The Annals of Statistics
1992, Vol. 20, No. 3, 1441–1453

AN IMPROVED SEQUENTIAL PROCEDURE FOR ESTIMATING THE REGRESSION PARAMETER IN REGRESSION MODELS WITH SYMMETRIC ERRORS

BY T. N. SRIRAM

University of Georgia

A sequential procedure for estimating the regression parameter $\boldsymbol{\beta} \in R^k$ in a regression model with symmetric errors is proposed. This procedure is shown to have asymptotically smaller regret than the procedure analyzed by Martinsek when $\boldsymbol{\beta} = \mathbf{0}$, and the same asymptotic regret as that procedure when $\boldsymbol{\beta} \neq \mathbf{0}$. Consequently, even when the errors are normally distributed, it follows that the asymptotic regret can be negative when $\boldsymbol{\beta} = \mathbf{0}$. These results extend a recent work of Takada dealing with the estimation of the normal mean, to both regression and nonnormal cases.

1. Introduction. Consider the general linear model

$$y_i = X_i'\boldsymbol{\beta} + \varepsilon_i = \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i, \tag{1.1}$$

$i = 1, 2, \ldots,$ where each $X_i = (x_{i1}, \ldots, x_{ik})'$ is a known k-vector of design points, $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ is an unknown k-vector of parameters, $\varepsilon_1, \varepsilon_2, \ldots$ are i.i.d. with mean 0 and variance σ^2, and $y_1, y_2, \ldots$ are the observed responses. Define $\mathbf{X}_n$ for each n to be the $n \times k$ matrix with (i, j) entry x_{ij} and assume that for each n, $M_n = \mathbf{X}_n'\mathbf{X}_n$ is nonsingular. Then the least squares estimator of $\boldsymbol{\beta}$ based on the first n observations is

$$\hat{\boldsymbol{\beta}}_n = M_n^{-1}\mathbf{X}_n'\mathbf{y}_n = \left(\sum_{i=1}^{n} X_i X_i'\right)^{-1} \sum_{i=1}^{n} y_i X_i, \tag{1.2}$$

where $\mathbf{y}_n$ is the n-vector $(y_1, \ldots, y_n)'$. Suppose that one can stop observing the sequence $(X_1, y_1), (X_2, y_2), \ldots$ after any number of observations n and estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_n$, subject to the loss function

$$L_n = An^{-1}\big(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\big)' M_n\big(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}\big) + n, \qquad A > 0. \tag{1.3}$$

If σ is known and a fixed sample size n is used, then the risk

$$R_n = E(L_n) = Ak\sigma^2/n + n \tag{1.4}$$

is approximately minimized by $n_0 \approx (Ak)^{1/2}\sigma$, with corresponding minimum risk $R_{n_0} \approx 2(Ak)^{1/2}\sigma$. On the other hand, if σ is unknown, then n_0 cannot be used and there is no fixed sample size procedure that will achieve the risk $2(Ak)^{1/2}\sigma$. Motivated by the formula for n_0 and the idea of Robbins

Received June 1991; revised November 1991.
AMS 1980 *subject classifications*. Primary 62L12; secondary 60G40, 62J05.
Key words and phrases. Sequential procedure, regression, least squares estimate, regret, stopping rule.

(1959), Martinsek (1990) proposed the following stopping time: Let $\hat{\sigma}_n^2 = n^{-1}\sum_{i=1}^n (y_i - X_i'\hat{\boldsymbol{\beta}}_n)^2$ and define

$$T = T_A = \inf\{n \geq m_A : n \geq (Ak)^{1/2}\hat{\sigma}_n\}$$

$$= \inf\left\{n \geq m_A : n^{-1}\sum_{i=1}^n \varepsilon_i^2 - n^{-1}\left(\sum_{i=1}^n \varepsilon_i X_i'\right) M_n^{-1}\left(\sum_{i=1}^n X_i\varepsilon_i\right) \leq n^2/(Ak)\right\}, \tag{1.5}$$

where m_A may depend on A and $\hat{\sigma}_n^2$ is an estimator of σ^2 at each stage n. The parameter $\boldsymbol{\beta}$ is then estimated by $\hat{\boldsymbol{\beta}}_T$ and the risk of the sequential procedure $\hat{\boldsymbol{\beta}}_T$ is $R_T = E(L_T)$.

Martinsek (1990) showed that if the errors are symmetric, nonlattice with $E|\varepsilon_1|^{6r} < \infty$ for some $r > 1$, the initial sample size m_A satisfies $\delta A^{1/2}/\log(A) \leq m_A = o(A^{1/2})$, and the design points and the matrix M_n satisfy some mild conditions, then the regret

$$R_T - 2(Ak)^{1/2}\sigma = 2.75 - \frac{3}{4}E(\varepsilon_1/\sigma)^4 + \frac{1}{k}E\left\{\sum_{i=1}^T X_i'[\Sigma^{-1} - TM_T^{-1}]X_i\right\} + \frac{1}{k\sigma^2}E\left\{\left(\sum_{i=1}^T \varepsilon_i X_i'\right)[TM_T^{-1} - \Sigma^{-1}]\left(\sum_{i=1}^T X_i\varepsilon_i\right)\right\} + o(1) \tag{1.6}$$

as $A \to \infty$ [see Martinsek (1990), Theorem 2]. Note that the above result extends the work of Mukhopadhyay (1974) and Finster (1983) to the nonparametric case and also generalizes the results of Woodroofe (1977).

The possibility of improving the performance of a stopping rule at a fixed value of a parameter has been investigated recently by Takada (1992) for estimating the mean of a normal population. Takada (1992) constructed a stopping rule analogous to the rule proposed by Robbins (1959), where the unknown variance σ^2 is estimated by an improved estimator of σ^2 instead of the usual sample variance. He showed that the asymptotic regret of his sequential procedure is negative when the mean is 0 and 1/2 otherwise, thus improving the expansion obtained by Woodroofe (1977). Note that the improved estimator of σ^2 given in Takada (1992) is similar to the one proposed by Stein (1964) to establish the inadmissibility of the usual estimator of σ^2 in the normal case.

In this paper we use the technique developed by Takada (1992) and propose a stopping rule N for the sequential estimation problem in regression setup (1.1). Under the same regularity conditions as in Theorem 2 of Martinsek (1990), it is shown that the procedure N has asymptotically smaller regret than T defined in (1.5) when $\boldsymbol{\beta} = \mathbf{0}$, and the same asymptotic regret as T when $\boldsymbol{\beta} \neq \mathbf{0}$. This result extends the work of Takada (1992), not only to the regression setup but also to the nonnormal cases. In the case of normal errors, our result implies that the asymptotic regret of N is negative when $\boldsymbol{\beta} = \mathbf{0}$ and

$1/2$ when $\boldsymbol{\beta} \neq \mathbf{0}$, which improves the regret expansion obtained by Finster (1983) at $\boldsymbol{\beta} = \mathbf{0}$.

To this end, instead of $\hat{\sigma}_n^2$ in (1.5), consider an estimator ϕ_n of the form

$$\phi_n = n^{-1} \min\left\{ \sum_{i=1}^{n} \left(y_i - X_i'\hat{\boldsymbol{\beta}}_n\right)^2, \sum_{i=1}^{n} y_i^2/c_n \right\}, \tag{1.7}$$

where $c_n \geq 1$ is a sequence of constants. Now, let

$$N = \inf\left\{n \geq m_A \colon n \geq (Ak)^{1/2}\phi_n^{1/2}\right\} \tag{1.8}$$

and estimate $\boldsymbol{\beta}$ by $\hat{\boldsymbol{\beta}}_N$. In what follows we will assume that c_n is nonincreasing and

$$c_n = 1 + c/n + o(n^{-1}) \quad \text{as } n \to \infty, \tag{1.9}$$

where c is a nonnegative constant. The main result is given next.

THEOREM. *Assume that for some $C < \infty$, $|x_{ij}| \leq C$ for every i and j. Assume further that*

$$n[M_n/n - \Sigma] = O(1) \tag{1.10}$$

as $n \to \infty$, where Σ is nonsingular, that $E|\varepsilon_1|^{6r} < \infty$ for $r > 1$, ε_1^2 is nonlattice, ε_1 has a distribution that is symmetric about 0, *and that $\delta A^{1/2}/\log(A) \leq m_A = o(A^{1/2})$ as $A \to \infty$, for some $\delta > 0$. Then as $A \to \infty$,*

$$\begin{aligned} R_N - 2(Ak)^{1/2}\sigma = {} & -\frac{3}{4}E\left(\frac{\varepsilon_1^2}{\sigma^2} - 1\right)^2 + \frac{1}{k}E\left\{\sum_{i=1}^{N} X_i'\left[\Sigma^{-1} - NM_N^{-1}\right]X_i\right\} \\ & + \frac{1}{k\sigma^2}E\left\{\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\left[NM_N^{-1} - \Sigma^{-1}\right]\left(\sum_{i=1}^{N} X_i\varepsilon_i\right)\right\} \\ & + \begin{cases} E(k^{-1}\chi_k^2 - 1)\max\{\chi_k^2, c\} + o(1), & \text{when } \boldsymbol{\beta} = \mathbf{0}, \\ 2 + o(1), & \text{when } \boldsymbol{\beta} \neq \mathbf{0}, \end{cases} \end{aligned} \tag{1.11}$$

where χ_k^2 is a chi-square random variable (*r.v.*) *with k degrees of freedom and c is as in* (1.9).

REMARKS. For the case $\boldsymbol{\beta} \neq \mathbf{0}$, the regret expansion for N agrees with that of T obtained by Martinsek (1990). However, for $\boldsymbol{\beta} = \mathbf{0}$, since it follows easily that

$$E\left[k^{-1}\chi_k^2 - 1\right]\max\{\chi_k^2, c\} \to 0 \quad \text{as } c \to \infty, \tag{1.12}$$

the procedure N has asymptotically smaller regret than T [see (1.6)]. If $n[M_n/n - \Sigma]$ has a limit as $n \to \infty$, then the terms in (1.11) that involve the design matrix cancel, thus yielding a negative regret at $\boldsymbol{\beta} = \mathbf{0}$ for large c. In particular, suppose $k = 1$ and all $x_{i1} = 1$, so that $y_1, y_2, \ldots$ are i.i.d. with mean β_1 and variance σ^2. Then $M_n/n \equiv 1$, and in this case, even if the error distribution is not symmetric, it can be shown directly [assuming $E|y_1|^{6r} < \infty$

for $r > 1$, y_1 is nonlattice and $\delta A^{1/4} \le m_A = o(A^{1/2})$ as $A \to \infty$] that

$$R_N - 2A^{1/2}\sigma = 2E^2(Z)^3 - (3/4)E[Z^2 - 1]^2$$
$$+ \begin{cases} E(\chi_1^2 - 1)\max\{\chi_1^2, c\} + o(1), & \text{when } \beta_1 = 0, \\ 2 + o(1), & \text{when } \beta_1 \neq 0, \end{cases}$$

where $Z = (y_1 - \beta_1)/\sigma$ and this extends the result of Takada (1992) to the nonparametric case, and improves the expansion obtained by Martinsek (1983) at $\beta_1 = 0$. Moreover, in setup (1.1), if the errors are normally distributed, then it can be shown using an analog of Property 2 of Finster (1983), (1.11) and (1.12) that $R_N - 2(Ak)^{1/2}\sigma$ is negative ($-3/2$, for large c) when $\boldsymbol{\beta} = \mathbf{0}$ and $1/2$ when $\boldsymbol{\beta} \neq \mathbf{0}$.

The proof of the theorem is given in Section 3. In Section 2 we prove some preliminary results which will be used in Section 3. The method of proof of the main theorem is similar to that of Theorem 2 of Martinsek (1990). Therefore, we will only elaborate on those necessary steps that distinguish the stopping rule N from T.

2. Preliminaries. In this section we prove several lemmas that will be used in the proof of the main theorem. We will label some repeatedly used assumptions as follows:

$$\text{(A1)} \qquad |x_{ij}| \le C < \infty \quad \text{for all } i, j;$$

$$\text{(A2)} \qquad M_n/n \to \Sigma \quad \text{as } n \to \infty, \text{ where } \Sigma \text{ is positive definite.}$$

Without loss of generality, take $\sigma = 1$.

LEMMA 1. *Assume* (A1) *and* (A2). *For N defined in* (1.8), *if $m_A \ge \delta A^{1/2}/\log(A)$ for some $\delta > 0$ and A sufficiently large, then for all $r > 0$,*

$$\text{(2.1)} \qquad \left\{\left[(Ak)^{1/2}/N\right]^r : A > 0\right\} \text{ is uniformly integrable } (u.i.).$$

PROOF. Let

$$\text{(2.2)} \qquad N_2 = \inf\left\{n \ge m_A : n \ge (Ak)^{1/2}\left(\sum_{i=1}^{n} y_i^2/nc_n\right)^{1/2}\right\}.$$

Then by (1.8), (1.5) and (2.2) it suffices to show that

$$\text{(2.3)} \qquad \left\{\left[(Ak)^{1/2}/T\right]^r : A > 0\right\} \text{ and } \left\{\left[(Ak)^{1/2}/N_2\right]^r : A > 0\right\} \text{ are u.i.}$$

The first assertion in (2.3) follows from Lemma 1 of Martinsek (1990) and the second assertion follows from Lemma 3 of Chow and Yu (1981). Hence the result. □

LEMMA 2. *If* $E|\varepsilon_1|^{2r} < \infty$ *for* $r \ge 1$ *and* $m_A = O(A^{1/2})$ *as* $A \to \infty$, *then*

$$\left\{\left[N/(Ak)^{1/2}\right]^r : A \ge 1\right\} \text{ is u.i.} \tag{2.4}$$

PROOF. Since $N \le T$ and $\{[T/(Ak)^{1/2}]^r: A \ge 1\}$ is u.i. [see display (2.7) in Martinsek (1990)], we have the required result. □

Results from nonlinear renewal theory [see Lai and Siegmund (1977, 1979) and Woodroofe (1982)] are given next. The stopping time N may be written using a Taylor expansion as

$$N = \inf\left\{n \ge m_A: n - (1/2)\sum_{i=1}^{n}\left(\varepsilon_i^2 - 1\right) + \xi_n \ge (Ak)^{1/2}\right\}, \tag{2.5}$$

where

$$\xi_n = (1/2)\gamma_n + (3/8)\lambda_n^{-5/2} n(\phi_n - 1)^2 \tag{2.6}$$

with

$$\gamma_n = \max\left\{\left(\sum_{i=1}^{n}\varepsilon_i X_i'\right)M_n^{-1}\left(\sum_{i=1}^{n}X_i\varepsilon_i\right), \sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}y_i^2/c_n\right\} \tag{2.7}$$

and λ_n is a random variable satisfying $|\lambda_n - 1| < |\phi_n - 1|$.

LEMMA 3. *Assume* (A1) *and* (A2). *If* $E|\varepsilon_1|^s < \infty$ *for some* $s > 4$, *then*

$$\{\xi_n: n \ge 1\} \text{ is slowly changing.} \tag{2.8}$$

PROOF. Since $\xi_n/n \to 0$ a.s., it follows easily that $n^{-1}\max_{1\le j\le n}|\xi_j| \to 0$ a.s. as $n \to \infty$. To show that $\{\xi_n: n \ge 1\}$ is uniformly continuous in probability (u.c.i.p.), we first show that $\{\gamma_n: n \ge 1\}$ is u.c.i.p. Let

$$A_n = \left\{\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}y_i^2/c_n \le \left(\sum_{i=1}^{n}\varepsilon_i X_i'\right)M_n^{-1}\left(\sum_{i=1}^{n}X_i\varepsilon_i\right)\right\}$$

and write

$$\gamma_n = \left(\sum_{i=1}^{n}\varepsilon_i X_i'\right)M_n^{-1}\left(\sum_{i=1}^{n}X_i\varepsilon_i\right)I_{A_n} + \left\{\sum_{i=1}^{n}\varepsilon_i^2 - \sum_{i=1}^{n}y_i^2/c_n\right\}I_{\bar{A}_n}, \tag{2.9}$$

where $\bar{A}_n$ is the complement of A_n. For any $\boldsymbol{\beta}$, write

$$\begin{aligned}\sum_{i=1}^{n} \varepsilon_i^2 - \sum_{i=1}^{n} y_i^2/c_n \\ = \left[n(c_n - 1)/c_n\right] \sum_{i=1}^{n} \varepsilon_i^2/n - 2c_n^{-1}\boldsymbol{\beta}' \sum_{i=1}^{n} X_i\varepsilon_i - c_n^{-1}\boldsymbol{\beta}' M_n \boldsymbol{\beta}.\end{aligned} \tag{2.10}$$

Now, if $E|\varepsilon_1|^r < \infty$ for $r \geq 2$, then by (2.10) we have for $\boldsymbol{\beta} \neq \mathbf{0}$ that

$$\begin{aligned} P(\bar{A}_n) &\leq P\left\{n^{-1} \sum_{i=1}^{n} \varepsilon_i^2 > (nc_n)^{-1} \sum_{i=1}^{n} y_i^2\right\} \\ &\leq P\left\{2n^{-1}\left|\boldsymbol{\beta}' \sum_{i=1}^{n} X_i\varepsilon_i\right| > (2n)^{-1}\boldsymbol{\beta}' M_n \boldsymbol{\beta}\right\} \\ &\quad + P\left\{n^{-2} \sum_{i=1}^{n} \varepsilon_i^2[n(c_n - 1)/c_n] > (2nc_n)^{-1}\boldsymbol{\beta}' M_n \boldsymbol{\beta}\right\} \\ &= O(n^{-r/2}) \quad \text{as } n \to \infty, \end{aligned} \tag{2.11}$$

where we used that

$$\boldsymbol{\beta}' \sum_{i=1}^{n} X_i\varepsilon_i = \sum_{j=1}^{k} \beta_j \left(\sum_{i=1}^{n} x_{ij}\varepsilon_i \right),$$

the Markov inequality, Theorem 10.3.2 and the proof of Corollary 10.3.2 of Chow and Teicher (1978), (A1), (A2) and $|n(c_n - 1)/c_n|$ is bounded in n. So, by Lemma 2 of Martinsek (1990), $\{\gamma_n I_{A_n}: n \geq 1\}$ is u.c.i.p. When $\boldsymbol{\beta} = \mathbf{0}$, by (2.10), (1.9) and SLLN, we have $\{\gamma_n I_{\bar{A}_n}: n \geq 1\}$ is u.c.i.p. For $\boldsymbol{\beta} \neq \mathbf{0}$, note that if $E|\varepsilon_1|^s < \infty$ for $s > 4$, then by the Borel–Cantelli lemma and (2.11), $nI_{\bar{A}_n} \to 0$ a.s. as $n \to \infty$. Now, restrict the terms on the right side of (2.10) to the set $\bar{A}_n$ and use (A1), (A2) and the proof of Lemma 2 of Martinsek (1990) to get $\{\gamma_n I_{\bar{A}_n}: n \geq 1\}$ is u.c.i.p. That $\{\xi_n: n \geq 1\}$ is u.c.i.p. now follows from the identity

$$n\phi_n = \sum_{i=1}^{n} \varepsilon_i^2 - \gamma_n, \tag{2.12}$$

the above arguments, Example 1.8 of Woodroofe (1982) and standard arguments. □

From Lemma 3 above and Theorem 1 of Lai and Siegmund (1977), it follows that if the distribution of ε_1^2 is nonlattice, then the overshoot

$$U_A = N - (1/2) \sum_{i=1}^{N} (\varepsilon_i^2 - 1) + \xi_N - (Ak)^{1/2} \to_D U \tag{2.13}$$

as $A \to \infty$, for some random variable U. Moreover, by the arguments similar to (3.4) of Martinsek (1990) and (2.1), we have that

$$\{U_A^r\colon A > 0\} \text{ is u.i. for all } r > 0. \tag{2.14}$$

LEMMA 4. *Assume* (A1), (A2) *and the condition on the initial sample* m_A *of the theorem. If* $E|\varepsilon_1|^{4p+\varepsilon} < \infty$ *for* $p \ge 1$ *and some* $\varepsilon > 0$, *then for* γ_n *defined in* (2.7)

$$\{|\gamma_N|^p\colon A \ge 1\} \textit{ is u.i.} \tag{2.15}$$

PROOF. Recall the identity (2.9). If $E|\varepsilon_1|^{2p+\varepsilon} < \infty$ for $p \ge 1$ and some $\varepsilon > 0$, then by (2.1), (2.4) and Lemma 5 of Chow and Yu (1981), we have

$$\left\{|\gamma_N|^p I_{A_N}\colon A \ge 1\right\} \text{ is u.i.} \tag{2.16}$$

See display (3.6) of Martinsek (1990), for example. As for the uniform integrability of $\{|\gamma_N|^p I_{\bar{A}_N}\colon A \ge 1\}$, first we obtain a rate at which $P(\bar{A}_N)$ tends to 0 (for $\boldsymbol{\beta} \ne \mathbf{0}$) as $A \to \infty$. But, when $\boldsymbol{\beta} = \mathbf{0}$, if $E|\varepsilon_1|^{2p+\varepsilon} < \infty$ for $p \ge 1$ and some $\varepsilon > 0$, then by (2.10) and (1.9)

$$E\left|\sum_{i=1}^N \varepsilon_i^2 - \sum_{i=1}^N y_i^2/c_N\right|^{p+\varepsilon/2} \le BE\left\{\sup_{n \ge m_A}\left|\sum_{i=1}^n \varepsilon_i^2/n\right|^{p+\varepsilon/2}\right\} < \infty \tag{2.17}$$

by Theorem 10.3.3 of Chow and Teicher (1978). When $\boldsymbol{\beta} \ne \mathbf{0}$, argue as in (2.11) to get

$$\begin{aligned} P(\bar{A}_N) &\le P\left\{2N^{-1}\left|\boldsymbol{\beta}'\sum_{i=1}^N X_i\varepsilon_i\right| > (2N)^{-1}\boldsymbol{\beta}' M_N\boldsymbol{\beta}\right\} \\ &\quad + P\left\{N^{-2}\sum_{i=1}^N \varepsilon_i^2[N(c_N-1)/c_N] > (2Nc_N)^{-1}\boldsymbol{\beta}' M_N\boldsymbol{\beta}\right\} \\ &= \text{(i)} + \text{(ii)}. \end{aligned} \tag{2.18}$$

By Markov inequality, c_s-inequality and (A2), for some generic constant K_s, we have

$$\begin{aligned} \text{(i)} &\le K_s \sup_{n\ge 1}\left[(2n)(\boldsymbol{\beta}' M_n\boldsymbol{\beta})^{-1}\right]^{2s}\sum_{j=1}^k |\beta_j|^{2s} E\left|\left(\sum_{i=1}^N x_{ij}\varepsilon_i\right)/N\right|^{2s} \\ &= O(A^{-s/2}) \end{aligned} \tag{2.19}$$

if $E|\varepsilon_1|^{2s+\varepsilon} < \infty$, $s \ge 1$, where we used (2.1), that

$$\left\{\left|\sum_{i=1}^N x_{ij}\varepsilon_i/\sqrt{A^{1/2}}\right|^{2p}\colon A \ge 1\right\} \text{ is u.i. } \left(\text{if } E|\varepsilon_1|^{2p} < \infty,\ p \ge 1\right), \tag{2.20}$$

which follows from (A1) and Lemma 5 of Chow and Yu (1981) and Hölder's inequality. A similar argument using Theorem 10.3.3 of Chow and Teicher (1978) and (1.9) yields (ii) $= O(A^{-s/2})$ (if $E|\varepsilon_1|^{2s+\varepsilon} < \infty$, $s \geq 1$). Hence, if $E|\varepsilon_1|^{2s+\varepsilon} < \infty$, $s \geq 1$, then (for $\boldsymbol{\beta} \neq \mathbf{0}$)

$$P(\bar{A}_N) = O(A^{-s/2}) \quad \text{as } A \to \infty. \tag{2.21}$$

Now, consider the expression on the right side of (2.10) with n replaced by N. By (A2), Hölder's inequality, (1.9), Lemma 2 and (2.21), we have

$$\begin{aligned} E|c_N^{-1}\boldsymbol{\beta}' M_N \boldsymbol{\beta}|^p I_{\bar{A}_N} &\leq \sup_{n \geq 1} \left|(nc_n)^{-1}\boldsymbol{\beta}' M_n \boldsymbol{\beta}\right|^p E|N|^p I_{\bar{A}_N} \\ &\leq K A^{p/2} E^{1/2} |N/A^{1/2}|^{2p} P^{1/2}(\bar{A}_N) = o(1) \end{aligned} \tag{2.22}$$

if we take $s = 2p + \varepsilon$ in (2.21). The rest of the terms in (2.10) are handled similarly using (2.17), (2.20) and (2.21) to get $\{|\gamma_N|^p I_{\bar{A}_N}: A \geq 1\}$ is u.i. □

LEMMA 5. *Assume* (A1), (A2) *and the condition on the initial sample* m_A *of the theorem. If* $E|\varepsilon_1|^{4p+\varepsilon} < \infty$ *for* $p \geq 1$ *and some* $\varepsilon > 0$, *then for* ξ_n *defined in* (2.6)

$$\{|\xi_N|^p: A \geq 1\} \textit{ is u.i.} \tag{2.23}$$

PROOF. By Lemma 4 it suffices to show that $\{|\lambda_N^{-5/2} N(\phi_N - 1)^2|^p: A \geq 1\}$ is u.i. Since $\phi_N \geq (N-1)N^{-1}\phi_{N-1} > 2^{-1}\phi_{N-1} \geq 2^{-1}(N-1)^2/(Ak)$ we have $\lambda_N^{-5/2} \leq O(1)\{1 + (Ak)^{5/2}/N^5\}$ and hence by (2.1), all positive powers of $\lambda_N^{-5/2}$ are u.i. Now, from (2.12)

$$\begin{aligned} N(\phi_N - 1)^2 &= N^{-1}\left(\sum_{i=1}^{N} \varepsilon_i^2 - \gamma_N - N\right)^2 \\ &\leq 2(A^{1/2}/N)\left[\sum_{i=1}^{N} (\varepsilon_i^2 - 1)/\sqrt{A^{1/2}}\right]^2 + 2\gamma_N^2/N. \end{aligned} \tag{2.24}$$

By (2.9), (2.16) and (2.10) (splitting the cases $\boldsymbol{\beta} = \mathbf{0}$ and $\boldsymbol{\beta} \neq \mathbf{0}$) and using (2.1), Theorem 10.3.3 of Chow and Teicher (1978), (1.9), (2.20), (A1), (A2) and (2.21), we get

$$\left\{|A^{1/2}/N|^{5p} |\gamma_N/\sqrt{N}|^{2p}: A \geq 1\right\} \text{ is u.i.} \tag{2.25}$$

If $E|\varepsilon_1|^{4p+\varepsilon} < \infty$, $p \geq 1$ and some $\varepsilon > 0$. The required result now follows from (2.24), (2.1) and Lemma 5 of Chow and Yu (1981) and the above arguments. □

It is possible to use the method of proof of (1.17) in Sriram (1988), (2.1), arguments similar to (2.25) and Lemma 5 of Chow and Yu (1981) and show that

$$\left\{\left|\left[N - (Ak)^{1/2}\right]/\sqrt{N}\right|^{p} : A \geq 1\right\} \text{ is u.i.,} \tag{2.26}$$

if $E|\varepsilon_1|^{2p+\varepsilon} < \infty$, $p \geq 1$. From this, the identity in (2.13), (2.14), (2.1) and that $\{|\Sigma_1^N(\varepsilon_i^2 - 1)/\sqrt{N}|^p : A \geq 1\}$ is u.i. (if $E|\varepsilon_1|^{2p+\varepsilon} < \infty$, $p \geq 1$) it follows that

$$\left\{\left|\xi_N/\sqrt{N}\right|^{p} : A \geq 1\right\} \text{ is u.i. } \left(\text{if } E|\varepsilon_1|^{2p+\varepsilon} < \infty,\ p \geq 1\right). \tag{2.27}$$

Also, from (2.7), that $\{\gamma_n : n \geq 1\}$ is u.c.i.p. (2.21), (2.10) of Martinsek (1990), (1.9) and Anscombe's theorem,

$$\gamma_N \to_D \begin{cases} \max\{\chi_k^2, c\}, & \text{when } \boldsymbol{\beta} = \mathbf{0}, \\ \chi_k^2, & \text{when } \boldsymbol{\beta} \neq \mathbf{0}, \end{cases} \tag{2.28}$$

where χ_k^2 is a chi-square r.v. with k degrees of freedom. Hence, it follows from the identity and the result in (2.13), (2.14), Wald's lemma, (2.28), that $\lambda_N^{-5/2} N(\phi_N - 1)^2 \to_D E(\varepsilon_1^2 - 1)^2 \chi_1^2$, and Lemma 5 that

$$\begin{aligned} EN - (Ak)^{1/2} &= EU - (3/8)E(\varepsilon_1^2 - 1)^2 \\ &\quad - \begin{cases} (1/2)E\max\{\chi_k^2, c\} + o(1), & \text{when } \boldsymbol{\beta} = \mathbf{0}, \\ (k/2) + o(1), & \text{when } \boldsymbol{\beta} \neq \mathbf{0}, \end{cases} \end{aligned} \tag{2.29}$$

where the expansion for $\boldsymbol{\beta} \neq \mathbf{0}$, agrees with (3.10) of Martinsek (1990) for the stopping T defined in his paper.

3. Main result.

PROOF OF THE THEOREM. Hölder's inequality is used many times in the proof, but an explicit mention of it will be suppressed. Now, argue exactly as in Martinsek (1990) [see (3.17) to (3.20)], using Wald's equation for the second moment of a stopped Martingale, (1.10), (2.1) and (2.20) to get

$$\begin{aligned} R_N - 2(Ak)^{1/2} &= E\left[AN^{-1}\left(\sum_{i=1}^{N} \varepsilon_i X_i'\right) M_N^{-1}\left(\sum_{i=1}^{N} X_i \varepsilon_i\right)\right] + EN - 2(Ak)^{1/2} \\ &= E\left[(AN^{-2} - k^{-1})\left(\sum_{i=1}^{N} \varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N} X_i \varepsilon_i\right)\right] \\ &\quad + k^{-1}E\left\{\sum_{i=1}^{N} X_i'\left[\Sigma^{-1} - NM_N^{-1}\right]X_i\right\} \\ &\quad + k^{-1}E\left\{\left(\sum_{i=1}^{N} \varepsilon_i X_i'\right)\left[NM_N^{-1} - \Sigma^{-1}\right]\left(\sum_{i=1}^{N} X_i \varepsilon_i\right)\right\} \\ &\quad + 2E\left\{N - (Ak)^{1/2}\right\} + o(1) \end{aligned} \tag{3.1}$$

as $A \to \infty$. It remains to analyze the first term on the right side of (3.1), which may be written, after some algebraic manipulations, as

$$\begin{aligned} &E\left[(AN^{-2}-k^{-1})\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right] \\ (3.2)\qquad &= 2k^{-1}E\left\{N^{-1}\left[(Ak)^{1/2}-N\right]\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right\} \\ &\quad + k^{-1}E\left\{N^{-2}\left[(Ak)^{1/2}-N\right]^2\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right\} \\ &= \text{(i)} + \text{(ii)}. \end{aligned}$$

By (2.1), the identity in (2.13), (2.14), (2.27) and Lemma 5 of Chow and Yu (1981)

$$(3.3)\qquad \begin{aligned} \text{(ii)} &= (1/4)A^{-1}k^{-2}E\left\{\left[\sum_{i=1}^{N}(\varepsilon_i^2-1)\right]^2\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right\} \\ &\quad + o(1) \end{aligned}$$

as $A \to \infty$. Once again, by the identity in (2.13)

$$(3.4)\qquad \begin{aligned} (k/2)\text{(i)} &= E\left\{N^{-1}\left[-U_A-(1/2)\sum_{i=1}^{N}(\varepsilon_i^2-1)+\xi_N\right]\right. \\ &\qquad \left.\times\left(\sum_{i=1}^{N}\varepsilon_i X_1'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right\}. \end{aligned}$$

Since $\{n^{-1}(\sum_{i=1}^{n}\varepsilon_i X_i')\Sigma^{-1}(\sum_{i=1}^{n}X_i\varepsilon_i): n \geq 1\}$ is slowly changing [see Lemma 2 of Martinsek (1990)], it can be shown along the lines of the lemma in Martinsek (1983) that U_A and $\{N^{-1}(\sum_{i=1}^{N}\varepsilon_i X_i')\Sigma^{-1}(\sum_{i=1}^{N}X_i\varepsilon_i)\}$ are asymptotically independent as $A \to \infty$. Since $N^{-1}(\sum_{i=1}^{N}\varepsilon_i X_i')\Sigma^{-1}(\sum_{i=1}^{N}X_i\varepsilon_i) \to_D \chi_k^2$ (by Anscombe's theorem), it follows from (2.13), (2.14), (2.20) and (2.1) that

$$(3.5)\qquad -E\left\{N^{-1}U_A\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right\} = -kE(U)+o(1)$$

as $A \to \infty$. Further, by (2.6) and (2.12)

$$\begin{aligned} &E\left\{N^{-1}\xi_N\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right)\right\} \\ (3.6)\qquad &= E\left\{(1/2)N^{-1}\gamma_N + (3/8)N^{-1}\lambda_N^{-5/2}\left[\sum_{i=1}^{N}(\varepsilon_i^2-1)/\sqrt{N}-\gamma_N/\sqrt{N}\right]^2\right\} \\ &\quad \times\left(\sum_{i=1}^{N}\varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N}X_i\varepsilon_i\right) \\ &= (\text{i}_\text{a}) + (\text{i}_\text{b}). \end{aligned}$$

By (1.10), that $\{|\gamma_N/N|^2 : A \geq 1\}$ is u.i. [similar to (2.25)], (2.1), and (2.20), we get

$$
\text{(3.7)} \qquad
\begin{aligned}
(\mathrm{i_a}) &= (1/2) E\gamma_N \left(\sum_{i=1}^{N} \varepsilon_i X_i' \right) M_N^{-1} \left(\sum_{i=1}^{N} \varepsilon_i X_i \right) + o(1) \\
&= \begin{cases} (1/2) E \max\{\chi_k^2, c\}\chi_k^2 + o(1), & \text{when } \boldsymbol{\beta} = \mathbf{0}, \\ (1/2)(k^2 + 2k) + o(1), & \text{when } \boldsymbol{\beta} \neq \mathbf{0}, \end{cases}
\end{aligned}
$$

where we used (2.7), (2.9), and (2.10) of Martinsek (1990), (1.9), (2.21), Anscombe's theorem, Lemma 4 and (2.20): By (2.1), Lemma 5 of Chow and Yu (1981), (2.20) and (2.25), we get

$$
\text{(3.8)} \qquad
\begin{aligned}
(\mathrm{i_b}) = (3/8)(Ak)^{-1} E \left\{ \left[\sum_{i=1}^{N} (\varepsilon_i^2 - 1) \right]^2 \left(\sum_{i=1}^{N} \varepsilon_i X_i' \right) \Sigma^{-1} \left(\sum_{i=1}^{N} X_i \varepsilon_i \right) \right\} \\
+ o(1).
\end{aligned}
$$

As for the remaining expression on the right side of (3.4), use arguments similar to (3.26), (3.18) and (3.27) in Martinsek (1990), (1.10), (2.1) and Lemma 5 of Chow and Yu (1981) to get

$$
\begin{aligned}
&- (1/2) E \left\{ N^{-1} \sum_{i=1}^{N} (\varepsilon_i^2 - 1) \left(\sum_{i=1}^{N} \varepsilon_i X_i' \right) \Sigma^{-1} \left(\sum_{i=1}^{N} X_i \varepsilon_i \right) \right\} \\
&\quad = -(1/2) E \left\{ N^{-1} \sum_{i=1}^{N} (\varepsilon_i^2 - 1) \left[\left(\sum_{i=1}^{N} \varepsilon_i X_i' \right) \Sigma^{-1} \left(\sum_{i=1}^{N} X_i \varepsilon_i \right) \right. \right. \\
&\qquad\qquad \left. \left. - \sum_{i=1}^{N} (X_i' \Sigma^{-1} X_i) \right] \right\} + o(1) \\
\text{(3.9)} \qquad &\quad = -(1/2)(Ak)^{-1/2} \left\{ \sum_{i=1}^{N} (\varepsilon_i^2 - 1) \left[\left(\sum_{i=1}^{N} \varepsilon_i X_i' \right) \Sigma^{-1} \left(\sum_{i=1}^{N} X_i \varepsilon_i \right) \right. \right. \\
&\qquad\qquad \left. \left. - \sum_{i=1}^{N} (X_i' \Sigma^{-1} X_i) \right] \right\} \\
&\qquad + (1/4)(Ak)^{-1} E \left\{ \left[\sum_{i=1}^{N} (\varepsilon_i^2 - 1) \right]^2 \left(\sum_{i=1}^{N} \varepsilon_i X_i' \right) \Sigma^{-1} \left(\sum_{i=1}^{N} X_i \varepsilon_i \right) \right\} \\
&\qquad - (1/4) A^{-1/2} k^{1/2} E \left\{ \left[\sum_{i=1}^{N} (\varepsilon_i^2 - 1) \right]^2 \right\} + o(1),
\end{aligned}
$$

where we used the identity in (2.13), (2.14), (2.27), (2.20), (2.1), Lemma 5 of Chow and Yu (1981) and arguments similar to (3.28) in Martinsek (1990) to get the last equality. Now, combine (3.1) to (3.9) to get

$$
\begin{aligned}
R_N - 2(Ak)^{1/2} &= 2E\{N - (Ak)^{1/2}\} + k^{-1}E\left\{\sum_{i=1}^{N} X_i'[\Sigma^{-1} - NM_N^{-1}]X_i\right\} \\
&\quad + k^{-1}E\left\{\left(\sum_{i=1}^{N} \varepsilon_i X_i'\right)[NM_N^{-1} - \Sigma^{-1}]\left(\sum_{i=1}^{N} X_i\varepsilon_i\right)\right\} \\
&\quad + (3/2)A^{-1}k^{-2}E\left\{\left[\sum_{i=1}^{N} (\varepsilon_i^2 - 1)\right]^2 \left(\sum_{i=1}^{N} \varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N} X_i\varepsilon_i\right)\right\} \\
&\quad - (1/2)A^{-1/2}k^{-1/2}E\left\{\left[\sum_{i=1}^{N} (\varepsilon_i^2 - 1)\right]^2\right\} - 2E(U) \qquad (3.10) \\
&\quad - A^{-1/2}k^{-3/2}E\left\{\sum_{i=1}^{N} (\varepsilon_i^2 - 1)\left[\left(\sum_{i=1}^{N} \varepsilon_i X_i'\right)\Sigma^{-1}\left(\sum_{i=1}^{N} X_i\varepsilon_i\right)\right.\right. \\
&\qquad\qquad \left.\left. - \sum_{i=1}^{N} (X_i'\Sigma^{-1}X_i)\right]\right\} \\
&\quad + \begin{cases} k^{-1}E\max\{\chi_k^2, c\}\chi_k^2 + o(1), & \text{when } \boldsymbol{\beta} = \mathbf{0}, \\ (k+2) + o(1), & \text{when } \boldsymbol{\beta} \neq \mathbf{0}. \end{cases}
\end{aligned}
$$

From (2.29), Lemma 3 of Martinsek (1990) applied to N in place of T, (3.10), Wald's identity for the second moment and Lemma 2,

$$
\begin{aligned}
R_N - 2(Ak)^{1/2} &= k^{-1}E\left\{\sum_{i=1}^{N} X_i'[\Sigma^{-1} - NM_N^{-1}]X_i\right\} \\
&\quad + k^{-1}E\left\{\left(\sum_{i=1}^{N} \varepsilon_i X_i'\right)[NM_N^{-1} - \Sigma^{-1}]\left(\sum_{i=1}^{N} X_i\varepsilon_i\right)\right\} \\
&\quad + (3/2)E(\varepsilon_1^2 - 1)^2 - (1/2)E(\varepsilon_1^2 - 1)^2 - 2E(U) \\
&\quad - E(\varepsilon_1^2 - 1)^2 + 2E(U) - (3/4)E(\varepsilon_1^2 - 1)^2 \\
&\quad + \begin{cases} E(k^{-1}\chi_k^2 - 1)\max\{\chi_k^2, c\} + o(1), & \text{when } \boldsymbol{\beta} = \mathbf{0}, \\ 2 + o(1), & \text{when } \boldsymbol{\beta} \neq \mathbf{0}. \end{cases}
\end{aligned}
$$

The result in (1.11) of the theorem now follows easily. □

Acknowledgments. I would like to thank the two referees and an Associate Editor for a careful reading of the paper.

REFERENCES

CHOW, Y. S. and TEICHER, H. (1978). *Probability Theory: Independence, Interchangeability, Martingales*. Springer, New York.

CHOW, Y. S. and YU, K. F. (1981). On the performance of a sequential procedure for the estimation of the mean. *Ann. Statist.* **9** 184–189.

FINSTER, M. (1983). A frequentistic approach to sequential estimation in the general linear model. *J. Amer. Statist. Assoc.* **78** 403–407.

LAI, T. L. and SIEGMUND, D. (1977). A nonlinear renewal theory with applications to sequential analysis. I. *Ann. Statist.* **5** 946–954.

LAI, T. L. and SIEGMUND, D. (1979). A nonlinear renewal theory with applications to sequential analysis. II. *Ann. Statist.* **7** 60–76.

MARTINSEK, A. T. (1983). Second order approximation to the risk of a sequential procedure. *Ann. Statist.* **11** 827–836.

MARTINSEK, A. T. (1990). Sequential point estimation in regression models with nonnormal errors. *Sequential Anal.* **9** 243–268.

MUKHOPADHYAY, N. (1974). Sequential estimation of regression parameters in Gauss–Markoff setup. *J. Indian Statist. Assoc.* **12** 39–43.

ROBBINS, H. (1959). Sequential estimation of the mean of a normal population. In *Probability and Statistics The Harald Cramér Volume* (U. Grenander, ed.) 235–245. Almquist and Wiksell, Uppsala.

SRIRAM, T. N. (1988). Sequential estimation of the autoregressive parameter in a first order autoregressive process. *Sequential Anal.* **7** 53–74.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

TAKADA, Y. (1992). A sequential procedure with asymptotically negative regret for estimating the normal mean. *Ann. Statist.* **20** 562–569.

WOODROOFE, M. (1977). Second order approximations for sequential point and interval estimation. *Ann. Statist.* **5** 984–995.

WOODROOFE, M. (1982). *Nonlinear Renewal Theory in Sequential Analysis*. SIAM, Philadelphia.

DEPARTMENT OF STATISTICS
UNIVERSITY OF GEORGIA
ATHENS, GEORGIA 30602