The Annals of Statistics
1987, Vol. 15, No. 3, 1031–1049

A LARGE DEVIATION RESULT FOR PARAMETER ESTIMATORS AND ITS APPLICATION TO NONLINEAR REGRESSION ANALYSIS[1]

BY ARTHUR SIEDERS AND KACHA DZHAPARIDZE

Delft University of Technology, Delft, and Center for Mathematics and Computer Science, Amsterdam

Elaborating on the work of Ibragimov and Has'minskii (1981) we prove a law of large deviations (LLD) for M-estimators, i.e., those estimators which maximize a functional, continuous in the parameter, of the observations. This LLD is applied, using the results of Petrov (1975), to the problem of parametrical nonlinear regression in the situation of discrete time, independent errors and regression functions which are continuous in the parameter. This improves a result of Prakasa Rao (1984).

1. Introduction. The main results of this paper are Theorems 3.1 and 3.2, which establish an LLD for the least-squares estimator of a nonlinear regression parameter. The proofs rely on Theorem 2.1, which is a generalization of Theorem 1.5.1 of Ibragimov and Has'minskii (1981). In order to understand why generalization is desirable, consider the following nonlinear regression model for the observations $X^n := X_1, X_2, \ldots, X_n$:

$$X_t = f_t(\theta) + \varepsilon_t, \qquad t = 1, 2, \ldots, n, \tag{1.1}$$

where the f_t are known continuous functions on a parameter set $\Theta \subset \mathbb{R}^k$, the ε_t are independent, not necessarily identically distributed, errors with zero expectation, and $\theta \in \Theta$ is the true value of the parameter, which is to be estimated by some functional $\hat{\theta}_n(X_1, X_2, \ldots, X_n)$.

If the distributions F_t of the ε_t are known, then we can construct a family of measures $\{\mathbb{P}_\theta^{(n)}, \theta \in \Theta\}$ on a suitable space of events $\{\mathscr{X}^{(n)}, \mathscr{U}^{(n)}\}$, define the family of statistical experiments $\{\mathscr{X}^{(n)}, \mathscr{U}^{(n)}, \mathbb{P}_\theta^{(n)}\}$, $n = 1, 2, \ldots,$ and proceed as in Ibragimov and Has'minskii (1981) in order to describe the asymptotic behavior of the maximum likelihood estimator $\hat{\theta}_n^{\mathrm{ML}}$.

For instance, we can apply Theorem 1.5.1 of Ibragimov and Has'minskii (1981), which states that a law of large deviations [i.e., an (exponential) inequality for the probability of a large deviation of the estimator $\hat{\theta}_n^{\mathrm{ML}}$ from the true value θ] holds if the normalized likelihood ratio $Z_{n,\theta}(u)$ satisfies two conditions, which, roughly stated, are that, for n large enough (ε small enough, in the formulation of the theorem, put $\varepsilon := 1/n$), $Z_{n,\theta}(u)$ is, in expectation, sufficiently continuous in u and that $\mathbb{E}Z_{n,\theta}(u)^{1/2}$ decreases exponentially as $|u| \to \infty$.

Received December 1985; revised September 1986.

[1]This research was supported by the Netherlands Foundation of Mathematics SMC with financial aid from the Netherlands Organization for the Advancement of Pure Research (Z.W.O.).

AMS 1980 *subject classifications*. Primary 60F10, 62F12, 62J02.

Key words and phrases. M-estimation, large deviations, rate of convergence, least-squares, nonlinear regression, Michaelis–Menten model.

However, if the distributions F_t are unknown, $\hat{\theta}_n^{\mathrm{ML}}$ is not defined. In this case, one often resorts to the so-called least-squares estimator $\hat{\theta}_n^{\mathrm{LS}}$, which minimizes the residual sum of squares

$$(1.2)\qquad Q_n(X^n, \theta) := \sum_{t \le n} (X_t - f_t(\theta))^2.$$

The properties of $\hat{\theta}_n^{\mathrm{LS}}$ can be investigated if one restricts the F_t to a sufficiently "nice" class $\{F_t\}$. We claim that Theorem 1.5.1 of Ibragimov and Has'minskii (1981), although formulated for the maximum likelihood scheme, can provide a valuable tool here. In the theory of M-estimators the idea has been developed [see, for instance, Serfling (1980)] that the classical maximum likelihood theory can be extended to estimators maximizing some other functional of the observations. Indeed, inspection of the proof of the previously mentioned theorem reveals that it continues to hold if the likelihood is replaced by some other θ-continuous $\mathbb{P}_\theta^{(n)}$-a.s. positive functional $C_n(X^n, \theta)$, which we shall call an M-functional.

We shall try to apply this generalized version of Theorem 1.5.1 to the LS-estimator for the model given by (1.1), which maximizes the M-functional

$$(1.3)\qquad C_n(X^n, \theta) := \exp\left(-\tfrac{1}{2} \sum_{t \le n} (X_t - f_t(\theta))^2\right),$$

which is, of course, the likelihood if the ε_t are i.i.d. standard normal. Theorem 1.5.1 (and our Theorem 2.1) express the large deviation properties of the estimator in the normalized ratio $Z_{n,\theta}(u)$ and not directly in $C_n(X^n, \theta)$ (the reason for this lies in the application of Lemma A.2). Therefore we define, for some choice of norming constants ϕ_n,

$$(1.4)\qquad Z_{n,\theta}(u) := C_n(X^n, \theta + \phi_n u)/C_n(X^n, \theta).$$

Unfortunately, it turns out that it is not at all easy to formulate conditions on the family of regressors $\{f_t(\theta), \theta \in \Theta\}$ and the class of distributions $\{F_t\}$ of ε_t which guarantee that the $Z_{n,\theta}(u)$ defined by (1.3) and (1.4) satisfies the conditions of the generalized theorem described above. It is perhaps for this reason that Prakasa Rao (1984) restricts himself to the case that the ε_t are i.i.d. Gaussian and the dimension k of Θ is equal to 1. The main difficulty inherent in Theorem 1.5.1 seems to be that its Hölder condition (1) is quite difficult to verify, as its authors, in their comment on Theorem 1.5.1, implicitly admit, especially if the dimension k of Θ is > 1. On page 56 of Ibragimov and Has'minskii (1981), a theorem is announced which concerns the case $k > 1$ (Theorem 1.5.8). The proof, however, is valid only for $k = 1$, and extension to the case $k > 1$ is not obvious. Less powerful, but more sound methods all require considerable manipulation, even in the Gaussian situation, cf. Ingster (1984), page 1179, and Ibragimov and Has'minskii (1981), Lemma 3.5.2 on page 202ff.

These observations motivated us to look for an LLD in the spirit of Theorem 1.5.1, which would not only apply to a much broader class of estimators than just ML, but which would also be more flexible in its conditions. This effort resulted in Theorem 2.1 of this paper, which we apply, in Section 3, to the nonlinear regression problem. For statistical applications of LD theorems we refer the reader to Theorem 3.1.3 of Ibragimov and Has'minskii (1981), which may give an idea of the possibilities.

Dzhaparidze (1986) used a rudimentary form of Theorem 2.1 to infer about intensity parameters of counting processes. Another study on Theorem 1.5.1 was recently made by Vostrikova (1984), who gives conditions for an LLD for Bayesian and ML-estimators in terms of variation distance and predictable terms. Large deviation results for M-estimators in an i.i.d. setting were recently obtained by Kester (1985).

2. A law of large deviations. Consider a family of statistical experiments $\mathscr{E}^{(\varepsilon)} = \{\mathscr{X}^{(\varepsilon)}, \mathscr{U}^{(\varepsilon)}, \mathbb{P}_\theta^{(\varepsilon)}; \theta \in \Theta\}$, where the $\mathbb{P}_\theta^{(\varepsilon)}$ are not necessarily of known form (see Section 1). The parameter set Θ is a Borel subset of k-dimensional Euclidean space. We shall consider M-estimators maximizing an M-functional C_ε: $\mathscr{X}^{(\varepsilon)} \times \Theta \to [0, \infty)$, which is assumed to be, for all $X^\varepsilon \in \mathscr{X}^{(\varepsilon)}$, a positive continuous function of θ and, for each $\theta \in \Theta$, a measurable functional of X^ε.

Throughout we assume that, for all $\theta \in \Theta$ and $\mathbb{P}_\theta^{(\varepsilon)}$-almost all X^ε, a solution $\hat{\theta}_\varepsilon$ to the equation

$$C_\varepsilon(X^\varepsilon, \hat{\theta}_\varepsilon) = \sup_{\theta \in \Theta} C_\varepsilon(X^\varepsilon, \theta) \tag{2.1}$$

exists (this is certainly true if Θ is compact). On the basis of the existence assumption we may demonstrate that a measurable functional $\hat{\theta}_\varepsilon$: $\mathscr{X}^{(\varepsilon)} \to \Theta$ exists which is a solution of (2.1). This is worked out in Lemma A.1 in the Appendix. So we assume henceforth that $\hat{\theta}_\varepsilon$ is measurable.

All our results are of asymptotic nature, i.e., they are valid for ε small enough and R large enough, where $\varepsilon \to 0$ describes the approach of the limit experiment $\mathscr{E}^{(0)}$ and R describes the normalized deviation of the estimator $\hat{\theta}_\varepsilon$ from the true value θ.

Let, for each ε and $\theta \in \Theta$, $\phi(\varepsilon, \theta)$ be a nonsingular $k \times k$ matrix and define the normalized M-ratio

$$Z_{\varepsilon,\theta}(u) := Z_{\varepsilon,\theta}(X^\varepsilon, u) = C_\varepsilon(X^\varepsilon, \theta + \phi(\varepsilon, \theta)u)/C_\varepsilon(X^\varepsilon, \theta), \tag{2.2}$$

which, for fixed observation X^ε, is a continuous, nonnegative finite function on the set $U_{\varepsilon,\theta} := \phi(\varepsilon, \theta)^{-1}(\Theta - \theta)$. Define $\Gamma_{\varepsilon,\theta,R} := \overline{U}_{\varepsilon,\theta} \cap \{u\colon R \le |u| \le R + 1\}$. We define the following sets of functions [compare Ibragimov and Has'minskii (1981), Chapter 1.5, page 41].

G is the set of all functions $g_\varepsilon(\cdot)$ possessing the following properties:

(1) for fixed ε, $g_\varepsilon(\cdot)$ is a function on $[0, \infty)$ monotonically increasing to infinity;
(2) for any $N > 0$,

$$\lim_{\substack{R \to \infty \\ \varepsilon \to 0}} R^N \exp(-g_\varepsilon(R)) = 0. \tag{2.3}$$

Let K be a measurable subset of Θ, then $\mathbf{H}_K$ is the set of all functions $\eta_{\varepsilon,\theta}(\cdot)$

possessing the following properties:

(1) for fixed ε and $\theta \in \Theta$, $\eta_{\varepsilon,\theta}(\cdot)$ is a function $U_{\varepsilon,\theta} \to (0,\infty)$;
(2) there exists a polynomial pol(R) in R such that, for ε small enough and R sufficiently large,

$$\sup_{\theta \in K;\ u \in \Gamma_{n,\theta,R}} \eta_{\varepsilon,\theta}(u)^{-1} \le \mathrm{pol}(R). \tag{2.4}$$

Let, for each ε and θ, $\tilde{\zeta}_{\varepsilon,\theta}$: $[0,\infty) \to \mathbb{R}$ be a monotonically nondecreasing continuous function and define the random functional

$$\zeta_{\varepsilon,\theta}(u) := \tilde{\zeta}_{\varepsilon,\theta}(Z_{\varepsilon,\theta}(u)). \tag{2.5}$$

The main result of this section is the following theorem, which gives sufficient conditions, in terms of the functionals $\zeta_{\varepsilon,\theta}(u)$, for an LLD to hold for $\hat{\theta}_\varepsilon$.

THEOREM 2.1. (a) *Let the functionals $\zeta_{\varepsilon,\theta}(u)$ possess the following properties: Given a measurable subset $K \subset \Theta \subset \mathbb{R}^k$, there correspond to it numbers m and α, where $m \ge \alpha > k$, functions $g_\varepsilon \in \mathbf{G}$ and $\eta_{\varepsilon,\theta} \in \mathbf{H}_K$, and a polynomial* pol$_K(R)$ *in R such that, for all ε small and R large enough, the following conditions hold*:

$$\mathbb{E}_\theta^{(\varepsilon)}|\zeta_{\varepsilon,\theta}(u) - \zeta_{\varepsilon,\theta}(v)|^m \le |u-v|^\alpha \mathrm{pol}_K(R), \tag{M.1}$$

for all $\theta \in K$ and u and $nv \in \Gamma_{\varepsilon,\theta,R}$;

$$\mathbb{P}_\theta^{(\varepsilon)}\{\zeta_{\varepsilon,\theta}(u) - \zeta_{\varepsilon,\theta}(0) \ge -\eta_{\varepsilon,\theta}(u)\} \le \exp(-g_\varepsilon(R)), \tag{M.2}$$

for all $\theta \in K$ and $u \in \Gamma_{\varepsilon,\theta,R}$.

Then the following uniform LLD holds: There exist positive constants B_0 and b_0, such that, for all ε small and H large enough,

$$\sup_{\theta \in K} \mathbb{P}_\theta^{(\varepsilon)}\Big\{\big|\phi(\varepsilon,\theta)^{-1}(\hat{\theta}_\varepsilon - \theta)\big| \ge H\Big\} \le B_0 \exp(-b_0 g_\varepsilon(H)).$$

The constant b_0 can be made arbitrarily close (from below) to $(\alpha - k)/(\alpha - k + mk)$ by choosing B_0 large enough.

(b) *The conclusion of part* (a) *continues to hold if* (M.1) *is replaced by the following condition* (M.1δ):

(M.1δ) (M.1) *holds for all $\theta \in K$ and $u, v \in \Gamma_{\varepsilon,\theta,R}$ satisfying $|u-v| \le \delta$, where δ is a fixed positive constant,*

provided one of the two following (weak) assumptions is satisfied:

(M.1′) Θ *is a convex set,*

(M.1″) $\mathbb{E}_\theta^{(\varepsilon)}|\zeta_{\varepsilon,\theta}(u)|^m \le \mathrm{pol}_K(R)$ *for all $\theta \in K$ and $u \in \Gamma_{\varepsilon,\theta,R}$.*

REMARKS.

1. For applications in the method of Ibragimov and Has'minskii (1981), the set K is chosen to be compact. For the preceding theorem this is not essential.

2. Theorem 1.5.1 of Ibragimov and Has'minskii (1981) follows from the preceding theorem by choosing $\zeta_{\varepsilon,\theta}(u) := Z_{\varepsilon,\theta}(u)^{1/m}$ and $\eta_{\varepsilon,\theta}(u) = \frac{1}{2}$. In particular, condition (2) of 1.5.1 implies (M.2) by Markov's inequality and condition (1) implies (M.1).
3. Compare also the conditions of Vostrikova (1984), Theorems 1 and 3.
4. If, for some θ, $\phi(\varepsilon, \theta) \to 0$ in operator norm as $\varepsilon \to 0$, then this θ is weakly consistently estimated by $\hat{\theta}_\varepsilon$.

The proof of Theorem 2.1 proceeds via a number of propositions. The reader is advised to consult the proof of Theorem 1.5.1 of Ibragimov and Has'minskii (1981), as our proof follows the same line. To avoid tedious repetitions, we assume at each stage of the proof that an initial choice of sufficiently small ε and sufficiently large R (or H) has been made.

PROPOSITION 2.2. *If there exist constants B and b such that*

$$\sup_{\theta \in K} \mathbb{P}_\theta^{(\varepsilon)}\Big\{ \sup_{u \in \Gamma_{\varepsilon,\theta,R}} \zeta_{\varepsilon,\theta}(u) \geq \zeta_{\varepsilon,\theta}(0)\Big\} \leq B\exp(-bg_\varepsilon(R)), \tag{2.6}$$

then (i) *the assertion of Theorem* 2.1 *holds, and* (ii) *the constant b_0 there can be chosen arbitrarily close (from below) to b.*

PROOF. Ibragimov and Has'minskii (1981), Chapter 1.5, page 42, prove a similar, but less precise, statement in (5.4). We apply Lemma A.2 (Appendix) and estimate its right-hand side. For any small positive δ, one has, using the monotonicity of g_ε and $\tilde{\zeta}$,

$$\begin{aligned} \mathbb{P}_\theta^{(\varepsilon)}\Big\{ \sup_{\substack{|u|\geq H \\ u \in U_{\varepsilon,\theta}}} Z_{\varepsilon,\theta}(u) \geq 1\Big\} &\leq B \sum_{r=0}^{\infty} \exp(-bg_\varepsilon(r+H)) \\ &= B\exp(-b_0 g_\varepsilon(H)) \sum_{r=0}^{\infty} \exp(-b\delta g_\varepsilon(H+r)), \end{aligned} \tag{2.7}$$

where $b_0 := b(1-\delta)$. The sum on the right-hand side is finite: Relation (2.3) says that in the limit, $R^N \exp(-g_\varepsilon(R)) \leq 1$ for all N, so put $N = 2/\delta b$. Then $\exp(-b\delta g_\varepsilon(R)) \leq R^{-2}$. □

PROPOSITION 2.3. *Condition* (M.1δ) *together with either condition* (M.1′) *or* (M.1″) *implies condition* (M.1).

PROOF.
Case 1. (M.1δ) and (M.1′) ⇒ (M.1). From the convexity of Θ follows that any u and v in $\Gamma_{\varepsilon,\theta,R}$ may be connected by a path in $\Gamma_{\varepsilon,\theta,R}$ consisting of linear segments of length $\leq \delta$, where the number of segments does not exceed $C\delta^{-1}|u-v|$ and C is a fixed constant not depending on θ or R. To all the segments (M.1δ) is applied; by Minkowski's inequality for integrals it then

follows that

$$(2.8)\qquad \left(\mathbb{E}|\zeta_{\varepsilon,\theta}(u)-\zeta_{\varepsilon,\theta}(v)|^m\right)^{1/m} \le C\delta^{-1}|u-v|\delta^{\alpha/m}\mathrm{pol}_K(R)^{1/m},$$

which leads to (M.1) because, as u and $v \in \Gamma_{\varepsilon,\theta,R}$,

$$|u-v| \le |u-v|^{\alpha/m}(2(R+1))^{1-\alpha/m},$$

where the second factor is absorbed by the polynomial pol_K.

Case 2. (M1.δ) and (M.1″) $\Rightarrow$ (M.1). From (M.1″) follows, using Minkowski's inequality again, that the left-hand side of (M.1) is bounded by $2^m\mathrm{pol}_K(R)$, which, for any u, v such that $|u-v|>\delta$, is bounded by $|u-v|^\alpha 2^m\delta^{-\alpha}\mathrm{pol}_K(R)$. □

Proof of Theorem 2.1. By Proposition 2.3 it suffices to prove only part (a). By Proposition 2.2 we need only prove relation (2.6). We subdivide the section $\{u\colon R \le |u| \le R+1\}$ into N regions, each with diameter at most h. Such a subdivision can be accomplished such that the number of regions is bounded by

$$(2.9)\qquad N \le \mathrm{const.}(k)(R+1)^{k-1}h^{-k},$$

where const.(k) is a constant depending only on k. This subdivision induces a partition of $\Gamma_{\varepsilon,\theta,R}$ in at most N sets; denote this partition by

$$(2.10)\qquad \Gamma_{\varepsilon,\theta,R} = \Gamma^{(1)}_{\varepsilon,\theta,R} \cup \Gamma^{(2)}_{\varepsilon,\theta,R} \cup \cdots \cup \Gamma^{(N')}_{\varepsilon,\theta,R},$$

where $N' \le N$, and choose in each member $\Gamma^{(i)}_{\varepsilon,\theta,R}$ a point u_i. Then

$$(2.11)\qquad \mathbb{P}^{(\varepsilon)}_\theta\Big\{\sup_{\Gamma_{\varepsilon,\theta,R}} \zeta_{\varepsilon,\theta}(u) \ge \zeta_{\varepsilon,\theta}(0)\Big\} \le P_1 + P_2,$$

where P_1 and P_2 are given by

$$(2.12)\qquad \begin{aligned} P_1 &:= \sum_{j=1}^{N'} \mathbb{P}^{(\varepsilon)}_\theta\{\zeta_{\varepsilon,\theta}(u_j) - \zeta_{\varepsilon,\theta}(0) \ge -\eta_{\varepsilon,\theta}(u_j)\},\\ P_2 &:= \mathbb{P}^{(\varepsilon)}_\theta\Big\{\max_{|u-v|\ge h}|\zeta_{\varepsilon,\theta}(u)-\zeta_{\varepsilon,\theta}(v)| \ge \inf_{\Gamma_{\varepsilon,\theta,R}}\eta_{\varepsilon,\theta};\ u,v\in\Gamma_{\varepsilon,\theta,R}\Big\}. \end{aligned}$$

From condition (M.2) and the inequality (2.9) we have immediately

$$(2.13)\qquad P_1 \le \mathrm{const.}(k)(R+1)^{k-1}h^{-k}\exp - g_\varepsilon(R).$$

The second term P_2 is bounded as follows. Throughout the argument we let pol(R) denote any (not necessarily always the same) polynomial in R, the coefficients of which may depend on α, k, m and pol_K but not on ε, R, θ, u and v. Now, let u_0 be any point in $\Gamma_{\varepsilon,\theta,R}$ and consider the random function $\zeta_{\varepsilon,\theta}(u) - \zeta_{\varepsilon,\theta}(u_0)$ on the closed set $\Gamma_{\varepsilon,\theta,R}$. Now apply it to Lemma A.3 in the Appendix. By assumption, ζ is continuous in u and hence it has a measurable and separable version [see Neveu (1970) for the notion of separability]. Put

$$(2.14)\qquad C(u) := \max\{1, |u-u_0|^\alpha\}\mathrm{pol}_K(R),$$

then $C(u)$ is bounded by pol(R), as u and $u_0 \in \Gamma_{\varepsilon,\theta,R}$. With this choice of $C(u)$, conditions (1) and (2) of the lemma are fulfilled due to condition (M.1) of Theorem 2.1. It then follows from this lemma and Markov's inequality that

$$P_2 \le h^{(\alpha-k)/m}\mathrm{pol}(R), \tag{2.15}$$

where we have used the property (2.4) of $\eta_{\varepsilon,\theta}^{-1}$ to be polynomially bounded in u. Putting the inequalities (2.11), (2.13) and (2.15) together we have

$$\begin{aligned}\mathbb{P}_\theta^{(\varepsilon)}\{\sup \zeta_{\varepsilon,\theta}(u) \ge \zeta_{\varepsilon,\theta}(0)\} \le h^{-k}\mathrm{pol}(R)\exp(-g_\varepsilon(R))\\ + h^{(\alpha-k)/m}\mathrm{pol}(R).\end{aligned} \tag{2.16}$$

Now we put $h := \exp(Cg_\varepsilon(R))$, where the constant C should be chosen such that no one tail in (2.16) dominates the other. This leads to

$$C = -m/(\alpha - k + mk). \tag{2.17}$$

The final result (2.6) follows from (2.16), (2.17) and the property (2.3) of $\exp g_\varepsilon$ to dominate any polynomial. The statement concerning b_0 is now obvious from the second part of Proposition 2.2. We remark that Ibragimov and Has'minskii (1981) use, instead of (2.9), the inequality $N \le = \mathrm{const.}(k)(R+1)/h^{k-1}$, which we were unable to verify. Of course, this would lead to another bound for b_0 in Theorem 2.1. □

3. Nonlinear least-squares regression with independent errors. Let Θ be a Borel subset of $\mathbb{R}^k$ and let $f_t(\theta)$ be a continuous deterministic function from Θ to $\mathbb{R}$ for each $t \in \mathbb{N}$; all our results can easily be generalized to the case of a deterministic triangular design array $(t_1^{(n)}, t_2^{(n)}, \dots, t_n^{(n)};\ n \in \mathbb{N})$. We consider the nonlinear regression model

$$X_t = f_t(\theta) + \varepsilon_t, \qquad t = 1, 2, \dots, n, \tag{3.1}$$

where $X^n := X_1, X_2, \dots, X_n$ are the observed random variables and $\{\varepsilon_t,\ t \in \mathbb{N}\}$ is a sequence of real independent random variables with expectation zero.

The least-squares estimator $\hat\theta_n$ (which we assume to exist; see Section 2 and Lemma A.1) maximizes the functional

$$C_n(X^n, \theta) := \exp -\tfrac{1}{2}\sum_{t\le n}(X_t - f_t(\theta))^2. \tag{3.2}$$

Given a sequence of nonsingular matrix norming factors $\phi_n(\theta)$ we define the ratio

$$\begin{aligned} Z_{n,\theta}(u) &:= C_n(X^n, \theta + \phi_n(\theta)u)/C_n(X^n,\theta)\\ &= \exp \sum_{t\le n} d_{tn\theta}(u)\varepsilon_t - \tfrac{1}{2}\sum_{t\le n} d_{tn\theta}(u)^2,\end{aligned} \tag{3.3}$$

where

$$d_{tn\theta}(u) := f_t(\theta + \phi_n(\theta)u) - f_t(\theta). \tag{3.4}$$

Because of the many practical applications of the model (3.1), the various properties of the least-squares estimator, such as strong or weak consistency, asymptotic normality and large deviation behavior, have been studied extensively. See, e.g., van de Geer (1986), Ivanov (1976), Läuter (1985), Prakasa Rao (1984) and Wu (1981). All these authors restrict themselves to the case that the errors ε_t are independent and identically distributed.

We shall study the large deviation probability of the least-squares estimator in the case of independent errors. To this end, we stipulate the following assumptions which allow us to apply Theorem 2.1.

Assume that, for some Borel subset K of Θ, there exist functions $g_n(R) \in \mathrm{G}$, positive constants $\gamma > 0$, $\Lambda_1 \in (0, \infty]$, $\delta \in (0, \frac{1}{2})$, $\kappa > 0$, $\rho \in (0, 1]$ and a polynomial $\mathrm{pol}(R)$ such that, for all n and R large enough, the following inequalities hold:

(N.1) for all $t \in \mathbb{N}$ and $|\lambda| \leq \Lambda_1$ (note that $\Lambda_1 = \infty$ is allowed)

$$\mathbb{E} \exp(\lambda \varepsilon_t) \leq \exp\left(\tfrac{1}{2}\gamma\lambda^2\right);$$

(N.2) for all $\theta \in K$ and $u, v \in \Gamma_{n,\theta,R}$, where $|u - v| \leq \kappa$, one has

$$\sum_{t \leq n} \left[f_t(\theta + \phi_n(\theta)u) - f_t(\theta + \phi_n(\theta)v)\right]^2 \leq |u - v|^{2\rho} \mathrm{pol}(R)$$

and

$$\sum_{t \leq n} \left[f_t(\theta + \phi_n(\theta)u) - f_t(\theta)\right]^2 \leq \mathrm{pol}(R);$$

(N.3) for all $\theta \in K$ and $u \in \Gamma_{n,\theta,R}$ one has

$$\sum_{t \leq n} \left[f_t(\theta + \phi_n(\theta)u) - f_t(\theta)\right]^2 \geq \Delta_n(\theta, u) g_n(R),$$

where

$$\Delta_n(\theta, u) := \max\left\{2\gamma\delta^{-2}, 2\Lambda_1^{-1}\delta^{-1} \max_n(\theta, u)\right\}$$

and

$$\max_n(\theta, u) := \max\left\{ \left| f_t(\theta + \phi_n(\theta)u) - f_t(\theta)\right|;\ t = 1, 2, \ldots, n\right\}.$$

The following theorem seems to us an instructive example of the application of the very general Theorem 2.1.

THEOREM 3.1. *Let, for some $K \subset \Theta$ and suitably chosen normings $\phi_n(\theta)$, assumptions* (N.1)–(N.3) *be fulfilled. Then the following LLD holds: There exist constants B_0 and b_0, such that, for all n and H large enough,*

$$\sup_{\theta \in K} \mathbb{P}_\theta^{(n)}\left\{ \left| \phi_n(\theta)^{-1}\left(\hat{\theta}_n - \theta\right)\right| \geq H\right\} \leq B_0 \exp(-b_0 g_n(H)).$$

Moreover, for any $\beta > 0$ we can choose B_0 such that

$$b_0 \geq \rho(\rho + k)^{-1} - \beta. \tag{3.5}$$

Before proving this theorem, let us discuss the significance of conditions (N.1)–(N.3) and the relation they bear to known results concerning the behavior of the least-squares estimator.

Condition (N.1) prescribes that the tails of the ε_t should be uniformly "thin." The uniformity is evident in the i.i.d. case. If the ε_t are, e.g., Gaussian or bounded, then (N.1) holds with $\Lambda_1 = \infty$; in that case Δ_n in (N.3) is constant and $|f_t(\theta + \phi_n(\theta)u) - f_t(\theta)|$ may increase unboundedly in t.

Condition (N.2) is a Hölder-type continuity condition on the parametrization $\theta \to f(\theta)$. It is directly related to condition (M.1) of Theorem 2.1. This assures that the regression functions do not behave too wildly in θ, so that uniform estimates can be obtained. Compare, e.g., Lemma 3 of Jennrich (1969), Condition III of Ivanov (1976), Assumption A(ii) of Wu (1981) and condition (2.5) of Prakasa Rao (1984), which are of a similar nature. It is easy to construct an example, where the regression functions $f_t(\theta)$ are not everywhere continuous in θ but still an LLD holds. Therefore, we mention the approach of van de Geer (1986) to impose entropy instead of continuity conditions; compare also our inequality (2.9) and Lemma A of Wu (1981).

Condition (N.3) prescribes the rate of asymptotic separation. Asymptotic separation (the regression functions keep enough apart to be statistically distinguishable) is a necessary condition for consistent estimation; see Wu (1981), Theorem 1. It may be interesting to note that asymptotic separation may be viewed as a form of continuity of the inverse of the parametrization, i.e., of the map $f(\theta) \to \theta$: If θ and $\theta' := \theta + \phi_n(\theta)u$ are "apart," i.e., if $|\phi_n(\theta)^{-1}(\theta - \theta')| \geq R$, then $f(\theta)$ and $f(\theta')$ are also apart in the sense of condition (N.3). Logically, this is equivalent to a form of continuity. In Jennrich (1969), the separation condition is that of existence of the tail cross products (see also his Lemma 3). In Wu (1981), this seems to be his (complicated) condition A(i). In the same line lie the conditions of Ivanov (1976) (Condition III), Prakasa Rao (1984), condition (2.6), and Läuter (1985), condition (12) to Theorem 1.

Proof of Theorem 3.1. The proof consists of checking conditions (M.1) and (M.2) to Theorem 2.1 with $\xi(Z) := \log Z$. We assume that an initial choice of sufficiently large n and R has been made. Let, throughout, $u, v \in \Gamma_{n,\theta,R}$, $|u - v| \leq \kappa$ and $\theta \in K$.

First we check condition (M.1).

Condition (N.2) may be expressed in the $d_{tn\theta}(u)$, as defined in (3.4):

$$\sum_{t \leq n} |d_{tn\theta}(u) - d_{tn\theta}(v)|^2 \leq |u - v|^{2\rho}\operatorname{pol}(R) \tag{3.6}$$

and

$$\sum_{t \leq n} d_{tn\theta}(u)^2 \leq \operatorname{pol}(R). \tag{3.7}$$

Note that from (3.7) it follows that (3.6) holds also, if $|u - v| > \kappa$. In fact, (3.7) gives

$$\sum_{t \leq n} |d_{tn\theta}(u) - d_{tn\theta}(v)|^2 \leq 2\operatorname{pol}(R) \leq 2|u - v|^{2\rho}\kappa^{-2\rho}\operatorname{pol}(R),$$

where the factor $\kappa^{-2\rho}$ is absorbed by the polynomial pol(R). From (3.3) we have, choosing $\zeta_{n,\theta}(u) := \log Z_{n,\theta}(u)$,

$$\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v) = \sum_{t\le n} A_t\varepsilon_t - B_t, \tag{3.8}$$

where

$$\begin{aligned} A_t &:= d_{tn\theta}(u) - d_{tn\theta}(v),\\ 2B_t &:= d_{tn\theta}(u)^2 - d_{tn\theta}(v)^2. \end{aligned} \tag{3.9}$$

Note that, by Lemma 5 in Chapter 3.4 of Petrov (1975), condition (N.1) implies the existence and boundedness, uniform in t, of moments of all order m of ε_t. Hence, using the independence of the ε_t, condition (N.1) and $E\varepsilon_t = 0$, we find, for all even $m \ge 2$,

$$\mathbb{E}|\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v)|^m \le \text{const.}(m) \sum_{l, l_1,\dots,l_s}^{*} \prod_{i=1}^{s}\left(\sum_1^n A_t^{l_i}\right)\left(\sum_1^n B_t\right)^l, \tag{3.10}$$

where * denotes summation over all positive even $l_1, l_2, \dots, l_s \ge 2$ and even $l \ge 0$ (where $s \ge 0$) having sum m. We have the following estimates:

$$\begin{aligned} \left|2\sum_1^n B_t\right| &\le \sum_1^n |d_{tn\theta}(u) - d_{tn\theta}(v)||d_{tn\theta}(u) + d_{tn\theta}(v)|\\ &\le \left(\sum_1^n |d_{tn\theta}(u) - d_{tn\theta}(v)|^2 \sum_1^n |d_{tn\theta}(u) + d_{tn\theta}(v)|^2\right)^{1/2}\\ &\le |u - v|^{\rho}\operatorname{pol}(R), \end{aligned} \tag{3.11}$$

where we have used Cauchy–Schwarz, the inequality $(a + b)^2 \le 2a^2 + 2b^2$, the fact that $u, v \in \Gamma_{n,\theta,R}$ by assumption and inequalities (3.6) and (3.7). We also have, for l even and ≥ 2, using (3.6) again,

$$0 \le \sum_{t\le n} A_t^l \le \left|\sum_{t\le n} A_t^2\right|^{l/2} \le |u - v|^{\rho l}\operatorname{pol}(R). \tag{3.12}$$

Consequently, (3.10) becomes, using (3.11) and (3.12),

$$\mathbb{E}|\zeta_{n,\theta}(u) - \zeta_{n,\theta}(v)|^m \le |u - v|^{\rho m}\operatorname{pol}(R). \tag{3.13}$$

If we choose m even and larger than k/ρ, (3.13) fulfills condition (M.1) of Theorem 2.1, with the constant $\alpha = \rho m$.

Now we check condition (M.2). We shall write, for simplicity of notation, $d_t := d_{tn\theta}(u)$ and $\max|d_t| := \max\{|d_{tn\theta}(u)|;\ t = 1, 2, \dots, n\}$. Choose

$$\eta_{n,\theta}(u) := \left(\tfrac{1}{2} - \delta\right) \sum_{t\le n} d_{tn\theta}(u)^2. \tag{3.14}$$

By condition (N.3), one has the inequality

$$\sum_{t\le n} d_{tn\theta}(u)^2 \ge 8\gamma g_n(R), \tag{3.15}$$

which shows that $\eta_{n,\theta}(u) \in \mathbf{H}_K$ because, as follows from (2.4), $g_n(R)^{-1} \le 1$ for

n and R sufficiently large. By (3.8), (3.9) and (3.14) and Lemma A.4 in the Appendix

$$\mathbb{P}_\theta^{(n)}\{\zeta_{n,\theta}(u) - \zeta_{n,\theta}(0) \geq -\eta_{n,\theta}(u)\} = \mathbb{P}_\theta^{(n)}\Big\{\sum_{t\leq n} d_t\varepsilon_t \geq \delta \sum_{t\leq n} d_t^2\Big\} \tag{3.16}$$
$$\leq \exp\Big(-\sum_{t\leq n} d_t^2/\Delta_n\Big),$$

where $\Delta_n(\theta, u)$ is defined in condition (N.3).

It remains to apply the inequality of (N.3) to (3.16), which yields

$$\mathbb{P}_\theta^{(n)}\{\zeta_{n,\theta}(u) - \zeta_{n,\theta}(0) \geq -\eta_{n,\theta}(u)\} \leq \exp(-g_n(R)), \tag{3.17}$$

thus fulfilling condition (M.2) of Theorem 2.1.

The last step consists of the verification of the theorem's statement (3.5) concerning b_0. This is easily accomplished by choosing $\alpha = \rho m$ and letting $m \to \infty$. □

We have formulated conditions (N.2) and (N.3) in the spirit of Ibragimov and Has'minskii (1981) and our Theorem 2.1. This has allowed a direct application of this theorem. From Theorem 3.1 we now deduce a slightly weaker theorem of friendlier appearance, which seems to suffice for many applications. To this end, we make the following observations:

1. Problems might occur if, for some θ and u, $\Delta_n(\theta, u)$ would increase to infinity in n. For it follows from (N.2) and (N.3) that $g_n(R) \leq \mathrm{pol}(R)/\Delta_n(\theta, u)$; if $\Delta_n \to \infty$, then condition (2.3) on the set $\mathbf{G}$ would be violated. Fortunately, one also has
$$\max_n (\theta, u) \leq \Big(\sum_{t\leq n} [f_t(\theta + \phi_t(\theta)u) - f_n(\theta)]^2\Big) \leq (\mathrm{pol}(R))^{1/2},$$
for all $\theta \in K$ and $u \in \Gamma_{n,\theta,R}$ by (N.2) and (N.3), so that Δ_n is bounded in n.
2. One might argue that Theorem 3.1 is of little value in applications because, in practice, one never knows the exact value of Λ_1. Indeed, when analyzing real data, we may as well set $\Lambda_1 = \infty$; the meaning of condition (N.1) is, of course, that it gives the theorem a certain robustness: nothing terrible happens when $\Lambda_1 < \infty$.
3. In practice, the constant ρ will usually be equal to 1 [a counterexample is provided by $f_t(\theta) = \theta^p$, $0 < p < 1$ and $\Theta = [-1, 1]$; the reparametrization $\theta^p =: \tau$ makes $\rho = 1$ again].
4. The polynomial pol(R) seems to be unimportant in applications; however, it saved us the two extra constants m_1 and M_1 used in Theorem 1.5.1 of Ibragimov and Has'minskii (1981).
5. Finally, a natural choice for the function $g_n(R)$ seems to be a quadratic function and for K we might [out of the context of Ibragimov and Has'minskii (1981)], as well choose the set Θ. To obtain simple conditions, we restrict ourselves to the case that ϕ_n does not depend on θ.

These considerations have motivated the following theorem:

THEOREM 3.2. *Let, for a suitable sequence of normalizing matrices ϕ_n, the following conditions be fulfilled*:

(N.1′) *For some* γ, *condition* (N.1) *holds with* $\Lambda_1 = \infty$.
(N.4) *Let there exist positive constants* D_1 *and* D_2 *such that, for all* $\theta, \theta' \in \Theta$ *and n large enough*,

$$D_1|\phi_n^{-1}(\theta - \theta')|^2 \le \sum_{t \le n} [f_t(\theta) - f_t(\theta')]^2 \le D_2|\phi_n^{-1}(\theta - \theta')|^2.$$

Then the following LLD holds for the LS estimator $\hat{\theta}_n$: *There exist constants* B_0 *and* b, *such that, for all n and H large enough*,

$$\sup_{\theta \in \Theta} \mathbb{P}_\theta^{(n)}\left\{\left|\phi_n^{-1}(\hat{\theta}_n - \theta)\right| \ge H\right\} \le B_0 \exp(-bH^2).$$

Moreover, for any $\beta > 0$ *we can choose* B_0 *such that*

$$b \ge D_1/(8\gamma(1 + k)) - \beta.$$

PROOF. To apply Theorem 3.1, let us verify its conditions. (N.1) holds by assumption; by (N.4), (N.2) holds with $\rho = 1$ and $\mathrm{pol}(R) = D_2$. By (N.4) and (N.1), (N.3) holds for any $\delta \in (0, \frac{1}{2})$, with the choice $\Delta_n := 2\gamma\delta^{-2}$ and $g_n(R) := (D_1/2\gamma\delta^{-2})R^2$. Now apply Theorem 3.1 and let $\delta \to \frac{1}{2}$. □

Theorem 3.2 extends a result of Ivanov (1976), namely his LD Lemma 1. It generalizes the result of Prakasa Rao (1984). His theorem follows immediately from ours. In Section 4 we give an example to show that our generalization is not void.

4. Examples and concluding remarks. In this section, we present some examples of the application of Theorem 3.2. Recall that two sequences of positive numbers (a_n) and (b_n) are called (asymptotically) equivalent (write $a_n \asymp b_n$) if there exist positive constants C_1 and C_2 such that $C_1 b_n \le a_n \le C_2 b_n$ for all n (large enough). In the same manner, we call a parametrized family of positive sequences $\{(a_n(\theta));\ \theta \in \Theta\}$ (asymptotically) uniformly equivalent to a positive sequence (b_n) if there exist positive constants C_1 and C_2 such that, for all n (large enough), the inequality $C_1 b_n \le a_n(\theta) \le C_2 b_n$ holds. We shall write $a_n(\theta) \asymp b_n$ (uniformly in θ). These definitions can, in an obvious manner, be generalized to sequences of positive definite symmetric matrices $(A_n;\ n = 1, 2, \dots)$. We say that $A_n \ge B_n$ if the difference is a positive semidefinite matrix.

Examples 1 and 2 are provided by the Michaelis–Menten model, which is used to describe the relation between the velocity v of an enzyme reaction and the concentration c of the substrate. The parameters are M, the maximal reaction velocity, and K, the chemical affinity. The parameter set Θ of the (K, M) is a bounded open set in the positive quadrant. The model is

$$v(c; K, M) = \frac{Mc}{K + c}. \tag{4.1}$$

We shall consider fixed designs c given by concentrations $c_1, c_2, \dots, c_n$, where $c_n \to 0$ as $n \to \infty$. At each concentration c_t an independent measurement of the velocity is taken, giving the data $X_1, X_2, \dots, X_n$:

$$X_t = v_t(K, M) + \varepsilon_t = \frac{Mc_t}{K + c_t} + \varepsilon_t, \tag{4.2}$$

where the ε_t are independent centered errors satisfying condition (N.1′) of Theorem 3.2 for some γ.

EXAMPLE 1. Consider the following simple model, which is obtained from (4.1) by assuming that K/M is known (put $K/M = 1$, without loss of generality) and putting $c_t = t^{-1/4}$. This model can be written as

$$f_t(\theta) := \frac{1}{K^{-1} + t^{1/4}}, \qquad t = 1, 2, 3 \dots. \tag{4.3}$$

Note that, for this model, the conditions of Jennrich (1969), Ivanov (1976) and, in particular, Prakasa Rao (1984), do not hold.

One has

$$\sum_{t \le n} \left(f_t(K) - f_t(K') \right)^2 = |K^{-1} - K'^{-1}|^2 C_n(K, K'), \tag{4.4}$$

where

$$C_n(K, K') := \sum_{t \le n} 1 / \left[(K^{-1} + t^{1/4})(K'^{-1} + t^{1/4}) \right]^2 \tag{4.5}$$

and it is easily shown that the sequence $C_n(K, K') \asymp \log n$, uniformly in K, K'. It follows in particular that, for n large enough (as usual),

$$\sum_{t \le n} \left(f_t(K) - f_t(K') \right)^2 \ge D_1 |K - K'|^2 \log n, \tag{4.6}$$

where D_1 can be chosen arbitrarily close (from below) to $1/(\sup K)^4$. Now we can apply Theorem 3.2, which yields

$$\sup_{K \in \Theta} \mathbb{P}_K^{(n)} \left\{ (\log n)^{1/2} |\hat{K}_n - K| \ge H \right\} \le B_0 \exp(-bH^2), \tag{4.7}$$

where b can be chosen arbitrarily close (from below) to $1/16\gamma(\sup K)^4$. We remark that, in the case of i.i.d. disturbances ε_t, the strong consistency of the LS-estimator for this model can be demonstrated by Theorem 3 of Wu (1981). By Theorem 5 of the same author, it is also asymptotically normal:

$$(\log n)^{1/2} \left(\hat{K}_n - K \right) \to \mathcal{N}(0, \sigma^2 K^4), \tag{4.8}$$

where σ^2 is the variance of the i.i.d. ε_t.

Of course, the results (4.7) and (4.8) do not imply each other. But information on the quality of our bound $1/16\gamma(\sup K)^4$ for b can be obtained by considering the following quantity [compare Sievers' definition of the inaccuracy rate; see Kester (1985), Chapter 1, Definition 1.1]:

$$b_1(\theta) := \liminf_{n \to \infty,\, H \to \infty} -H^{-2} \log \mathbb{P}_\theta^{(n)} \left\{ (\log n)^{1/2} |\hat{\theta}_n - \theta| \ge H \right\}.$$

From (4.7) it follows that $b_1(K) \geq 1/16\gamma(\sup K)^4$, whereas (4.8) yields $b_1(K) = \frac{1}{2}\sigma^2 K^4$. In the case that the ε_t are Gaussian, γ equals σ^2 and the bound $1/16\gamma$ is at most a factor 8 $(\sup K)^4/(\inf K)^4$ too pessimistic. This is a consequence of the approximations made in Lemma A.3 and the proof of Theorem 2.1.

Our bound may be improved by using the apparently more natural parametrization $L := K^{-1}$. Then (4.6) continues to hold with K replaced by L and D_1 arbitrarily close to 1. Consequently, (4.7) and (4.8) yield $b_1(L) \geq 1/16\gamma$ and $b_1(L) = 1/2\sigma^2$, respectively. Our bound is then a factor 8 too pessimistic, uniformly over Θ.

EXAMPLE 2. Now we consider the model (4.1) in its full generality. One has

$$v_t(K', M') - v_t(K, M) = a_t(M'/K' - M/K) + b_t(M' - M), \tag{4.9}$$

where

$$\begin{aligned} a_t(K, K') &:= KK'c_t/(K + c_t)(K' + ct), \\ b_t(K, K') &:= c_t^2/(K + c_t)(K' + c_t), \end{aligned} \tag{4.10}$$

which suggests the reparametrization $(K, M) \to (L, M)$ with $L := M/K$ (compare $L := 1/K$ in Example 1). Note that the transform of Θ is again bounded and open in the positive quadrant. Putting

$$B_n(K, K') := \begin{bmatrix} \sum_{t \leq n} a_t(K, K')^2 & \sum_{t \leq n} a_t(K, K')b_t(K, K') \\ \sum_{t \leq n} a_t(K, K')b_t(K, K') & \sum_{t \leq n} b_t(K, K')^2 \end{bmatrix} \tag{4.11}$$

and $\Delta := \operatorname{col}\{L' - L, M' - M\}$ we have

$$\sum_{t \leq n} [v_t(K', M') - v_t(K, M)]^2 = \Delta^T B_n(K, K')\Delta. \tag{4.12}$$

Now we make the following assumptions on the design sequence:

$$\sum_1^\infty c_t^4 = \infty, \tag{4.13}$$

$$\liminf_{n \to \infty} r_n > \sup_\Theta (K_1/K_2)^2, \tag{4.14}$$

where r_n is defined by

$$r_n := \sum_{t \leq n} c_t^2 \sum_{t \leq n} c_t^4 \Big/ \Big(\sum_{t \leq n} c_t^3 \Big)^2. \tag{4.15}$$

Observe that these assumptions are easily checked if, e.g., $c_t \simeq t^{-p}$. In the case that $0 < p < 1/4$, the left-hand side of (4.14) is equivalent to $1 + 1/(1 - 2p)(1 - 4p)$; hence (4.14) can be fulfilled by choosing p close enough to $1/4$. Assumption (4.14) is always fulfilled if $p = 1/4$.

We show that under assumptions (4.13) and (4.14) the family $B_n(K, K')$ is uniformly equivalent. First, note that

$$\begin{aligned} a_t &= c_t + O(c_t^2), \\ b_t &= c_t^2/KK' + O(c_t^3), \end{aligned} \tag{4.16}$$

where all our Landau symbols are valid uniformly over the range of (K, K'). Next apply Lemma A.5(ii): The traces and determinants mentioned in this lemma can be expressed as quotients of sequences s_n defined by

$$\begin{aligned} s_n(K_1, K_2, K_3, K_4) := &\sum_{t\le n} a_t(K_1, K_2)^2 \sum_{s\le n} b_s(K_3, K_4)^2 \\ &- \sum_{t\le n} a_t(K_1, K_2) b_t(K_1, K_2) \sum_{s\le n} a_s(K_3, K_4) b_s(K_3, K_4), \end{aligned} \tag{4.17}$$

for various values of the parameters K_i. Hence, it suffices that these sequences be uniformly equivalent.

Using (4.13) and (4.16) it follows that

$$\sum_{t\le n} a_t^2 = \sum_{t\le n} c_t^2(1 + o_n(1)) \tag{4.18}$$

and the like for Σb_t^2 and $\Sigma a_t b_t$. This leads to

$$\begin{aligned} s_n(K_1, K_2, K_3, K_4) = &\Big(\sum_{t\le n} c_t^3/K_3K_4 \Big)^2 \big(r_n(1 + o_n(1)) \\ &-(K_3K_4/K_1K_2)(1 + o_n(1))\big), \end{aligned} \tag{4.19}$$

and together with (4.14) uniform equivalence follows: Fixing arbitrary values of K and K', say K_0 and K_0', we have, uniformly,

$$B_n(K, K') \asymp B_n(K_0, K_0'), \tag{4.20}$$

whence condition (N.4) holds for some choice of constants D_1 and D_2 [which can be obtained from Lemma A.5(ii)] and $\phi_n := B_n(K_0, K_0')^{-1/2}$. Application of Theorem 3.2 yields

$$\sup \mathbb{P}_{K,M}^{(n)}\left\{ |\phi_n^{-1}\mathrm{col}\{\hat{L} - L, \hat{M} - M\}| \ge H \right\} \le B_0 \exp(-bH^2), \tag{4.21}$$

where b can be chosen arbitrarily close (from below) to $D_1/24\gamma$. A similar inequality can be derived for the pair of estimators $(\hat{K}, \hat{M})$ but, as in Example 1, the bounds for b are of poorer quality.

EXAMPLE 3. Consider the linear model

$$X_t = \theta + \varepsilon_t, \qquad t = 1, 2, \ldots, n, \tag{4.22}$$

where the ε_t are i.i.d. standard normal variables. One immediately obtains

$$\mathbb{P}_\theta^{(n)}\{ n^{1/2}|\hat{\theta}_n - \theta| \ge H \} \le (2/\pi)^{1/2} \exp(-bH^2/2). \tag{4.23}$$

For b we can take any value ≤ 1. Theorem 3.2 allows us to take any $b < 1/16$, which is a factor 16 too pessimistic. No other estimator can improve the value $b = 1$ [see Kester (1985), Chapter 2, Example 1.1].

In Section 3, we applied the very general Theorem 2.1 to the problem of least-squares estimation. It would be nice to try our method on other M-estimators, e.g., the Huber estimators in nonlinear regression, i.e., estimators maximizing a functional of the form

$$C_n(X^n, \theta) := -\sum_{t \le n} \Psi(X_t - f_t(\theta)) \tag{4.24}$$

and to compare our bound for b with the exact rate of convergence obtained by Kester (1985) in the case that ε_t are i.i.d. and θ is a location parameter, i.e., $f_t(\theta) = \theta$. For details see Kester [(1985), Chapter 2.4b, Theorem 4.2].

However, we wish to point out that there are also situations where our Theorems 2.1 and 3.1 do not apply. For instance, consider the power model $f_t(\theta) = t^{-\theta}$, $\theta \in \Theta := [0, a]$, where $a \le \frac{1}{2}$. This model is also discussed by Wu (1981), who shows that the LS-estimator is strongly consistent.

Our theorems do not apply because the rate of growth (in n) of $\Sigma_{t \le n}(f_t(\theta) - f_t(\theta'))^2$ depends on θ and θ', whereas our theory assumes a uniform growth rate in n. Hence, a suitable norming $\phi_n(\theta)$ does not exist for this example [Has'minskii (1986)]. An extension of Theorem 2.1 to a theorem with more flexible normings would meet this difficulty and would also contribute to Ibragimov and Has'minskii's theory.

APPENDIX

In this appendix, we list the lemmata we used in the paper.

Lemma A.1. *Let $(\mathscr{X}, \mathscr{U})$ be a measurable space and let $\{\mathbb{P}_\theta;\ \theta \in \Theta\}$ be a family of probability measures on $(\mathscr{X}, \mathscr{U})$, where Θ is a Borel subset of $\mathbb{R}^k$. Let C be a real function from $\mathscr{X} \times \Theta$ to $[0, \infty)$ which is, for all $X \in \mathscr{X}$, a positive continuous function of θ and, for each $\theta \in \Theta$, a $(\mathscr{U}, \mathscr{B})$-measurable function of X. Finally, let Θ^0 be a subset of Θ which has a countable subset D which is dense in Θ^0. Then the following assertions hold*:

(i) *the random variable $S(X) := \sup_{\theta \in \Theta^0} C(X, \theta)$ is $\mathscr{U}$-measurable*;

(ii) *if Θ is compact then, for any X, the equation in t,*

$$\sup_{\theta \in \Theta} C(X, \theta) = C(X, t), \tag{A.1}$$

has a solution [which we denote $\hat{\theta}(X)$], which is $\mathscr{U}$-measurable;

(iii) *if, for arbitrary (noncompact) Θ the existence of a solution to* (A.1) *is assumed, then there exists a measurable version $\hat{\theta}(X)$ of this solution.*

Proof. (i) See Schmetterer (1974), Chapter 5.3, Lemma 3.2, page 307. We observe that any subset Θ^0 of $\mathbb{R}^k$ has a countable subset D which is dense in the closure Θ^0.

(ii) See Schmetterer (1974), Chapter 5.3, Lemma 3.3, page 307ff. or Jennrich (1969), Lemma 2.

(iii) The set Θ is Borel, whence it is possible to approximate it by an increasing sequence of compact sets $K_i \uparrow \Theta$. Let $\Theta(X)$ be the set of the θ solving (A.1). Let $i^*(X)$ be the smallest i such that $K_i \cap \Theta(X) \neq \varnothing$. Then i^* is finite by assumption; it is also measurable, which can be seen as follows.

Let D be a countable dense subset of Θ. Then the event $\{i^* > n\}$ can be written as

$$\left\{X: \bigcap_{i=1}^{n} \bigcup_{k=1}^{\infty} \bigcup_{\tau \in D} \sup_{\theta \in K_i} C(X, \theta) \leq C(X, \tau) - k^{-1}\right\},$$

which is clearly measurable by part (i) of this lemma. Then

$$\sup_{K_{i^*}} C(X, \theta) = \sup_{\Theta} C(X, \theta) \tag{A.2}$$

and also, because the K_i are compact, the equation in t

$$\sup_{K_i} C(X, \theta) = C(X, t) \tag{A.3}$$

has a measurable solution $t = \hat{\theta}_i(X)$ for each i, as is seen by application of part (ii) of this lemma. Combining (A.2) and (A.3) it follows that $\hat{\theta}_{i^*}(X)$ provides a solution to (A.1), which is measurable because i^* is measurable. □

LEMMA A.2. *Let the quantities C, Z, $\hat{\theta}_\varepsilon$, etc., be defined as in Section 2. Then the following inequality holds*:

$$\mathbb{P}_\theta^{(\varepsilon)}\left\{\left|\phi(\varepsilon, \theta)^{-1}\left(\hat{\theta}_\varepsilon - \theta\right)\right| \geq H\right\} \leq \mathbb{P}_\theta^{(\varepsilon)}\left\{\sup_{\substack{|u| \geq H \\ u \in U_{\varepsilon, \theta}}} Z_{\varepsilon, \theta}(u) \geq 1\right\}. \tag{A.4}$$

PROOF. See Ibragimov and Has'minskii (1981), Chapter 1.5 and Wu (1981), Lemma 1. □

LEMMA A.3. *Let $\zeta(u)$ be a real-valued function defined on a closed subset Γ of the Euclidean space $\mathbb{R}^k$, which is measurable and separable. Let the following condition be fulfilled: There exists numbers $m \geq \alpha > k$ and a function $C: \mathbb{R}^k \to \mathbb{R}$, bounded on compact sets, such that for all $u, v \in \Gamma$*

(i) $$\mathbb{E}|\zeta(u)|^m \leq C(u),$$

(ii) $$\mathbb{E}|\zeta(u) - \zeta(v)|^m \leq C(u)|u - v|^\alpha.$$

Then a.s. the realizations of $\zeta(u)$ are continuous functions on Γ. Moreover, set

$$\omega(h, \zeta, L) := \sup|\zeta(u) - \zeta(v)|,$$

where the sup *is taken over all $u, v \in \Gamma$ with $|u - v| \leq h$, $|u| \leq L$, $|v| \leq L$. Then*

$$\mathbb{E}\omega(h, \zeta, L) \leq B\left(\sup_{|u| \leq L} C(u)\right)^{1/m} L^{k/m} h^{(\alpha - k)/m},$$

where B is a constant depending on m, α and k.

PROOF. See Ibragimov and Has'minskii (1981), page 372ff., where in (8) L^k should be replaced by $L^{k/m}$ (printing error). □

LEMMA A.4. *Let $Y_1, Y_2, \dots, Y_n$ be independent random variables. Let $d_1, \dots, d_n$ be reals and let $S_n := \sum_{i \le n} d_i Y_i$. Suppose there exist positive constants γ_i, $i = 1, 2, \dots, n$, and Λ_1 (Λ_1 possibly ∞) such that, for all $\lambda \in [-\Lambda_1, \Lambda_1]$ and $t = 1, 2, \dots, n$, one has*

$$\mathbb{E} \exp(\lambda Y_t) \le \exp\left(\tfrac{1}{2}\gamma_t \lambda^2\right). \tag{A.5}$$

Write $G := \sum_{i \le n} \gamma_i d_i^2$ and $\Lambda := \Lambda_1 / \max\{|d_1|, \dots, |d_n|\}$. Then

$$\mathbb{P}\{S_n \ge x\} \le \exp\left(-\min\{x^2/2G, \Lambda x/2\}\right). \tag{A.6}$$

The same inequalities hold if we replace S_n by $-S_n$.

PROOF. This lemma is a simple extension of Theorem 16 of Petrov (1975), Chapter 3.4. □

LEMMA A.5. *Let $\{\Psi_n,\ n \in \mathbb{N}\}$ be a sequence of positive definite symmetric matrices and let $\mathscr{M} := \{M_n(K)\colon K \in \mathbf{K},\ n \in \mathbb{N}\}$ be a family of sequences of positive definite symmetric matrices indexed by the parameter K. For all K in $\mathbf{K}$ define the sequence*

$$R_n(K) := \Psi_n^{-1/2} M_n(K) \Psi_n^{-1/2}, \qquad n \in \mathbb{N}. \tag{A.7}$$

Then the following assertion holds: The family $\mathscr{M}$ is uniformly equivalent (for a definition see Section 4) to the sequence Ψ_n iff there exists an interval $I := [\alpha, \beta]$, with $\beta > \alpha > 0$, such that for all $n \in \mathbb{N}$ and all $K \in \mathbf{K}$, the spectrum of $R_n(K)$ is contained in the interval I.

REMARKS. (i) For Ψ_n we may always take $M_n(K_0)$, where K_0 is an arbitrary, but fixed, element in $\mathbf{K}$;

(ii) if all $M_n(K)$ are of size 2×2, then it is also necessary and sufficient that the trace and determinant of $R_n(K)$ remain in some fixed positive interval for all n and K. In fact, one has, for any K,

$$\left(\inf_{\mathbf{K}} \det R_n(K)/\operatorname{tr} R_n(K)\right)\Psi_n \le M_n(K) \le \left(\sup_{\mathbf{K}} \operatorname{tr} R_n(K)\right)\Psi_n. \tag{A.8}$$

PROOF. If $\mathscr{M} \simeq \Psi_n$, then there exists an $\alpha > 0$ such that, for all K and n,

$$\alpha\Psi_n \le M_n(K) \le \beta\Psi_n. \tag{A.9}$$

Now let x be any eigenvector of $R_n(K)$ and sandwich (A.9) between $\Psi_n^{-1/2}x$ and its transpose; this yields $\alpha \le \lambda \le \beta$, where λ is the eigenvalue belonging to x. On the other hand, from eigenvectors of $R_n(K)$ one may form an orthonormal basis of $\mathbb{R}^n$ so the converse reasoning also holds. □

Acknowledgments. We acknowledge C. Scheffer for his helpful advice and L. Lekx for her careful manipulation of the text. We thank the referee, whose remarks have substantially improved the paper.

REFERENCES

DZHAPARIDZE, K. O. (1986). On asymptotic inference about intensity parameters of a counting process. Report, Centre for Mathematics and Computer Science, Amsterdam.

HAS'MINSKII, R. Z. (1986). Personal communication.

IBRAGIMOV, I. A. and HAS'MINSKII, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*. Springer, New York.

INGSTER, YU. I. (1984). Asymptotic regularity of a family of measures corresponding to a Gaussian random process which contains a white noise component for a parametric family of spectral densities. *J. Soviet Math.* **25** 1165–1181.

IVANOV, A. V. (1976). An asymptotic expansion for the distribution of the least-squares estimator of the nonlinear regression parameter. *Theory Probab. Appl.* **21** 557–570.

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

KESTER, A. D. M. (1985). Some large deviation results in statistics. CWI Tract 18, Centre for Mathematics and Computer Science, Amsterdam.

LÄUTER, H. (1985). Strong consistency of the least squares estimator in nonlinear regression. Preprint, Akad. der Wissensch. der DDR, Berlin.

NEVEU, J. (1970). *Calcul des Probabilités*. Masson, Paris.

PETROV, V. V. (1975). *Sums of Independent Random Variables*. Springer, Berlin.

PRAKASA RAO, B. L. S. (1984). On the exponential rate of convergence of the least squares estimator in the nonlinear regression model with Gaussian errors. *Statist. Probab. Lett.* **2** 139–142.

SCHMETTERER, L. (1974). *Introduction to Mathematical Statistics*. Springer, Berlin.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

VAN DE GEER, S. (1986). On rates of convergence in least squares estimation. Report, Centre for Mathematics and Computer Science, Amsterdam.

VOSTRIKOVA, L. JU. (1984). On criteria for $c(n)$-consistency of estimators. *Stochastics* **11** 265–290.

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

DEPARTMENT OF MATHEMATICS
DELFT UNIVERSITY OF TECHNOLOGY
JULIANALAAN 132
2628 BL DELFT
THE NETHERLANDS

CENTRE FOR MATHEMATICS AND COMPUTER SCIENCE
KRUISLAAN 413
1098 SJ AMSTERDAM
THE NETHERLANDS