The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1725–1734

LINEAR SPACES AND UNBIASED ESTIMATION[1]

BY JUSTUS SEELY[2]

Iowa State University

1. Introduction. In this paper some general results are obtained on unbiased estimation when the choice of estimators is restricted to a finite-dimensional linear space $\bar{\mathscr{A}}$. The results concern mainly necessary and sufficient conditions for existence of unbiased estimators within $\bar{\mathscr{A}}$ and procedures for obtaining such estimators when they exist. At the outset the approach is from a coordinate-free (Kruskal [2]) viewpoint; but then it is found useful, and for many situations natural, to use a fixed reference set of spanning elements for particular subspaces. Identical analogues with estimation procedures in what is commonly referred to as linear model theory will be seen to exist. Much of the formulation has been motivated by problems especially relevant in the study of a general mixed linear model $Y = X\beta + e$ where the random vector Y has expectation $X\beta$ and covariance matrix

$$E[ee'] = \sum_{i=1}^{m} \nu_i V_i.$$

Concerning this model attention is given in a second paper to quadratic estimation, i.e., $\bar{\mathscr{A}} = \{Y'AY: A \text{ real and symmetric}\}$, of parametric functions of the form $\sum_{i \leq j} \lambda_{ij}\beta_i\beta_j + \sum_k \lambda_k \nu_k$ with special emphasis on parametric functions of the form $\sum_k \lambda_k \nu_k$.

2. Preliminary framework. Terminology and elementary properties associated with linear spaces (vector spaces) are utilized in the sequel. Much of the notation and terminology is similar to that used in Chapter II of Wilansky [3] and in Halmos [1]. Since only finite-dimensional real linear spaces are considered, we adopt the convention that whenever a linear space is referred to it is assumed to be a finite-dimensional real linear space. In addition to vector space notions, the usual notations and ideas of elementary set theory are also employed.

Concerning vector space notions, we mention a few at this point. If $\mathscr{A}$ and $\mathscr{B}$ are non-empty subsets of a vector space $\mathscr{L}$, then $\mathscr{A} + \mathscr{B}$ denotes the set $\{a+b: a \in \mathscr{A}, b \in \mathscr{B}\}$. When the set $\mathscr{A}$ consists of a single element a, we abbreviate $\mathscr{A} + \mathscr{B}$ by $a + \mathscr{B}$. The linear span of a non-empty subset $\mathscr{A}$ of $\mathscr{L}$ is denoted by $\operatorname{sp} \mathscr{A}$. If $\mathscr{A}$ and $\mathscr{B}$ are disjoint subspaces of $\mathscr{L}$, i.e., their intersection is the null vector, then we denote the sum of $\mathscr{A}$ and $\mathscr{B}$ by $\mathscr{A} \oplus \mathscr{B}$. The null vector is denoted by 0, and even though the symbol 0 is used for other purposes, it should be clear from the context what 0 represents.

Received January 26, 1970.

[1] This work was prepared partly in connection with work on Contract AF33(615) 68C1343, Office of Aerospace Research, United States Air Force, Wright-Patterson Air Force Base, Ohio and an NSF Traineeship.

[2] Now at Oregon State University.

We assume that $\mathscr{P} = \{P\}$ is a family of probability measures with an associated measurable space $(\mathscr{U}, S)$ and that $\bar{\mathscr{A}}$ is a linear space of real-valued random variables such that the expectation of each random variable in $\bar{\mathscr{A}}$ exists with respect to all $P \in \mathscr{P}$. The form of $P \in \mathscr{P}$, or even the complete structure of $\mathscr{P}$, need not be known completely; however, we do assume that Ω is some given parameter space and that E is a function on $\bar{\mathscr{A}} \times \Omega$, called the expectation function, from which the expectation of an element in $\bar{\mathscr{A}}$ may be characterized over the class $\mathscr{P}$. Formally we assume that to each $P \in \mathscr{P}$ there is a $\theta \in \Omega$ and to each $\theta \in \Omega$ there is a $P \in \mathscr{P}$ such that the following expression is true:

$$\bar{a} \in \bar{\mathscr{A}} \Rightarrow E_P[\bar{a}] = E[\bar{a} \mid \theta], \tag{2.1}$$

where E_P denotes expectation with respect to the measure P. Note that (2.1) insures for a fixed θ that the function $E[\cdot \mid \theta]$ has all the usual properties associated with expectations. In essence: $\bar{\mathscr{A}}$ is a linear space (finite-dimensional) of random variables from $\mathscr{U}$ into the real line R^1; Ω is some set of parameters which may be given either explicitly or implicitly; interest is in obtaining unbiased estimators from $\bar{\mathscr{A}}$ for functions from Ω into R^1; and the possible expectations of any random variable in $\bar{\mathscr{A}}$ may be described via the parameter space Ω and the function E.

For the linear space of random variables $\bar{\mathscr{A}}$ and a parametric function g, i.e., a function from Ω into R^1, let $\bar{\mathscr{A}}_g$ denote the collection of unbiased estimators for g that are in $\bar{\mathscr{A}}$ and let $\bar{\mathscr{A}}_0$ denote $\bar{\mathscr{A}}_g$ when g is the identically zero function. Note that $\bar{\mathscr{A}}_0$ is a linear space and that $\bar{\mathscr{A}}_g$ is an affine set provided it is non-empty; in fact, $\bar{t} \in \bar{\mathscr{A}}_g$ implies that $\bar{\mathscr{A}}_g = \bar{t} + \bar{\mathscr{A}}_0$.

Definition 1. A parametric function g is said to be $\bar{\mathscr{A}}$-estimable if and only if $\bar{\mathscr{A}}_g$ is non-empty.

Definition 1 is essentially a standard one. For instance, in a fixed linear model $Y = X\beta + e$ with $\bar{\mathscr{A}} = \{a'Y : a \in R^n\}$ the term $\bar{\mathscr{A}}$-estimable corresponds to linearly estimable or if $\bar{\mathscr{A}} = \{Y'AY : A = A'\}$ the term $\bar{\mathscr{A}}$-estimable corresponds to quadratically estimable.

For the space $\bar{\mathscr{A}}$ it is convenient to let the random variables in $\bar{\mathscr{A}}$ have an explicit inner product representation. Without loss of generality let $\bar{\mathscr{A}}$ be specified in the form $\bar{\mathscr{A}} = \{(a, Y) : a \in \mathscr{A}\}$ where $(\mathscr{A}, (\cdot, \cdot))$ is a finite-dimensional inner product space and Y is a random variable from $\mathscr{U}$ into $\mathscr{A}$. This method is similar (we do not insist that Y range over the entire space $\mathscr{A}$) to that employed by Kruskal [2]. Without loss of generality it could be assumed that $\operatorname{sp}\{Y(u) : u \in \mathscr{U}\} = \mathscr{A}$; but this restriction would exclude several useful representations sometimes employed in the usual linear model. By not using the restriction $\operatorname{sp}\{Y(u) : u \in \mathscr{U}\} = \mathscr{A}$, the representation (a, Y) is not necessarily unique, i.e., $(a, Y) = (b, Y)$ does not imply that $a = b$. For a parametric function g let $\mathscr{A}_g$ denote the set of elements $a \in \mathscr{A}$ such that (a, Y) is an unbiased estimator for g and for $\mathscr{B} \subset \mathscr{A}$ let $\bar{\mathscr{B}}$ denote the collection of random variables $\{(b, Y) : b \in \mathscr{B}\}$. Note that the usage of $\bar{\mathscr{A}}_g$ is consistent with previous usage, i.e.,

$$\mathscr{A}_g = \{a : (a, Y) \in \bar{\mathscr{A}}_g\} \quad \text{and} \quad \bar{\mathscr{A}}_g = \{(a, Y) : a \in \mathscr{A}_g\},$$

where $\bar{\mathscr{A}}_g$ in the first expression denotes the collection of all unbiased estimators for g that are members of $\bar{\mathscr{A}}$.

DEFINITION 2. For each $\theta \in \Omega$ let μ_θ be the unique element in $\mathscr{A}$ such that $E[(a, Y) \mid \theta] = (a, \mu_\theta)$ for all $a \in \mathscr{A}$ and let $\mathscr{E} = \text{sp}\{\mu_\theta : \theta \in \Omega\}$.

In Definition 2 the quantity $E[(a, Y) \mid \theta]$, considered as a function on $\mathscr{A}$ with θ fixed, is a linear functional on $\mathscr{A}$ and so the existence and uniqueness of μ_θ is assured. Also, note that μ_θ may be regarded as the expectation of Y in the sense that $E[(a, Y) \mid \theta] = (a, E[Y \mid \theta]) = (a, \mu_\theta)$.

Let $(\mathscr{L}, \langle \cdot, \cdot \rangle)$ be a finite-dimensional inner product space and let $\mathbf{T}$ be a linear operator from $\mathscr{L}$ into $\mathscr{A}$. The range of $\mathbf{T}$ is denoted by $\underline{R}(\mathbf{T})$, the null space by $\underline{N}(\mathbf{T})$, and the adjoint by $\mathbf{T}^*$. By the adjoint of $\mathbf{T}$ we mean the linear operator $\mathbf{T}^*$ from $\mathscr{A}$ into $\mathscr{L}$ which takes an element $a \in \mathscr{A}$ into the unique element $\mathbf{T}^*a$ in $\mathscr{L}$ that satisfies the condition $\langle \mathbf{T}^*a, \rho \rangle = (a, \mathbf{T}\rho)$ for all $\rho \in \mathscr{L}$. We denote the rank of $\mathbf{T}$ by $\underline{r}(\mathbf{T})$, the nullity by $\underline{n}(\mathbf{T})$, and the orthogonal complement of a non-empty subset $\mathscr{B}$ of $\mathscr{L}$ by $\mathscr{B}^\perp$. In the event that several inner products are considered simultaneously on the same linear space, we shall take care to explain exactly what the symbol $\mathscr{B}^\perp$ denotes. For matrices the same notation is used as that just described for linear operators, except that a prime is used to denote the transpose of a matrix or vector.

To facilitate computational procedures and to represent certain parametric functions explicitly, it is convenient to have another representation for μ_θ. Suppose that $(\mathscr{H}, \langle \cdot, \cdot \rangle)$ is an inner product space and that $\mathbf{H}$ is a linear operator from $\mathscr{H}$ into $\mathscr{A}$ such that $\mathscr{E} \subset \underline{R}(\mathbf{H})$. For each $\theta \in \Omega$ let $\xi_\theta \in \mathscr{H}$ be such that $\mu_\theta = \mathbf{H}\xi_\theta$ and let $\Omega_H = \{\xi_\theta : \theta \in \Omega\}$. Note that in general ξ_θ is not unique; however, for any particular problem we assume that one set of elements $\{\xi_\theta\}$ has been specified and remains fixed throughout. Representing μ_θ in the fashion just described will frequently be both useful and natural in the sequel; thus, we adopt the convention that whenever reference is made to a $\mu_\theta = \mathbf{H}\xi_\theta$ representation without referring to a particular circumstance, it will be understood that $(\mathscr{H}, \langle \cdot, \cdot \rangle)$, $\mathbf{H}$, and a set of elements $\{\xi_\theta\}$ are defined as in the beginning of this paragraph.

3. $\bar{\mathscr{A}}$-estimability—general results. Let $\mathscr{G}$ denote the entire collection of $\bar{\mathscr{A}}$-estimable functions. From Definition 1 it is clear that a parametric function g is $\bar{\mathscr{A}}$-estimable if and only if there exists an $a \in \mathscr{A}$ such that

$$\theta \in \Omega \Rightarrow E[(a, Y) \mid \theta] = (a, \mu_\theta) = g(\theta);$$

i.e., if and only if there exists an $F \in \mathscr{A}^\#$ ($\mathscr{A}^\#$ is the collection of linear functionals on $\mathscr{A}$) such that $F(\mu_\theta) = g(\theta)$ for all θ. It follows that

$$\mathscr{G} = \{g : g(\theta) = (a, \mu_\theta), a \in \mathscr{A}\}.$$

Since some elements $a \in \mathscr{A}$ may describe the same parametric function g it is possible that $\mathscr{G}$ may be characterized through a subset of $\mathscr{A}$. The following theorem gives the conditions on such a subset.

THEOREM 1. *A necessary and sufficient condition for the collection* $\mathscr{G}$ *to be equal to* $\{(b, \mu_\theta): b \in \mathscr{B}\}$ *is that* $\mathscr{B} + \mathscr{E}^\perp = \mathscr{A}$.

PROOF. Suppose that $\mathscr{G}$ is as stated and that $a \in \mathscr{A}$, then there must exist $b \in \mathscr{B}$ such that $(a, \mu_\theta) = (b, \mu_\theta)$ for all $\theta \in \Omega$. Therefore, $a - b \in \mathscr{E}^\perp$ which implies that $a = b + f$ for some $f \in \mathscr{E}^\perp$. Conversely suppose that $\mathscr{B} + \mathscr{E}^\perp = \mathscr{A}$ and that $a \in \mathscr{A}$. Let $a = b' + f$ where $b' \in \mathscr{B}$ and $f \in \mathscr{E}^\perp$, then $(a, \mu_\theta) = (b', \mu_\theta)$ for all θ and thus the desired result follows.

Note that $\mathscr{G}$ is a vector space and that Theorem 1 implies $\dim \mathscr{E} = \dim \mathscr{G}$ ($\dim \mathscr{C}$ means the dimension of the subspace $\mathscr{C}$). Also, several other conclusions can be drawn from Theorem 1. For example, the subspace $\mathscr{A}_0$, i.e., the subspace corresponding to the collection of zero estimators $\bar{\mathscr{A}}_0$, is $\mathscr{E}^\perp$. Further if g is $\bar{\mathscr{A}}$-estimable and $a \in \mathscr{A}_g$, then $\mathscr{A}_g = a + \mathscr{A}_0$.

EXAMPLE 1. Suppose that Y is an $n \times 1$ random vector with expectation $X\beta$ where X is a given $n \times p$ matrix and it is known that the parameter vector β satisfies the equations $\Delta\beta = 0$. Let $\mathscr{A} = R^n$ and let $(\cdot, \cdot)$ denote the usual inner product on R^n so that $\bar{\mathscr{A}} = \{a'Y: a \in R^n\}$, i.e., $\bar{\mathscr{A}}$ denotes the class of linear estimators. Since our concern is only with the expectation of random variables of the form $a'Y$, it is clear that we may take Ω as $\underline{N}(\Delta)$. It follows immediately that

(3.1)
(a) $\beta \in \Omega \Rightarrow \mu_\beta = X\beta$.
(b) $\mathscr{E} = \{X\beta: \Delta\beta = 0\}$.
(c) $\mathscr{G} = \{a'X\beta: a \in R^n\}$.
(d) $\dim \mathscr{E} = \underline{r}(X) - \dim[\underline{R}(X') \cap \underline{R}(\Delta')]$.

The number of linearly independent estimable (i.e., $\bar{\mathscr{A}}$-estimable) functions is given by (3.1.d). In linear model terminology the expression in (3.1.d) says that the dimension of $\mathscr{E}$ is equal to the rank of X minus the number of linearly independent estimable restrictions. Suppose that A is a $p \times q$ matrix such that $\underline{N}(\Delta) \subset \underline{R}(A)$, then $\mathscr{E} \subset \underline{R}(XA)$ so that Theorem 1 implies

$$\mathscr{G} = \{(XA\rho, X\beta): \rho \in R^q\}.$$

Of the possible choices for the matrix A, the special cases $A = I$ and $\underline{R}(A) = \underline{N}(\Delta)$ should be noted. In the latter case observe that $A = I - \Delta^- \Delta$ will suffice for any g-inverse Δ^- of Δ and that $\underline{R}(A) = \underline{N}(\Delta)$ implies $\mathscr{E} = \underline{R}(XA)$ and $\mathscr{E}^\perp = \mathscr{A}_0 = \underline{N}(A'X')$.

For the remainder of this section suppose that μ_θ is described by a $\mu_\theta = \mathbf{H}\xi_\theta$ representation and that $\mathbf{W}$ is a linear operator from a finite-dimensional inner product space $(\mathscr{W}, (\cdot, \cdot)^*)$ into $(\mathscr{A}, (\cdot, \cdot))$. When μ_θ is described via a $\mu_\theta = \mathbf{H}\xi_\theta$ representation, the parametric functions usually of interest are of the form $\langle \lambda, \xi_\theta \rangle$. Suppose that $\mathbf{W}$ is such that $\underline{R}(\mathbf{W}) + \mathscr{A}_0 = \mathscr{A}$, then from Theorem 1 it follows that a parametric function $\langle \lambda, \xi_\theta \rangle$ is in $\mathscr{G}$ if and only if there exists a $\rho \in \mathscr{W}$ such that

$$\theta \in \Omega \Rightarrow \langle \lambda, \xi_\theta \rangle = (\mathbf{W}\rho, \mu_\theta) = \langle \mathbf{H}^*\mathbf{W}\rho, \xi_\theta \rangle;$$

that is, if and only if $\lambda - \mathbf{H}^*\mathbf{W}\rho \in \Omega_H{}^\perp$ for some $\rho \in \mathscr{W}$. We state this last condition in the following theorem.

THEOREM 2. *Suppose μ_θ is described by a $\mu_\theta = \mathbf{H}\xi_\theta$ representation and that $\mathbf{W}$ is a linear operator such that $\underline{R}(\mathbf{W}) + \mathscr{A}_0 = \mathscr{A}$, then a parametric function $\langle \lambda, \xi_\theta \rangle$ is $\bar{\mathscr{A}}$-estimable if and only if $\lambda \in \underline{R}(\mathbf{H}^*\mathbf{W}) + \Omega_H{}^\perp$.*

COROLLARY 2.1. *Under the conditions in Theorem 2 a parametric function $\langle \lambda, \xi_\theta \rangle$ is $\bar{\mathscr{A}}$-estimable if and only if there exists a $\lambda_1 \in \underline{R}(\mathbf{H}^*\mathbf{W})$ such that $\langle \lambda_1, \xi_\theta \rangle = \langle \lambda, \xi_\theta \rangle$ for all $\theta \in \Omega$.*

From Corollary 2.1 it is clear that $\underline{R}(\mathbf{H}^*)$ (i.e., $\mathbf{W} = \mathbf{I}$) contains all the information regarding $\bar{\mathscr{A}}$-estimability in the sense that $g \in \mathscr{G}$ if and only if $g(\theta) = \langle \lambda, \xi_\theta \rangle$ for some $\lambda \in \underline{R}(\mathbf{H}^*)$. Corollary 2.1 also implies it is not necessarily true that if $\langle \lambda, \xi_\theta \rangle$ is $\bar{\mathscr{A}}$-estimable then $\lambda \in \underline{R}(\mathbf{H}^*)$.

COROLLARY 2.2. *If $\mathbf{W}$ is a linear operator into $\mathscr{A}$ and ρ and λ are such that $\mathbf{H}^*\mathbf{W}\rho = \lambda$, then the random variable $(\mathbf{W}\rho, Y)$ is an unbiased estimator for the parametric function $\langle \lambda, \xi_\theta \rangle$.*

COROLLARY 2.3. *If $\mathbf{W}$ is a linear operator such that $\underline{R}(\mathbf{W}) + \mathscr{A}_0 = \mathscr{A}$ and if $\Omega_H{}^\perp \subset \underline{R}(\mathbf{H}^*\mathbf{W})$, then a parametric function $\langle \lambda, \xi_\theta \rangle$ is $\bar{\mathscr{A}}$-estimable if and only if there exists a ρ such that $\mathbf{H}^*\mathbf{W}\rho = \lambda$.*

COROLLARY 2.4. *A sufficient condition for $\mathbf{W}$ to satisfy the relationship $\underline{R}(\mathbf{W}) + \mathscr{A}_0 = \mathscr{A}$ is that $\underline{R}(\mathbf{H}^*\mathbf{W}) = \underline{R}(\mathbf{H}^*)$. Moreover, this condition, i.e., $\underline{R}(\mathbf{H}^*\mathbf{W}) = \underline{R}(\mathbf{H}^*)$, is both necessary and sufficient when $\underline{R}(\mathbf{H}) = \mathscr{E}$.*

EXAMPLE 2. Consider the same situation and notation as in Example 1. Let $\mathscr{H} = R^p$ with the usual inner product, let $\mathbf{H}\delta = X\delta$ for all $\delta \in R^p$, and let $\xi_\beta = \beta$ for all $\beta \in \Omega$. It follows that $X\beta$ is a $\mu_\beta = \mathbf{H}\xi_\beta$ representation and that $\Omega_H = \underline{N}(\Delta)$. Let A be any $p \times q$ matrix such that $\underline{N}(\Delta) \subset \underline{R}(A)$, then with $\mathbf{W} = XA$ Theorem 2 and Corollaries 2.1 and 2.2 imply the following statements:

(3.2) (a) $\lambda'\beta$ is estimable if and only if $\lambda \in \underline{R}(X'XA) + \underline{R}(\Delta')$.

(b) $\lambda'\beta$ is estimable if and only if there exists a $\lambda_1 \in \underline{R}(X'XA)$ such that $\lambda'\beta = \lambda_1'\beta$ for all $\beta \in \Omega$.

(c) If ρ and λ are such that $X'XA\rho = \lambda$, then $(XA\rho, Y)$ is an unbiased estimator for $\lambda'\beta$.

If Y has a covariance structure of the form $\sigma^2 I$ and A is selected such that $\underline{R}(A) = \underline{N}(\Delta)$, then it may be noted that the estimator in (3.2.c) is in fact the best linear unbiased estimator for the parametric function $\lambda'\beta$.

Noting the similarities thus far with estimability in linear model theory, e.g., Examples 1 and 2, we now consider finding a random variable $\hat{\xi}$ from $\mathscr{U}$ into $\mathscr{H}$ such that $\langle \lambda, \hat{\xi} \rangle$ is an unbiased estimator for an $\bar{\mathscr{A}}$-estimable $\langle \lambda, \xi_\theta \rangle$. One possible way to proceed is via Corollary 2.2. If $\lambda = \mathbf{H}^*\mathbf{W}\rho$, then $(\mathbf{W}\rho, Y)$ is an unbiased estimator for $\langle \lambda, \xi_\theta \rangle$. Thus, if $\lambda = \mathbf{H}^*\mathbf{W}\rho$ and $\hat{\xi}$ is such that $\mathbf{W}^*Y = \mathbf{W}^*\mathbf{H}\hat{\xi}$, then

$$\langle \lambda, \hat{\xi} \rangle = (\rho, \mathbf{W}^*\mathbf{H}\hat{\xi})^* = (\rho, \mathbf{W}^*Y)^* = (\mathbf{W}\rho, Y).$$

A necessary and sufficient condition for the existence of a random variable $\hat{\xi}$ from $\mathscr{U}$ into $\mathscr{H}$ such that $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^* Y$ is

$$\text{(3.3)} \qquad \operatorname{sp}\{\mathbf{W}^* Y(u) : u \in \mathscr{U}\} \subset \underline{R}(\mathbf{W}^* \mathbf{H}).$$

If the condition $\operatorname{sp}\{Y(u): u \in \mathscr{U}\} = \mathscr{A}$ holds, then this last condition is equivalent to $\underline{R}(\mathbf{W}^*) = \underline{R}(\mathbf{W}^*\mathbf{H})$. We summarize in the following theorem and corollary.

THEOREM 3. *Let* $\mathbf{W}$ *be a linear operator such that* (3.3) *is satisfied*, e.g., $\underline{R}(\mathbf{W}^*) = \underline{R}(\mathbf{W}^*\mathbf{H})$. *To any random variable* $\hat{\xi}$ *such that* $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^* Y$ *and to any* $\lambda \in \underline{R}(\mathbf{H}^*\mathbf{W})$ *the random variable* $\langle \lambda, \hat{\xi} \rangle$ *is an unbiased estimator for* $\langle \lambda, \xi_\theta \rangle$.

COROLLARY 3.1. *Suppose that* $\mathbf{W}$ *is a linear operator such that*

(3.4)
(a) *condition* (3.3) *is satisfied,*
(b) $\underline{R}(\mathbf{W}) + \mathscr{A}_0 = \mathscr{A}$, *and*
(c) $\Omega_H^{\perp} \subset \underline{R}(\mathbf{H}^*\mathbf{W})$;

then $\langle \lambda, \hat{\xi} \rangle$ *is an unbiased estimator for each* $\mathscr{A}$*-estimable* $\langle \lambda, \xi_\theta \rangle$ *whenever* $\hat{\xi}$ *is such that* $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^* Y$.

EXAMPLE 3. Consider the same situation and notation as in Example 1 and Example 2. Let $\mathbf{W} = XA$ where A is any matrix such that $\underline{N}(\Delta) \subset \underline{R}(A)$. Since $\underline{r}(A'X') = \underline{r}(A'X'X)$ it follows that (3.3) is satisfied. Thus, if $\hat{\beta}$ is such that $A'X'X\hat{\beta} = A'X'Y$ and $\lambda \in \underline{R}(X'XA)$, then Theorem 3 implies that $\lambda'\hat{\beta}$ is an unbiased estimator for $\lambda'\beta$. As noted in Example 2, if the covariance structure of Y is of the form $\sigma^2 I$ and A is such that $\underline{R}(A) = \underline{N}(\Delta)$, then $\lambda'\hat{\beta}$ is actually the best linear unbiased estimator for $\lambda'\beta$.

If in Corollary 2.2 there exist ρ_1 and ρ_2 such that $\mathbf{H}^*\mathbf{W}\rho_1 = \mathbf{H}^*\mathbf{W}\rho_2 = \lambda$, then $(\mathbf{W}\rho_1, Y)$ and $(\mathbf{W}\rho_2, Y)$ are both unbiased estimators for $\langle \lambda, \xi_\theta \rangle$. Although both are unbiased estimators for the same parametric function, it is not necessarily true that $(\mathbf{W}\rho_1, Y) = (\mathbf{W}\rho_2, Y)$, i.e., for a given $\mathbf{W}$ in Corollary 2.2 unbiased estimators are not in general unique. However, if $\underline{R}(\mathbf{W}) \cap \underline{N}(\mathbf{H}^*) = \{0\}$, then $\mathbf{H}^*\mathbf{W}\rho_1 = \mathbf{H}^*\mathbf{W}\rho_2$ implies that $\mathbf{W}(\rho_1 - \rho_2) \in \underline{N}(\mathbf{H}^*)$ so that $\mathbf{W}\rho_1 = \mathbf{W}\rho_2$. Thus, for a given $\mathbf{W}$ such that $\underline{R}(\mathbf{W}) \cap \underline{N}(\mathbf{H}^*) = \{0\}$, an unbiased estimator in Corollary 2.2 is unique. Note that this uniqueness is dependent upon $\mathbf{W}$; that is, different $\mathbf{W}$'s do not in general yield the same estimators. As a final point note that $\underline{R}(\mathbf{W}) \cap \underline{N}(\mathbf{H}^*) = \{0\}$ is equivalent to having $\underline{R}(\mathbf{W}^*\mathbf{H}) = \underline{R}(\mathbf{W}^*)$ which implies Condition 3.3. Thus, if $\mathbf{W}$ is selected in Theorem 3 or Corollary 3.1 such that $\underline{R}(\mathbf{W}^*\mathbf{H}) = \underline{R}(\mathbf{W}^*)$, then Condition (3.3) is satisfied and the estimator $\langle \lambda, \hat{\xi} \rangle$ is unique.

4. $\mathscr{A}$-estimability—main results. We consider in this section $\mathscr{A}$-estimability for parametric functions of the form $\sum_{i=1}^{M} \lambda_i \xi_i(\theta)$ when μ_θ has the representation

$$\text{(4.1)} \qquad \theta \in \Omega \Rightarrow \mu_\theta = \textstyle\sum_{i=1}^{M} \xi_i(\theta) b_i.$$

In this representation $\mathscr{B} = \{b_1, b_2, \cdots, b_M\}$ is a known set of elements in $\mathscr{A}$ and $\xi_1, \xi_2, \cdots, \xi_M$ are parametric functions. A characterization as given in (4.1) may be obtained from any finite set of elements which contains a spanning set for $\mathscr{E}$;

however, in this section we assume the collection $\mathscr{B}$ is such that for any set of real numbers $\{\alpha_i\}$ the following condition is satisfied:

$$(4.2) \qquad \sum_{i=1}^{M}\alpha_i\xi_i(\theta) = 0 \qquad \text{for all} \quad \theta\in\Omega \Rightarrow \alpha_i = 0, \quad i = 1, 2, \cdots, M.$$

Note that any representation of the form in (4.1) may always be reparametrized to satisfy (4.2). In the event that one does not wish to reparametrize a representation as in (4.1) when (4.2) is not satisfied, the techniques in Section 3 may still be employed for estimability considerations regarding parametric functions of the form $\sum_{i=1}^{M}\lambda_i\xi_i(\theta)$. For instance, in Example 1 denote the columns of X by $x_1, \cdots, x_p$, then $\mu_\beta = \sum_{i=1}^{p}\beta_i x_i$ is a representation as in (4.1) and when $\Delta \neq 0$ it is easily seen that (4.2) is not satisfied. However, as can be seen from Examples 1, 2 and 3 estimability considerations for parametric functions of the form $\lambda'\beta$ may be treated by the results in Section 3.

In many problems the form of μ_θ in (4.1) is a natural one and in many of these situations Condition 4.2 will be satisfied. From (4.1) a very natural linear operator for a $\mu_\theta = \mathbf{H}\xi_\theta$ representation may be defined. For this representation induced by (4.1) the condition in (4.2) implies that $\operatorname{sp}\Omega_H = \mathscr{H}$. The significance of $\operatorname{sp}\Omega_H = \mathscr{H}$ is partially evidenced in Section 3 (e.g., $\Omega_H{}^\perp = \{0\}$) and is exhibited further in the present section.

It was noted in Section 3 that a parametric function g is $\mathscr{A}$-estimable if and only if there exists an $F\in\mathscr{A}^\#$ such that $F(\mu_\theta) = g(\theta)$ for all $\theta\in\Omega$. Thus, a parametric function $\sum_i\lambda_i\xi_i(\theta)$ is $\mathscr{A}$-estimable if and only if there exists an $F\in\mathscr{A}^\#$ such that

$$\sum_{i=1}^{M}\xi_i(\theta)F(b_i) = F(\mu_\theta) = \sum_{i=1}^{M}\lambda_i\xi_i(\theta)$$

for all $\theta\in\Omega$. From this expression and (4.2) the following theorem may be stated.

THEOREM 4. *Assuming a representation as in* (4.1) *and that Condition* 4.2 *is true, the existence of an* $F\in\mathscr{A}^\#$ *such that* $F(b_i) = \lambda_i$ *for* $i = 1, 2, \cdots, M$ *is both a necessary and sufficient condition for* $\sum_{i=1}^{M}\lambda_i\xi_i(\theta)$ *to be* $\mathscr{A}$*-estimable.*

COROLLARY 4.1. *The parametric function* $\xi_k(\theta)$ *is* $\mathscr{A}$*-estimable if and only if* $b_k \notin \operatorname{sp}\{b_i : i \neq k\}$.

COROLLARY 4.2. *Each* $\xi_i(\theta)$ *is* $\mathscr{A}$*-estimable if and only if* $\mathscr{B}$ *forms a basis for* $\mathscr{E}$.

Corollary 4.2 follows easily from Corollary 4.1 which may be obtained from Theorem 4. To see that Theorem 4 implies Corollary 4.1 note that $\xi_k(\theta)$ is $\mathscr{A}$-estimable if and only if there exists an $F\in\mathscr{A}^\#$ such that $F(b_k) = 1$ and $F(b_i) = 0$ for $i \neq k$.

To utilize the results in the previous section, consider the following convenient $\mu_\theta = \mathbf{H}\xi_\theta$ representation induced by (4.1). Let $\mathscr{H} = R^M$, let $\langle\cdot,\cdot\rangle$ be the usual inner product on R^M, and define $\mathbf{H}$ from R^M into $\mathscr{A}$ by

$$\rho = (\rho_1, \rho_2, \cdots, \rho_M)' \in R^M \Rightarrow \mathbf{H}\rho = \sum_{i=1}^{M}\rho_i b_i.$$

Define $\xi_\theta = (\xi_1(\theta), \cdots, \xi_M(\theta))'$ for each $\theta\in\Omega$, then $\sum_{i=1}^{M}\xi_i(\theta)b_i = \mathbf{H}\xi_\theta$ for all $\theta\in\Omega$. Observe that $\Omega_H{}^\perp = \{0\}$, $\underline{R}(\mathbf{H}) = \mathscr{E}$, $\underline{N}(\mathbf{H}^*) = \mathscr{A}_0$, and that the linear operator

$\mathbf{H}^*$ from $\mathscr{A}$ into R^M takes an element $a \in \mathscr{A}$ into the vector $\mathbf{H}^*a$ in R^M with element i equal to (b_i, a) for each $i = 1, 2, \cdots, M$. In the remainder of this section $\mathbf{H}$ is defined as just described.

We illustrate the utility of $\mathbf{H}$ by using Corollaries 2.3 and 3.1 with $\mathbf{W} = \mathbf{H}$. Both corollaries require the form of $\mathbf{H}^*\mathbf{H}$ which may be conveniently described by an $M \times M$ matrix. To see this let $\rho \in R^M$, then

$$\begin{aligned}
(4.3) \qquad \mathbf{H}^*\mathbf{H}\rho &= \textstyle\sum_{i=1}^M \rho_i \mathbf{H}^* b_i \\
&= \textstyle\sum_{i=1}^M \rho_i \begin{bmatrix} (b_1, b_i) \\ \vdots \\ (b_M, b_i) \end{bmatrix} \\
&= \begin{bmatrix} (b_1, b_1) \cdots (b_1, b_M) \\ \vdots \qquad\qquad \vdots \\ (b_M, b_1) \cdots (b_M, b_M) \end{bmatrix} \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_M \end{bmatrix},
\end{aligned}$$

where the last expression denotes usual matrix multiplication. Denote by $\mathbf{H}^*\mathbf{H}$ the $M \times M$ matrix in (4.3), then Corollary 2.3 says that $\langle \lambda, \xi_\theta \rangle$ is $\bar{\mathscr{A}}$-estimable if and only if there exists ρ such that $\mathbf{H}^*\mathbf{H}\rho = \lambda$. In fact if $\mathbf{H}^*\mathbf{H}\rho = \lambda$, then by Corollary 2.2 the random variable

$$(4.4) \qquad (\mathbf{H}\rho, Y) = \textstyle\sum_{i=1}^M \rho_i (b_i, Y)$$

is an unbiased estimator for $\langle \lambda, \xi_\theta \rangle$. Further, Corollary 3.1 with $\mathbf{W} = \mathbf{H}$ implies that if $\lambda \in \underline{R}(\mathbf{H}^*\mathbf{H})$ and $\hat{\xi}$ is such that

$$(4.5) \qquad \mathbf{H}^*\mathbf{H}\hat{\xi} = \begin{bmatrix} (b_1, b_1) \cdots (b_1, b_M) \\ \vdots \qquad\qquad \vdots \\ (b_M, b_1) \cdots (b_M, b_M) \end{bmatrix} \begin{bmatrix} \hat{\xi}_1 \\ \vdots \\ \hat{\xi}_M \end{bmatrix} = \begin{bmatrix} (b_1, Y) \\ \vdots \\ (b_M, Y) \end{bmatrix} = \mathbf{H}^* Y,$$

then $\langle \lambda, \hat{\xi} \rangle$ is in $\bar{\mathscr{A}}$ and is an unbiased estimator for $\langle \lambda, \xi_\theta \rangle$.

EXAMPLE 4. Let Y denote an $n \times 1$ random vector with mean $\beta\mathbf{1}$ where β is an unknown parameter and $\mathbf{1}$ is a vector of ones. Suppose the covariance matrix of Y exists and is of the form $vV + \sigma^2 I$ where V is a known symmetric matrix and v and σ^2 are unknown parameters. Let Ω denote the subset of R^3 that describes the range of the unknown parameter vector $\theta = (\beta, v, \sigma^2)'$ and assume there is not a restriction on the parameters of the form $\alpha_1\beta^2 + \alpha_2 v + \alpha_3\sigma^2 = 0$ for some α_1, α_2, α_3 not all zero. Let $\mathscr{A}$ denote the linear space of $n \times n$ symmetric matrices and define $(A, B) = \operatorname{tr}(AB)$ for $A, B \in \mathscr{A}$, then $(\mathscr{A}, (\cdot\,,\cdot))$ is a finite-dimensional inner product space. Denote the random variable YY' which takes its values in $\mathscr{A}$ by U, then $\bar{\mathscr{A}} = \{(A, U) : A \in \mathscr{A}\} = \{Y'AY : A \in \mathscr{A}\}$ is the linear space of quadratic estimators. It follows that

$$(4.6) \qquad \text{(a)}\ \mu_\theta = \beta^2 J + vV + \sigma^2 I \quad (J = \mathbf{1}\mathbf{1}') \qquad \text{and}$$

$$\text{(b)}\ \mathscr{E} = \operatorname{sp}\{J, V, I\}.$$

The representation in (4.6.a) is of the form in (4.1) and the condition on the parameters assumed above implies that (4.2) is satisfied. Thus, the results of this section may be immediately applied. For instance, from Corollary 4.1 it follows that v is $\mathscr{A}$-estimable if and only if V is not a linear combination of J and I. The forms of $\mathbf{H}^*\mathbf{H}$ and $\mathbf{H}^*U$ for the linear operator $\mathbf{H}$ induced by (4.6.a) are

$$\mathbf{H}^*\mathbf{H} = \begin{bmatrix} n^2 & \mathbf{1}'V\mathbf{1} & n \\ \mathbf{1}'V\mathbf{1} & \operatorname{tr}(V^2) & \operatorname{tr}(V) \\ n & \operatorname{tr}(V) & n \end{bmatrix}, \qquad \mathbf{H}^*U = \begin{bmatrix} Y'JY \\ Y'VY \\ Y'Y \end{bmatrix}.$$

With these expressions one may use any of the results stated in this section. For example, the parametric function $\lambda_1\beta^2+\lambda_2 v+\lambda_3\sigma^2$ is $\mathscr{A}$-estimable if and only if the vector $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ is in the row space of the 3×3 matrix $\mathbf{H}^*\mathbf{H}$.

Under the transformation $\mathbf{H}$ a parametric function $\langle\lambda, \xi_\theta\rangle$ is equal to $\sum_i\lambda_i\xi_i(\theta)$. For some purposes concerning $\mathscr{A}$-estimability interest is in a certain subset of the ξ_i's. Thus, it seems natural to ask if we can find an operator $\mathbf{W}$ such that $\lambda \in \underline{R}(\mathbf{H}^*\mathbf{W})$ is both a necessary and sufficient condition for $\langle\lambda, \xi_\theta\rangle$ to be $\mathscr{A}$-estimable when certain λ_i's are zero. The next theorem (i.e., Theorem 5) is probably the most interesting result in this section and its usefulness is illustrated in a following paper which concerns quadratic estimability in a mixed linear model situation. In Theorem 5 and the corollaries it is assumed that $\mathscr{B}_0 = \{b_i : i \in S_0\}$ and $\mathscr{B}_1 = \{b_i : i \in S_1\}$ where S_0 and S_1 are disjoint sets with union equal to the first M integers.

THEOREM 5. *Let* $\mathbf{W}$ *be a linear operator such that* $\underline{R}(\mathbf{W})+\mathscr{A}_0 = \mathscr{B}_0^\perp$. *A necessary and sufficient condition for* $g(\theta) = \sum_{i\in S_1}\lambda_i\xi_i(\theta)$ *to be* $\mathscr{A}$*-estimable is the existence of a* ρ *such that* $\mathbf{H}^*\mathbf{W}\rho = \lambda$ *where* $\lambda_i = 0$ *for each* $i \in S_0$.

PROOF. Sufficiency follows from Corollary 2.2. Conversely if $g(\theta)$ is $\mathscr{A}$-estimable, then there exists an a such that $\lambda = \mathbf{H}^*a$ and such that

$$(a, \mu_\theta) = \textstyle\sum_i\xi_i(\theta)(a, b_i) = \sum_{i\in S_1}\lambda_i\xi_i(\theta) = g(\theta)$$

for all $\theta \in \Omega$. Thus by Condition 4.2 it follows that

$$\begin{aligned} (a, b_i) &= \lambda_i \qquad &\text{for } i \in S_1 \\ &= 0 \qquad & i \in S_0; \end{aligned}$$

and so $a \in \mathscr{B}_0^\perp$. Now let $a = \mathbf{W}\rho+f$ where $f \in \mathscr{A}_0$, then $\lambda = \mathbf{H}^*a = \mathbf{H}^*\mathbf{W}\rho+\mathbf{H}^*f = \mathbf{H}^*\mathbf{W}\rho$. Thus, the proof is complete.

COROLLARY 5.1. *Let* $\mathbf{W}$ *be a linear operator such that* $\underline{R}(\mathbf{W}) \oplus \mathscr{A}_0 = \mathscr{B}_0^\perp$. *If* $\langle\lambda, \xi_\theta\rangle = \sum_{i\in S_1}\lambda_i\xi_i(\theta)$ *is* $\mathscr{A}$*-estimable then* $\langle\lambda, \hat{\xi}\rangle$ *is an unbiased estimator for* $\langle\lambda, \xi_\theta\rangle$ *provided that* $\hat{\xi}$ *satisfies* $\mathbf{W}^*\mathbf{H}\hat{\xi} = \mathbf{W}^*Y$.

COROLLARY 5.2. *The number of linearly independent* $\mathscr{A}$*-estimable functions of the form* $g(\theta) = \sum_{i\in S_1}\lambda_i\xi_i(\theta)$ *is equal to* $\underline{r}(\mathbf{H}) - \dim[\operatorname{sp}\mathscr{B}_0] = \dim[\operatorname{sp}\mathscr{B}_1] - \dim[\operatorname{sp}\mathscr{B}_0 \cap \operatorname{sp}\mathscr{B}_1] = \dim[\underline{R}(\mathbf{H}) \cap \mathscr{B}_0^\perp]$.

COROLLARY 5.3. *Suppose that* $\mathbf{W}$ *is a linear operator such that* $\underline{R}(\mathbf{W})+\underline{N}(\mathbf{H}^*)\subset\mathscr{B}_0^{\perp}$ *but not equal to* $\mathscr{B}_0^{\perp}$. *For any* $a\in\mathscr{B}_0^{\perp}$ *such that* $a\notin\underline{R}(\mathbf{W})+\underline{N}(\mathbf{H}^*)$ *the parametric function* $\sum_{i\in S_1}(a, b_i)\xi_i(\theta)$ *is* $\bar{\mathscr{A}}$*-estimable.*

In Corollary 5.3 note that under the stated assumptions such an a always exists; furthermore, for such an a it is also true that $(a, b_i)\neq 0$ for at least one $i\in S_1$. Thus, in Theorem 5 the statement $\underline{R}(\mathbf{W})+\mathscr{A}_0=\mathscr{B}_0^{\perp}$ cannot be weakened to any form of inclusion in $\mathscr{B}_0^{\perp}$.

Acknowledgment. I wish to express my appreciation to Professor George Zyskind for his encouragement and guidance during the research and writing for my Ph.D. thesis of which this work is a part.

REFERENCES

[1] HALMOS, P. R. (1958). *Finite-dimensional Vector Spaces*, 2nd ed. Van Nostrand, Princeton.

[2] KRUSKAL, W. (1961). The coordinate-free approach to Gauss-Markov estimation, and its applications to missing and extra observations. *Fourth Berkeley Symp. Math. Statist. Prob.* **1** 435–451.

[3] WILANSKY, A. (1964). *Functional Analysis.* Blaisdell, New York.