SEQUENTIAL INFERENCE PROCEDURES OF STEIN'S TYPE FOR A CLASS OF MULTIVARIATE REGRESSION PROBLEMS

By Shoutir Kishore Chatterjee

University of Calcutta

Summary. In this paper the case of multivariate regression with *stochastic* predictors is considered, the joint distribution of the predictors being unknown, and the conditional distribution of the predictand given the predictors being normal with an unknown standard deviation. Sequential procedures of Stein's [13] type, terminating with probability one, are developed to obtain tests, confidence regions, and point estimates for the regression parameters. For the tests, the power function does not depend on the unknown distribution of the predictors or any nuisance parameters; for the confidence regions, the "span" is fixed and known; and for the point estimates, the expected loss, for a particular type of loss function, is a known constant. The procedure is subsequently modified to get more useful and "efficient" tests and estimates. Some study of the distribution and expectation of the sample size is also made for the sequential procedures developed.

1. Introduction. In many cases of testing parametric hypotheses, the power functions of the usual tests, based on a fixed number of observations, involve certain nuisance parameters and are thus not completely known. In such cases, the question naturally arises, whether tests with completely known power[1] exist at all. For testing the normal mean, when the standard deviation is unknown, a partial answer to this question was given by Dantzig [8], who showed that for the normal mean no non-trivial tests, based on a fixed number of observations and with power completely known, exist. Thereafter, Stein [13] proposed a two-sample (i.e., two-step sequential) test with completely known power for the same problem and extended these results to the case of the linear regression set-up with *non-stochastic* predictors, the predicted variables corresponding to the different predictor levels being independent, homoscedastic, and normal. This two-sample procedure of Stein was found to yield a confidence region with known "span", for the parameters concerned. Lehmann [10] considered the more general problem of estimation of the location parameter of a distribution involving an unknown scale parameter, and proved that, in case of a fixed number of observations, there are no interval estimates of bounded length, and point estimates of bounded expected loss for a loss function of the form

$$\varphi(|\hat{\theta} - \theta|), \tag{1.1}$$

where θ is the location parameter, $\hat{\theta}$ its estimate, and $\varphi(x)$ is increasing in (0,

Received March 27, 1961; revised January 19, 1962.

[1] Hereafter a test with completely known power will imply a test whose power function does not involve any nuisance parameters.

∞). He also showed that for the normal mean Stein's procedure yields a point estimate with known expected loss, when the loss is of the form (1.1).

There are two important types of general linear regression set-up to which Stein's two-step procedure cannot be easily extended. In the first of these, the predicted variables are assumed to be correlated, having an unknown dispersion matrix. A particular case of this type is the inference problem concerning the mean vector of a multinormal population with unknown dispersion matrix, and in this case, a two-step procedure yielding tests and estimates of the above types has been developed in [2], [3] and [4]. The other type of regression set up relates to the situation where the predictors are stochastic, and we are sampling from a population in which the predictors and the predictand are jointly distributed. This case arises in practice, where predictors cannot be easily or exactly controlled and we have to sample direct from the population, without any prior division of the population into subpopulations giving rise to so-called arrays. Here we propose to construct tests, confidence regions, and point estimates of the above types in the latter case. For point estimation, we take the loss in estimating a multiple parameter $\boldsymbol{\theta}$ (a column vector) by $\hat{\boldsymbol{\theta}}$ as

$$\varphi\{(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})'\boldsymbol{\Lambda}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})\} \tag{1.2}$$

where $\boldsymbol{\Lambda}$ is a positive definite matrix, and $\varphi(x)$ is nondecreasing in $(0, \infty)$. Obviously, (1.2) is a generalization of the form (1.1) to the multiparametric case.

We start with a $(p + 1)$-variate population represented by the joint density function

$$f(x_1, \cdots, x_p)(2\pi\sigma^2)^{-\frac{1}{2}} \exp[-(1/2\sigma^2)(y - \alpha - \beta_1 x_1 - \cdots - \beta_p x_p)^2], \tag{1.3}$$

where f is any p-variate density function (not involving $\alpha, \beta_1, \cdots, \beta_p$, or σ), and the range of y for every fixed $(x_1, \cdots, x_p)$ is $(-\infty, \infty)$. This form of density function obviously implies that, for conditionally fixed $x_1, \cdots, x_p$ (the predictors), y (the predictand) is normally distributed with a linear regression

$$E(y \mid x_1, \cdots, x_p) = \alpha + \beta_1 x_1 + \cdots + \beta_p x_p \tag{1.4}$$

and a constant standard deviation (s.d.) σ. Unless otherwise stated, we shall always assume that $\alpha, \beta_1, \cdots, \beta_p, \sigma$ and the density function f are all unknown.

The particular case of a single predictor $(p = 1)$ which is comparatively easy has been already considered in [6]. Here we take the more complicated case of a general p. It can be seen that, for the present set up, due to our lack of knowledge regarding σ and the distribution of the x's, the test, confidence region, and point estimate for any set of regression parameters, usually constructed from a fixed number of observations, do not satisfy respectively the requirements of known power, "span", and expected loss. In fact, as shown in [5], the results of Dantzig [8], Stein [13], and Lehmann [10] regarding the non-existence of tests and estimates of the desired types, based on a fixed number of observations, can be extended to the present set-up as well, and therefore, we have to adopt a sequential approach to obtain such tests and estimates.

In this paper we specifically consider two broad types of inference regarding the regression parameters of (1.4):

(i) inference on some or all of the regression coefficients $\beta_1, \cdots, \beta_p$, and

(ii) inference on the regression constant α. However, other types of inference, such as, inference on the linear function $\alpha + \sum_1^p \beta_i x_i'$ of the parameters (the predicted value), for given $x_1', \cdots, x_p'$, can also be dealt with by methods similar to that given in the paper (see Section 7). For each of the above two cases, we first derive a sequential procedure which yields tests and estimates of the desired types. Subsequently, we improve upon these procedures to derive tests and estimates which utilize a greater amount of information and are, at the same time, practically more useful.

Before concluding this section, we explain here a phrase we shall often use in connection with confidence regions. For a multiple set of parameters, a confidence region will be said to have *a fixed ellipsoidal contour*, when its form is that of a given ellipsoid with its center unspecified.

2. Some preliminaries. In this paper we shall use t_ν, χ_r^2 and F_{n_1,n_2} as generic notations for Student's t with d.f. (degrees of freedom) ν, central χ^2 with d.f. r, and Fisher's F with d.f. n_1, n_2, respectively. For any quadratic form or symmetric matrix, we shall abbreviate the terms "positive definite" and "positive semidefinite" by p.d. and p.s.d. respectively. A vector will be denoted by a boldface, lower case, and a matrix, by a boldface, capital letter. Where necessary, the order of a matrix will be indicated at the top right-hand corner of the matrix symbol.

$\mathbf{I}^{r\times r}$ will as usual denote the $r \times r$ identity matrix. For $r < s$, we shall write

$$\mathbf{I}^{r\times s} = (\mathbf{I}^{r\times r}\mathbf{O}^{r\times(s-r)})$$

and $\mathbf{I}^{s\times r}$ for its transpose. For any r, s, $\mathbf{1}^{r\times s}$ will denote the $r \times s$ matrix all whose elements are unity.

We shall write

$$\mathbf{x}' = (x_1, \cdots, x_p), \qquad \boldsymbol{\beta}' = (\beta_1, \cdots, \beta_p),$$

where $x_1, \cdots, x_p$ are the predictors, and $\beta_1, \cdots, \beta_p$ are the regression coefficients. For any l observations on $\mathbf{x}$: $\mathbf{x}_k$, $k = 1, 2, \cdots, l$, we shall write

$$\bar{\mathbf{x}}_l = \frac{1}{l}\sum_1^l \mathbf{x}_k, \qquad \mathbf{S}_l^{p\times p} = \sum_{k=1}^l (\mathbf{x}_k - \bar{\mathbf{x}}_l)(\mathbf{x}_k - \bar{\mathbf{x}}_l)' = \sum_{k=1}^l \mathbf{x}_k \mathbf{x}_k' - l\bar{\mathbf{x}}_l \bar{\mathbf{x}}_l'. \tag{2.1}$$

$$\tilde{\mathbf{x}}_k' = (1, \mathbf{x}_k'), \qquad \tilde{\mathbf{S}}_l^{(p+1)\times(p+1)} = \sum_{k=1}^l \tilde{\mathbf{x}}_k \tilde{\mathbf{x}}_k' = \begin{pmatrix} l & \sum_1^l \mathbf{x}_k' \\ \sum_1^l \mathbf{x}_k & \sum_1^l \mathbf{x}_k \mathbf{x}_k' \end{pmatrix} \tag{2.2}$$

For any set of $l(>p)$ observations, we shall always take $\mathbf{S}_l$ and $\tilde{\mathbf{S}}_l$ as p.d. This will be so with probability one because of the continuity of the distribution of $\mathbf{x}$. If, corresponding to $\mathbf{x}_k$, the y-observation is y_k, $k = 1, 2, \cdots$, we shall denote

$$(2.3) \qquad \bar{y}_l = (1/l)\sum_1^l y_k, \quad \boldsymbol{\omega}_l = \sum_1^l y_k(\mathbf{x}_k - \bar{\mathbf{x}}_l).$$

For all the sequential procedures to be considered, to start with, we shall take an initial sample of a fixed size $n_0 \geqq p + 2$:

$$(2.4) \qquad \mathbf{x}_k, y_k, k = 1, 2, \cdots, n_0.$$

We shall write

$$(2.5) \qquad (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_0}) = \mathbf{X}_{n_0}^{p\times n_0}$$

Let $\bar{\mathbf{x}}_{n_0}$, $\mathbf{S}_{n_0}$, $\bar{y}_{n_0}$ and $\boldsymbol{\omega}_{n_0}$ be defined by (2.1) and (2.3). Also let a_{n_0}, $\mathbf{b}_{n_0}$, and $\hat{\sigma}^2$ be respectively the usual estimates of α, $\boldsymbol{\beta}$, and σ^2, based on (2.4). Then as is well known

$$(2.6) \qquad \mathbf{S}_{n_0}\mathbf{b}_{n_0} = \boldsymbol{\omega}_{n_0}, \qquad a_{n_0} = \bar{y}_{n_0} - \mathbf{b}'_{n_0}\bar{\mathbf{x}}_{n_0},$$

and for fixed $\mathbf{X}_{n_0}$, $\mathbf{b}_{n_0}$ is distributed as $N(\boldsymbol{\beta}, \mathbf{S}_{n_0}^{-1}\sigma^2)$, independently of $\bar{y}_{n_0}$. Also, for fixed $\mathbf{X}_{n_0}$, $(n_0 - p - 1)\hat{\sigma}^2/\sigma^2$ is distributed as $\chi^2_{n_0-p-1}$, independently of $\bar{y}_{n_0}$, $\mathbf{b}_{n_0}$.

Let $\mathbf{T}_{n_0}^{p\times p}$ be defined by

$$(2.7) \qquad \mathbf{T}'_{n_0}\mathbf{S}_{n_0}^{-1}\mathbf{T}_{n_0} = \mathbf{I}, \quad \text{i.e.,} \quad \mathbf{T}_{n_0}\mathbf{T}'_{n_0} = \mathbf{S}_{n_0}.$$

We shall write

$$(2.8) \qquad \mathbf{g}_{n_0} = \mathbf{T}'_{n_0}\mathbf{b}_{n_0}$$

Then, for fixed $\mathbf{X}_{n_0}$, $\mathbf{g}_{n_0}$ will be distributed as $N(\mathbf{T}'_{n_0}\beta, \mathbf{I}\sigma^2)$, independently of $\bar{y}_{n_0}$ and $\hat{\sigma}^2$.

Before concluding this section we put forth two results, which will be useful afterwards.

LEMMA 2.1. *If*

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix}$$

is a partitioned square matrix, where $\mathbf{A}_{11}$ *is square,* $\mathbf{A}_{22}$ *is symmetric and non-singular, and* $\mathbf{A}'_{12} = \mathbf{A}_{21}$, *and if* $\mathbf{B}$ *is the non-singular matrix of the same order defined by*

$$\mathbf{B} = \begin{pmatrix} \mathbf{I} & -\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ \mathbf{O} & \mathbf{I} \end{pmatrix},$$

then

$$\mathbf{BAB}' = \begin{pmatrix} \mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21} & \mathbf{O} \\ \mathbf{O} & \mathbf{A}_{22} \end{pmatrix}.$$

The result is given in [1], p. 343, where **A** is p.d. But it can be proved in the same way, under the conditions stated here.

LEMMA 2.2. *Given two* p.d. *matrices* $\mathbf{A}^{q\times q}$ *and* $\mathbf{B}^{q\times q}$, *if* $\mathbf{A} - \mathbf{B}$ *is* p.d., *then* $\mathbf{z}'\mathbf{A}\,\mathbf{z} > \mathbf{z}'\mathbf{B}\,\mathbf{z}$ *for any q-vector* $\mathbf{z} \neq \mathbf{o}$, *and the ellipsodial region* $\mathbf{z}'\mathbf{A}\,\mathbf{z} \leqq f^2$ *lies completely within the ellipsodial region* $\mathbf{z}'\mathbf{B}\,\mathbf{z} \leqq f^2$ *for any* $f^2 > 0$.

The first assertion is evident. The second assertion follows from the first, since, if $\mathbf{z}'\mathbf{B}\mathbf{z} < \mathbf{z}'\mathbf{A}\mathbf{z}$ for any q-vector $\mathbf{z} \neq \mathbf{o}$, then $\mathbf{z}'\mathbf{A}\mathbf{z} \leqq f^2$ implies $\mathbf{z}'\mathbf{B}\mathbf{z} \leqq f^2$.

3. Sequential procedure for inferring on some or all of the regression coefficients $\beta_1, \beta_2, \cdots, \beta_p$. We first derive some results that will be necessary.

LEMMA 3.1. *For a sequence* $\{\mathbf{x}_k\}$ *of real p-vectors, let* $\mathbf{S}_l$ *be as defined by* (2.1). *Then, if for a given symmetric matrix* $\mathbf{\Gamma}^{p\times p}$ *and some* l, $\mathbf{S}_l - \mathbf{\Gamma}$ *is* p.d., *so would be* $\mathbf{S}_{l+1} - \mathbf{\Gamma}$.

PROOF. It is easily seen that $\mathbf{S}_{l+1} = \mathbf{S}_l + \boldsymbol{\zeta}_l\boldsymbol{\zeta}_l'$, where[2]

$$\boldsymbol{\zeta}_l = \{l(l+1)\}^{-\frac{1}{2}}\Big\{\sum_1^l \mathbf{x}_k - l\mathbf{x}_{l+1}\Big\}.$$

Hence, as the sum of a p.d. and a p.s.d. matrix is p.d., the lemma follows.

LEMMA 3.2. *Let, for a sequence of independent observations* $\{\mathbf{x}_k\}$ *on a random p-vector* $\mathbf{x}$, $\mathbf{S}_l$ *be defined as in* (2.1), *and let* $\mathbf{\Gamma}^{p\times p}$ *be a fixed* p.d. *or* p.s.d. *matrix and* $l_0 > p$ *a given integer. Then* $\mathbf{S}_l - \mathbf{\Gamma}$ *is* p.d. *for some* $l \geqq l_0$ *with probability one, if and only if*

$$\lim_{l\to\infty} P\{\rho_l \geqq 1\} = 0, \tag{3.1}$$

where ρ_l *is the largest root of the equation in* ρ

$$|\mathbf{\Gamma} - \rho\mathbf{S}_l| = 0. \tag{3.2}$$

PROOF. The lemma will be proved, if we can show

$$P\{\mathbf{S}_l - \mathbf{\Gamma} \text{ being not p.d. for all } l \geqq l_0\} = 0 \tag{3.3}$$

if and only if (3.1) holds.

Now, by Lemma 3.1 the sets

$$[\{\mathbf{x}_k\} : \mathbf{S}_l - \mathbf{\Gamma} \text{ not p.d.}], \qquad l = l_0, l_0 + 1, \cdots,$$

form a non-decreasing sequence, and, as for such a sequence the measure of the limit is the limit of the measure, we can write (3.3) equivalently as

$$\lim_{l\to\infty} P\{\mathbf{S}_l - \mathbf{\Gamma} \text{ being not p.d.}\} = 0. \tag{3.4}$$

But, from a well known result of matrix algebra (see, e.g., [9], Theorem 48, p. 151), it follows that $\mathbf{S}_l - \mathbf{\Gamma}$ is not p.d., i.e., $\mathbf{\Gamma} - \mathbf{S}_l$ is not negative definite, if and only if ρ_l, the largest root of (3.2), satisfies $\rho_l \geqq 1$. Therefore (3.3) and (3.1) are equivalent.

[2] In a different context, Robbins [11] applies the same transformation on the observations of a single variate.

Note. The least integer $l \geqq l_0$, for which $\mathbf{S}_l - \mathbf{\Gamma}$ is p.d., is a random variable depending on l_0, $\mathbf{\Gamma}$, and $\{\mathbf{x}_k\}$.

In this and the following section, we shall make the following assumption.

ASSUMPTION. The distribution of the set of predictors $\mathbf{x}$ is such that the condition (3.1) holds for any p.d. or p.s.d. matrix $\mathbf{\Gamma}$.

This will ensure termination with probability 1 of the sequential procedures for inferring about the regression coefficients. It is, however, not very restrictive, as may be seen from the following lemma.

LEMMA 3.3. *Under the set-up of Lemma* 3.2, *a sufficient condition for* (3.1) *to hold for any* $\mathbf{\Gamma}$ *is the existence of the dispersion matrix of* $\mathbf{x}$.

PROOF. Let the dispersion matrix of $\mathbf{x}$ exist and be $\mathbf{\Sigma}$ and the mean vector (which then necessarily exists) be $\mathbf{m}$. By a straightforward multivariate generalization of the Weak Law of Large Numbers, as $l \rightarrow \infty$

$$(1/l)\sum_1^l \mathbf{x}_k \xrightarrow{P} \mathbf{m}, \quad (1/l)\sum_1^l \mathbf{x}_k\mathbf{x}_k' \xrightarrow{P} \mathbf{\Sigma} + \mathbf{m}\mathbf{m}'$$

the sign $\xrightarrow{P}$ denoting convergence in probability of a random matrix. Hence, by a well-known result (see Cramér [7], Section 20.6) on the stochastic convergence of rational functions of stochastically convergent sequences, from (2.1), as $l \rightarrow \infty$,

$$(1/l)\mathbf{S}_l \xrightarrow{P} \mathbf{\Sigma},$$

and hence $l\mathbf{S}_l^{-1} \xrightarrow{P} \mathbf{\Sigma}^{-1}$. Therefore, denoting trace by tr, as $l \rightarrow \infty$

$$\text{tr}\,\mathbf{\Gamma}\mathbf{S}_l^{-1} \xrightarrow{P} 0. \tag{3.5}$$

But, as $\mathbf{\Gamma}$ is, at worst, p.s.d., from (3.2),

$$0 \leqq \rho_l \leqq \text{tr}\,\mathbf{\Gamma}\mathbf{S}_l^{-1}. \tag{3.6}$$

From (3.5) and (3.6), we get $\rho_l \xrightarrow{P} 0$ as $l \rightarrow \infty$ and this implies (3.1)

The sampling rule. We now come to the sequential procedure for inferring on any q $(1 \leqq q \leqq p)$ of the regression coefficients and we take these, without any loss of generality, to be $\beta_1, \cdots, \beta_q$. In what follows, we shall write

$$\boldsymbol{\beta}_{(q)}' = (\beta_1, \cdots, \beta_q), \tag{3.7}$$

$\boldsymbol{\beta}_{(q)}$ reducing to $\boldsymbol{\beta}$ or β_1 when $q = p$ or 1.

We start with an integer $n_0 \geqq p + 2$, a constant $z > 0$, and a p.d. matrix $\mathbf{\Lambda}^{q\times q} = (\lambda_{ij})$, all at our choice, and define

$$\mathbf{\Gamma}^{p\times p} = \begin{pmatrix} \mathbf{\Lambda} & \mathbf{O}^{q\times(p-q)} \\ \mathbf{O}^{(p-q)\times q} & \mathbf{O}^{(p-q)\times(p-q)} \end{pmatrix}, \qquad \text{if } q < p;$$

$$= \mathbf{\Lambda}, \qquad \text{if } q = p.$$

Then $\mathbf{\Gamma}$ is p.d. or p.s.d. according as $q < p$ or $q = p$.

Next, taking an initial sample of size n_0 denoted by (2.4), we determine the quantities mentioned in Section 2. We then make $n - n_0$ further observations according to the following rule.

We go on making random selection of units (numbered $n_0 + 1, n_0 + 2, \cdots$) from the population and observing the $\mathbf{x}$-values $\mathbf{x}_k$, $k = n_0 + 1, n_0 + 2, \cdots$. For each $l \geqq n_0 + q$, we calculate $\mathbf{S}_l$ given by (2.1) and see if $\mathbf{S}_l - \hat{\sigma}^2 z^{-1}\mathbf{\Gamma}$ is p.d. (As $\mathbf{S}_l$ is p.d., and $\mathbf{\Gamma}$ is of the above form, in practice, to judge the positive-definiteness of $\mathbf{S}_l - \hat{\sigma}^2 z^{-1}\mathbf{\Gamma}$, we need see the positivity of only its determinant and the $q - 1$ principal minors obtained by deleting the 1st row and column, first 2 rows and columns, $\cdots$, and the first $q - 1$ rows and columns respectively). If $\mathbf{S}_l - \hat{\sigma}^2 z^{-1}\mathbf{\Gamma}$ is p.d., we stop with the lth unit; otherwise, we take the $(l + 1)$th unit and proceed likewise. The terminal value of l being n, we complete the sampling by observing y_k corresponding to the kth unit, $k = n_0 + 1, \cdots, n$.

The sampling rule being as above, the sample-size n is the least integer subject to

$$(3.9) \qquad n \geqq n_0 + q, \qquad \mathbf{S}_n - (\hat{\sigma}^2/z)\mathbf{\Gamma} \quad \text{is p.d.}$$

To understand the meaning of (3.9), let us consider the particular case $q = p$. Here (3.9) becomes: $n \geqq n_0 + p$, $\mathbf{S}_n - \hat{\sigma}^2 z^{-1}\mathbf{\Gamma}$ is p.d. By Lemma 2.2, $\mathbf{S}_n - \hat{\sigma}^2 z^{-1}\mathbf{\Lambda}$ is p.d., if and only if the ellipsoidal region $(\boldsymbol{\beta} - \mathbf{b}_n)'\mathbf{S}_n(\boldsymbol{\beta} - \mathbf{b}_n) \leqq p\hat{\sigma}^2 F_\epsilon$ in the $\boldsymbol{\beta}$ space, lies completely within the ellipsoidal region $\boldsymbol{\beta}'\mathbf{\Lambda}\boldsymbol{\beta} \leqq pzF_\epsilon$ when placed at the same center, whatever be $\mathbf{b}_n$ and F_ϵ. Now, if F_ϵ is the upper $100\epsilon\%$ point of the F_{p,n_0-p-1}-distribution, and $\mathbf{b}_n$ is the estimate of $\boldsymbol{\beta}$ obtained in the usual way from the n observations, the first ellipsoidal region will represent a 100 $(1 - \epsilon)\%$ confidence region for $\boldsymbol{\beta}$. (This will be seen to be rigorously true from the reasoning of Section 4). As z and $\mathbf{\Lambda}$ are given, obviously the second ellipsoidal region is given. Thus, the sampling rule means: "Go on sampling till the sample size $n(\geqq n_0 + p)$ is such that the confidence region for $\boldsymbol{\beta}$ (obtained as above) based on n observations lies completely within a given fixed ellipsoid, when placed at the same center." A similar meaning can be attributed to the sampling rule in the general case $1 \leqq q \leqq p$.

In what follows, we shall write $\mathbf{X}_{n_0}$ for the matrix (2.5) and also,

$$(3.10) \qquad \mathbf{X}_{n-n_0}^{p\times(n-n_0)} = (\mathbf{x}_{n_0+1}, \mathbf{x}_{n_0+2}, \cdots, \mathbf{x}_n),$$

$$(3.11) \qquad \mathbf{X}_\infty^{p\times\infty} = (\mathbf{x}_1, \mathbf{x}_2, \cdots).$$

$\mathbf{X}_\infty$ represents in matric form the infinite sequence of observations of which only the first n are actually taken.

We now prove the following.

THEOREM 3.1. *Under our assumption regarding the* $\mathbf{x}$*-distribution, the above sequential procedure terminates with probability one.*

PROOF. $\hat{\sigma}^2$ is distributed independently of $\mathbf{X}_{n_0}$, and hence, independently of $\mathbf{X}_\infty$. Therefore, given $\hat{\sigma}^2$, $\mathbf{X}_\infty$ still represents a sequence of independent observations on $\mathbf{x}$. Hence, by our assumption regarding the $\mathbf{x}$-distribution and Lemma 3.2, for every given $\hat{\sigma}^2$, the probability that for some $l \geqq n_0 + q$ $\mathbf{S}_l - \hat{\sigma}^2 z^{-1}\mathbf{\Gamma}$ is p.d. is unity. It follows that, unconditionally also, the probability of this event, which is same as the probability of termination of the above sequential procedure, is unity.

By the note under Lemma 3.2, here the value of the sample-size n depends on the particular $\mathbf{X}_\infty$ encountered and $\hat{\sigma}^2$, besides the initial constants n_0, z and $\mathbf{\Lambda}$.

After completing the sampling, we define

$$\mathbf{H}^{(n-n_0+p+1)\times(p+1)} = \begin{pmatrix} n_0^{\frac{1}{2}} & n_0^{\frac{1}{2}}\bar{\mathbf{x}}'_{n_0} \\ \mathbf{o}^{p\times 1} & \mathbf{T}'_{n_0} \\ \mathbf{1}^{(n-n_0)\times 1} & \mathbf{X}'_{n-n_0} \end{pmatrix}. \tag{3.12}$$

Then, by the rule of multiplication of partitioned matrices, we get using (2.1) and (2.7),

$$\mathbf{H}'\mathbf{H} = \begin{pmatrix} n & \sum_1^n \mathbf{x}'_k \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k\mathbf{x}'_k \end{pmatrix} = \tilde{\mathbf{S}}_n . \tag{3.13}$$

We next choose a matrix $\mathbf{C}^{(n-n_0+p+1)\times q}$ subject to

$$\mathbf{C}'\mathbf{H} = (\mathbf{o}^{q\times 1}, \mathbf{I}^{q\times q}) \tag{3.14}$$

$$\mathbf{C}'\mathbf{C} = (z/\hat{\sigma}^2)\mathbf{\Lambda}^{-1} \tag{3.15}$$

according to some predetermined rule depending on $\hat{\sigma}^2$, n, and $\mathbf{H}$. That such a $\mathbf{C}$ can be found under the conditions (3.9) will be shown in the appendix (see Theorem A.1).

Let a $(n - n_0 + p + 1)$-vector be defined by

$$\mathbf{n}' = (n_0^{\frac{1}{2}}\bar{y}_{n_0}, \mathbf{g}'_{n_0}, y_{n_0+1}, \cdots, y_n), \tag{3.16}$$

where $\mathbf{g}_{n_0}$ is given by (2.8). Also, let

$$\hat{\boldsymbol{\beta}}_{(q)}^{q\times 1} = \mathbf{C}'\mathbf{n}. \tag{3.17}$$

Then we have the following theorem.

THEOREM 3.2. *Given $z > 0$ and the* p.d. *matrix $\mathbf{\Lambda}^{q\times q}$, if $\hat{\boldsymbol{\beta}}_{(q)}$ is defined by* (3.16) *and* (3.17), *where $\mathbf{C}$ satisfies the conditions* (3.14) *and* (3.15), *and n is the least integer subject to* (3.9), *the statistic*

$$(1/zq)(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)})'\mathbf{\Lambda}(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)}) \tag{3.18}$$

has the distribution of F_{q,n_0-p-1}. In the particular case $q = 1$, if we take $\mathbf{\Lambda} = \lambda_{11} = 1$, the statistic

$$z^{-\frac{1}{2}}(\hat{\beta}_1 - \beta_1) \tag{3.19}$$

has the distribution of t_{n_0-p-1}.

PROOF. If $\mathbf{X}_\infty$ and $\hat{\sigma}^2$ are fixed, n becomes fixed. Therefore, from (3.16) and the results of Section 2 it follows that, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\mathbf{n}$ is distributed as a multinormal vector with dispersion matrix $\sigma^2\mathbf{I}^{(n-n_0+p+1)\times(n-n_0+p+1)}$ and mean vector given by

$$E(\mathbf{n}' \mid \mathbf{X}_\infty, \hat{\sigma}^2) = (\alpha, \beta_1, \cdots, \beta_p)\mathbf{H}'.$$

Hence, as, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\mathbf{C}$ also is fixed, from (3.17) we see that, given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\hat{\boldsymbol{\beta}}_{(q)}$ has a multinormal distribution with

$$E(\hat{\boldsymbol{\beta}}'_{(q)} \mid \mathbf{X}_\infty, \hat{\sigma}^2) = (\alpha, \beta_1, \cdots, \beta_p)\mathbf{H}'\mathbf{C} = \boldsymbol{\beta}'_{(q)} \tag{3.20}$$

[by (3.14)] and dispersion matrix

$$\sigma^2\mathbf{C}'\mathbf{C} = z(\sigma^2/\hat{\sigma}^2)\mathbf{\Lambda}^{-1}. \tag{3.21}$$

Hence, it follows that, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$

$$\begin{aligned}(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)})'[z(\sigma^2/\hat{\sigma}^2)\mathbf{\Lambda}^{-1}]^{-1}(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)}) &= \\ (1/z)(\hat{\sigma}^2/\sigma^2)(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)})'\mathbf{\Lambda}(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)})&\end{aligned} \tag{3.22}$$

is distributed as a χ^2_q. So, when $\mathbf{X}_\infty$ only is fixed, (3.22) is distributed as a χ^2_q independently of $\hat{\sigma}^2$. Therefore remembering the distribution of $\hat{\sigma}^2$ for fixed $\mathbf{X}_\infty$, we see that, for fixed $\mathbf{X}_\infty$, and hence, unconditionally also, the statistic (3.18) is distributed as F_{q,n_0-p-1}.

In the particular case $q = 1$, $\boldsymbol{\beta}_{(q)}$ reduces to β_1, and $\hat{\boldsymbol{\beta}}_{(q)}$ to, say, $\hat{\beta}_1$. There, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$,

$$(\hat{\sigma}/\sigma)(\lambda_{11}/z)^{\frac{1}{2}}(\hat{\beta}_1 - \beta_1)$$

is distributed as $N(0, 1)$. If we put $\lambda_{11} = 1$, (which, for $q = 1$, we shall hereafter always do by incorporating λ_{11} with $1/z$) then, reasoning as before, we see that the statistic (3.19) is distributed as t_{n_0-p-1} unconditionally. This completes the proof.

Theorem 3.2 can be utilized for testing and estimating $\boldsymbol{\beta}_{(q)}$.

Test. We first consider the test for the hypothesis $H^0: \boldsymbol{\beta}_{(q)} = \boldsymbol{\beta}^0_{(q)}$. By Theorem 3.2 under H^0

$$F = (1/zq)(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}^0_{(q)})'\mathbf{\Lambda}(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}^0_{(q)}) \tag{3.23}$$

is distributed as F_{q,n_0-p-1}. For any $\epsilon(0 < \epsilon < 1)$, let F_ϵ be the point in the F_{q,n_0-p-1}-distribution with upper tail probability ϵ. Then, rejecting H^0 whenever

$$F > F_\epsilon, \tag{3.24}$$

we get a test for H^0 at level of significance ϵ. That the test should correspond to the upper tail of the distribution of F is suggested by the inequality

$$E(F \mid \boldsymbol{\beta}_{(q)}) > E(F \mid \boldsymbol{\beta}^0_{(q)}) \quad \text{for} \quad \boldsymbol{\beta}_{(q)} \neq \boldsymbol{\beta}^0_{(q)}$$

which can be easily proved. We now consider some properties of the test corresponding to the rejection rule (3.24).

We first show that *for any alternative hypothesis H specifying $\boldsymbol{\beta}_{(q)}$ the power of the test is completely known.* For this, we derive the non-null distribution under H of the statistic F given by (3.23).

Given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\hat{\boldsymbol{\beta}}_{(q)}$ has a multinormal distribution with mean vector and dispersion matrix given by (3.20) and (3.21) respectively. Hence, reasoning just as for Theorem 3.2, it follows that, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $q\hat{\sigma}^2F/\sigma^2$ is distributed as a non-central χ^2 with d.f. q and non-centrality parameter $q\hat{\sigma}^2\Delta/\sigma^2$, where

$$\Delta = \frac{1}{zq}(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}^0_{(q)})'\mathbf{\Lambda}(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}^0_{(q)}). \tag{3.25}$$

Also, for fixed $\mathbf{X}_\infty$, $(n_0 - p - 1)\hat{\sigma}^2/\sigma^2$ is distributed as a $\chi^2_{n_0-p-1}$. Combining these two facts, it directly follows that, given $\mathbf{X}_\infty$, F has the density function

$$\sum_{r=0}^{\infty} \frac{\Delta^r}{r\,!}\left(\frac{q}{n_0 - p - 1}\right)^{\frac{1}{2}(q+4r)} \frac{\Gamma\{\frac{1}{2}(n_0 + q - p + 4r - 1)\}}{\Gamma\{\frac{1}{2}(q + 2r)\}\Gamma\{\frac{1}{2}(n_0 - p - 1)\}} \cdot \frac{F^{\frac{1}{2}(q+2r-2)}}{\left\{1 + \dfrac{q(\Delta + F)}{n_0 - p - 1}\right\}^{\frac{1}{2}(n_0+q-p+4r-1)}} \tag{3.26}$$

which evidently must represent the unconditional density of F as well.

The distribution of $qF/(n_0 - p - 1)$, as easily derived from (3.26), is really identical with the non-central distribution given by Stein [13] in the case of testing the general linear hypothesis. However, Stein gives the distribution in an integral form. In a recent paper [12], Ruben, while arriving at Stein's test by a different approach, gives the distribution in the series form as above.

The distribution (3.26) involves the parameter Δ given by (3.25), without depending on any nuisance parameter, or, the unknown $\mathbf{x}$-distribution. Therefore, for any alternative hypothesis H specifying $\boldsymbol{\beta}_{(q)}$ the power of the test corresponding to the rejection rule (3.24) is completely known. Integrating the series (3.26) term by term between the limits (F_ϵ, ∞) we obtain, after a little reduction, the following form for the power function:

$$1 - \left(1 + \frac{q}{n_0 - p - 1}\right)^{-\frac{1}{2}(n_0-p-1)} \sum_{r=0}^{\infty} \frac{1}{r!} \frac{\Gamma\{\frac{1}{2}(n_0 - p - 1) + r\}}{\Gamma\{\frac{1}{2}(n_0 - p - 1)\}} \cdot \left(\frac{q\Delta}{n_0 - p - 1 + q\Delta}\right)^r \cdot I_{\frac{qF_\epsilon}{n_0-p-1+q(\Delta+F_\epsilon)}}\left(\frac{1}{2}q + r, \frac{1}{2}(n_0 - p - 1) + r\right),$$

where I denotes the usual incomplete beta-function ratio.

The power function is monotonic increasing in Δ. This may be seen by writing the power as the expectation of

$$P\{F > F_\epsilon \mid \mathbf{X}_\infty, \hat{\sigma}^2, \boldsymbol{\beta}_{(q)}\} = P\{q(\hat{\sigma}^2/\sigma^2)F > q(\hat{\sigma}^2/\sigma^2)F_\epsilon \mid \mathbf{X}_\infty, \hat{\sigma}^2, \boldsymbol{\beta}_{(q)}\} \tag{3.27}$$

over $\mathbf{X}_\infty$ and $\hat{\sigma}^2$. Now, from what has been said above, (3.27) would be a function of the non-centrality parameter $q\hat{\sigma}^2\Delta/\sigma^2$ of the non-central χ^2, $q\hat{\sigma}^2F/\sigma^2$. Then, from the well-known fact that the upper tail area of the non-central χ^2-distribution beyond a fixed point increases with the non-centrality parameter, it follows that, for every fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, (3.27) would increase with Δ, and hence, so would the power. Reasoning in the same way it can also be seen that we can make the power as close to ϵ or 1 as we please by making Δ sufficiently small or large.

In practice, choice of $\mathbf{\Lambda}$ would determine the form of the 'distance function' Δ, and hence, the shape of equidetectibility contours of the test, which from (3.25) are seen to be ellipsoids of the same family. After choosing $\mathbf{\Lambda}$ and n_0 in a particular problem of testing, we can choose z to make Δ assume any value on a given ellipsoid

$$(3.28) \qquad (\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}^0_{(q)})'\mathbf{\Lambda}(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}^0_{(q)}) = f^2,$$

and thus make the power on the ellipsoid (3.28) equal to any previously specified level $\epsilon' > \epsilon$. Thereby, because of the monotonicity property mentioned above, the power at all points outside (3.28) would be greater than ϵ'.

By Theorem 3.2, in the particular case $q = 1$, to test $H^0:\beta_1 = \beta_1^0$ we can alternatively use the statistic

$$(3.29) \qquad t = z^{-\frac{1}{2}}(\hat{\beta}_1 - \beta_1^0),$$

which is distributed as t_{n_0-p-1} under H_0. Under an alternative hypothesis H specifying β_1, t would be obviously distributed as

$$t_{n_0-p-1} + z^{-\frac{1}{2}}(\beta_1 - \beta_1^0).$$

For formulating the test, we can take either a single tail or both tails of the t_{n_0-p-1}-distribution, according as alternatives on a single side or both sides of β_1^0 are of interest. The power function can always be expressed as an incomplete integral of the t_{n_0-p-1}-distribution. Thus, in the two-sided case, the critical region is: $|t| > t_\epsilon$, where t_ϵ corresponds to the $100\epsilon\%$ level of significance of t. Denoting by $p_{n_0-p-1}(t)$ the density function of t_{n_0-p-1}, the power function of this test would be

$$1 - \int_{-t_\epsilon - z^{-\frac{1}{2}}(\beta_1-\beta_1^0)}^{t_\epsilon + z^{-\frac{1}{2}}(\beta_1-\beta_1^0)} p_{n_0-p-1}(t)\, dt.$$

The power functions of the one-sided tests can similarly be written down. In all cases, it can be seen that the power is completely known and increases with the 'distance' of the alternative value from the null point β_1^0 on the admissible side (or sides) of it. Control of power, in the same sense as before, is also possible by the choice of z.

Confidence region. To derive a confidence region for $\boldsymbol{\beta}_{(q)}$, we see from Theorem 3.2 that, F_ϵ being same as before, the fixed-contour confidence region

$$(3.30) \qquad (\boldsymbol{\beta}_{(q)} - \hat{\boldsymbol{\beta}}_{(q)})'\mathbf{\Lambda}(\boldsymbol{\beta}_{(q)} - \hat{\boldsymbol{\beta}}_{(q)}) \leqq zqF_\epsilon$$

for $\boldsymbol{\beta}_{(q)}$ would have confidence coefficient $1 - \epsilon$. Also, here we can choose beforehand $\mathbf{\Lambda}$ and z (for the given n_0), so that the region (3.30) has the contour of any specified q-dimensional ellipsoid. In the particular case $q = 1$, we can put the estimate of β_1 in the form of a confidence interval, the fixed length of the interval being adjustable by the choice of z.

Point estimate. For point estimation of $\boldsymbol{\beta}_{(q)}$ given a loss function of the form (1.2) (wherein $\boldsymbol{\theta}$ now stands for $\boldsymbol{\beta}_{(q)}$ and $\mathbf{\Lambda}^{q \times q}$ is a p.d. matrix), we can start with any z, n_0 and the $\mathbf{\Lambda}$ *involved in the loss function*, and then adopt the sequential procedure described above. Then taking $\hat{\boldsymbol{\beta}}_{(q)}$ given by (3.17) as the point estimate of $\boldsymbol{\beta}_{(q)}$ by Theorem 3.2, the expected loss comes out as

$$(3.31) \quad E\varphi\{(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)})'\mathbf{\Lambda}(\hat{\boldsymbol{\beta}}_{(q)} - \boldsymbol{\beta}_{(q)})\} = E\varphi\{zqF_{q,\, n_0-p-1}\}.$$

We assume that for the given $\varphi(x)$ the expectation (3.31) exists. Then $\varphi(x)$ being non-decreasing in $(0, \infty)$, it may be readily seen that, as $z > 0$ increases, the expected loss (3.31) increases continuously from $\varphi(0)$ to $\varphi(\infty)$. Therefore, starting with an appropriate z, we can make the expected loss of the proposed point estimate attain any desired intermediate value.

4. Improved procedure for inferring on the regression coefficients. The sequential procedure developed in the preceding section is designed to yield a test with known power, a confidence region with fixed ellipsoidal "contour", and a point estimate with known expected loss for the set of regression coefficients.

As in Stein's [13] two-step procedure for the parameters of the linear regression set-up, the above procedure seems to waste some information in the process of meeting those objectives. With a view to utilizing some more information in his case, Stein suggested a modified two-step procedure, which cuts down somewhat on the cost of sampling and provides tests and estimates which are better as regards performance and, at the same time, are simpler to apply than those given by his original procedure. But the power of the test and the expected loss of the point estimate given by the modified procedure involve the unknown scale parameter; also, the confidence region (with a predetermined confidence coefficient) given by it has a random "contour". In this section, we shall develop in an analogous manner a more economical sequential procedure, which would yield a test, a confidence region, and a point estimate for $\boldsymbol{\beta}_{(q)}$ superior to those of Section 3 as regards performance. Also, the tests and estimates of this section will be much simpler than those of Section 3 from the view point of application. But the power of the test and the expected loss of the point estimate will not be completely known, nor will the "contour" of the confidence region be fixed.

We start with the same n_0, z, and $\mathbf{\Lambda}$ as in Section 3 and, after taking a first sample of size n_0 as there, take $n - n_0$ further observations by a similar sequential procedure, n here being the least integer subject to

$$(4.1) \qquad n \geqq n_0, \qquad \mathbf{S}_n - (\hat{\sigma}^2/z)\mathbf{\Gamma} \quad \text{is p.d.}$$

Comparing (3.9) and (4.1), we see that in the present sequential procedure we are allowed to take $n = l$ even for $n_0 \leqq l < n_0 + q$, provided $\mathbf{S}_l - (\hat{\sigma}^2/z)\mathbf{\Gamma}$ is p.d. Termination with probability 1 of the sequential procedure is assured as before.

Evidently, the present procedure is more economical than that of the preceding section in terms of the A.S.N. We now show that it yields tests and estimates better than those given by the latter. For this we first make a reduction of the conditions (4.1). Let

$$\mathbf{S}_n^{-1} = (S^{ij\cdot n})_{i,j=1,\cdots,p}\,,\ \mathbf{S}_{n(q)}^{q\times q} = [(S^{ij\cdot n})_{i,j=1,\cdots,q}]^{-1} \tag{4.2}$$

and let us partition

$$\mathbf{S}_n^{p\times p} = \begin{pmatrix} \mathbf{S}_{11\cdot n}^{q\times q} & \mathbf{S}_{12\cdot n}^{q\times(p-q)} \\ \mathbf{S}_{21\cdot n}^{(p-q)\times q} & \mathbf{S}_{22\cdot n}^{(p-q)\times(p-q)} \end{pmatrix}.$$

Then, using the well known result (see Anderson [1], p. 42) $\mathbf{S}_{n(q)} = \mathbf{S}_{11\cdot n} - \mathbf{S}_{12\cdot n}\mathbf{S}_{22\cdot n}^{-1}\mathbf{S}_{21\cdot n}$, by Lemma 2.1, we get

$$\mathbf{B}\mathbf{S}_n\mathbf{B}' = \begin{pmatrix} \mathbf{S}_{n(q)} & \mathbf{O}^{q\times(p-q)} \\ \mathbf{O}^{(p-q)\times q} & \mathbf{S}_{22\cdot n} \end{pmatrix}, \tag{4.3}$$

where

$$\mathbf{B} = \begin{pmatrix} \mathbf{I}^{q\times q} & -\mathbf{S}_{12\cdot n}\mathbf{S}_{22\cdot n}^{-1} \\ \mathbf{O}^{(p-q)\times q} & \mathbf{I}^{(p-q)\times(p-q)} \end{pmatrix}. \tag{4.4}$$

Also, by (3.8) and (4.4),

$$\mathbf{B}\mathbf{\Gamma}\mathbf{B} = \mathbf{\Gamma}. \tag{4.5}$$

From (4.3) and (4.5), it follows that there is a non-singular matrix $\mathbf{B}$, for which $\mathbf{B}(\mathbf{S}_n - z^{-1}\hat{\sigma}^2\mathbf{\Gamma})\mathbf{B}'$ is

$$\begin{pmatrix} \mathbf{S}_{n(q)} & \mathbf{O}^{q\times(p-q)} \\ \mathbf{O}^{(p-q)\times q} & \mathbf{S}_{22\cdot n} \end{pmatrix} - (\hat{\sigma}^2/z)\mathbf{\Gamma} = \begin{pmatrix} \mathbf{S}_{n(q)} - (\hat{\sigma}^2/z)\mathbf{\Lambda}, & \mathbf{O}^{q\times(p-q)} \\ \mathbf{O}^{(p-q)\times q} & \mathbf{S}_{22\cdot n} \end{pmatrix}. \tag{4.6}$$

As $\mathbf{S}_{22\cdot n}$ is p.d., (4.6) is p.d., if and only if $\mathbf{S}_{n(q)} - (\hat{\sigma}^2/z)\mathbf{\Lambda}$ is p.d. Therefore, we can rewrite (4.1) as

$$n \geqq n_0 \qquad \mathbf{S}_{n(q)} - (\hat{\sigma}^2/z)\mathbf{\Lambda} \quad \text{is p.d.} \tag{4.7}$$

Now, going back to the inference procedure, let $\mathbf{b}_n^{p\times 1}$ be defined by

$$\mathbf{S}_n\mathbf{b}_n = \boldsymbol{\omega}_n\,, \tag{4.8}$$

where n is the sample size for the present procedure, and $\boldsymbol{\omega}_n$ is given by (2.3). Let us consider the distribution of $\mathbf{b}_n$, given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$. By definition,

$$\boldsymbol{\omega}_n = \boldsymbol{\omega}_{n_0} + n_0\bar{y}_{n_0}(\bar{\mathbf{x}}_{n_0} - \bar{\mathbf{x}}_n) + \sum_{n_0+1}^{n} y_k(\mathbf{x}_k - \bar{\mathbf{x}}_n). \tag{4.9}$$

In (4.9), for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\boldsymbol{\omega}_{n_0}$ is distributed as $N(\mathbf{S}_{n_0}\boldsymbol{\beta}, \sigma^2\mathbf{S}_{n_0})$ independently of $\bar{y}_{n_0}$, which has the usual normal distribution. Also, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, n is fixed, and y_k, $k = n_0 + 1, \cdots, n$ are as usual normally distributed, independently of each other and, of course, $\boldsymbol{\omega}_{n_0}$, $\bar{y}_{n_0}$. From these facts, it directly follows that, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\boldsymbol{\omega}_n$ is distributed as $N(\mathbf{S}_n\boldsymbol{\beta}, \sigma^2\mathbf{S}_n)$. Hence, by virtue of (4.8), we see that the distribution of $\mathbf{b}_n$, given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$ is $N(\boldsymbol{\beta}, \sigma^2\mathbf{S}_n^{-1})$. Therefore given $\mathbf{X}_\infty$, $\mathbf{b}_n$ is distributed as $N(\boldsymbol{\beta}, \sigma^2\mathbf{S}_n^{-1})$ independently of $\hat{\sigma}^2$.

Now, let us write

$$(4.10)\quad \mathbf{b}_n' = (b_{1\cdot n}, b_{2\cdot n}, \cdots, b_{p\cdot n}), \quad \mathbf{b}_{n(q)}' = (b_{1\cdot n}, b_{2\cdot n}, \cdots, b_{q\cdot n})$$

Then clearly, for fixed $\mathbf{X}_\infty$, $\mathbf{b}_{n(q)}$ is distributed as $N(\boldsymbol{\beta}_{(q)}, \sigma^2\mathbf{S}_{n(q)}^{-1})$ independently of $\hat{\sigma}^2$, where $\mathbf{S}_{n(q)}$ is defined by (4.2). From this and the distribution of $\hat{\sigma}^2$ given $\mathbf{X}_\infty$, it follows

THEOREM 4.1. *$\mathbf{b}_{n(q)}$ being defined by* (4.8) *and* (4.10), *where n is the least integer subject to* (4.1), *and $\mathbf{S}_{n(q)}$ is given by* (4.2),

$$(1/q\hat{\sigma}^2)(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)})'\mathbf{S}_{n(q)}(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)})$$

is distributed (*unconditionally*) *as F_{q,n_0-p-1}. In the particular case $q = 1$,*

$$(b_{1\cdot n} - \beta_1)/[\hat{\sigma}(S^{11\cdot n})^{\frac{1}{2}}]$$

is distributed (*unconditionally*) *as t_{n_0-p-1}.*

On the basis of this theorem, we can construct a test for $H^0: \boldsymbol{\beta}_{(q)} = \boldsymbol{\beta}_{(q)}^0$ and a confidence region for $\boldsymbol{\beta}_{(q)}$. We first show that this test would be superior to that developed in Section 3.

Let

$$(4.11)\qquad F' = (1/q\hat{\sigma}^2)(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)}^0)'\mathbf{S}_{n(q)}(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)}^0),$$

and let F_ϵ denote the point in F_{q,n_0-p-1} - distribution with upper tail ϵ. Then, at significance level ϵ, the test for H^0, based on Theorem 4.1, would consist in rejecting H^0, whenever

$$(4.12)\qquad F' > F_\epsilon .$$

For any alternative hypothesis specifying $\boldsymbol{\beta}_{(q)}$, the power of this test would be the expectation of

$$(4.13)\quad P\{F' > F_\epsilon \mid \mathbf{X}_\infty, \hat{\sigma}^2, \boldsymbol{\beta}_{(q)}\} = P\{q(\hat{\sigma}^2/\sigma^2)F' > q(\hat{\sigma}^2/\sigma^2)F_\epsilon \mid \mathbf{X}_\infty, \hat{\sigma}^2, \boldsymbol{\beta}_{(q)}\}$$

over $\mathbf{X}_\infty$ and $\hat{\sigma}^2$. The power of the test corresponding to (3.24) was seen in (3.27) to be the expectation of

$$(4.14)\qquad P\{q(\hat{\sigma}^2/\sigma^2)F > q(\hat{\sigma}^2/\sigma^2)F_\epsilon \mid \mathbf{X}_\infty, \hat{\sigma}^2, \boldsymbol{\beta}_{(q)}\}.$$

Now, from (4.11) and what has been said above, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $q(\hat{\sigma}^2/\sigma^2)F'$ is distributed as a non-central χ^2 with d.f. q and noncentrality parameter

$$(4.15)\qquad (1/\sigma^2)(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}_{(q)}^0)'\mathbf{S}_{n(q)}(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}_{(q)}^0).$$

Also, in Section 3 it was seen that, under the same condition, $q(\hat{\sigma}^2/\sigma^2)F$ has a non-central χ^2-distribution with the same d.f. and noncentrality parameter

$$(4.16)\qquad (\hat{\sigma}^2/z\sigma^2)(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}_{(q)}^0)'\mathbf{\Lambda}(\boldsymbol{\beta}_{(q)} - \boldsymbol{\beta}_{(q)}^0).$$

From (4.7) and Lemma 2.2, it follows that, for any $\boldsymbol{\beta}_{(q)} \neq \boldsymbol{\beta}_{(q)}^0$, (4.15) is greater than (4.16). By a well-known property of the non-central χ^2-distribution, this implies that, for any alternative hypothesis, (4.13) exceeds (4.14) for every fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$. Hence, the test corresponding to (4.12) is uniformly more powerful than that corresponding to (3.24).

Here, in the particular case $q = 1$, we can alternatively base our test for $H^0:\beta_1 = \beta_1^0$ on $t' = (b_{1\cdot n} - \beta_1^0)/[\hat{\sigma}(S^{11\cdot n})^{\frac{1}{2}}]$ and the t_{n_0-p-1} - distribution. As in this case, (4.7) implies $(S^{11\cdot n})^{-1} > z^{-1}\hat{\sigma}^2$, considering conditional powers given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, it can be seen that any two-sided or one-sided test based on t' would be uniformly more powerful than the corresponding test based on (3.29), on the admissible side (or sides) of β_1^0. However, as is easy to see, for the tests of this section, the power function involves both the nuisance parameter σ and the unknown distribution of $\mathbf{x}$ and is thus not evaluable.

Coming to estimation, the $100(1 - \epsilon)$ % confidence region based on Theorem 4.1, is

$$(\boldsymbol{\beta}_{(q)} - \mathbf{b}_{n(q)})'\mathbf{S}_{n(q)}(\boldsymbol{\beta}_{(q)} - \mathbf{b}_{n(q)}) \leqq q\hat{\sigma}^2F_\epsilon, \tag{4.18}$$

and the confidence region (3.30) can be written as

$$(\hat{\sigma}^2/z)(\boldsymbol{\beta}_{(q)} - \hat{\boldsymbol{\beta}}_{(q)})'\mathbf{\Lambda}(\boldsymbol{\beta}_{(q)} - \hat{\boldsymbol{\beta}}_{(q)}) \leqq q\hat{\sigma}^2F_\epsilon. \tag{4.19}$$

From (4.7) and Lemma 2.2, it follows that the region (4.18) lies within (4.19) when placed at the same center and is thus "narrower" than the latter. However, the "contour" of (4.18) depends on the observations taken and therefore varies randomly.

For point estimation of $\boldsymbol{\beta}_{(q)}$, we can adopt the procedure of this section with the $q \times q$ matrix involved in the loss of the form (1.2) as $\mathbf{\Lambda}$ and take $\mathbf{b}_{n(q)}$ as the estimate of $\boldsymbol{\beta}_{(q)}$. As, in (1.2), $\varphi(x)$ is nondecreasing in $(0, \infty)$, by (4.7) and Lemma 2.2, the resulting expected loss

$$\begin{aligned} &E\varphi\{(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)})'\mathbf{\Lambda}(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)})\} \\ &\qquad \leqq E\varphi\{(z/\hat{\sigma}^2)(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)})'\mathbf{S}_{n(q)}(\mathbf{b}_{n(q)} - \boldsymbol{\beta}_{(q)})\}. \end{aligned} \tag{4.20}$$

The equality in (4.20) holds only in the trivial case when $\varphi(x)$ is a constant. Now, by Theorem 4.1, the right-hand side of (4.20) is $E\varphi\{z\cdot F_{q,n_0-p-1}\}$, which, as shown in (3.31), is the expected loss of the point estimate $\hat{\boldsymbol{\beta}}_{(q)}$ proposed in Section 3. Thus, $\mathbf{b}_{n(q)}$ has smaller expected loss than $\hat{\boldsymbol{\beta}}_{(q)}$ as a point estimate of $\boldsymbol{\beta}_{(q)}$. But of course the expected loss of $\mathbf{b}_{n(q)}$ (the left-hand side of (4.20)) depends on σ and the $\mathbf{x}$-distribution and so is not evaluable.

5. Sequential procedures for inferring on the regression constant α. We first note some results similar to those considered at the beginning of Section 3.

LEMMA 5.1. *For a sequence $\{\mathbf{x}_k\}$ of real p-vectors, let $\tilde{\mathbf{S}}_l^{(p+1)\times(p+1)}$ be defined as in* (2.2). *Then, if for a given symmetric matrix $\tilde{\mathbf{\Gamma}}^{(p+1)\times(p+1)}$ and some l, $\tilde{\mathbf{S}}_l - \tilde{\mathbf{\Gamma}}$ is* p.d., *the same would be true for $\tilde{\mathbf{S}}_{l+1} - \tilde{\mathbf{\Gamma}}$.*

PROOF. As, by definition, $\tilde{\mathbf{S}}_{l+1} = \tilde{\mathbf{S}}_l + \tilde{\mathbf{x}}_{l+1}\tilde{\mathbf{x}}_{l+1}'$, the lemma follows directly.

LEMMA 5.2. *Let, for a sequence of independent observations $\{\mathbf{x}_k\}$ on a random* *p-vector* $\mathbf{x}$, $\tilde{\mathbf{S}}_l$ *be defined as in* (2.2), *and let* $\tilde{\mathbf{\Gamma}}^{(p+1)\times(p+1)}$ *be a fixed* p.d. *or* p.s.d. *matrix and* $l_0 > p$ *a given integer. Then* $\tilde{\mathbf{S}}_l - \tilde{\mathbf{\Gamma}}$ *is* p.d. *for some* $l \geqq l_0$ *with probability one, if and only if*

$$\lim_{l\to\infty} P\{\tilde{\rho}_l \geqq 1\} = 0, \tag{5.1}$$

where $\tilde{\rho}_l$ *is the largest root of the equation in* ρ

$$|\tilde{\mathbf{\Gamma}} - \rho\tilde{\mathbf{S}}_l| = 0.$$

Proof. The proof using Lemma 5.1 closely follows that of Lemma 3.2.

Note. The least integer $l \geqq l_0$ for which $\tilde{\mathbf{S}}_l - \tilde{\mathbf{\Gamma}}$ is p.d. is determined by l_0, $\tilde{\mathbf{\Gamma}}$ and $\{\mathbf{x}_k\}$.

To ensure termination with probability 1 of the sequential procedures to be developed in this section, we shall assume the following

Assumption. The distribution of the set of predictors $\mathbf{x}$ is such that, for every matrix $\tilde{\mathbf{\Gamma}}^{(p+1)\times(p+1)}$ for which the element at the upper left-hand corner is positive and all other elements are zero, condition (5.1) holds.

Lemma 5.3. *Under the set-up of Lemma 5.2, a sufficient condition for* (5.1) *to hold for any* p.d. *or* p.s.d. $\tilde{\mathbf{\Gamma}}$ *is that the dispersion matrix exists for* $\mathbf{x}$.

Proof. If the dispersion matrix of $\mathbf{x}$ exists, using the notation for Lemma 3.3 and by a similar argument, it follows that

$$(1/l)\tilde{\mathbf{S}}_l \xrightarrow{P} \begin{pmatrix} 1 & \mathbf{m}' \\ \mathbf{m} & \mathbf{\Sigma} \end{pmatrix}.$$

Hence, we get $\tilde{\rho}_l \xrightarrow{P} 0$, which implies condition (5.1). The above assumption also is realized if the dispersion matrix exists for $\mathbf{x}$.

We now give a sequential procedure for inferring about the regression constant α. As in Section 3, we start with $n_0 \geqq p + 2$, and $z > 0$ and define the p.s.d. matrix

$$\tilde{\mathbf{\Gamma}}^{(p+1)\times(p+1)} = \begin{pmatrix} 1 & \mathbf{0}^{1\times p} \\ \mathbf{0}^{p\times 1} & \mathbf{O}^{p\times p} \end{pmatrix}. \tag{5.2}$$

Then, as before, after determining the quantities mentioned in Section 2 from an initial sample (2.4) of size n_0, we sequentially take $n - n_0$ further observations $\mathbf{x}_k$, y_k, $k = n_0 + 1, \cdots, n$, where n is the least integer subject to

$$n \geqq n_0 + 1, \qquad \tilde{\mathbf{S}}_n - (\hat{\sigma}^2/z)\tilde{\mathbf{\Gamma}} \quad \text{is p.d.} \tag{5.3}$$

Under our assumption regarding the $\mathbf{x}$-distribution, the termination with probability 1 of this sequential procedure is assured by Lemma 5.2 and the sample size n depends on, besides n_0 and z, the particular $\mathbf{X}_\infty$ (defined by (3.11)) encountered and $\hat{\sigma}^2$.

Let $\mathbf{H}$ be as defined by (3.10) and (3.12) (where n, of course, is as here). We next determine a $\tilde{\mathbf{c}}^{(n-n_0+p+1)\times 1}$ subject to

$$\tilde{\mathbf{c}}'\mathbf{H} = \mathbf{I}^{1\times(p+1)} \tag{5.4}$$

$$\tilde{\mathbf{c}}'\tilde{\mathbf{c}} = (z/\hat{\sigma}^2) \tag{5.5}$$

according to some predetermined rule depending on $\hat{\sigma}^2$, n, and $\mathbf{H}$. In the appendix, we shall show (Theorem A.2) that such a $\tilde{\mathbf{c}}$ can be determined, if conditions (5.3) hold. Now consider,

$$\hat{\alpha} = \tilde{\mathbf{c}}'\mathbf{n}, \tag{5.6}$$

where $\mathbf{n}$ is as in (3.16) (n being as here). Then, reasoning as in Section 3 and remembering (5.4) and (5.5), we see that, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, $\hat{\alpha}$ is distributed as $N(\alpha, z\sigma^2/\hat{\sigma}^2)$. Hence we get the following

THEOREM 5.1. *For a given z, if $\hat{\alpha}$ is defined by* (5.6), *where $\tilde{\mathbf{c}}$ satisfies* (5.4) *and* (5.5), *and n is the least integer subject to* (5.3), *then $z^{-\frac{1}{2}}(\hat{\alpha} - \alpha)$ is distributed (unconditionally) as t_{n_0-p-1}.*

On the basis of this, we can construct a test for $H^0:\alpha = \alpha^0$ and a confidence interval for α in the usual manner. The power of the test would be a known function of $z^{-\frac{1}{2}}(\alpha - \alpha^0)$ and would be controllable by the choice of z. The length of the confidence interval would be a known constant multiple of $z^{\frac{1}{2}}$ and would therefore be adjustable by the choice of z. Further, proceeding as above, if we take $\hat{\alpha}$ as a point estimate of α, the expected loss, for a loss function of the type (1.2), would be known and adjustable by the choice of z.

We now indicate how an improved procedure of the type given in Section 4 can be developed in this case also. Starting with the same n_0, z, and $\tilde{\mathbf{\Gamma}}$ as above and taking a first sample of size n_0, here we take sequentially a second sample of size $n - n_0$, where n is the least integer subject to

$$n \geqq n_0 \qquad \tilde{\mathbf{S}}_n - (\hat{\sigma}^2/z)\tilde{\mathbf{\Gamma}} \quad \text{is p.d.,} \tag{5.7}$$

termination with probability 1 of the procedure being guaranteed as before.

Comparison of (5.3) and (5.7) reveals that the A.S.N. of the present procedure would be slightly lesser than that of the preceeding. Before coming to the test and estimates resulting from it, we note that by Lemma 2.1 the positive-definiteness of $\tilde{\mathbf{S}}_n - z^{-1}\hat{\sigma}^2\tilde{\mathbf{\Gamma}}$ is equivalent to that of

$$\begin{pmatrix} \mathbf{S}_n & \mathbf{l}^{(p+1)\times 1} \\ \mathbf{l}^{1\times(p+1)} & (z/\hat{\sigma}^2) \end{pmatrix} = \begin{bmatrix} n & \sum_1^n \mathbf{x}_k' & 1 \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k \mathbf{x}_k' & \mathbf{0}^{p\times 1} \\ 1 & \mathbf{0}^{1\times p} & z \end{bmatrix}. \tag{5.8}$$

After pre- and post-multiplication by the non-singular matrix

$$\begin{pmatrix} 1 & \mathbf{0}^{1\times p} & 0 \\ -\bar{\mathbf{x}}_n & \mathbf{I}^{p\times p} & \mathbf{0}^{p\times 1} \\ -(1/n) & \mathbf{0}^{1\times p} & 1 \end{pmatrix}$$

and its transpose respectively, the right-hand side of (5.8) becomes (because of (2.1))

$$\begin{pmatrix} n & \mathbf{0}^{1\times p} & 0 \\ \mathbf{0}^{p\times 1} & \mathbf{S}_n & -\bar{\mathbf{x}}_n \\ 0 & -\bar{\mathbf{x}}_n' & (z/\hat{\sigma}^2) - (1/n) \end{pmatrix}. \tag{5.9}$$

As $\mathbf{S}_n$ is p.d., (5.9) is so, if and only if

$$\begin{vmatrix} \mathbf{S}_n & -\bar{\mathbf{x}}_n \\ -\bar{\mathbf{x}}_n' & (z/\hat{\sigma}^2) - (1/n) \end{vmatrix} > 0.$$

Therefore, (5.7) can be rewritten as

$$n \geqq n_0, \qquad (1/n) + \bar{\mathbf{x}}_n'\mathbf{S}_n^{-1}\bar{\mathbf{x}}_n < (z/\hat{\sigma}^2). \tag{5.10}$$

After completing the sampling, let us define $\mathbf{b}_n$ by (4.8) (n being as here), and write

$$a_n = \bar{y}_n - \mathbf{b}_n'\bar{\mathbf{x}}_n . \tag{5.11}$$

Now writing $\bar{y}_n = (1/n)(n_0\bar{y}_{n_0} + \sum_{n_0+1}^{n} y_k)$, and then arguing as in Section 4, from (5.11) we deduce that, for fixed $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, a_n is distributed as $N(\alpha, \sigma^2\{(1/n) + \bar{\mathbf{x}}_n'\mathbf{S}_n^{-1}\bar{\mathbf{x}}_n\})$. Hence we get the following

THEOREM 5.2. *a_n being defined by (5.11) and (4.8), where n is the least integer subject to (5.7),*

$$(a_n - \alpha)/\hat{\sigma}\left\{\frac{1}{n} + \bar{\mathbf{x}}_n'\,\mathbf{S}_n^{-1}\bar{\mathbf{x}}_n\right\}^{\frac{1}{2}}$$

is distributed (unconditionally) as t_{n_0-p-1}.

As n satisfies (5.10), we can show that the test based on Theorem 5.2 is uniformly more powerful than the corresponding test based on Theorem 5.1 (by considering conditional powers given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$), and the confidence interval based on the former is narrower than the corresponding confidence interval based on the latter. As a point estimate of α, a_n can similarly be shown to have a smaller expected loss than $\hat{\alpha}$, for a loss of the type (1.2).

6. Distribution and expectation of sample-size. In this section we show how rough approximations to the distribution functions of the sample-size n for the sequential procedures developed above can be obtained. The exact distributions are derived under the assumption of normality of $\mathbf{x}$, only for the case of inference regarding a single regression coefficient and inference regarding the regression constant α, in the former case, close bounds to the expectation of n (A.S.N.) being also available. For all these studies, we consider only the original procedures (the procedure of Section 3 and the first procedure of Section 5), corresponding results for the improved procedures requiring little modifications.

First, consider the procedure of Section 3. As here n is the least integer subject to (3.9), by Lemma 3.1, we get $P\{n < n_0 + q\} = 0$, and for $l \geqq n_0 + q$,

$$P\{n \leqq l\} = \text{Prob. } \mathbf{S}_l - (\hat{\sigma}^2/z)\mathbf{\Gamma} \quad \text{being p.d.} \tag{6.1}$$

Now, remembering that $\mathbf{\Gamma}$ is defined by (3.8), and carrying out a reduction similar to that of Section 4, we get, for $l \geqq n_0 + q$,

$$P\{n \leqq l\} = \text{Prob. } (1/l)\mathbf{S}_{l(q)} - (\hat{\sigma}^2/zl)\mathbf{\Lambda} \quad \text{being p.d.,} \tag{6.2}$$

where $S_{l(q)}$ is defined by

$$\mathbf{S}_l^{-1} = (S^{ij\cdot l})_{i,j=1,\cdots,p}, \qquad \mathbf{S}_{l(q)} = [(S^{ij\cdot l})_{i,j=1,\cdots,q}]^{-1}. \tag{6.3}$$

Now, let the dispersion matrix of the $\mathbf{x}$-distribution exist and be $\mathbf{\Sigma}$, and let

$$\mathbf{\Sigma}^{-1} = (\sigma^{ij})_{i,j=1,\cdots,p}, \qquad \mathbf{\Sigma}_{(q)} = [(\sigma^{ij})_{i,j=1,\cdots,q}]^{-1}. \tag{6.4}$$

Then, as by the multivariate generalization of the weak law of large numbers $l^{-1}\mathbf{S}_l \xrightarrow{P} \mathbf{\Sigma}$ we have, by a well-known result (see [7] §20.6),

$$(1/l)\mathbf{S}_{l(q)} \xrightarrow{P} \mathbf{\Sigma}_{(q)}. \tag{6.5}$$

Because of (6.5), for large l, in (6.2) we can replace $l^{-1}\mathbf{S}_{l(q)}$ by $\mathbf{\Sigma}_{(q)}$ as an approximation. Therefore, for a large $l \geqq n_0 + q$,

$$P\{n \leqq l\} \simeq \text{Prob. } \mathbf{\Sigma}_{(q)} - (\hat{\sigma}^2/zl)\mathbf{\Lambda} \quad \text{being p.d.} \tag{6.6}$$

Writing τ_m for the minimum root of the equation in τ

$$|\mathbf{\Sigma}_{(q)} - \tau\mathbf{\Lambda}| = 0, \tag{6.7}$$

from (6.6) we have (see [9], Theorem 48, p. 151), for large $l \geqq n_0 + q$,

$$P\{n \leqq l\} \simeq P\{(\hat{\sigma}^2/zl) < \tau_m\}. \tag{6.8}$$

Now, as $(n_0 - p - 1)\ \hat{\sigma}^2/\sigma^2$ is distributed as $\chi^2_{n_0-p-1}$, from (6.1) and (6.8) we can state

$$\begin{aligned} &P\{n < n_0 + q\} = 0, \\ &P\{n \leqq l\} \simeq P\left\{\chi^2_{n_0-p-1} < zl(n_0 - p - 1)\frac{\tau_m}{\sigma^2}\right\}, \qquad \text{for } l \geqq n_0 + q, \end{aligned} \tag{6.9}$$

the approximation in (6.9) being close for large l.

In the particular case $q = 1$, an exact expression for the distribution function can be derived if we assume that the distribution of $\mathbf{x}$ is normal with dispersion matrix as before $\mathbf{\Sigma}$. Then, as in this case we take $\mathbf{\Lambda} = 1$, for $l \geqq n_0 + 1$, (6.2) can be written as

$$P\{n \leqq l\} = P\{1/S^{11\cdot l} > \hat{\sigma}^2/z\}. \tag{6.10}$$

Now, it is known (see e.g., [1] p. 85) that, under normality of $\mathbf{x}$, $\sigma^{11}/S^{11\cdot l}$ is distributed as a χ^2_{l-p}. Also, $(n_0 - p - 1)\hat{\sigma}^2/\sigma^2$ is distributed as a $\chi^2_{n_0-p-1}$ independently of $\mathbf{X}_\infty$, and hence, of $\sigma^{11}/S^{11\cdot l}$. From these facts we can express (6.10) as a double integral, and by reducing through repeated integration, we get

$$\begin{aligned} &P\{n < n_0 + 1\} = 0 \\ &P\{n \leqq l\} = 1 - I_{G^2}[\tfrac{1}{2}(l - p), \tfrac{1}{2}(n_0 - p - 1)] \quad \text{for } l \geqq n_0 + 1, \end{aligned} \tag{6.11}$$

where

$$G^2 = \sigma^2\sigma^{11}/z(n_0 - p - 1) + \sigma^2,$$

and I denotes the usual incomplete beta function ratio.

Close bounds, for the A.S.N. of the sequential procedure for inferring about β_1, can be obtained by comparing the distribution (6.11) with that of the sample-size n' of a sequential procedure for inferring about the bivariate regression coefficient, (the case $p = 1$) which has been developed in [6]. Referring to [6], it can be seen that the distribution of $n' + p - 1$ is same as the distribution of n given by (6.11), provided for the former we take the pilot sample-size and predictor variance as $n_0 - p + 1$ and $1/\sigma^{11}$ respectively. Therefore, we can obtain bounds for $E(n)$ from the bounds of $E(n')$ given in [6], by introducing these adjustments. (The bounds for $E(n')$ have been obtained by an indirect method and that is not extensible here). Thereby we get for $E(n)$,

$$\text{upper bound} = n_0 + 1 + (\sigma^2\sigma^{11}/z)I_{G^2}[\tfrac{1}{2}(n_0 - p), \tfrac{1}{2}(n_0 - p + 1)] - (n_0 - p)I_{G^2}[\tfrac{1}{2}(n_0 - p + 2), \tfrac{1}{2}(n_0 - p - 1)],$$

$$\text{lower bound} = \text{upper bound} - I_{G^2}[\tfrac{1}{2}(n_0 - p), \tfrac{1}{2}(n_0 - p - 1)].$$

As regards inference about α, for the first procedure of Section 5, n is the least integer subject to (5.3). Therefore, by Lemma 5.1, the distribution of n is given by $P\{n < n_0 + 1\} = 0$, and for $l \geqq n_0 + 1$,

$$P\{n \leqq l\} = \text{Prob. } \tilde{\mathbf{S}}_l - (\hat{\sigma}^2/z)\tilde{\mathbf{\Gamma}} \text{ being p.d.} \tag{6.12}$$

Reducing as shown in the same section, for $l \geqq n_0 + 1$,

$$P\{n \leqq l\} = P\{[1 + l\bar{\mathbf{x}}_l'\mathbf{S}_l^{-1}\bar{\mathbf{x}}_l]^{-1} > (\hat{\sigma}^2/zl)\}. \tag{6.13}$$

Now, let $\mathbf{m}$ and $\mathbf{\Sigma}$ denote the mean vector and dispersion matrix of $\mathbf{x}$. Then, in (6.13) we can replace $l\bar{\mathbf{x}}_l'\mathbf{S}_l^{-1}\bar{\mathbf{x}}_l$ by its probabilistic limit $\mathbf{m}'\mathbf{\Sigma}^{-1}\mathbf{m}$, as an approximation, and work out $P\{n \leqq l\}$, for $l \geqq n_0 + 1$, from the distribution of $\hat{\sigma}^2$.

Otherwise, if we assume normality of $\mathbf{x}$, we have $l(l-1)\bar{\mathbf{x}}_l'\mathbf{S}_l^{-1}\bar{\mathbf{x}}_l$ distributed as a non-central Hotelling's T^2, with d.f.p., $l - p$, and non-centrality parameter $\mu^2 = l\mathbf{m}'\mathbf{\Sigma}\mathbf{m}$, independently of $\hat{\sigma}^2$. Hence, after some reduction the probability (6.13) can be expressed as the sum of the convergent series

$$\{2^{\frac{1}{2}(n_0-p-1)}\Gamma[\tfrac{1}{2}(n_0 - p - 1)]\}^{-1}e^{-\frac{1}{2}\mu^2}\sum_{r=0}^{\infty}(1/r!)(\mu^2/2)^r[\beta(\tfrac{1}{2}(l-p), \tfrac{1}{2}p + r)]^{-1} \cdot \int_0^1 v^{\frac{1}{2}(l-p)}(1-v)^{r-\frac{1}{2}}\,dv\int_0^{v/d^2} e^{-\frac{1}{2}u}u^{\frac{1}{2}(n_0-p-3)}\,du,$$

where $d^2 = \sigma^2/z(n_0 - p - 1)l$.

7. Concluding remarks. So far in this paper, we have considered only sequential procedures suitable for inferring about (i) some or all of the regression coefficients $\beta_1, \beta_2, \cdots, \beta_p$, and (ii) the regression constant α. However, it should be possible to tackle any problem of inference, regarding a set of

$r\ (1 \leqq r \leqq p + 1)$ independent linear functions of $\alpha, \beta_1, \beta_2, \cdots, \beta_p$, by methods similar to those described above. The principle would be to generate r linear functions of the y-observations, such that their conditional distribution, given $\mathbf{X}_\infty$ and $\hat{\sigma}^2$, is multinormal, with the given parametric functions as means, and with $(\sigma^2/\hat{\sigma}^2)$-times a given p.d. matrix as dispersion matrix. The sample size should be so taken that this may be possible. In particular, many such problems of inference will be reducible, by a simple transformation of the predictors, to either of the two types (i) and (ii) mentioned above. There, working with the transformed predictors, we can directly follow the appropriate precedure as described in Sections 3–5. Such a case at hand is the problem of inference on the predicted value $Y' = \alpha + \beta_1 x_1' + \cdots + \beta_p x_p'$, given $x_1', \cdots x_p'$. For this, we can make the transformation $u_1 = x_1 - x_1', \cdots, u_p = x_p - x_p'$ in (1.3). Thereby, we get the joint density function of $u_1, \cdots, u_p, y$ as

$$(7.1) \quad g(u_1, \cdots, u_p)(2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-(1/2\sigma^2)(y - Y' - \beta_1 u_1 - \cdots - \beta_p u_p)^2\right],$$

where g denotes the joint density function of $u_1, \cdots, u_p$. Obviously, Y' occurs as the regression constant in (7.1), and therefore, to infer about Y', we can apply the procedure of Section 5, taking $u_1, \cdots, u_p$ as the predictors.

Acknowledgment. The author is deeply grateful to Mr. H. K. Nandi, Department of Statistics, Calcutta University, who suggested the problem and also offered help and guidance throughout the investigation. Also, thanks are due to the referee for the improved presentation of the paper.

APPENDIX

In this appendix we shall derive certain results regarding the existence of the coefficient matrix $\mathbf{C}$ of Section 3 and the coefficient vector $\bar{\mathbf{c}}$ of Section 5, subject to the respective sets of conditions.

We first prove the following lemma.

Lemma A.1. *If* $\mathbf{A}^{m\times r}$ *is a matrix such that* $\mathbf{A}'\mathbf{A}$ *is non-singular,* $\mathbf{d}^{r\times 1}$ *is a given vector, and* $f^2 > 0$ *is a given number, then a* $\mathbf{c}^{m\times 1}$, *subject to*

$$(\text{A.1}) \qquad \mathbf{c}'\mathbf{A} = \mathbf{d}',$$

$$(\text{A.2}) \qquad \mathbf{c}'\mathbf{c} = f^2,$$

can be determined, if and only if, either

$$(\text{A.3}) \qquad m = r, \qquad \begin{vmatrix} \mathbf{A}'\mathbf{A} & \mathbf{d} \\ \mathbf{d}' & f^2 \end{vmatrix} = 0,$$

or

$$(\text{A.4}) \qquad m \geqq r + 1, \qquad \begin{vmatrix} \mathbf{A}'\mathbf{A} & \mathbf{d} \\ \mathbf{d}' & f^2 \end{vmatrix} \geqq 0.$$

Proof. As $\mathbf{A}'\mathbf{A}$ is non-singular, $\mathbf{A}^{m\times r}$ must be of rank r. Therefore $m \geqq r$, and the set of r linear equations (A.1) (in the components of $\mathbf{c}$) are consistent.

Now, if $m = r$, the equations (A.1) have a unique solution for $\mathbf{c}$, and a vector satisfying both (A.1) and (A.2) would be determinable, if and only if this solution satisfies (A.2). This means

$$\text{(A.5)} \qquad \mathbf{d}' (\mathbf{A}'\mathbf{A})^{-1} \mathbf{d} = f^2;$$

(A.5) is equivalent to the second relation of (A.3).

If $m \geqq r + 1$, (A.1) will have an infinite number of solutions, and $\mathbf{c}'\mathbf{c}$, subject to (A.1), will assume all positive values above a minimum. Therefore, a $\mathbf{c}$ satisfying (A.1) and (A.2) would be determinable, if f^2 is not less than this minimum. By Lagrange's method of undetermined multipliers, it can be seen that the $\mathbf{c}$, minimizing $\mathbf{c}'\mathbf{c}$ subject to (A.1), must be of the form

$$\text{(A.6)} \qquad \mathbf{c} = \mathbf{A}\mathbf{k},$$

where $\mathbf{k}^{r \times 1}$ is determined by (A.1). Eliminating $\mathbf{k}$ from (A.1) and (A.6),

$$\text{Min}_{(\text{A}.1)}\, \mathbf{c}'\mathbf{c} = \mathbf{d}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{d}$$

Hence, for $m \geqq r + 1$, a $\mathbf{c}$ satisfying both (A.1) and (A.2) can be found, if and only if

$$\mathbf{d}'(\mathbf{A}'\mathbf{A})^{-1}\mathbf{d} \geqq f^2.$$

This is equivalent to the second relation of (A.4)

We now prove the following theorem regarding the determinability of the coefficient matrix $\mathbf{C}$ of Section 3.

THEOREM A.1. *For any given n, a matrix $\mathbf{C}^{(n-n_0+p+1)\times q}$ satisfying the relations* (3.14) *and* (3.15) *can be determined, if and only if the conditions* (3.9) *hold.*

PROOF. For convenience, let us write

$$\text{(A.7)} \qquad (z/\hat{\sigma}^2)\mathbf{A}^{-1} = \mathbf{W} = (w_{ij})_{i,j=1,\cdots,q}\,.$$

For all r, $1 < r < q$, we shall write

$$\text{(A.8)} \qquad (w_{ij})_{i,j=1,\cdots,r} = (w_{ij})_r\,.$$

Also, let $\mathbf{C}^{(n-n_0+p+1)\times q} = (\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_q)$. Then the conditions (3.14) and (3.15) can be rewritten as

$$\text{(A.9)} \quad \begin{array}{lll} \text{(A.9.1)} & \mathbf{c}_1'\mathbf{H} = (0, 1, 0, 0, \cdots, 0)^{1\times(p+1)}, & \mathbf{c}_1'\mathbf{c}_1 = w_{11}, \\ \text{(A.9.2)} & \mathbf{c}_2'(\mathbf{H}, \mathbf{c}_1) = (0, 0, 1, 0, \cdots 0, w_{21})^{1\times(p+2)}, & \mathbf{c}_2'\mathbf{c}_2 = w_{22}, \\ & \cdots \qquad \cdots & \cdots \\ \text{(A.9.p)} & \mathbf{c}_q'(\mathbf{H}, \mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_{q-1}) & \\ & \quad = (0, 0, 0, \cdots, 0, 1, w_{q1}, w_{q2}, \cdots, w_{q\cdot q-1})^{1\times(p+q)}, & \\ & & \mathbf{c}_q'\mathbf{c}_q = w_{qq}\,. \end{array}$$

We first consider the determination of a $\mathbf{c}_1^{(n-n_0+p+1)\times 1}$ subject to (A.9.1). As by (3.13) $\mathbf{H}'\mathbf{H}$ is non-singular, we can apply Lemma A.1. Thereby, because of (3.13), we get that a $\mathbf{c}_1$ satisfying (A.9.1) can be determined, if and only if, either

$$n = n_0, \begin{vmatrix} n & \sum_1^n \mathbf{x}_k' & 0 \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k \mathbf{x}_k' & \mathbf{I}^{p\times 1} \\ 0 & \mathbf{I}^{1\times p} & w_{11} \end{vmatrix} = 0, \tag{A.10}$$

or

$$n \geqq n_0 + 1, \begin{vmatrix} n & \sum_1^n \mathbf{x}_k' & 0 \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k \mathbf{x}_k' & \mathbf{I}^{p\times 1} \\ 0 & \mathbf{I}^{1\times p} & w_{11} \end{vmatrix} \geqq 0. \tag{A.11}$$

Now, as the distribution of $\mathbf{x}$ is continuous, for any n the determinant in (A.10) and (A.11) can be zero with probability zero, so that we can neglect the possibility (A.10) and the equality in the second relation of (A.11). Therefore, a $\mathbf{c}_1$ satisfying (A.9.1) can be determined, if and only if

$$n \geqq n_0 + 1, \begin{vmatrix} n & \sum_1^n \mathbf{x}_k' & 0 \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k \mathbf{x}_k' & \mathbf{I}^{p\times 1} \\ 0 & \mathbf{I}^{1\times p} & w_{11} \end{vmatrix} > 0. \tag{A.12}$$

As the $(p + 1) \times (p + 1)$ matrix $\tilde{\mathbf{S}}_n$ formed by the first $p + 1$ rows and columns of the determinant in (A.12) is p.d., (A.12) can be rewritten as

$$n \geqq n_0 + 1, \begin{pmatrix} n & \sum_1^n \mathbf{x}_k' & 0 \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k \mathbf{x}_k' & I^{p\times 1} \\ 0 & I^{1\times p} & w_{11} \end{pmatrix} \text{ is p.d.} \tag{A.13}$$

Now suppose, (A.13) holds, and a $\mathbf{c}_1$ subject to (A.9.1) has been determined. We next find what further conditions must hold, so that a $\mathbf{c}_2$ satisfying (A.9.2) (with respect to this $\mathbf{c}_1$) can be found. By (A.13),

$$\begin{pmatrix} \mathbf{H}' \\ \mathbf{c}_1' \end{pmatrix} (\mathbf{H} \quad \mathbf{c}_1) = \begin{pmatrix} n & \sum_1^n \mathbf{x}_k' & 0 \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k \mathbf{x}_k' & \mathbf{I}^{p\times 1} \\ 0 & \mathbf{I}^{1\times p} & w_{11} \end{pmatrix}$$

is non-singular. Therefore, by Lemma A.1, as before neglecting cases with probability zero, we find that a $\mathbf{c}_2$ subject to (A.9.2) can be found, if and only if

$$n \geqq n_0 + 2, \quad \begin{vmatrix} n & \sum_1^n \mathbf{x}_k' & \mathbf{o}^{1\times 2} \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k\mathbf{x}_k' & \mathbf{I}^{p\times 2} \\ \mathbf{o}^{2\times 1} & \mathbf{I}^{2\times p} & (w_{ij})_2 \end{vmatrix} > 0, \tag{A.14}$$

where $(w_{ij})_2$ is defined by (A.8).

Combining (A.13) and (A.14) we see that $\mathbf{c}_1$, $\mathbf{c}_2$ subject to (A.9.1) and (A.9.2) can be found, if and only if

$$n \geqq n_0 + 2, \quad \begin{pmatrix} n & \sum_1^n \mathbf{x}_k' & \mathbf{o}^{1\times 2} \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k\mathbf{x}_k' & \mathbf{I}^{p\times 2} \\ \mathbf{o}^{2\times 1} & \mathbf{I}^{2\times p} & (w_{ij})_2 \end{pmatrix} \text{ is p.d.} \tag{A.15}$$

Proceeding step by step exactly in the same manner, it follows that $\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_q$ subject to (A.9) can be determined, if and only if

$$n \geqq n_0 + q, \quad \begin{pmatrix} n & \sum_1^n \mathbf{x}_k' & \mathbf{o}^{1\times q} \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k\mathbf{x}_k' & \mathbf{I}^{p\times q} \\ \mathbf{o}^{q\times 1} & \mathbf{I}^{q\times p} & \mathbf{W} \end{pmatrix} \text{ is p.d.} \tag{A.16}$$

Now, for the non-singular matrix

$$\mathbf{D} = \begin{pmatrix} 1 & \mathbf{o}^{1\times p} & \mathbf{o}^{1\times q} \\ -\bar{\mathbf{x}}_n & \mathbf{I}^{p\times p} & \mathbf{O}^{p\times q} \\ \mathbf{o}^{q\times 1} & \mathbf{O}^{q\times p} & \mathbf{I}^{q\times q} \end{pmatrix},$$

we have by (2.1),

$$\mathbf{D} \begin{pmatrix} n & \sum_1^n \mathbf{x}_k' & \mathbf{o}^{1\times q} \\ \sum_1^n \mathbf{x}_k & \sum_1^n \mathbf{x}_k\mathbf{x}_k' & \mathbf{I}^{p\times q} \\ \mathbf{o}^{q\times 1} & \mathbf{I}^{q\times p} & \mathbf{W} \end{pmatrix} \mathbf{D}' = \begin{pmatrix} n & \mathbf{o}^{1\times p} & \mathbf{o}^{1\times q} \\ \mathbf{o}^{p\times 1} & \mathbf{S}_n & \mathbf{I}^{p\times q} \\ \mathbf{o}^{q\times 1} & \mathbf{I}^{q\times p} & \mathbf{W} \end{pmatrix}. \tag{A.17}$$

Therefore, the positive-definiteness of the matrix in (A.16) is equivalent to the positive-definiteness of the matrix on the right-hand side of (A.17), or similarly, of the $(p + q) \times (p + q)$ matrix

$$\begin{pmatrix} \mathbf{S}_n & \mathbf{I}^{p\times q} \\ \mathbf{I}^{q\times p} & \mathbf{W} \end{pmatrix}. \tag{A.18}$$

Now, by Lemma 2.1, (A.7), and (3.8), by pre- and post-multiplying the matrix (A.18) by a non-singular matrix and its transpose respectively, we get

$$\begin{pmatrix} \mathbf{S}_n - \mathbf{I}^{p\times q}\mathbf{W}^{-1}\mathbf{I}^{q\times p} & \mathbf{o}^{p\times q} \\ \mathbf{o}^{q\times p} & \mathbf{W} \end{pmatrix} = \begin{pmatrix} \mathbf{S}_n - \frac{\hat{\sigma}^2}{z}\mathbf{\Gamma} & \mathbf{o} \\ \mathbf{o} & \mathbf{W} \end{pmatrix}. \tag{A.19}$$

Therefore the positive-definiteness of (A.18) and (A.19) is equivalent. Further, as by definition, $\mathbf{W}$ is p.d., (A.19) is p.d., if and only if $\mathbf{S}_n - z^{-1}\hat{\sigma}^2\mathbf{\Gamma}$ is true. Thus, the conditions (A.16) reduce to the conditions (3.9), so that the theorem is proved.

We next prove the following theorm regarding the determinability of the coefficient vector $\tilde{\mathbf{c}}$ of Section 5.

THEOREM A.2. *For a given* n, *a vector* $\tilde{\mathbf{c}}^{(n-n_0+p+1)\times 1}$ *satisfying the relations* (5.4) *and* (5.5) *can be determined, if and only if the conditions* (5.3) *hold.*

PROOF. As $\mathbf{H}'\mathbf{H}$ given by (3.13) is non-singular, applying Lemma A.1 and neglecting cases with probability zero as in the proof of the preceding theorem, we see that $\tilde{\mathbf{c}}$ subject to (5.4) and (5.5) can be determined, if and only if

$$n \geqq n_0 + 1, \begin{vmatrix} \tilde{\mathbf{S}}_n & \mathbf{I}^{(p+1)\times 1} \\ \mathbf{I}^{1\times(p+1)} & (z/\hat{\sigma}^2) \end{vmatrix} > 0. \tag{A.20}$$

As $\tilde{\mathbf{S}}_n$ is p.d., we can write (A.20) equivalently as

$$n \geqq n_0 + 1, \begin{pmatrix} \tilde{\mathbf{S}}_n & \mathbf{I}^{(p+1)\times 1} \\ \mathbf{I}^{1\times(p+1)} & \frac{z}{\hat{\sigma}^2} \end{pmatrix} \text{ is p.d.} \tag{A.21}$$

Now applying Lemma 2.1, (A.21) and (5.3) are seen to be equivalent, so that the theorem is proved.

REFERENCES

[1] ANDERSON, T. W. (1958). *Introduction to Multivariate Statistical Analysis*. Wiley, New York.

[2] CHATTERJEE, S. K. (1959). On an extension of Stein's two-sample procedure to the multinormal problem. *Calcutta Statist. Assn. Bull.* **8** 121–148.

[3] CHATTERJEE, S. K. (1959). Some further results on the multinormal extension of Stein's two sample procedure. *Calcutta Statist. Assn. Bull.* **9** 20–28.

[4] CHATTERJEE, S. K. (1960). Bounded-risk estimates for the multinormal mean vector. *Proc. 47th Session Indian Science Congress*, Pt. III (Abstracts). p. 18

[5] CHATTERJEE, S. K. (1960). Some sequential inference procedures with "nuisance-param-

eter-free" performance in multivariate problems. Thesis (unpublished) accepted by the Calcutta University.

[6] CHATTERJEE, S. K. (1960). Sequential tests for the bivariate regression parameters with known power and related estimation procedures, *Calcutta Statist. Assn. Bull.* **10** 19–34.

[7] CRAMÉR, H. (1946). *Mathematical Methods of Statistics*, Princeton Univ. Press.

[8] DANTZIG, G. B. (1940). On the non-existence of tests of "Student's" hypothesis having power functions independent of σ. *Ann. Math. Statist.* **11** 186–192.

[9] FERRAR, W. L. (1957). *Algebra*. Oxford Univ. Press.

[10] LEHMANN, E. L. (1950). *Theory of Estimation*, notes recorded by C. Blyth, Associated Students' Store, Univ. of California.

[11] ROBBINS, H. (1959). Sequential estimation of the mean of a normal population, 235–245. *Probability and Statistics, The Harald Cramér Volume*. Wiley, New York.

[12] RUBEN, H. (1961). Studentization of two-stage sample means from normal populations with unknown common variance. *Sankhyā* **23** 231–250.

[13] STEIN, C. (1945). Two-sample test of a linear hypothesis whose power is independent of the variance. *Ann. Math. Statist.* **16** 243–258.