The Annals of Statistics
1980, Vol. 8, No. 5, 1057–1064

A NOTE ON STRONG CONSISTENCY OF LEAST SQUARES ESTIMATORS IN REGRESSION MODELS WITH MARTINGALE DIFFERENCE ERRORS

BY PAUL I. NELSON

Bucknell University and The Pennsylvania State University

Conditions for the pointwise consistency of weighted least squares estimators from multivariate regression models with martingale difference errors are given in terms of the relative rates at which certain quadratic forms diverge to infinity.

1. Introduction. This note concerns the pointwisè consistency of weighted least squares estimators from multivariate regression models with dependent error terms. The focus is on models where the error terms are martingale differences.

Let $\{y_t\}_{t=0}^{\infty}$ be an integrable real $r \times 1$ vector valued discrete parameter stochastic process whose distribution depends on an unknown $p \times 1$ vector of parameters θ^*, which is assumed to be an interior point of an open subset $\Theta = \{\theta\}$ of Euclidean p-space. Let $E_\theta(\cdot)$ and $E_\theta(\cdot|\cdot)$ denote expectation and conditional expectation respectively under the distribution determined by θ. When $\theta = \theta^*$ (the "true" value) the subscript will be omitted. A regression model may be created by writing for $t = 0, 1, \cdots$,

$$\begin{aligned} (1.1) \qquad y_{t+1} &= E_\theta(y_{t+1}|F_t) + (y_{t+1} - E_\theta(y_{t+1}|F_t)) \\ &\equiv g(\theta, F_t) + u_{t+1}(\theta), \end{aligned}$$

where F_t is a subsigma field generated by an arbitrary subset of $\{y_j, j \leqslant t\}$ for $t \geqslant 0$. The least squares estimators considered here are obtained by minimizing, with respect to θ, the weighted sum of squares.

$$Q_n(\theta) = \Sigma_{t=0}^{n-1}(y_{t+1} - g(\theta, F_t))'S_t(y_{t+1} - g(\theta, F_t)),$$

where $\{S_t\}$, called the "weights sequence", consists of $r \times r$ positive semidefinite symmetric matrices which are free of θ and with S_t measurable with respect to $F_t, t \geqslant 0$.

Call the sequence $\{\hat{\theta}_n\}$ "a strongly consistent sequence of least squares estimators on the event A" if $\{\hat{\theta}_n\} \to \theta^*$ a.e. on A; and for some sequence of weights $\{S_t\}$, for any $\varepsilon > 0$ there is a positive integer n_0 such that for all $n \geqslant n_0, Q_n(\hat{\theta}_n)$ is a local minimum of $Q_n(\cdot)$ on part of A with probability at least $P(A) - \varepsilon$. The existence of such a sequence follows from

$$(1.2) \quad \liminf_{\delta \to 0} \liminf_{n\to\infty} \inf_{\|\theta-\theta^*\|=\delta}(Q_n(\theta) - Q_n(\theta^*)) > 0 \qquad \text{a.e. on } A,$$

which implies that a.e. on A $Q_n(\cdot)$ ultimately in n (random) has a local minimum in any neighborhood of θ^*. The strongly consistent least squares estimators can then

Received September 1978; February 1979.
AMS 1970 *subject classifications.* Primary 62J05; secondary 62M10.
Key words and phrases. Consistency, least squares, time series, martingale.

be constructed by using Egoroff's theorem as in Theorem 2.1 and its corollary of Klimko and Nelson (1978).

Klimko and Nelson (1978) obtained (1.2) in the unweighted univariate case by using the Taylor expansion

$$Q_n(\theta) = Q_n(\theta^*) + (\theta - \theta^*)' \frac{\partial Q_n(\theta^*)}{\partial \theta} + \tfrac{1}{2}(\theta - \theta^*)' \frac{\partial^2 Q_n(\theta^*)}{\partial \theta^2}(\theta - \theta^*) + R_n(\theta) \tag{1.3}$$

to find conditions under which the remainder $R_n\{\theta\}$ is negligible and such that there is a sequence $\{k_n\}$ of positive values converging to zero for which

$$\lim_{n\to\infty} k_n \frac{\partial Q_n(\theta^*)}{\partial \theta} = 0^{p\times 1} \text{ a.e.}, \tag{1.4}$$

$$\lim_{n\to\infty} k_n \frac{\partial^2 Q_n(\theta^*)}{\partial \theta^2} = V^{p\times p} \text{ a.e.}, \tag{1.5}$$

a positive definite matrix of constants. Theorem 2.1 of this paper extends this approach, which was mainly designed to handle stationary ergodic processes $\{y_t\}$; first, by replacing (1.5) with a condition concerning the relative rates at which the eigenvalues of a sequence of matrices, whose elements are sums of products of partial derivatives of the regression functions $\{g(\theta, F_t)\}$ with respect to θ, diverge to infinity. The condition is given in Lemma 2.1. It is seen to hold in Example 2.1 for a linear regression model to which the results of Klimko and Nelson (1978) do not apply since (1.5) fails to hold for any sequence $\{k_n\}$. In addition, when the error terms $\{u_t(\theta^*)\}$ are martingale differences, (1.4) is replaced here by using a result of Neveu (1965, page 150) which implies the convergence to zero of a martingale divided by the sum of the conditional variances of its difference sequence a.e. on the set where that sum diverges. Some necessary conditions for the existence of strongly consistent least squares estimators are also given in Theorem 2.1. A feature of the approach taken here is that it allows consistency to hold with probability less than one. This is convenient, for example, in working with branching processes where asymptotic results do not hold on the set of extinction. See Example 2.2.

Jennrich (1969) investigated consistency for univariate nonlinear nonrandom regression functions with i.i.d. error terms. His approach avoids the differentiability and remainder assumptions of (1.3) by getting at (1.2) through an argument that amounts to an examination of the terms of the decomposition

$$Q_n(\theta) = Q_n(\theta^*) + 2\Sigma_{t=0}^{n-1}(g(\theta^*, F_t) - g(\theta, F_t))' S_t u_{t+1}(\theta^*) + \Sigma_{t=0}^{n-1}(g(\theta^*, F_t) - g(\theta, F_t))' S_t (g(\theta^*, F_t) - g(\theta, F_t)). \tag{1.6}$$

In lieu of (1.5), Jennrich assumes that $(1/n)$ times the last term in (1.6) converges uniformly in θ to a function of θ which has a unique minimum at θ^* and uses a special case of the aforementioned martingale convergence theorem of Neveu to show that $(1/n)$ times the cross product term of (1.6) converges to zero a.e., which is the analogue of (1.4).

Many others have worked on the problem of consistency for least squares estimators in a wide variety of models. We mention only a few of the recent papers in this area. Dunsmuir and Hannan (1976) and Nicholls (1976) deal with stationary vector valued autoregressive moving average processes. Robinson (1972) considers asymptotically stable vector valued nonlinear regression functions which are functions of exogeneous variables only and have a stationary error sequence. Phillips (1976) compares least squares estimators with what are called quasimaximum likelihood estimators in similar models with independent error terms. Robinson (1977) treats univariate processes which are quadratic moving averages. The results of Anderson and Taylor (1976a,b) and Lai, Robbins and Wei (1978) are compared to those presented here directly after the proof of Theorem 2.1.

2. Consistency. Henceforth, let A denote an event and assume that the components $g_m(\theta, F_t)$, $m = 1, 2, \cdots, r$ of the regression vectors $g(\theta, F_t)$ are a.e. continuously differentiable with respect to $\theta, \theta \in \Theta, t \geqslant 0$. Define a.e.

$$\nabla_t^{r\times p} = \left(\frac{\partial g_m(\theta^*, F_t)}{\partial \theta_j}\right)_{m\leqslant r, j\leqslant p} \qquad t \geqslant 0.$$

Let $d = d(\theta) = \theta - \theta^*$ for a general $\theta \in \Theta$. With $R_n(\theta)$ an appropriate remainder term, expansion (1.3) may be written as

$$\begin{aligned}
(2.1)\qquad Q_n(\theta) &= Q_n(\theta^*) + (\theta - \theta^*)'\frac{\partial Q_n(\theta^*)}{\partial \theta} \\
&\quad + (\theta - \theta^*)'(\Sigma_{t=0}^{n-1} \nabla t' S_t \nabla_t)(\theta - \theta^*) + R_n(\theta) \\
&= Q_n(\theta^*) - 2(\theta - \theta^*)'\Sigma_{t=0}^{n-1} \nabla_t' S_t u_{t+1}(\theta^*) \\
&\quad + (\theta - \theta^*)'(\Sigma_{t=0}^{n-1} \nabla_t' S_t \nabla_t)(\theta - \theta^*) + R_n(\theta) \\
&= Q_n(\theta^*) - 2\Sigma_{t=0}^{n-1} v_t(d) + \Sigma_{t=0}^{n-1} b_t^2(d) + R_n(\theta) \\
&= Q_n(\theta^*) + \Sigma_{t=0}^{n-1} b_t^2(d)\big(-2\Sigma_{t=0}^{n-1} v_t(d)/ \\
&\quad \Sigma_{t=0}^{n-1} b_t^2(d) + 1 + R_n(\theta)/\Sigma_{t=0}^{n-1} b_t^2(d)\big) \\
&\qquad\qquad \text{a.e. on } \{\Sigma_{t=0}^{n-1} b_t^2(d) > 0\},
\end{aligned}$$

where we have set

$$v_t(d) \equiv d' \nabla_t' S_t u_{t+1}(\theta^*), \qquad t \geqslant 0,$$

$$b_t^2(d) \equiv d' \nabla_t' S_t \nabla_t d, \qquad t \geqslant 0.$$

In view of the preceding discussion, (2.1) and the following conditions imply (1.2) and hence the existence of a sequence of strongly consistent least squares estimators.

$$(2.2)\qquad \liminf_{n\to\infty} \Sigma_{t=0}^{n-1} b_t^2(d) > 0 \text{ a.e. on } A \text{ for all } d \neq 0$$

$$(2.3)\qquad \limsup_{\delta\to 0} \limsup_{n\to\infty} \sup_{\|d\|=\delta} \Sigma_{t=0}^{n-1} v_t(d)/\Sigma_{t=0}^{n-1} b_t^2(d) \leqslant 0 \text{ a.e. on } A,$$

$$(2.4)\qquad \liminf_{\delta\to 0} \liminf_{n\to\infty} \inf_{\|d\|=\delta} R_n(\theta)/\Sigma_{t=0}^{n-1} b_t^2(d) \geqslant 0 \text{ a.e. on } A.$$

A martingale convergence theorem will be used to obtain (2.3) for fixed d and a constraint on the rate of growth of quadratic forms in the sequence of $p \times p$ matrices $\{\Sigma_{t=0}^{n-1} \nabla_t' S_t \nabla_t\}$ will be used to make the limit uniform in $\delta = \|\theta - \theta^*\|$.

Let

$$e_n(\text{min. (max.)}) = \text{min. (max.) eigenvalue} \, \Sigma_{t=0}^{n-1} \nabla_t' S_t \nabla_t, \qquad n \geq 1,$$

$$\rho_n = e_n(\text{max.})/e_n(\text{min.}), \qquad n \geq 1.$$

LEMMA 2.1. *If either of the following holds,*

(i) $\Sigma_{t=0}^{n-1} v_t(d)/\Sigma_{t=0}^{n-1} b_t^2(d) \to 0$ *for all* $d \neq 0$ *and* $\limsup_{n\to\infty} \rho_n < \infty$ a.e. *on* A,

(ii) $\Sigma_{t=0}^{n-1} v_t(d)/e_n(\text{min.}) \to 0$ a.e. *on* A *for all* $d \neq 0$,

then, (2.3) *holds.*

PROOF. Given $\delta > 0$, let $d(j)$ denote the $p \times 1$ vector with δ in the jth place and zeros elsewhere. Then, for all d, $\|d\| = \delta$; a.e. on A

$$\begin{aligned}
|\Sigma_{t=0}^{n-1} v_t(d)|/\Sigma_{t=0}^{n-1} b_t^2(d) &\leq |\Sigma_{t=0}^{n-1} v_t(d)|/e_n(\text{min})\,\delta^2 \\
&\leq \Sigma_{j=1}^{p} |\Sigma_{t=0}^{n-1} v_t(d(j))|/e_n(\text{min.})\,\delta^2 \\
&\leq \rho_n \Sigma_{j=1}^{p} |\Sigma_{t=0}^{n-1} v_t(d(j))|/e_n(\text{max.})\,\delta^2 \\
&\leq (2 \limsup \rho_n) \Sigma_{j=1}^{p} |\Sigma_{t=0}^{n-1} v_t(d(j))|/e_n(\text{max.})\,\delta^2 \\
&\leq (2 \limsup \rho_n) \Sigma_{j=1}^{p} |\Sigma_{t=0}^{n-1} v_t(d(j))|/|\Sigma_{t=0}^{n-1} b_t^2(d(j))|,
\end{aligned}$$

where the next to last inequality holds ultimately in n. If (ii) holds uniform convergence follows from the second inequality; if (i) holds it follows from the last.

For linear regression models

$$\frac{\partial^2 Q_n(\theta^*)}{\partial \theta^2} = 2\Sigma_{t=0}^{n-1} \nabla_t' S_t \nabla_t \text{ a.e.}$$

In such cases (1.5) implies that k_n times the eigenvalues of $\Sigma_{t=0}^{n-1} \nabla_t' S_t \nabla_t$ converge to positive constants and hence that $\limsup_{n\to\infty} \rho_n < \infty$ a.e. The following example shows that the converse implication is not valid.

EXAMPLE 2.1. Let $y_t = \theta_1 x_{t1} + \theta_2 x_{t2} + u_t, t = 1, 2, \cdots$, be a univariate ($r = 1$) linear model where the error sequence $\{u_t\}$, whose distribution is free of θ, and the σ fields $\{F_t\}$ are such that $E(u_t | F_{t-1}) = 0$ a.e., $t = 1, 2, \cdots,$. Take $x_{11} = 1$ and set $x_{t1} = 0$ if t is even and $x_{t1} = (2^{t-1} - \Sigma_{j=1}^{t-1} x_{j1}^2)^{\frac{1}{2}}$ if t is odd, $t \geq 2$. Let $x_{t2} = 0$ if t is odd and $x_{t2} = (2^{t-1} - \Sigma_{j=1}^{t-1} x_{j2}^2)^{\frac{1}{2}}$ if t is even, $t \geq 1$. Then (taking $S_t = 1, t \geq 1$)

$$\Sigma_{t=1}^{n} \nabla_t' \nabla_t = \begin{pmatrix} \Sigma_{t=1}^{n} x_{t1}^2 & 0 \\ 0 & \Sigma_{t=1}^{n} x_{t2}^2 \end{pmatrix},$$

where $\Sigma_{t=1}^{n} x_{t1}^2 = 2^{n-1}, \Sigma_{t=1}^{n} x_{t2}^2 = 2^{n-2}$ if $n \geq 3$ is odd and $\Sigma_{t=1}^{n} x_{t1}^2 = 2^{n-2}, \Sigma_{t=1}^{n} x_{t2}^2 = 2^{n-1}$ if $n \geq 2$ is even. Clearly, $\rho_n = 2, n \geq 3$ but (1.5) fails to hold for any sequence $\{k_n\}$. Of course, as a practical matter, this particular model would be analyzed as two separate problems.

When $F_t = \sigma(y_0, y_1, \cdots y_t)$, $t \geq 0$, $\{u_t(\theta^*)\}$ is a sequence of martingale differences. With this choice of $\{F_t\}$, called the martingale case, the following elementwise integrability assumptions will be needed.

$$(2.5) \quad u_t(\theta^*)u_t'(\theta^*)\varepsilon L^1, \qquad \nabla_t{}'S_t\nabla_t\,\varepsilon L^1, \qquad \nabla_t{}'S_t u_{t+1}(\theta^*)u'_{t+1}(\theta^*)S_t\nabla_t\,\varepsilon L^1, \qquad t \geq 0.$$

Now, whenever (2.5) holds, define the $r \times r$ conditional covariance matrices by

$$\Omega_t = E(u_{t+1}(\theta^*)u'_{t+1}(\theta^*)|F_t), \qquad t \geq 0.$$

The main result on consistency can now be given.

THEOREM 2.1. *In the martingale case, suppose that the integrability conditions* (2.5) *hold. We then have the following partial dichotomy: a strongly consistent sequence of least squares estimators*

(a) *exists* a.e. *on the event A if*

(i) $\{\min.$ *eigenvalue* $\Sigma_{t=0}^{n-1}\nabla_t{}'S_t\nabla_t\} \to \infty$ a.e. *on* A,

(ii) $h_1 \equiv \inf_{t\to\infty}\{\min.$ *eigenvalue* $B_t{}'\Omega_t B_t\} > 0$ a.e. *on* A,

$h_2 \equiv \sup_{t\to\infty}\{\max.$ *eigenvalue* $B_t{}'\Omega_t B_t\} < \infty$ a.e. *on* A,

where $S_t = B_t B_t{}', t \geq 0$.

(iii) $\limsup_{n\to\infty}\rho_n < \infty$ a.e. *on* A,

(iv) (2.4) *holds*;

(b) *does not exist* a.e. *if*

(v) $\sup_{n\to\infty}\{\min.$ *eigenvalue* $E(\Sigma_{t=0}^{n-1}\nabla_t{}'S_t\nabla_t)\} < \infty$,

(vi) $\sup_{t\to\infty}\{\max.$ *eigenvalue* $B_t{}'\Omega_t B_t\} \leq k$ a.e., *k a positive constant*,

(vii) *the regression functions* $\{g(\theta, F_t)\}$ *are componentwise* a.e. *twice continuously differentiable*, $t \geq 0$, *and* $\sup_{\theta\varepsilon\Theta}\sup_{n\to\infty}|(\partial^2 Q_n(\theta))/(\partial\theta_i\partial\theta_j)| < \infty$ a.e., $i,j = 1, 2, \cdots, p$.

(viii) (2.2) *holds* a.e.

PROOF. In view of Lemma 2.1, to prove (a) it suffices to show that (i) and (ii) imply that for all $d \neq 0$, $\Sigma_{t=0}^{n-1}v_t(d)/\Sigma_{t=0}^{n-1}b_t^2(d) \to 0$ a.e. on A. Since for all d, d' $\nabla_t{}'S_t$ is measurable with respect to $F_t, t \geq 0$, $f_n(d) \equiv \Sigma_{t=0}^{n-1}v_t(d) = \Sigma_{t=0}^{n-1}d'\nabla_t{}'S_t u_{t+1}(\theta^*)$, $n \geq 1$ is a martingale. Therefore,

$$(2.6) \qquad f_n(d)/\Sigma_{t=0}^{n-1}E(v_{t+1}^2(d)|F_t) \to 0$$

for all $d \neq 0$ a.e. on the set where $\Sigma_{t=0}^{\infty}E(v_{t+1}^2(d)|F_t) = \infty$ (Neveu (1965), page 150). Since for all d,

$$\Sigma_{t=0}^{n-1}E(v_{t+1}^2(d)|F_t) = d'\Sigma_{t=0}^{n-1}\nabla_t{}'S_t\Omega_t S_t\nabla_t d \geq (d'\Sigma_{t=0}^{n-1}\nabla_t{}'S_t\nabla_t d)h_1$$
$$= \Sigma_{t=0}^{n-1}b_t^2(d)h_1 \geq h_1\|d\|^2 \min.\text{eigenvalue}\,(\Sigma_{t=0}^{n-1}\nabla_t{}'S_t\nabla_t),$$

and

$$\Sigma_{t=0}^{n-1}E(v_{t+1}^2(d)|F_t) \leq h_2\Sigma_{t=0}^{n-1}b_t^2(d)\text{ a.e. on }A,$$

(i), (ii) and (2.6) imply that

$$f_n(d)/\Sigma_{t=0}^{n-1} b_t^2(d) \to 0 \text{ a.e. on } A \text{ for all } d \neq 0.$$

This completes the proof of (a).

To prove (b), suppose that a strongly consistent sequence $\{\hat{\theta}_n\}$ of least squares estimators exists a.e. Expand the $p \times 1$ vector $\partial Q_n(\hat{\theta}_n)/\partial\theta$ in a Taylor series about θ^* to obtain

$$\frac{\partial Q_n(\hat{\theta}_n)}{\partial\theta} = \frac{\partial Q_n(\theta^*)}{\partial\theta} + \frac{\partial^2 Q_n(\bar{\theta}_n)}{\partial\theta^2}(\hat{\theta}_n - \theta^*), \tag{2.7}$$

where $\bar{\theta}_n$ is some vector between $\hat{\theta}_n$ and θ^*. Ultimately in n(random) $(\partial Q_n(\hat{\theta}_n))/(\partial\theta) = 0^{p\times 1}$ a.e. Since the elements of $(\partial^2 Q_n(\bar{\theta}_n))/(\partial\theta_2)$ are a.e. bounded by assumption and $\hat{\theta}_n \to \theta^*$, a.e. (2.7) implies

$$\lim_{n\to\infty} \frac{\partial Q_n(\theta^*)}{\partial\theta} = 0^{p\times 1} \text{ a.e.} \tag{2.8}$$

The proof proceeds by showing that (v), (vi) and (viii) lead to a contrary implication.

Assumptions (v) and (vi) imply the existence of a $p \times 1$ vector $d^0 = \theta - \theta^* \neq 0$ such that

$$\sup_{n\to\infty} E\left(d^{0\prime}\Sigma_{t=0}^{n-1} \nabla_t' S_t \Omega_t S_t \nabla_t d^0\right) \leqslant k \sup_{n\to\infty} E\left(d^0 \Sigma_{t=0}^{n-1} \nabla_t' S_t \nabla_t d^0\right) < \infty.$$

Therefore, $\{f_n(d^0)\}$ is an L^2 – bounded martingale. Hence, $\{f_n(d^0)\}$ is uniformly integrable and $f(d^0) = \lim_{n\to\infty} f_n(d^0)$ exists a.e. with

$$E\left(f(d^0)|F_n\right) = f_n(d^0) \text{ a.e.}, \qquad n \geqslant 1, \tag{2.9}$$

Neveu (1965, Proposition IV. 5.6.). Since from (viii) $\{f_n(d^0)\}$ is not a.e. a sequence of zeros, (2.9) implies that $P(f(d^0) \neq 0) > 0$. But, $\{f_n(d^0)\} = -\frac{1}{2}d^{0\prime}(\partial Q_n(\theta^*))/(\partial\theta)$ is a linear combination of the elements of $(\partial Q_n(\theta^*))/(\partial\theta)$, $n \geqslant 1$. Therefore, $P(\lim_{n\to\infty} (\partial Q_n(\theta^*))/(\partial\theta) = 0^{p\times 1}) < 1$, which contradicts (2.8) and completes the proof of (b).

To compare Theorem 2.1 to some related results, first consider the univariate ($r = 1$) multiple regression model $Y = X_n\theta + \varepsilon, Y = (y_1, y_2, \cdots, y_n)', X_n$ a nxp matrix of constants of rank $p, n > p, \varepsilon \sim \text{MVN}(0, V_n)$ with V_n a known positive definite matrix. Take $y_0 \equiv 0$. When $S_t = \Omega_t^{-1}$, which is free of θ here, $t \geqslant 1$, the sum of squares given in (1.1) becomes $Q_n(\theta) = (Y - X_n\theta)'V_n^{-1}(Y - X_n\theta)$ and hence $\hat{\theta}_n = (X_n'V_n^{-1}X_n)^{-1}X_n'V_n^{-1}Y$, the Gauss-Markov estimator. Theorem 2.1 implies that $\hat{\theta}_n \to \theta^*$ a.e. if min. eigenvalue $X_n'V_n^{-1}X_n \to \infty$ and $\limsup_{n\to\infty}$ {max. eigenvalue $X_n'V_n^{-1}X_n$/min. eigenvalue $X_n'V_n^{-1}X_n\} < \infty$; and $\hat{\theta}_n \not\to \theta^*$ a.e. if $\sup_{n\to\infty}$ {max. eigenvalue $X_n'V_n^{-1}X_n\} < \infty$. For this particular model with $V_n = \sigma^2 I^{n\times n}$, and $S_t = (1)$, $t \geqslant 1$, Anderson and Taylor (1976a) obtain sharper results by showing that {min. eigenvalue $\Sigma_{t=1}^n \nabla_t' \nabla_t\} \to \infty$ is necessary and sufficient for the strong consistency of $\{\hat{\theta}\}$. Lai, Robbins and Wei (1978) show that this condition is sufficient for strong consistency when the assumptions that $\varepsilon \sim \text{MVN}(0, \sigma^2 I^{n\times n})$ are weakened to $\sup_t E(\varepsilon_t^2) < \infty$ and $\{\varepsilon_t\}$ are martingale differences.

Theorem 2.1 extends the results of Anderson and Taylor (1976b) who, for model (1.1) in the martingale case, assume that (i) holds; that the regression functions $\{g(\theta, F_t)\}$ are a.e. linear in $\theta, t \geqslant 0 (R_n(\theta) = 0$ a.e.), $\Omega_t = \Omega$ a.e. independent of $t \geqslant 0$, positive definite and $S_t = I^{r\times r}, t \geqslant 0$ ((ii) holds); and that $\{\rho_n\}$ is uniformly bounded in n with probability one ((iii) holds).

EXAMPLE 2.1. Let $\{y_t\}$ be a super critical single type ($r = 1$) branching process whose branching distribution has unknown mean θ, $\theta > 1$, and finite variance σ^2. See Jagers (1975) for a discussion of such processes. Let A be the complement of the set where the process becomes extinct. The process is Markov and we take $F_t = \sigma(y_t)$, $t \geqslant 0$. Then, $g(\theta, F_t) \equiv E_\theta(y_{t+1}|F_t) = \theta y_t$ a.e. and $\Omega_t = E((y_{t+1} - \theta y_t)^2|F_t) = \sigma^2 y_t$ a.e., $t \geqslant 0$. Let the weight $S_t^{1\times 1} = 1/y_t$ if $y_t > 0$ and 0 elsewhere, $t \geqslant 0$. Then $\Sigma_{t=0}^{n-1}\nabla_t' S_t \nabla_t = \Sigma_{t=0}^{n-1} y_t = \Sigma_{t=0}^{n-1}\theta^t y_t/\theta^t \to \infty$ a.e. on A since $\{y_t/(\theta^t)\}$ is a nonnegative martingale which converges to a positive random variable a.e. on A. That is, (i) of Theorem 2.1 holds. Since $\Omega_t = \sigma^2 S_t^{-1}$ a.e. on A and $p = 1$ conditions (ii) and (iii) hold trivially. Since $\nabla_t' S_t \nabla_t = y_t$ and $\nabla_t' S_t u_{t+1}(\theta) u'_{t+1}(\theta) S_t \nabla_t = (y_{t+1} - \theta y_t)^2$ a.e. on A, $t \geqslant 0$, the integrability assumptions of the martingale case hold. Condition (2.4) holds since $R_n(\theta) \equiv 0$. Therefore, Theorem 2.1 implies that the weighted least squares estimator $\hat{\theta} = \Sigma_{t=1}^{n} y_t / \Sigma_{t=0}^{n-1} y_t$ converges to θ a.e. on A. It is interesting to note that $\hat{\theta}$ is also the maximum likelihood estimator for a wide variety of offspring distributions (Feigin (1977)) whose consistency is already known. Extensions of both approaches to estimation of the r^2 elements of the matrix M of conditional means in a positively regular super critical multitype branching process have not yet been worked out. One problem here is an appropriate parameterization since the elements of M are constrained by the condition that its Perron-Frobenius root be greater than one. See Mode (1971).

Acknowledgment. I would like to thank the referee whose suggestions led to a greatly improved presentation.

REFERENCES

[1] ANDERSON, T. W. and TAYLOR, J. B. (1976a). Strong consistency of least squares estimates in normal linear regression. *Ann. Statist.* **4** 788–790.

[2] ANDERSON, T. W. and TAYLOR, J. B. (1976b). Conditions for strong consistency of least squares estimates in linear models. Technical report No. 213, The Economic Series, Stanford Univ.

[3] DUNSMUIR, W. and HANNAN, E. J. (1976). Vector linear time series models. *Adv. Appl. Probability* **8** 339–364.

[4] FEIGIN, P. D. (1977). A note on maximum likelihood estimation for simple branching process. *Austral. J. Statist.* **19** 152–154.

[5] JAGERS, P. (1975). *Branching Processes with Biological Applications.* Wiley, New York.

[6] JENNRICH, R. J. (1969). Asymptotic properties of nonlinear least squares estimators. *Ann. Math. Statist.* **40** 533–543.

[7] KLIMKO, L. A. and NELSON, P. I. (1978). On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6** 629–642.

[8] LAI, T. L., ROBBINS, H. and WEI, C. Z. (1978). Strong consistency of least squares estimates in multiple regression. *Proc. Nat. Acad. Sci. U.S.A.* **75** 3034–3036.

[9] MODE, C. J. (1971). *Multitype Branching Processes.* American Elsevier, New York.
[10] NEVEU, J. (1965). *Mathematical Foundations of the Calculus of Probability*. Holden Day, San Francisco.
[11] NICHOLLS, D. F. (1976). The efficient estimation of vector linear time series models. *Biometrika* **63** 381–390.
[12] PHILLIPS, P. C. B. (1976). On the minimum distance estimator and the quasimaximum likelihood estimator. *Econometrics* **44** 449–460.
[13] ROBINSON, P. M. (1972). Nonlinear regression for multiple time series. *J. Appl. Probability* **9** 758–768.
[14] ROBINSON, P. M. (1977). The estimation of a nonlinear moving average model. *Stochastic Processes Appl.* **5** 81–90.

DEPARTMENT OF STATISTICS
PENNSYLVANIA STATE UNIVERSITY
219 POND LABORATORY
UNIVERSITY PARK, PENNSYLVANIA 16802