SEQUENTIAL SHRINKAGE ESTIMATION IN THE GENERAL LINEAR MODEL

by

T. N. Sriram and Arup Bose
Purdue University

Technical Report #87-23

Department of Statistics
Purdue University

June 1987

SEQUENTIAL SHRINKAGE ESTIMATION IN THE GENERAL LINEAR MODEL

by

T. N. SRIRAM and ARUP BOSE
PURDUE UNIVERSITY

ABSTRACT

The paper considers the estimation of the slope parameter $\beta \in \mathbb{R}^k$ for $k \geq 3$, in a general linear model. A class of James-Stein estimator is proposed and is compared with the least squares estimator under an appropriate stopping rule. It is shown that the sequential James-Stein estimator dominates the sequential least squares estimator. Furthermore, under mild regularity conditions, a second order asymptotic risk expansion for the sequential James-Stein estimator is obtained.

Keywords and Phrases: James-Stein estimators; regression parameters; risk expansion; reverse submartingales.

1. Introduction. Let $Y_1, Y_2, \ldots$ be a sequence of independent vector-valued observations generated via the general linear model

$$Y_i = X_i\beta + \epsilon_i, \quad i = 1, 2, \ldots, \tag{1.1}$$

where $\beta \in \mathbb{R}^k (k \geq 3)$, is unknown and $\{\epsilon_i,\ i \geq 1\}$ is a sequence of independent and identically distributed (i.i.d.) real-valued vectors having normal distribution with mean zero and covariance $\sigma^2 I$. Throughout this paper it will be assumed that $\{X_i,\ i \geq 1\}$ is a sequence of i.i.d. $m \times k$ random matrices and that they are jointly independent of ϵ_i's. Furthermore, assume also that, with probability one each X_i has minimum rank $r \geq 1$. If $r < k$, assume that the distribution of X_i is continuous on the matrices of rank less than k so that for $n \geq p = [k/r]$, the matrix $M_n = \sum_{i=1}^{n} X_i' X_i$ is invertible with probability one. Here $[\cdot]$ represents the greatest integer.

Suppose one may choose a sample of size n and wish to estimate β by an estimator $\delta_n = \delta_n(Y_1, \ldots, Y_n)$, with loss

$$L_n = L_n(\beta, \delta_n) = n^{-1}(\beta - \delta_n)' M_n (\beta - \delta_n) + cn, \tag{1.2}$$

where $c(> 0)$ is the cost per observation. The case when δ_n is the least squares estimator

$$\hat{\beta}_n = M_n^{-1} \left(\sum_{i=1}^{n} X_i' Y_i \right), \tag{1.3}$$

was considered by Finster (1983) (with a slightly more general loss than in (1.2)), where the problem was to choose an appropriate sample size that will minimize the risk $R_n = EL_n$. The following arguments given in Finster (1983) show that, if σ^2 is unknown, there is no fixed sample size procedure that minimizes R_n. Suppose $\sigma^2 (> 0)$ is known, then the risk

$$\begin{aligned} R_n &= n^{-1} E[E\{(\beta - \hat{\beta}_n)' M_n (\beta - \hat{\beta}_n) | X_1, \ldots, X_n\}] + cn \\ &= n^{-1} k\sigma^2 + cn \end{aligned} \tag{1.4}$$

is minimized by an integer adjacent to

$$N = (k\sigma^2/c)^{1/2}, \tag{1.5}$$

with corresponding minimum risk

$$R_N \simeq 2cN. \tag{1.6}$$

However, if σ^2 is unknown, then the best fixed sample size N cannot be used. For this case, Finster (1983) proposed a stopping rule T, given by

$$T = \inf\{n \geq \eta\colon\ n \geq (k/c)^{1/2} \hat{\sigma}_n \ell_n\}, \tag{1.7}$$

where $\hat{\sigma}_n^2 = (nm-k)^{-1} \sum_{i=1}^{n} |Y_i - X_i\hat{\beta}_n|^2$, $\eta(\geq p)$ is the initial sample size, $\ell_n \geq 1$, $n \geq 1$ and $\ell_n \downarrow 1$ as $n \to \infty$. Here $|Y|^2 = Y'Y$ for any vector Y.

This type of stopping rule was first introduced by Robbins (1959) for the sequential estimation of the mean of a normal population with unknown variance. Using the sequential procedure $\hat{\beta}_T$ to estimate β, Finster (1983) showed that, when $\eta > \max(p, (k+1)/m)$, the "regret"

$$R_T - R_N = 2^{-1}mc + o(c), \qquad \text{as } c \to 0, \tag{1.8}$$

where $R_T = EL_T$. The proof of regret expansion in (1.8) involves the use of non-linear renewal theory, which was first formulated by Lai and Siegmund (1977, 1979). For a good exposition of the literature see chapters 4 and 10 in Woodroofe (1982).

In this paper we consider a class δ_n^b of James-Stein (J-S) estimators defined by

$$\delta_n^b = \hat{\beta}_n - \frac{bs_n^2}{(\hat{\beta}_n - \beta_0)'M_n(\hat{\beta}_n - \beta_0)}(\hat{\beta}_n - \beta_0) \qquad b > 0, \tag{1.9}$$

where $s_n^2 = \alpha\hat{\sigma}_n^2$, $\alpha(> 0)$ is a suitable constant to be chosen later and $\beta_0 \in \mathbb{R}^k$ is the known point towards which we shrink. In what follows, first we establish the risk dominance of the class of estimators δ_T^b over $\hat{\beta}_T$ for each fixed c and appropriate choice of b's and α's. Secondly, we also obtain a second order asymptotic risk expansion for δ_T^b, under mild regularity conditions on the design matrices. At this point it must be mentioned that the results of the type mentioned above have been obtained by Nickerson (1987) for the sequential shrinkage estimation of β. The results obtained here extend Nickerson (1987)'s work. Furthermore, the method of proof of crucial lemmas used here are different from that of the ones in Nickerson (1987). The remarks following Theorems 1 and 2 below clarify the extent of generalization.

The following theorems establish the risk dominance of δ_T^b over $\hat{\beta}_T$ and asymptotic risk expansion for δ_T^b. The proof of Theorems 1 and 2 will be given in sections 2 and 3 respectively. Before we state the theorems, we need to introduce the following notations. Let

$$U_n = U_n(\beta_0) = (\hat{\beta}_n - \beta_0)'M_n(\hat{\beta}_n - \beta_0), \quad Z_i = X_i(\beta - \beta_0), \quad \text{for } i \geq 1, \tag{1.10}$$

$$\tilde{Z}_i = Z_i'Z_i,$$

$$\lambda_n = (\beta - \beta_0)'M_n(\beta - \beta_0)/2\sigma^2 = \sum_{i=1}^{n} \tilde{Z}_i/2\sigma^2,$$

$$\lambda = (\beta - \beta_0)'(EX_1'X_1)(\beta - \beta_0)(< \infty)$$

$$I(\lambda_n) = (2\sigma^2)^{-1}ne^{-\lambda_n}\sum_{r=0}^{\infty}(r + k/2 - 1)^{-1}\lambda_n^r/r!,$$

and

$$\alpha_0 = (\eta m - k)(\eta m - k + 2)^{-1}$$
$$R_T^b = EL_T(\beta, \delta_T^b).$$

Theorem 1. If $\eta > \max(p, 2k/m, k/m + 1)$

$$R_T^b < R_T \tag{1.11}$$

for every $b \in (0, 2(k-2))$ and $\alpha \in [\frac{1}{2}, \alpha_0]$.

Remark 1. Nickerson (1987) considers the case when in model (1.1), $m = 1$ and X_i's are non-random $1 \times k$ vectors. In this case, $\max(p, 2k/m, k/m + 1) = 2k$ and hence we need $\eta \geq 2k$, which agrees with Nickerson's condition (see Nickerson (1987), Theorem 2.1).

The multiplier α of $\hat{\sigma}_n^2$ in the J-S estimator δ_n^b makes s_n^2 a biased estimator of σ^2. The need for a biased estimator was originally observed by Nickerson (1987), who uses $\alpha = k/(k+2)$, and this bias tends to zero as $k \to \infty$. Our range of values for α not only includes $k/(k+2)$ (for $\eta \geq 2k/m$) but also has the additional property that the bias can be made as small as we please by choosing a large enough initial sample size. This reduction in bias with respect to initial sample size seems more reasonable than the reduction of bias by increasing the dimension k (or m), especially if one has a problem of estimation of β of a fixed dimension.

Theorem 2. Assume for $\epsilon > 0$, $P[\tilde{Z}_i \leq x] = 0(x^\epsilon)$ as $x \to 0$. Let $\eta > \max(p, (k+1)/m, 1/\epsilon)$. Then for $\beta \neq \beta_0$,

$$R_T^b - R_T = \alpha^2(\lambda k)^{-1}\sigma^2 b[b - 2\alpha^{-1}(k-2)]c + o(c), \text{ as } c \to 0, \tag{1.12}$$

while for $\beta = \beta_0$,

$$R_T^b - R_T = c^{1/2}\alpha^2(k-2)^{-1}k^{-1/2}\sigma[b - 2\alpha^{-1}(k-2)]b + o(c^{1/2}), \text{ as } c \to 0. \tag{1.13}$$

Remark 2. This is also a generalization of Theorem 3.2 of Nickerson (1987), for similar reasons given in the first paragraph of remark 1. The i.i.d. structure of X_i's is needed only for Lemma 3 below to hold. If X_i's are non-random, then the proof of Lemma 3 becomes redundant, if we assume $M_n/n \to M(>0)$. Hence the proof of Theorem 2 remains valid for non-random X_i's.

2. Fixed c domination. The proof of Theorem 1 is along the lines of the proof of Theorem 2.1 in Nickerson (1987). The following lemma is useful for the proof of Theorem 1.

Lemma 1. Assume the model (1.1). Then

$$E\{U_n^{-1}|X_1,\ldots,X_n\} = n^{-1}I(\lambda_n) \text{ almost surely (a.s.).} \tag{2.1}$$

and

$$n^{-1}I(\lambda_n) \text{ is nonincreasing in } n. \tag{2.2}$$

Proof. Given $X_1, X_2, \ldots, X_n$, $M_n^{1/2}(\hat{\beta}_n - \beta_0) \sim N(M_n^{1/2}(\beta - \beta_0), \sigma^2 I)$. Therefore, U_n/σ^2 has a noncentral chi-square distribution with the noncentrality parameter λ_n and k degrees of freedom.

Assertion (2.1) follows from direct computions using the density function of noncentral chi-square random variable (r.v.).

As for assertion (2.2), use the fact that λ_n is nondecreasing in n to conclude that the distribution function of the noncentral chi-square r.v. is decreasing in n (see Fabian and Hannan (1985), Theorem 31, p. 118). That $n^{-1}(\lambda_n)$ is nonincreasing in n now follows easily. Hence the lemma. □

Proof of Theorem 1. Algebraic manipulations yield

$$\begin{aligned} R_T^b - R_T &= -2bET^{-1}s_T^2U_T^{-1}(\hat{\beta}_T - \beta_0)'M_T(\hat{\beta}_T - \beta) \\ &\quad + b^2ET^{-1}s_T^4U_T^{-1} \\ &= (i) + (ii). \end{aligned} \tag{2.3}$$

Consider (ii). Write

$$\begin{aligned} b^{-2}(ii) &= \sum_{n=\eta}^{\infty} E\{n^{-1}s_n^4U_n^{-1}1_{\{T=n\}}\} \\ &= \sum_{n=\eta}^{\infty} E[E\{n^{-1}s_n^4U_n^{-1}1_{\{T=n\}}|X_1,\ldots,X_n\}] \\ &= \sum_{n=\eta}^{\infty} E[E\{n^{-1}s_n^4 1_{\{T=n\}}|X_1,\ldots,X_n\}E\{U_n^{-1}|X_1,\ldots X_n\}], \end{aligned} \tag{2.4}$$

where the last equality was obtained using the conditional independence of $\sigma_\eta^2, \ldots \sigma_n^2$ and $\hat{\beta}_n$ on $\{T = n\}$, given $X_1, \ldots, X_n$ (see Finster (1983), p. 405). From (2.1),

$$(ii) = b^2ET^{-2}s_T^4I(\lambda_T). \tag{2.5}$$

Argue as in (2.4) to get

$$\begin{aligned} (i) = -2b\sum_{n=\eta}^{\infty} E[&E\{n^{-1}s_n^2 1_{\{T=n\}}|X_1,\ldots,X_n\} \\ &\times E\{U_n^{-1}(\hat{\beta}_n - \beta_0)'M_n(\hat{\beta}_n - \beta)|X_1,\ldots,X_n\}]. \end{aligned} \tag{2.6}$$

Let $\tilde{Y} = M_n^{1/2}(\hat{\beta}_n - \beta_0) = (\tilde{Y}_1, \ldots, \tilde{Y}_k)'$ and $\underset{\sim}{\theta} = M_n^{1/2}(\beta - \beta_0) = (\theta_1, \ldots, \theta_k)'$. Then

$$(2.7) \qquad \begin{aligned} U_n^{-1}(\hat{\beta}_n - \beta_0)' M_n (\hat{\beta}_n - \beta) &= \tilde{Y}'(\tilde{Y} - \theta)/\tilde{Y}'\tilde{Y} \\ &= \sum_{i=1}^{k} \tilde{Y}_i(\tilde{Y}_i - \theta_i) / \left(\sum_{j=1}^{k} \tilde{Y}_j^2 \right). \end{aligned}$$

Observe that, given $X_1, \ldots, X_n$, each $\tilde{Y}_i$ has a normal distribution with mean θ_i and variance σ^2. Now use Stein's Identity (see Stein (1981)) to get

$$(2.8) \qquad \begin{aligned} & E\{\tilde{Y}_i(\tilde{Y}_i - \theta_i) / \left(\sum_{j=1}^{k} Y_j^2 \right) | X_1, \ldots, X_n\} \\ &= \sigma^2 E\{\frac{\partial}{\partial \tilde{Y}_i}(\tilde{Y}_i / \sum_{j=1}^{k} \tilde{Y}_j^2) | X_1, \ldots, X_n\} \\ &= \sigma^2 E\{ \left(\sum_{j=1}^{k} \tilde{Y}_j^2 - 2\tilde{Y}_i^2 \right) / \left(\sum_{j=1}^{k} \tilde{Y}_j^2 \right)^2 | X_1, \ldots, X_n\}. \end{aligned}$$

From (2.6)–(2.8) and (2.1)

$$(2.9) \quad \begin{aligned} E\{U_n^{-1}(\hat{\beta}_n - \beta_0)' M_n (\hat{\beta}_n - \beta) | X_1, \ldots, X_n\} &= (k-2)\sigma^2 E\{U_n^{-1} | X_1, \ldots, X_n\} \\ &= (k-2)\sigma^2 n^{-1} I(\lambda_n). \end{aligned}$$

Therefore,

$$(2.10) \qquad (i) = -2b(k-2)\sigma^2 E T^{-2} s_T^2 I(\lambda_T).$$

From (2.3), (2.5) and (2.10)

$$(2.11) \qquad \begin{aligned} R_T^b - R_T &= b E T^{-1} I(\lambda_T) T^{-1} s_T^2 \{b s_T^2 - 2(k-2)\sigma^2\} \\ &= b\sigma^2 \sum_{n=\eta}^{\infty} E[n^{-1} I(\lambda_n) n^{-1} s_n^2 \{b s_n^2 \sigma^{-2} - 2(k-2)\} 1_{\{T=n\}}] \\ &< 2(k-2) b\sigma^2 \sum_{n=\eta}^{\infty} E[n^{-1} I(\lambda_n) n^{-1} s_n^2 \{s_n^2 \sigma^{-2} - 1\} 1_{\{T=n\}}] \end{aligned}$$

where the last inequality holds for every $b \in (0, 2(k-2))$. From this, it suffices to show that

$$(2.12) \qquad \sum_{n=\eta}^{\infty} E[n^{-1} I(\lambda_n) n^{-1} s_n^2 \{s_n^2 \sigma^{-2} - 1\} 1_{\{T=n\}}] \leq 0$$

Let n_0 be the smallest integer $\geq \eta$ such that $\alpha c n_0^2/k\ell_{n_0}^2 \geq \sigma^2$. Write

$$\begin{aligned}
\ell hs \text{ of } (2.12) &= \sum_{n=\eta}^{n_0-1} E[n^{-1}I(\lambda_n)n^{-1}s_n^2\{s_n^2\sigma^{-2}-1\}1_{\{T=n\}}] \\
&\quad + E[n_0^{-1}I(\lambda_{n_0})n_0^{-1}s_{n_0}^2\{s_{n_0}^2\sigma^{-2}-1\}1_{\{T\geq n_0\}}] \\
&\quad + \sum_{n=n_0}^{\infty} E\{[(n+1)^{-1}I(\lambda_{n+1})(n+1)^{-1}s_{n+1}^2\{s_{n+1}^2\sigma^{-2}-1\} \\
&\qquad - n^{-1}I(\lambda_n)n^{-1}s_n^2\{s_n^2\sigma^{-2}-1\}]1_{\{T\geq n+1\}}\} \\
&= (iii)+(iv)+(v). \qquad (2.13)
\end{aligned}$$

Consider (v). Let $n \geq n_0$. Use the fact that on $\{T \geq n+1\}$, $s_n^2 \geq \sigma^2$, together with (2.2) to get

$$\begin{aligned}
(v) \leq \sum_{n=n_0}^{\infty} & E\{(n+1)^{-1}I(\lambda_{n+1})[(n+1)^{-1}s_{n+1}^2\{s_{n+1}^2\sigma^{-2}-1\} \\
& - n^{-1}s_n^2\{s_n^2\sigma^{-2}-1\}]1_{\{T\geq n+1\}}\}. \qquad (2.14)
\end{aligned}$$

Now, by the classical Helmert-type transformation

$$(nm-k)\hat{\sigma}_n^2/\sigma^2 = V_0 + \sum_{i=1}^{n-p} W_i, \qquad (2.15)$$

where $V_0, W_1, W_2, \ldots$ are independent with V_0 having a χ^2_{pm-k} distribution and $W_i (i \geq 1)$ having χ^2_m distribution (see Finster (1983), display (3.1)). Moreover, the distribution of V_0, W_1, $W_2, \ldots$ are independent of $X_i, i \geq 1$ (see Finster (1983), p. 405).

Let $\underset{\sim}{X}_n = (X_1, \ldots, X_n)$ and $\underset{\sim}{W}_{n-p} = (V_0, W_1, \ldots W_{n-p})$, $n \geq 1$. Write the rhs of (2.14)

$$\begin{aligned}
&= \sum_{n=\eta}^{\infty} E(n+1)^{-1}I(\lambda_{n+1})E\{[(n+1)^{-1}s_{n+1}^2\{s_{n+1}^2\sigma^{-2}-1\} \\
&\qquad - n^{-1}s_n^2\{s_n^2\sigma^{-2}-1\}]1_{\{T\geq n+1\}}|\underset{\sim}{X}_{n+1}, \underset{\sim}{W}_{n-p}\}. \qquad (2.16)
\end{aligned}$$

By (2.15), the facts following that and letting $a = nm-k$ we get

(2.17)

$$E[(n+1)^{-1}s_{n+1}^2\{s_{n+1}^2\sigma^2-1\}1_{\{T\geq n+1\}}|\underset{\sim}{X}_{n+1},\underset{\sim}{W}_{n-p}]$$

$$= 1_{\{T\geq n+1\}}\{(n+1)^{-1}(a+m)^{-2}[a^2s_n^4+2a\alpha m\sigma^2 s_n^2+\alpha^2\sigma^4 m(m+2)]\sigma^{-2}$$

$$- (n+1)^{-1}(a+m)^{-1}(as_n^2+\alpha m\sigma^2)\}$$

$$= 1_{\{T\geq n+1\}}n^{-1}s_n^2\{s_n^2\sigma^{-2}-1\}+1_{\{T\geq n+1\}}\{[a^2(a+m)^{-2}(n+1)^{-1}-n^{-1}]s_n^4\sigma^{-2}$$

$$- [a\{a-m(2\alpha-1)\}(a+m)^{-2}(n+1)^{-1}-n^{-1}]s_n^2$$

$$- [(a+m)-\alpha(m+2)]\alpha m(n+1)^{-1}(a+m)^{-2}\sigma^2\}$$

$$= (A)+(B) \qquad \text{a.s.,}$$

where we used the fact that $\{T\geq n+1\}$ is measurable w.r.t. the σ-field generated by $\underset{\sim}{W}_{n-p}$. Note that the multiple of $1_{\{T\geq n+1\}}$ in (B) is a quadratic function of s_n^2 with a negative leading coefficient. Hence the maximum value of this function occurs at

$$\frac{(n+1)(a+m)^2-an[a-m(2\alpha-1)]}{2[(a+m)^2(n+1)-na^2]}\sigma^2<\sigma^2,$$

if $\alpha<3/2$. Since on $\{T\geq n+1\}$, $s_n^2\geq\sigma^2$, it follows that the

(2.18) $$(B)\leqq 1_{\{T\geq n+1\}}\{[a^2(a+m)^{-2}(n+1)^{-1}-n^{-1}]\sigma^2$$
$$- [a\{a-m(2\alpha-1)\}(a+m)^{-2}(n+1)^{-1}-n^{-1}]\sigma^2$$
$$- [(a+m)-\alpha(m+2)]\alpha m(n+1)^{-1}(a+m)^{-2}\sigma^2\}$$

$$= 1_{\{T\geq n+1\}}(a+m)^{-2}(n+1)^{-1}m\{(m+2)\alpha^2+\alpha(a-m)-a\}\sigma^2.$$

Let $\mathcal{G}(\alpha)=(m+2)\alpha^2+\alpha(a-m)-a$. Clearly, the function $\mathcal{G}$ has a minimum value at $\alpha_n=-\frac{(a-m)}{2(m+2)}(<0)$ and $\mathcal{G}(\alpha_n^0)=0$ for $\alpha_n^0=[\sqrt{(a-m)^2+4(m+2)a}-(a-m)]/2(m+2)$. Hence, for each n,

$$\mathcal{G}(\alpha)\leq 0 \quad \forall\, \alpha_n^1\leq\alpha\leq\alpha_n^0,$$

where $\alpha_n^1(<0)$ is the other zero of the function $\mathcal{G}$. Since α_n^0 increases with n, we get that

(2.19) $$\mathcal{G}(\alpha)\leq 0 \qquad \forall\, \alpha_n^1\leq\alpha\leq\min_n \alpha_n^0=\tilde{\alpha}$$

where $\tilde{\alpha}=\{\sqrt{[(\eta-1)m-k]^2+4(m+2)(\eta m-k)}-[(\eta-1)m-k]\}/2(m+2)$. Now combine (2.14)–(2.18) with (2.19) to get

(2.20) $$(v)\leq 0 \qquad \forall\, 0<\alpha\leq\tilde{\alpha}.$$

Next, note that $1_{\{T \geq \eta\}} = 1$ w.p.1. Use (2.15) to get

$$
\begin{aligned}
&E\eta^{-1}I(\lambda_\eta)E\{\eta^{-1}s_\eta^2\{s_\eta^2\sigma^{-2}-1\}|\underset{\sim}{X}_\eta\} \\
&\qquad = E\eta^{-1}I(\lambda_\eta)\eta^{-1}(\eta m-k)^{-1}\alpha\sigma^2[\alpha(\eta m-k+2)-(\eta m-k)] \\
(2.21) &\qquad \leq 0,
\end{aligned}
$$

since $\alpha \leq \alpha_0$, by assumption, where α_0 is as in (1.10). Thus, if $n_0 = \eta$ then $(iv) \leq 0$. It can be checked that $\tilde{\alpha} > \alpha_0$. Hence we need at least $\alpha \leq \alpha_0$ for the theorem to hold. Let $n_0 > \eta$. Consider (iv). Once again using (2.15) and arguments similar to (2.17) and letting $a_0 = n_0 m - k$ we get

$$
\begin{aligned}
(2.22)\ (iv) &= En_0^{-1}I(\lambda_{n_0})E[n_0^{-1}s_{n_0}^2\{s_{n_0}^2\sigma^{-2}-1\}1_{\{T\geq n_0\}}|\underset{\sim}{X}_{n_0},\underset{\sim}{W}_{n_0-1-p}] \\
&= En_0^{-1}I(\lambda_{n_0})E\{n_0^{-1}s_{n_0}^2\{s_{n_0}^2\sigma^{-2}-1\}1_{\{T\geq n_0\}}|\underset{\sim}{X}_{n_0},\underset{\sim}{W}_{n_0-1-p}\} \\
&= En_0^{-1}I(\lambda_{n_0})1_{\{T\geq n_0\}}\{n_0^{-1}a_0^{-2}[(a_0-m)^2s_{n_0-1}^4+2(a_0-m)\alpha m\sigma^2 s_{n_0-1}^2 \\
&\qquad + \alpha^2\sigma^4 m(m+2)]\sigma^{-2} - n_0^{-1}a_0^{-1}[(a_0-m)s_{n_0-1}^2+\alpha m\sigma^2]\} \\
&= En_0^{-1}I(\lambda_{n_0})\{b_{n_0}s_{n_0-1}^4\sigma^{-2} - c_{n_0}s_{n_0-1}^2 - d_{n_0}\sigma^2\}1_{\{T\geq n_0\}},
\end{aligned}
$$

where $b_{n_0} = n_0^{-1}a_0^{-2}(a_0-m)^2$, $c_{n_0} = n_0^{-1}a_0^{-2}(a_0-m)\{a_0-2\alpha m\}$ and $d_{n_0} = n_0^{-1}a_0^{-2}\alpha m$ $\{a_0-\alpha(m+2)\}$. Note that since $\alpha \geq 1/2$

$$c_{n_0} - b_{n_0} = n_0^{-1}a_0^{-2}(a_0-m)m\{1-2\alpha\} \leq 0$$

and that by arguments similar to (2.19)

$$
\begin{aligned}
d_{n_0} + (c_{n_0} - b_{n_0}) &= -n_0^{-1}a_0^{-2}m\{\alpha^2(m+2)+\alpha(a_0-2m)-(a_0-m)\} \\
&\geq 0,
\end{aligned}
$$

for all $\alpha_{n_0}^1 \leq \alpha \leq \alpha_{n_0}$, where both $\alpha_{n_0} = [\sqrt{(a_0-2m)^2+4(m+2)(a_0-m)} - (a_0-2m)]/2(m+2)$ and $\alpha_{n_0}^1$ are zeros of the quadratic function of α above. Since α_{n_0} is increasing in n_0 we have

$$d_{n_0} + (c_{n_0} - b_{n_0}) \geq 0 \qquad \text{for all } 0 < \alpha \leq \alpha^1,$$

where $\alpha^1 = \{\sqrt{[(\eta-2)m-k]^2+4(m+2)[(\eta-1)m-k]} - [(\eta-2)m-k]\}/2(m+2)$. Once again it can be checked that $\alpha^1 \geq \alpha_0$, since $\eta \geq k/m+1$ by assumption. Now, let

$$f_{n_0} = (n_0-1)(2b_{n_0}-c_{n_0}) = n_0^{-1}a_0^{-2}(n_0-1)(a_0-m)\{a_0-2m(1-\alpha)\}.$$

Clearly $f_{n_0} \in (0,1)$, since $n_0 > \eta \geq k/m+1$ and $0 \leq \alpha \leq \alpha_0 < 1$. Rewriting the last expression on the rhs of (2.22) we get

$$
\begin{aligned}
(2.23)\quad (iv) &= E[n_0^{-1}I(\lambda_{n_0})\{(n_0-1)(2b_{n_0}-c_{n_0})(n_0-1)^{-1}s_{n_0-1}^2(s_{n_0-1}^2\sigma^{-2}-1) \\
&\qquad + (c_{n_0}-b_{n_0})\{s_{n_0-1}^4\sigma^{-2}-2s_{n_0-1}^2+\sigma^2\} \\
&\qquad - (d_{n_0}+c_{n_0}-b_{n_0})\sigma^2\}1_{\{T\geq n_0\}}] \\
&\leq f_{n_0}E[n_0^{-1}I(\lambda_{n_0})\{(n_0-1)^{-1}s_{n_0-1}^2(s_{n_0-1}^2\sigma^{-2}-1)\}1_{\{T\geq n_0\}}].
\end{aligned}
$$

If $E\{(n_0-1)^{-1}s^2_{n_0-1}(s^2_{n_0-1}\sigma^{-2}-1)1_{\{T\geq n_0\}}|\underset{\sim}{X}n_0\}\leq 0$, then by the fact that $s^2_n<\sigma^2$ on the set $\{T=n\}$ for all $n\leq n_0-1$ we get that $(iii)+(iv)\leq 0$. In this case (2.12) would follow from the above work. If not, combining (iii) with extreme rhs of (2.23) and using $f_{n_0}\in(0,1)$ and (2.2) one gets

$$\begin{aligned}(iii)+\text{ rhs of (2.23) } &\leq \sum_{n=\eta}^{n_0-2} En^{-1}I(\lambda_n)n^{-1}s^2_n\{s^2_n\sigma^{-2}-1\}1_{\{T=n\}}\\ &+En_0^{-1}I(\lambda_{n_0})(n_0-1)^{-1}s^2_{n_0-1}(s^2_{n_0-1}\sigma^{-2}-1)1_{\{T\geq n_0-1\}}.\end{aligned} \tag{2.24}$$

Argue as in (2.22) and (2.23) and proceed inductively to get either the 2nd expression on the rhs of (2.24) ≤ 0 or finally end with

$$\begin{aligned}\ell\text{hs of (2.24)} &\leq En_0^{-1}I(\lambda_{n_0})\eta^{-1}s^2_\eta(s^2_\eta\sigma^{-2}-1)1_{\{T\geq\eta\}}\\ &\leq 0,\end{aligned}$$

by arguments similar to (2.21). Hence the theorem follows from (2.20) and the above arguments.

3. Asymptotic risk expansion. The proof of Theorem 2 depends on two lemmas, the first of which deals with the necessary and sufficient condition for the assumption of Theorem 2 to hold.

Lemma 2. Let $\xi_1,\xi_2,\ldots$ be sequence of non-negative i.i.d. r.v.'s. If for $\epsilon>0$,

$$P[\xi_1\leq x]=0(x^\epsilon),\qquad \text{as } x\to 0, \tag{3.1}$$

then for $s\geq 0$

$$\left\|1/\sum_{i=1}^{M}\xi_i\right\|_s<\infty\qquad\text{for all } M>s/\epsilon. \tag{3.2}$$

Conversely, if for $s\geq 0$ and $M\geq 1$, $\left\|1/\sum_{i=1}^{M}\xi_i\right\|_s<\infty$, then (3.1) holds for every $\epsilon\leq s/M$.

Proof. Let $\epsilon>0$. Assume (3.1) holds. Let $S_M=\sum_{i=1}^{M}\xi_i$. Write

$$\begin{aligned}E(1/S_M)^s &= \int_0^\infty P[(1/S_M)^s\geq x]dx\\ &=\int_0^\infty P[e^{-x^{1/s}S_M}\geq e^{-1}]dx\\ &\leq e\int_0^\infty E(e^{-x^{1/s}S_M})dx\\ &=e(\int_0^1+\int_1^\infty)\{Ee^{-x^{1/s}\xi_1}\}^M dx,\end{aligned} \tag{3.3}$$

where the inequality above is obtained using the Markov inequality. Clearly $\int_0^1 \{Ee^{-x^{1/s}\xi_1}\}^M dx \le 1$. For $1 < x < \infty$, write

$$
\begin{aligned}
Ee^{-x^{1/s}\xi_1} &= \int_0^\infty P[e^{-x^{1/s}\xi_1} \ge y]dy \\
&= \int_0^\infty P[\xi_1 \le -(\log y)/x^{1/s}]dy \\
&= x^{1/s} \int_0^\infty P[\xi_1 \le z]e^{-zx^{1/s}} dz \\
&\le Cx^{1/s} \int_0^\infty z^{\epsilon} e^{-zx^{1/s}} dz \qquad \text{(by assumption)} \\
&= C_1 x^{-(\epsilon/s)},
\end{aligned}
\tag{3.4}
$$

where the third equality above was obtained using change of variable and that $\xi_i \ge 0$. C and C_1 in (3.4) are constants. From (3.4)

$$
\begin{aligned}
\int_1^\infty \{Ee^{-x^{1/s}\xi_1}\}^M dx &\le C_1 \int_1^\infty x^{-M\epsilon/s} dx \\
&< \infty,
\end{aligned}
$$

for all $M > (s/\epsilon)$. This proves (3.2).

For the converse,

$$
\begin{aligned}
P^M[\xi_i \le x] &= P[\bigcap_{i=1}^{M} \{\xi_i \le x\}] \\
&\le P[(S_M/M) \le x] \\
&= P[(M/S_M) \ge \frac{1}{x}] \\
&\le x^s E(M/S_M)^s 1_{\{(M/S_M) \ge x^{-1}\}} \\
&= o(x^s), \text{ as } x \to 0.
\end{aligned}
$$

Hence the Lemma. □

Henceforth all unidentified limits are as $c \to 0$.

Lemma 3. Assume the model (1.1). Let $\beta \ne \beta_0$. Then

$$
I(\lambda_T) \longrightarrow \lambda^{-1} \quad \text{a.s.,} \tag{3.5}
$$

where $I(\lambda_T)$ is as in (1.10) with λ_n replaced by λ_T. Moreover, under the assumptions of Theorem 2, for some $\delta > 0$

$$
\sup_{c>0} ||I(\lambda_T)||_{1+\delta} < \infty. \tag{3.6}
$$

Hence, $\{[I(\lambda_T)]^{1+\delta'}, c > 0\}$ is uniformly integrable (U.I.) for all $\delta' \in (0, \delta)$.

Proof. For (3.5), write

$$I(\lambda_T) = (2\sigma^2)^{-1}(T\lambda_T^{-1})e^{-\lambda_T}\sum_{\ell=0}^{\infty} 2\ell(2\ell + k - 4)^{-1}\lambda_T^{\ell}/\ell!. \tag{3.7}$$

It follows from the strong law of large numbers and the fact that $T \to \infty$ that

$$(2\sigma^2)^{-1}(T\lambda_T^{-1}) \longrightarrow \lambda^{-1} \quad \text{a.s.} \tag{3.8}$$

Let $\Omega_1 = \{\lambda_T/T \to \lambda/2\sigma^2\}$. For every fixed $\omega \in \Omega_1$,

$$e^{-\lambda_T}(\omega)\sum_{\ell=0}^{\infty} 2\ell(2\ell + k - 4)^{-1}\lambda_T^{\ell}(\omega)/\ell! = E_\omega\{2P(2P + k - 4)^{-1}\}, \tag{3.9}$$

where P is a Poisson r.v. with $\lambda_T(\omega)$ as its mean. Note that P may be defined on a different probability space. Since $\lambda_T(\omega) \to \infty$ we have that $P \xrightarrow{\mathcal{D}} \infty$ (that is, for every $K \geq 0$, $Pr[P \leqq K] \to 0$ as $\lambda_T(\omega) \to \infty$). Therefore,

$$2P(2P + k - 4)^{-1} \xrightarrow{\mathcal{D}} 1. \tag{3.10}$$

Moreover,

$$0 \leq 2P(2P + k - 4)^{-1} \leq 2.$$

Hence,

$$E_\omega[2P(2P + k - 4)^{-1}] \to 1. \tag{3.11}$$

The required result now follows from (3.7) to (3.11). As for (3.6), observe that,

$$2\sigma^2 I(\lambda_T) \leq 2T\lambda_T^{-1},$$

since $2r + k - 2 \geq r + 1$ for $k \geq 3$. Therefore, it suffices to show that

$$\sup_{c>0} ||T/\lambda_T||_{1+\delta} < \infty. \tag{3.12}$$

From (1.10)

$$(2\sigma^2)^{-1}(n/\lambda_n) = n/\sum_{i=1}^{n} \tilde{Z}_i. \tag{3.13}$$

Since $\{\tilde{Z}_i; i \geq 1\}$ forms an i.i.d. sequence of r.v.'s, we have that $\{\ell^{-1}\sum_{i=1}^{\ell} \tilde{Z}_i, \mathcal{G}_\ell; \ell \geq 1\}$ is a reverse martingale with $\mathcal{G}_\ell = \sigma\{j^{-1}\sum_{i=1}^{j} \tilde{Z}_i; j \geq \ell\}$. Let $s = 1 + \delta$, where δ is chosen so that $\eta\epsilon > s$. This is possible since $\eta\epsilon > 1$ by assumption. An application of the maximal inequality for reverse submartingale yields

$$\left\|\sup_{n\geq\eta}(n/\sum_{i=1}^{n}\tilde{Z}_i)\right\|_s \leq [s/(s-1)]\left\|\eta/\sum_{i=1}^{\eta}\tilde{Z}_i\right\|_s < \infty, \tag{3.14}$$

by (3.2) and since $\eta > s/\epsilon$. Now, use (3.13) and (3.14) to get the required result. Hence the lemma. □

Proof of Theorem 2. Consider the case where $\beta \neq \beta_0$. Recall from (2.3), (2.5) and (2.10) that

$$\begin{aligned} R_T^b - R_T &= -2b(k-2)\sigma^2 ET^{-2}s_T^2 I(\lambda_T) + b^2 ET^{-2}s_T^4 I(\lambda_T) \\ &= (i) + (ii). \end{aligned}$$

In order to obtain the required second order risk expansion for δ_T^b, it suffices to show that

$$(i) = -2b(k-2)(\lambda k)^{-1}\sigma^2\alpha c + o(c) \tag{3.15}$$

$$(ii) = b^2(\lambda k)^{-1}\sigma^2\alpha^2 c + o(c). \tag{3.16}$$

Consider (ii). The fact that $Tc^{1/2} \longrightarrow k^{1/2}\sigma$ a.s. (see Finster (1983), Property 3), $\hat{\sigma}_T^2 \longrightarrow \sigma^2$ a.s. and (3.5) yields

$$(Tc^{1/2})^{-2}s_T^4 I(\lambda_T) \longrightarrow (\lambda k)^{-1}\sigma^2\alpha^2 \quad \text{a.s.}. \tag{3.17}$$

From (1.7) and since $\ell_n \geq 1$

$$\begin{aligned} (Tc^{1/2})^{-2}s_T^4 &\leq k^{-1}\alpha^2\hat{\sigma}_T^2 \\ &\leq k^{-1}\alpha^2 \sup_{n\geq\eta}\hat{\sigma}_n^2, \\ &\leq \text{(constant)} \times \sup_{n\geq\eta}(V_0 + \overline{W}_n), \end{aligned} \tag{3.18}$$

where the last inequality was obtained using (2.15). From (3.18), the fact that all powers of $\sup_{n\geq\eta}(V_0 + \overline{W}_n)$ are integrable (see Finster (1983), the proof of Property 4) and Hölder's inequality with $p = (1+\delta)/(1+\delta')$ for $\delta'\epsilon(0,\delta)$, we have that

$$\begin{aligned} \sup_{c>0}\|(Tc^{1/2})^{-2}\hat{\sigma}_T^4 I(\lambda_T)\|_{1+\delta'} &\leq \sup_{c>0}\|I(\lambda_T)\|_{1+\delta}^{\frac{1+\delta'}{1+\delta}}\|\sup_{n\geq\eta}(V_0 + \overline{W}_n)\|_q \\ &< \infty, \end{aligned} \tag{3.19}$$

by (3.6). Therefore

$$\{(Tc^{1/2})^{-2}\hat{s}_T^4 I(\lambda_T); c > 0\} \text{ is U.I.} \tag{3.20}$$

Hence, (3.16) follows from (3.17) and (3.20).

Consider (i). As in (3.17),

$$(Tc^{1/2})^{-2}s_T^2 I(\lambda_T) \longrightarrow \alpha(\lambda k)^{-1} \text{ a.s.} \tag{3.21}$$

Moreover, by (1.7)

$$(Tc^{1/2})^{-2}s_T^2 I(\lambda_T) \leq \alpha k^{-1} I(\lambda_T), \tag{3.22}$$

which is U.I. by Lemma 3. Thus, we obtain (3.15) from (3.21) and (3.22).

For the case when $\beta = \beta_0$, first observe that by Lemma 1

$$n^{-1}I(\lambda_n) = \sigma^{-2}(k-2)^{-1}, \qquad \forall\, n \geq 1. \tag{3.23}$$

Consequently,

$$R_T^b - R_T = -2bET^{-1}s_T^2 + b^2\sigma^{-2}(k-2)^{-1}ET^{-1}s_T^4.$$

Once again use the fact that $Tc^{1/2} \to k^{1/2}\sigma$ a.s. and $\hat{\sigma}_T^2 \to \sigma^2$ a.s., to get

$$(Tc^{1/2})^{-1}s_T^2 \to k^{-1/2}\alpha\sigma \text{ a.s.,} \tag{3.24}$$

and

$$(Tc^{1/2})^{-1}s_T^4 \to k^{-1/2}\alpha^2\sigma^3 \text{ a.s.} \tag{3.25}$$

From (1.7) and arguments similar to (3.18)

$$\begin{aligned}(Tc^{1/2})^{-1}s_T^2 &\leq \alpha k^{-1/2}\hat{\sigma}_T \\ &\leq \text{(constant)} \times \{\sup_{n\geq\eta}(V_0 + \overline{W}_n)\}^{1/2}\end{aligned} \tag{3.26}$$

and

$$\begin{aligned}(Tc^{1/2})^{-1}s_T^4 &\leq \text{(constant)} \times \hat{\sigma}_T^3 \\ &\leq \text{(constant)} \times \{\sup_{n\geq\eta}(V_0 + \overline{W}_n)\}^{3/2}.\end{aligned} \tag{3.27}$$

Now, once again recall the fact that all powers of $\sup_{n\geq\eta}(V_0 + \overline{W}_n)$ are integrable to conclude that both

$$\{(Tc^{1/2})^{-1}s_T^2; c > 0\} \text{ and } \{(Tc^{1/2})^{-1}s_T^4; c > 0\} \text{ are U.I..} \tag{3.28}$$

Combine (3.24) and (3.25) with (3.28) to get the required expansion in (1.13). Hence the theorem. □

ACKNOWLEDGEMENTS

We wish to thank Professor H. Rubin for suggesting the necessary and sufficient condition in Lemma 2. We also wish to thank Professor D. M. Nickerson for providing us his manuscript prior to its publication.

REFERENCES

[1] Fabian, V. and Hannan, J. (1985). *Introduction to Probability and Mathematical Statistics.* Wiley. New York.

[2] Finster, M. (1983). A frequentistic approach to sequential estimation in the general linear model. *Jour. Amer. Statist. Assoc.* **78**, 403–407.

[3] Lai, T. L., and Siegmund, D. (1977). A non-linear renewal theory with application to sequential analysis, I. *Ann. Statist.* **5**, 946–954.

[4] Lai, T. L., and Siegmund, D. (1979). Non-linear renewal theory with application to sequential analysis II. *Ann. Statist.* **7**, 60–76.

[5] Nickerson, D. M. (1987). Sequential shrinkage estimation of linear regression parameters. *Sequential Analysis* (To appear).

[6] Robbins, H. E. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics* (Harald Cramèr volume). Alamquist-Uppsala., 235–245.

[7] Stein, C. M. (1981). Estimation of the mean of a Multivariate normal distribution. *Ann. Statist.* **9**, 1135–1151.

[8] Woodroofe, M. (1982). *Nonlinear renewal theory in Sequential analysis.* SIAM. Philadelphia.