N^- \le \hat{\theta}_T(x) \le L_N^+ < \infty,$$

on Ω_0. For fixed $\omega \in \Omega_0$, put $M_N = \sup_{x \in J_N} |\theta_0(x)|$ and $\alpha_N(\omega) = |L_N^-| \vee |L_N^+| \vee M_N$. Then, for every T and every $\hat{\theta}_T \in \mathscr{G}_T(\omega_T)$,

$$-\infty < -\alpha_N(\omega) \le \hat{\theta}_T(x) \le \alpha_N(\omega) < \infty, \quad \text{for all } x \in J_N \tag{2.7}$$

and

$$-\infty < -\alpha_N(\omega) \le \theta_0(x) \le \alpha_N(\omega) < \infty, \quad \text{for all } x \in J_N. \tag{2.8}$$

If $\hat{\theta}_T$, $T = 1, 2, \ldots,$ is a sequence of least-squares estimators such that $\hat{\theta}_T \in \mathscr{G}_T$, let us put

$$\hat{\theta}_T^N(\omega)(x) = (-\alpha_N(\omega)) \vee \hat{\theta}_T(\omega_T)(x) \wedge \alpha_N(\omega) \tag{2.9}$$

and

$$\theta_0^N(x) = (-\alpha_N(\omega)) \vee \theta_0(x) \wedge \alpha_N(\omega). \tag{2.10}$$

Then, for each N and $\omega \in \Omega_0$, the $\hat{\theta}_T^N(\omega)$, $T = 1, 2, \ldots,$ and θ_0^N are uniformly bounded [by $\alpha_N(\omega)$] monotone increasing functions and, by virtue of (2.7) and (2.8)

$$\hat{\theta}_T^N(\omega)(x) = \hat{\theta}_T(\omega_T)(x) \quad \text{and} \quad \theta_0^N(x) = \theta_0(x), \tag{2.11}$$

for all $x \in J_N$.

Hence we have proved

PROPOSITION 1. *Under* (A.1)–(A.4), *any sequence* $\hat{\theta}_T$, $T = 1, 2, \ldots,$ *of monotone least-squares estimators in* $\mathscr{G}_T$ *coincides on* J_N *for each* N *and* $\omega \in \Omega_0$ *with sequences* $\hat{\theta}_T^N$, $T = 1, 2, \ldots,$ *of uniformly* (*in* T) *bounded monotone increasing functions.*

3. Proof of strong consistency. The considerations at the end of the previous section show that for fixed ω and on finite intervals in $\mathscr{X}$ we may work with the truncated least-squares estimates, which are uniformly bounded on the whole of $\mathscr{X}$. This suggests as an intermediate parameter set the class $\mathscr{M}_I$ of measurable increasing functions taking values in some finite interval I. In accordance with (2.1), the topology to be chosen on $\mathscr{M}_I$ should be such that it leads to convergence at all continuity points of the limit function. To achieve this we shall endow $\mathscr{M}_I$ with the topology of pointwise convergence on a *countable dense set D containing* Δ; i.e., a neighborhood base of a point $\tilde{\theta} \in \mathscr{M}_I$ is given by the sets

$$\mathscr{N} = \left\{\theta \in \mathscr{M}_I : \left|\theta(x_i) - \tilde{\theta}(x_i)\right| < \varepsilon,\ x_i \in D,\ i = 1, \ldots, k\right\}, \tag{3.1}$$

with $\varepsilon > 0$ and $k \in \mathbb{N}$. Call the resulting space Θ_I (or simply Θ, if the interval I is of no concern). The next lemma shows that (3.1) is indeed an appropriate topology in the sense required above. In fact, convergence holds even in a somewhat stronger sense than just pointwise at the continuity points of the limit function.

LEMMA 5. *Let $\theta_T \to \theta$ in Θ. Then for every continuity point x of θ and every sequence $x_n \to x$*

$$\theta_T(x_n) \to \theta(x) \quad \text{as } T \to \infty,\ n \to \infty.$$

PROOF. This is a standard result on monotone functions. □

LEMMA 6. *Θ is sequentially compact.*

PROOF. By constructing a diagonal sequence, as in the proof of the first Helly theorem. □

For the results to follow, recall some facts about the weak convergence of distribution functions. Suppose that the H_n are (not necessarily probability) distribution functions on Euclidean d-space converging weakly to some limit distribution function H, i.e., $H_n(x) \to H(x)$ for all continuity points x of H. Then the following assertions are true.

(a) For every bounded continuous function h and every finite H-continuity interval J

$$\int_J h(x)\, dH_n(x) \to \int_J h(x)\, dH(x) \quad \text{as } n \to \infty. \tag{3.2}$$

Hence, in particular, denoting by μ_n and μ the measures induced by H_n and H, resp., $\mu_n(J) \to \mu(J)$, for all such intervals, so that on J we have complete convergence.

(b) Let h_n, h be measurable functions such that $h_n(x_n) \to h(x)$ whenever $x_n \to x$ and $x \notin E$. Then, if $\mu(J \cap E) = 0$, for every bounded continuous function f

$$\int_J f(h_n(x))\, dH_n(x) \to \int_J f(h(x))\, dH(x) \quad \text{as } n \to \infty.$$

This follows from Theorem 5.5. in [1] applied to the measures $\mu_n|_J$ and $\mu|_J$ on the metric space J. In particular, if the h_n and h are bounded by some constant M, choosing $f(t)$ as in the proof of Theorem 5.2. in [1] will lead to

$$\int_J h_n(x)\, dH_n \to \int_J h(x)\, dH(x). \tag{3.3}$$

Let us return to the situation in assumption (A.4). If $F_{T'} \to F$ weakly, then we may extract (by the usual diagonal procedure) a further subsequence (T'') such that $F^c_{T''} \to F^c$ weakly and either

$$a_{iT''} = \int_{\{x_i\}} dF_{T''} \nearrow a_i \text{ or } \searrow a_i, \quad \text{as } T'' \to \infty \text{ for all } i = 1, 2, \dots, \tag{3.4}$$

if we let $\Delta = \{x_1, x_2, \dots\}$ for a moment. Since $\int_\Delta dF_{T''} \le 1$, it easily follows that $\Sigma a_i \le 1$. From Lemma 8 below it then follows that [writing $(n) = (T'')$ for simplicity]

$$\sum_{i=1}^\infty a_{in} \to \sum_{i=1}^\infty a_i \quad \text{as } n \to \infty,$$

and, as in the proof of Scheffé's theorem (cf. [1]) we may conclude further that even

$$\sum_{i=1}^{\infty} |a_{in} - a_i| \to 0 \quad \text{as } n \to \infty \tag{3.5}$$

holds. Define now a distribution function F^d by setting

$$F^d(x) = \sum_{\substack{x_i \le x \\ x_i \in \Delta}} a_i. \tag{3.6}$$

Then, with F_T^d defined as in Section 2, if the $h_n(x_i)$ are uniformly bounded in n, i and $h_n(x_i) \to h(x_i)$ for all $x_i \in \Delta$ as $n \to \infty$, it follows from (3.5) by some straightforward estimates that

$$\int h_n(x)\, dF_n^d \to \int h(x)\, dF^d \quad \text{as } n \to \infty. \tag{3.7}$$

Note that $F = F^c + F^d$.

LEMMA 7. *Let, for each ω, $(T'(\omega))$ be a subsequence of (T) such that*

—$F_{T'(\omega)}(\cdot, \omega) \to F(\cdot, \omega)$ *weakly;*
—$F_{T'(\omega)}^c(\cdot, \omega) \to F^c(\cdot, \omega)$ *weakly;*
—$\int_{\{x\}} dF_{T'}(\cdot, \omega)$ *converges in a monotone way* [*cf.* (3.4)] *for all* $x \in \Delta$.

Then, if $\theta_S \to \theta$ in Θ, $\tilde{\theta} \in \Theta$, and the J_N are defined as in Section 2,

$$\begin{aligned}&\lim_{S, T'(\omega) \to \infty} \frac{1}{T'(\omega)} \sum_{x_t \in J_N} \left|\theta_S(x_t) - \tilde{\theta}(x_t)\right|^2 \\ &\quad = \int_{J_N} \left|\theta(x) - \tilde{\theta}(x)\right|^2 F(dx, \omega)\end{aligned} \tag{3.8}$$

holds for all $N = 1, 2, \ldots,$ and all $\omega \in \Omega_0'$.

PROOF. Let the double sequence $(S, T'(\omega))$ be ordered into a linear sequence (n) by enumeration. Put $h_n(x) = |\theta_S(x) - \tilde{\theta}(x)|^2$, $h(x) = |\theta(x) - \tilde{\theta}(x)|^2$, $H_n = F_{T'}^c(\cdot, \omega)$, $H = F^c(\cdot, \omega)$. Then the left-hand side of (3.8) is

$$\int_{J_N} h_n(x)\, dF_n = \int_{J_N} h_n(x)\, dF_n^c + \int_{J_N} h_n(x)\, dF_n^d.$$

Since $h_n(x) 1_{J_N} \to h(x) 1_{J_N}(x)$ boundedly for all $x \in \Delta$, the last term converges to $\int_{J_N} h(x)\, dF^d$ by virtue of (3.7). As to the first term on the right-hand side, note that $h_n(x_n) \to h(x)$ for all continuity points x of h (cf. Lemma 5). Hence the set E in (b) above is contained in the set of discontinuity points of h, which is a countable union of monotone graphs (cf. [3]) and thus has F^c-measure 0 by (A.4). So (3.3) applies. □

LEMMA 8. *Let, for each $n = 1, 2, \ldots,$ $i = 1, 2, \ldots,$ a_{in} and a_i be nonnegative real numbers such that $\sum_i a_{in} < \infty$ for all n and $\sum_i a_i = a < \infty$. Then*

$a_{in} \nearrow a_i$ ($a_{in} \searrow a_i$) as $n \to \infty$ for all i implies that

$$\sum_i a_{in} \nearrow a, \qquad \sum_i a_{in} \searrow a.$$

PROOF. This is a consequence of the monotone convergence theorem using counting measure on $i = 1, 2, \ldots$. □

LEMMA 9. *Let $(T'(\omega))$ be as in Lemma 7. Then, under* (A.1)–(A.2) *and* (A.4), *there exists a set Ω' of μ-measure 1 such that, whenever $\theta_S \to \theta$ in Θ, for all $N = 1, 2, \ldots$,*

$$\lim_{S, T'(\omega) \to \infty} \frac{1}{T'(\omega)} \sum_{x_t \in J_N} \theta_S(x_t) \varepsilon_t(\omega) = 0,$$

for all $\omega \in \Omega'$.

PROOF. Let $\theta^{(i)}$, $i = 1, 2, \ldots$, be a countable dense set in Θ. Fix N. For each i, $\xi_t^{(i)} = \theta^{(i)}(x_t) 1_{[x_t \in J_N]} \varepsilon_t$ defines a martingale difference sequence such that $\sum_{t=1}^{\infty} E(|\xi_t^{(i)}|^2 | \mathscr{F}_{t-1})/t^2 < \infty$ a.e. (since $\theta^{(i)}$ is bounded), hence by Chow's theorem

$$\frac{1}{T} \sum_{t=1}^{T} \xi_t^{(i)} = \frac{1}{T} \sum_{x_t \in J_N} \theta^{(i)}(x_t) \varepsilon_t \to 0 \quad \text{a.e.,} \tag{3.9}$$

and we can find a set $\tilde{\Omega}$ of full measure such that (3.9) holds for all $\omega \in \tilde{\Omega}$, all $i = 1, 2, \ldots$, and all N.

If $\theta \in \Theta$ is arbitrary, then $\theta^{(i')} \to \theta$ for some subsequence (i') and

$$\begin{aligned}\left| \frac{1}{T'(\omega)} \sum_{x_t \in J_N} \theta(x_t) \varepsilon_t(\omega) \right| &\leq \left| \frac{1}{T'(\omega)} \sum_{x_t \in J_N} \theta^{(i')}(x_t) \varepsilon_t(\omega) \right| \\ &\quad + \left(\frac{1}{T'(\omega)} \sum_{x_t \in J_N} \left| \theta^{(i')}(x_t) - \theta(x_t) \right|^2 \right)^{1/2} \\ &\quad \times \left(\frac{1}{T'(\omega)} \sum_{t=1}^{T'(\omega)} \varepsilon_t^2(\omega) \right)^{1/2}.\end{aligned}$$

For $\omega \in \Omega_0'$ and i' and $T'(\omega)$ large enough, say $\geq n(\varepsilon)$, the second term of the right-hand side is smaller than ε by virtue of Lemma 7, and, for fixed $i' \geq n(\varepsilon)$, the first term can be made arbitrarily small for $\omega \in \tilde{\Omega}$ by virtue of (3.9). Hence,

$$\frac{1}{T'(\omega)} \sum_{x_t \in J_N} \theta(x_t) \varepsilon_t(\omega) \to 0,$$

for all $\omega \in \Omega' := \tilde{\Omega} \cap \Omega_0'$. The assertion follows now from the estimate

$$\begin{aligned}\left| \frac{1}{T'} \sum_{x_t \in J_N} \theta_S(x_t) \varepsilon_t \right| &\leq \left(\frac{1}{T'} \sum_{x_t \in J_N} |\theta_S(x_t) - \theta(x_t)|^2 \right)^{1/2} \\ &\quad \times \left(\frac{1}{T'} \sum_{t=1}^{T'} \varepsilon_t^2 \right)^{1/2} + \left| \frac{1}{T'} \sum_{x_t \in J_N} \theta(x_t) \varepsilon_t \right|,\end{aligned}$$

applying Lemma 7 again. □

REMARK 6. Note that Lemmas 7 and 9 are valid for *random* ($\mathscr{F}_{t-1}$-measurable) x_t and any *nonrandom* sequence $\theta_S \to \theta$, with the set Ω' on which the assertions of the lemmas hold true being *independent of the particular sequence* (θ_S). The basic content is that convergence holds uniformly in the parameter θ. Moreover, since $\Theta = \Theta_I \subset \Theta_{[-N, N]}$ for some N, Ω' may also be taken independent of I.

Fixing $\omega \in \Omega'$ and letting $(S) = (T'(\omega))$, $\theta_S(x) = \theta_{T'(\omega)}(x, \omega)$, we obtain

PROPOSITION 2. *Let $(T'(\omega))$ be as in Lemmas 7 and 9. If $\theta_{T'(\omega)}(\omega) \to \theta(\omega)$ in Θ for all $\omega \in \Omega'$ and $\tilde{\theta}(\omega) \in \Theta$, then*

$$\lim_{T'(\omega)\to\infty} \frac{1}{T'(\omega)} \sum_{x_t \in J_N} \left|\theta_{T'(\omega)}(x_t, \omega) - \tilde{\theta}(x_t, \omega)\right|^2$$

$$= \int_{J_N} \left|\theta(x, \omega) - \tilde{\theta}(x, \omega)\right|^2 F(dx, \omega)$$

and

$$\lim_{T'(\omega)\to\infty} \frac{1}{T'(\omega)} \sum_{x_t \in J_N} \theta_{T'(\omega)}(x_t, \omega)\varepsilon_t(\omega) = 0,$$

for all $\omega \in \Omega'$ and all N.

PROOF OF THEOREM 1. Let $(T'(\omega))$ be any subsequence of (T) satisfying the assumptions of Lemma 7 and $(1/T'(\omega))\sum_{t=1}^{T'(\omega)}\sigma_t^2(\omega) \to \sigma^2(\omega)$. Let Ω_1', Ω_0 and Ω' be defined as in Remarks 4 and 5 and Lemma 9 and let, for $\omega \in \Omega_0 \cap \Omega' \cap \Omega_1'$, $(T'(\omega)) = (T)$ for notational simplicity. Let $\hat{\theta}_T^N(\omega)$ and θ_0^N be defined as in Section 2 [cf. (2.9) and (2.10)]. For each N, the $\hat{\theta}_T^N(\omega)$, $T = 1, 2, \ldots,$ and θ_0^N are then in $\mathscr{M}_{I_N(\omega)}$ with $I_N(\omega) = [-\alpha_N(\omega), \alpha_N(\omega)]$. Let $T_i^N(\omega)$, $i = 1, 2, \ldots,$ be further subsequences such that $(T_i^{N'}(\omega)) \subset (T_i^N(\omega))$ for $N' \geq N$ and $\hat{\theta}_{T_i^N}^N(\omega) \to \hat{\theta}^N(\omega)$ in $\Theta = \Theta_{I_N(\omega)}$. Existence of such sequences is guaranteed by Lemma 6. Put

$$Q_T^N(\theta) = \frac{1}{T} \sum_{x_t \in J_N} \left|\theta_0^N(x_t) - \theta(x_t) + \varepsilon_t\right|^2,$$

$$Q^N(\theta) = \int_{J_N} \left|\theta_0^N(x) - \theta(x)\right|^2 dF(x).$$

It follows from Lemma 2 and Proposition 2 that, for fixed N,

$$\limsup_{i\to\infty} Q_{T_i^N}^N\left(\hat{\theta}_{T_i^N}^N\right) = Q^N(\hat{\theta}^N) + s_N$$

(cf. notation in proof of Lemma 2), with $s_N \to \sigma^2$ as $N \to \infty$. On the other hand, since $\hat{\theta}_{T_i^N}$ is a least-squares estimate,

$$\begin{aligned} Q_{T_i^N}(\theta_0) &= \frac{1}{T_i^N} \sum_{t=1}^{T_i^N} \varepsilon_t^2 \geq Q_{T_i^N}\left(\hat{\theta}_{T_i^N}\right) \\ &\geq \frac{1}{T_i^N} \sum_{x_t \in J_N} \left|\theta_0(x_t) - \hat{\theta}_{T_i^N}(x_t) + \varepsilon_t\right|^2 \\ &= Q_{T_i^N}^N\left(\hat{\theta}_{T_i^N}^N\right), \end{aligned}$$

where the last equality follows from (2.11). For $i \to \infty$

$$Q^N(\hat{\theta}^N) + s_N \leq \sigma^2. \tag{3.10}$$

Again by (2.11), for $N' \geq N$,

$$\hat{\theta}^N(x) = \hat{\theta}^{N'}(x),$$

at all continuity points $x \in J_N$ of $\hat{\theta}^N$ and at all $x \in \Delta \cap J_N$. By monotonicity, the continuity points (in J_N) of $\hat{\theta}^N$ and $\hat{\theta}^{N'}$ coincide. There exists then a monotone increasing function $\hat{\theta}$ (not necessarily bounded) such that for all N

$$Q^N(\hat{\theta}^N) = Q^N(\hat{\theta}) = \int_{J_N} |\theta_0(x) - \hat{\theta}(x)|^2 \, dF(x).$$

Moreover, taking the diagonal sequence $T_N = T_N^N$, $\hat{\theta}_{T_N}(x) = \hat{\theta}_{T_N}^N(x)$ for $x \in J_N$ implies that

$$\hat{\theta}_{T_N}(x) \to \hat{\theta}(x) \quad \text{as } N \to \infty, \tag{3.11}$$

at all continuity points x of $\hat{\theta}$ and all $x \in \Delta$. Hence, letting $N \to \infty$, it follows from (3.10) and $s_N \to \sigma^2$ that

$$\int_{\mathrm{int}(\mathscr{X}) \cup \Delta} |\theta_0(x) - \hat{\theta}(x)|^2 \, dF + \sigma^2 \leq \sigma^2,$$

from which

$$\int_{\mathrm{int}(\mathscr{X}) \cup \Delta} |\theta_0(x) - \hat{\theta}(x)|^2 \, dF = 0. \tag{3.12}$$

By virtue of (A.3) this implies that $\hat{\theta}(x) = \theta_0(x)$ at all continuity points $x \in \mathrm{int}(\mathscr{X})$ of $\hat{\theta}$. But these coincide (by monotonicity) with the continuity points of θ_0. Consequently, by (3.6), $\hat{\theta}_{T_N}(x) \to \theta_0(x)$ at interior continuity points of θ_0. Hence, we have shown that every subsequence (T') contains a further subsequence (T'') such that $\hat{\theta}_{T''}(x) \to \theta_0(x)$ at the interior continuity points of θ_0. This shows the assertion. □

REMARK 7. Let $\Delta' = \{x: \liminf_{T \to \infty} \int_{\{x\}} dF_T > 0\}$. Note that, since $\int_{\{x\}} dF > 0$ for $x \in \Delta'$, we get from (3.12) the additional information that

$$\hat{\theta}_T(x) \to \theta_0(x) \quad \text{at } x \in \Delta'. \tag{3.13}$$

If θ_0 is continuous on some finite interval $[a, b] \subset \mathrm{int}(\mathscr{X})$, convergence holds even uniformly on this interval.

COROLLARY 1. *Suppose that* θ_0 *is continuous on* $[a, b] \subset \operatorname{int}(\mathscr{X})$. *Then, under* (A.1)–(A.4), *every sequence* $\hat{\theta}_T$, $T = 1, 2, \dots$, *of least-squares estimators converges a.e. to* θ_0 *uniformly on* $[a, b]$.

PROOF. Use Lemma 5. □

COROLLARY 2. *Suppose that the* x_t *are i.i.d.* (*stationary and ergodic*) *with common distribution function F assigning positive measure to every nondegenerate interval, and its continuous part charging no monotone graph* (*being absolutely continuous*). *Then under* (A.1)–(A.2), *every sequence* $\hat{\theta}_T$, $T = 1, 2, \dots$, *of least-squares estimators is strongly consistent on* $\operatorname{int}(\mathscr{X})$. *Moreover, if the* (x_t, ε_t) *are stationary and* ε_t *i.i.d., then* (A.2) *may be weakened to* $\alpha = 1$.

PROOF. The first part follows from Theorems 1 and 2 together with Remark 3. For the second assertion, note that $\alpha > 1$ is only needed in Lemma 2 in order to prove (2.3). But in the indicated special case (2.3) follows from Theorem 2.19 in [5]. □

In [4] it is shown that for monotone θ_0 with $\sup_{|x|>M}|\theta_0'(x)| < 1$ for some M and i.i.d. ε_t with positive density the scheme

$$y_t = \theta_0(y_{t-1}) + \varepsilon_t$$

possesses a unique solution y_t with a.e. positive density such that (y_{t-1}, ε_t) is stationary ergodic, so that Corollary 2 applies. Actually, for jointly stationary ergodic (x_t, ε_t), no additional assumptions about the dependence structure of the error process need to be made, as the following corollary shows.

COROLLARY 3. *Suppose that* (x_t, ε_t) *is stationary ergodic with the distribution function of* x_t *having vanishing singular part,* $E(\varepsilon_t^2) < \infty$ *and* $E(g(x_t)\varepsilon_t) = 0$ *for every bounded measurable function g. Then strong consistency obtains.*

PROOF. Lemmas 1–3 are obtained from ergodicity. The predeterminedness of the x-process is exploited in the proof of Lemma 9. □

We want to compare the results obtained with those known in the literature. Let us first stress that our assumptions do not presume any a priori knowledge on continuity of θ_0 and that a wide range of dependency patterns of ε_t and x_t is admitted (cf. the example just mentioned). A comparison is best made by specializing to the case of independent ε_t and i.i.d. x_t (independent of the ε_t). As far as the ε_t are concerned, the best known results (cf. [14]) work with the condition

$$H(y) \to 0 \quad \text{as } y \to \infty \text{ and } \int_0^\infty y|dH(y)| < \infty, \tag{3.14}$$

where $H(y) = \sup_t P(|\varepsilon_t| \geq y)$. In the i.i.d. case this is apparently weaker than our condition (A.2), requiring only first moments. In [6], (3.13) is sufficient only in the case $d = 1$; for $d \geq 2$, the conditions imposed [cf. (22)] imply our (A.2) if $r > 1$ is taken large enough to make the Borel–Cantelli lemma work on (26).

As to the common distribution of the x_t in the i.i.d. case, it is required in [14] that the singular continuous part vanish. With the improvement provided by Smythe (cf. [10]) that no monotone graph be charged by the continuous part, this amounts to a case covered in our Theorem 3 [(A.3) is needed in all approaches]. A further improvement by Wright (cf. [15]) allows one to treat certain special cases not covered by our results. As for deterministic regressors, (A.2) is implied by (25) and (27) in [6] for $d \geq 2$. In the case $d = 1$, only (A.3) is needed in [6]. So there seems to be some redundancy in (A.4) for this case.

4. Truncated estimators. In practical applications, the values which the regressors may take are often confined to some compact interval $\mathscr{X} = [a, b]$ in Euclidean d-space. Theorem 1 guarantees consistency of the monotone least-squares estimator in the interior of $\mathscr{X}$. At the boundary, however, things will usually go wrong (unless there are enough observations in the sense that a or b are in Δ'), as the following counterexample in dimension 1 shows.

Counterexample. Let θ_0 be continuous on $[a, b]$ and let the ε_t be normal i.i.d. and the x_t numbers $< b$ satisfying (A.3) and (A.4). Denote $x_{t(T)} = \max\{x_t: 1 \leq t \leq T\}$. Then $\hat{\theta}_T(x_{t(T)}) \geq y_{t(T)}$ [and hence $\hat{\theta}_T(x) \geq y_{t(T)}$ for all $x_{t(T)} \leq x \leq b$]. Since $\limsup_{T \to \infty} \varepsilon_{t(T)} = \infty$ a.e., for a.e. ω there exists a subsequence $(T'(\omega))$ such that $\varepsilon_{t(T'(\omega))}(\omega) \nearrow \infty$. Consequently, $\limsup_{T \to \infty} \hat{\theta}_T(b) = \infty$, showing that *convergence of $\hat{\theta}_T(b)$ to $\theta_0(b)$ does not hold with probability* 1. A similar argument applies to the left endpoint. If the x_t are i.i.d. (with their common distribution not charging the endpoints) and independent of the ε_t, inconsistency at the endpoints remains true by virtue of Fubini's theorem.

The fact that the least-squares estimate systematically over- (under-)estimated the true value at the right (left) endpoint of the interval is well known to practitioners. The annoying thing is not so much that $\hat{\theta}_T$ is not consistent on the boundary itself, but rather that the impact of the right- (or left-)most observation leads to poor speed of convergence near the boundary. We shall therefore suggest a modification of the monotone least-squares estimator that guarantees consistency on the whole interval, including its boundary. To simplify the exposition, we shall only treat the case of a one-dimensional interval $\mathscr{X} = [a, b]$, but the results are easily carried over to the multidimensional setting. Let the ε_t be i.i.d., $E\varepsilon_t^2 < \infty$. Consider first deterministic regressors x_t satisfying (A.3) and (A.4). For fixed T, let Π_T contain the n_T largest observation points x_t, $1 \leq n_T \leq T$. Let

$$\mu_T = \frac{1}{n_T} \sum_{t \in \Pi_T} y_t$$

(neglecting ties for the moment). μ_T can be written in the form

$$\mu_T = \frac{1}{n_T} \sum_{t \in \Pi_T} \theta_0(x_t) + \sum_{t=1}^{T} a_{tT} \varepsilon_t, \tag{4.1}$$

where $a_{tT} = 1/n_T$ or $= 0$ according to whether $t \in \Pi_T$ or not. If the n_T, $T = 1, 2, \ldots,$ are chosen in such a way that, for some positive constant C,

$$n_T \geq C \cdot T^{1/2}, \tag{4.2}$$

then

$$\sum_{t=1}^{T} a_{tT}\varepsilon_t \to 0 \quad \text{a.e.}$$

(cf. [1], Theorem 4.1.5.) Let $\underline{x}_T = \min\{x_t\colon t \in \Pi_T\}$. Then since

$$\int_{[\underline{x}_T,\, b]} dF_T = \frac{n_T}{T},$$

it follows from (A.3) that $\underline{x}_T \to b$ as $T \to \infty$, provided the n_T are chosen in such a way that

$$\frac{n_T}{T} \to 0. \tag{4.3}$$

Then, if θ_0 is continuous at b,

$$\frac{1}{n_T} \sum_{t \in \Pi_T} \theta_0(x_t) \to \theta_0(b).$$

Hence,

$$\mu_T \to \theta_0(b) \quad \text{a.e.} \tag{4.4}$$

Consider now the following *truncated* monotone least-squares estimator

$$\tilde{\theta}_T(x) = \min\{\hat{\theta}_T(x), \mu_T\}.$$

Then $\tilde{\theta}_T$ is increasing, and

$$\tilde{\theta}_T(x) \to \theta_0(x), \tag{4.5}$$

at each continuity point x of θ_0, including $x = b$, i.e., *the truncated monotone least-squares estimator is strongly consistent on the interval* $(a, b]$.

If ties occur, there may arise the need to choose among different y_t-values belonging to one observation x_t, $t \in \Pi_T$. In this case, any choice independent of the actually observed y_t-values will do. Finally, for i.i.d. x_t independent of the ε_t, (4.5) will follow from the result for deterministic regressors together with Fubini's theorem.

Of course, a corresponding modification can be done at the left endpoint of $\mathscr{X}$, thus providing us with a modified least-squares estimator that is consistent uniformly on $[a, b]$. The choice of n_T [within the limits provided by (4.2) and (4.3)] will heavily influence the rate of convergence of $\tilde{\theta}_T$. A "good" choice will have to take into account the distribution of the x_t and—if known a priori—the rate of increase of θ_0 at the endpoints. In general, it will be left to practical experience to find an appropriate n_T.

REFERENCES

[1] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.

[2] Breiman, L. (1968). *Probability*. Addison-Wesley, Reading, Mass.

[3] Brunk, H. D., Ewing, G. M. and Utz, W. R. (1957). Minimizing integrals in certain classes of monotone functions. *Pacific J. Math.* 7 833–844.

[4] CHRISTOPEIT, N. (1983). Asymptotic properties of monotone least squares estimates. Technical report, Univ. Bonn.

[5] HALL, P. and HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application*. Academic, New York.

[6] HANSON, D. L., PLEDGER, G. and WRIGHT, F. T. (1973). On consistency in monotonic regression. *Ann. Statist.* **1** 401–421.

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

[8] KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrica* **29** 1–21.

[9] ROBERTSON, T. and WRIGHT, F. T. (1975). Consistency in generalized isotonic regression. *Ann. Statist.* **3** 350–362.

[10] SMYTHE, R. T. (1980). Maxima of partial sums and a monotone regression estimator. *Ann. Probab.* **8** 630–635.

[11] STOUT, W. F. (1974). *Almost Sure Convergence*. Academic, New York.

[12] TOSSTORFF, G. (1984). *Methoden der Geometrischen Datenanalyse und ihre Anwendung bei der Untersuchung des Entwicklungsprozesses*. Physica, Würzburg.

[13] TUKEY, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, Mass.

[14] WRIGHT, F. T. (1979). A strong law for variables indexed by a partial ordered set with applications to isotonic regression. *Ann. Probab.* **7** 109–127.

[15] WRIGHT, F. T. (1981). The empirical discrepancy of lower layers and a related law of large numbers. *Ann. Probab.* **9** 323–329.

INSTITUTE FOR ECONOMETRICS
AND OPERATIONS RESEARCH
UNIVERSITY OF BONN
ADENAUERALLEE 24-42
D-5300 BONN 1
WEST GERMANY