INVARIANT ESTIMATION WITH APPLICATION TO LINEAR MODELS

by

Mary Sue Younger

Dissertation submitted to the Graduate Faculty of the

Virginia Polytechnic Institute and State University

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Statistics

APPROVED:

Dr. Lawrence S. Mayer, Chairman
Department of Statistics

Dr. Raymond H. Myers
Department of Statistics

Dr. Boyd Harshbarger, Head
Department of Statistics

Dr. I. J. Good
Department of Statistics

Dr. Paul W. Hamelman
Department of Business Administration

Dr. David L. Klemmack
Department of Sociology

August, 1972

Blacksburg, Virginia

ACKNOWLEDGEMENTS

The author would like to express her appreciation to the following people for their help in the preparation of this dissertation:

To Dr. Boyd Harshbarger for his encouragement throughout the author's years of graduate study and for his securing the financial aid which made graduate study possible.

To Dr. Lawrence S. Mayer, who directed this research with tireless energy and without whose confidence, enthusiasm and guidance this dissertation would have never been undertaken.

To Dr. Raymond H. Myers for his invaluable suggestions while co-reading this manuscript and his continued concern and encouragement throughout the author's years of graduate study.

To the Faculty serving on the Committee for their interest and criticism.

To the author's colleagues in the College of Business for their interest and concern for the completion of this dissertation, and for their aid in having copies made.

To who did the computer work necessary to the examples presented.

Finally, to who typed this manuscript quickly and accurately, and who never became irritated at having to make so many corrections and revisions.

TABLE OF CONTENTS

LIST OF FIGURES AND TABLES

CHAPTER I

INTRODUCTION

The principle of adopting a group of transformations on the sample space in a statistical estimation problem is quite common. Usually the group structure adopted affects the problem in one of two ways: i) The group structure is isomorphic to the parameter space and is assumed to "carry" the estimation problem in a natural manner. Attention is restricted to estimators which are also "carried" by the group in a natural manner. ii) The group structure generates orbits on the parameter space and the problem is to estimate in which orbit a parameter lies. Attention is restricted to estimators which are unaffected by the group.

Estimation problems of type i) are called <u>equivariant estimation problems</u> and have been widely studied. For reviews of this work see Zacks (1971), Ferguson (1967), Fraser (1968), Pitman (1939), Brown (1966), and Younger (1969). Equivariant estimation procedures are usually adopted when it is impossible to find a "best" estimator within the class of all estimators; by restricting attention to the class of equivariant estimators an optimal estimator may appear. Much of the literature deals with the properties of the "best" equivariant estimator when considered as a member of the class of all estimators. Hora and Buehler (1966) show the relationship between the best fiducial and best equivariant estimators.

Estimation problems of type ii) are called <u>invariant estimation problems</u>. These have been considered by Wijsman (1965), Stein (1959), Kiefer (1957), Kudō (1955), Hall, Wijsman and Ghosh (1965), and Zacks (1971). Invariant estimation problems arise when the orbit of the

parameter under a certain group, but not the parameter itself, must be estimated. The group can be thought of as a "nuisance" group, since estimation is restricted to be modulo the group. The most notable results dealing with these problems include the Stein theorem relating invariance and sufficiency, as presented by Zacks (1971) or Hall, Wijsman and Ghosh (1965); Wijsman's (1965) results obtaining probability ratios of maximal invariants, utilizing invariant measures and differential manifolds; and the results in multivariate distribution theory using invariance under the orthogonal group presented by James (1954, 1964).

In this dissertation we study estimation problems which give rise to both types of group structures. In particular, the parameter space supports a nuisance group of transformations and the objective is to estimate the orbit in which the parameter lies, thus giving rise to an invariant estimation problem. In addition, a second group can be defined such that the set of ordered pairs from the two groups is given a group structure and becomes isomorphic to the parameter space, thus defining an equivariant estimation problem. Part of the goal will be to show that if the relationship between the two groups satisfies certain properties, then an estimator of orbits in the invariant estimation problem can be naturally obtained from the best estimator in the equivariant estimation problem.

The motivation for the problem studied here is the problem of estimating the parameters in the general linear model, the nuisance group being the group of scale changes on the dependent and independent variables. This problem is of interest since in many areas of research,

most notably in the social sciences, an estimate of the relationship between variables is required which is independent of the units in which the variables are measured. The invariant estimator commonly used is called a "beta coefficient" or "standardized regression coefficient," but there appears to have been no rigorous derivation of this measurement and no investigation of its properties.

In Chapter II is a review of the definitions and concepts from group theory necessary to the development of invariant estimators. While most of the notions presented can be found in any elementary text on group theory, some less conventional ideas are introduced in order to facilitate the application of group theory to statistical problems.

Chapter III considers the structure of the estimation problem, defining the parameter, sample, and estimation spaces of interest and the groups of transformations defined on them. Chapter IV develops the theoretic basis for invariant estimation of the orbit in which the parameter of interest lies.

Application of the theory developed in Chapters III and IV to the problem of invariant estimation in the linear model for the case in which the independent variables are nonrandom is contained in Chapter V. The primary result in Chapter V is that the invariant estimator obtained through group theoretic considerations is not the same as the commonly used "beta coefficient." Comparison of the properties of the two estimators in Chapter VI reveals that the estimator obtained here is superior to the usual estimator in terms of such criteria as consistency, simplicity of distribution, and magnitude of

bias. Chapter VII compares the proposed estimator to the fiducial and minimum risk equivariant estimators, showing that the proposed estimator is a simple function of the fiducial and minimum risk equivariant estimators.

It is shown in Chapter VIII that the results obtained for the case in which the independent variables are fixed carry over to the case in which the input matrix is stochastic. However, in this latter case, some justification can be made for the use of the usual "beta coefficient."

Invariance under nonsingular transformations is considered in Chapter IX. It is shown that the proposed theory can be straightforwardly applied and an invariant estimator obtained. Interpretation of the physical meaning of this invariant estimator appears to be obscure, however.

In the last chapter, we illustrate the application of the methods developed herein to actual research problems: one from a sociological study and another from an engineering experiment. A dramatic change in the results is apparent in the engineering example, and to a lesser extent in the sociology example, due to the poor correlations among variables in the latter problem.

CHAPTER II

REVIEW OF GROUP THEORY

The concept of a group of transformations is fundamental to the notion of invariance. In this chapter we present some of the rudiments of group theory; in particular we present those results which will be utilized in our approach to invariant estimation. For a more complete yet elementary discussion of group theory, the reader is referred to Herstein (1964); for a more advanced discussion, see Goldhaber and Ehrlich (1970). Proofs of well-known lemmas will not be given here.

We begin with the notions of a group and a subgroup:

Definition 2.1. Let K be a nonempty set and let $\circ$ be a binary operation on K; that is, $\circ$ is a mapping defined on $K \times K$. The pair $(K, \circ)$, abbreviated K, is a group provided the following four properties are satisfied:

i) K is closed under $\circ$; that is, if $a,b \in K$, then $a \circ b \in K$.

ii) $\circ$ is associative in K; that is, for $a,b,c \in K$, $(a \circ b) \circ c = a \circ (b \circ c)$.

iii) K contains an identity element; that is, there exists an element $e \in K$ such that for every element $a \in K$, $a \circ e = e \circ a = a$.

iv) Every element in K has an inverse which is also in K; that is, for every $a \in K$, there exists an element $a^{-1} \in K$ such that $a \circ a^{-1} = a^{-1} \circ a = e$.

From now on, we will write simply ab for $a \circ b$ if the operation is understood.

Definition 2.2. A group K is a group of transformations on a space Ω if for all $k \in K$, k is a map from Ω to Ω and if $a,b \in K$ and $\omega \in \Omega$ then $(ab)\omega = a(b\omega)$; that is, if the group operator is composition.

Definition 2.3. A group K is called Abelian if the operation in K is commutative; that is, if for every $a,b \in K$, $ab = ba$.

Definition 2.4. A subset K_1 of a group K is a subgroup of K if K_1 is itself a group under the operation defined in K.

Lemma 2.1. Necessary and sufficient conditions for K_1 to be a subgroup of K are:

i) K_1 must be closed under the operation in K; that is, if $a,b \in K_1$, then $ab \in K_1$.

ii) K_1 must contain inverses; that is, if $a \in K_1$, then $a^{-1} \in K_1$.

Lemma 2.2. If K_1 is a subgroup of a group K, then the identity element of K is in K_1 and is the identity element of K_1.

The concepts of a coset and a normal subgroup are central to our development of the theory of invariant estimation.

Definition 2.5. Let K_1 be a subgroup of a group K, and let k be an arbitrary element of K. Then the set

$$K_1k = \{k_1k: \ k_1 \in K_1\}$$

is a right coset of K_1 in K. Left cosets are defined analogously.

The class of all right cosets of K_1 in K will be denoted $C_R(K_1:K)$.

Lemma 2.3. If K_1 is a subgroup of a group K, and a and b are elements of K, then the following conditions are equivalent:

i) $a \in K_1 b$

ii) $K_1 a = K_1 b$

iii) $ab^{-1} \in K_1$.

Lemma 2.4. If K_1 is a subgroup of a group K, then the distinct right (left) cosets of K_1 in K partition K into equivalence classes, such that each element of K is in one and only one equivalence class and any element in an equivalence class can be reached from any other element in the same equivalence class via the operation of some $k_1 \in K_1$.

Definition 2.6. A subgroup K_1 of a group K is a normal subgroup if and only if kk_1k^{-1} is in K_1 for every $k_1 \in K_1$ and $k \in K$.

Lemma 2.5. If K_1 is a normal subgroup of a group K, then $kK_1 = K_1k$ for every $k \in K$. That is, if K_1 is a normal subgroup, then the right and left cosets of K_1 in K are identical.

We note that if K is an Abelian group, then all subgroups of K will be normal subgroups.

Lemma 2.6. If K_1 is a normal subgroup of a group K, then for every pair of cosets K_1a and K_1b, $(K_1a)(K_1b) = K_1ab$.

In essence Lemma 2.6 states that if K_1 is normal, then a natural

operation can be defined on $C_R(K_1:K)$. Note, however, that if K_1 is not normal, while the cosets of K_1 in K partition K, they do not inherit a natural operation and thus do not possess group structure.

Lemma 2.7. If K_1 is a normal subgroup of a group K, then $C_R(K_1:K)$, the set of all distinct right (or left) cosets of K_1 in K, is itself a group, called the quotient group of K modulo K_1 and denoted K/K_1.

The notion of an isomorphism is essential to the development of both equivariant and invariant estimation problems.

Definition 2.7. Let K and L be groups with operations $\circ$ and $*$, respectively. A homomorphism is a mapping μ of K into L such that for every a and b in K, $\mu(a \circ b) = \mu(a) * \mu(b)$. If this last relationship holds, μ is said to preserve the group operation.

Definition 2.8. A mapping from K onto L requires that every element of L be the image of at least one element of K. A mapping from K to L is a one-to-one mapping if each element of L is the image of at most one element of K.

Definition 2.9. An isomorphism of a group K onto a group L is a homomorphism which is a one-to-one mapping of K onto L. If an isomorphism of K onto L exists, we say that the two groups are isomorphic.

In order to apply results in group theory to statistical estimation problems, it is necessary to define isomorphisms between groups and spaces.

Definition 2.10. A group K of transformations on a space Ω is transitive on Ω if for all $a,b \in \Omega$ there exists an element $k \in K$ such that $a = k(b)$.

Definition 2.11. A group K of transformations on a space Ω is isomorphic to Ω if for all $a,b \in \Omega$ there exists one and only one $k \in K$ such that $a = k(b)$.

We note that i) the isomorphism defined in Definition 2.11 is not a group isomorphism since Ω is not a group; and ii) if K is isomorphic to Ω, then K is transitive on Ω, but not vice versa.

Definition 2.12. A group L of transformations on a space Ω is sub-isomorphic to Ω if there exists a group K isomorphic to Ω and L is isomorphic to a subgroup of K.

Invariant estimation problems require that the parameter space be partitioned into equivalence classes, or orbits:

Definition 2.13. If K is a group of transformations on a space Ω, then $\{\omega : \omega = k\omega_0$ for fixed $\omega_0 \in \Omega$ and any $k \in K\}$ is the orbit of ω_0 in Ω.

Lemma 2.8. The orbits in Ω partition Ω into disjoint exhaustive classes.

Proof. We must show that i) every $\omega \in \Omega$ belongs to at least one orbit and ii) that no $\omega \in \Omega$ belongs to two orbits. Let ω_0 be an arbitrary point in Ω. Since K is a group of transformations on Ω, the identity e is in K and thus ω_0 is in the orbit of ω_0. Clearly if

$\omega = k_0\omega_0$ for some $k_0 \in K$ the orbit of ω is the orbit of ω_0. Suppose ω in Ω is in the orbit of ω_0 and in the orbit of ω_1. Then $\omega = k_0\omega_0 = k_1\omega_1$ and thus $\omega_0 = k_0^{-1}k_1\omega_1$ so that the orbit of ω_0 in Ω is the same as the orbit of ω_1 in Ω. Therefore ω is in only one orbit.

Lemma 2.9. If K is a group transitive on Ω then Ω is a single orbit.

Proof. If K is transitive on Ω, then for a fixed ω_0 and an arbitrary ω in Ω, there exists an element $k \in K$ such that $\omega = k\omega_0$, so that ω is in the orbit of ω_0.

Definition 2.14. Let the equivalence relation "$\sim_K$" be defined on Ω by orbit membership; that is, if $\omega_1, \omega_2 \in \Omega$, then $\omega_1 \sim_K \omega_2$ if and only if there exists an element $k \in K$ such that $\omega_2 = k\omega_1$.

Definition 2.15. Let $S_K(\Omega)$ be the space of orbits of Ω; that is, $S_K(\Omega) = \Omega/\sim_K$.

Lemma 2.10. If the group K is isomorphic to Ω and L is a subgroup of K, then the cosets of L and the elements of $\Omega/\sim_L = S_L(\Omega)$ are in one-to-one correspondence.

Proof. Choose a point ω_0 in Ω to identify with e, the identity transformation in K. Consider the mapping $f : C_R(L:K) \to S_L(\Omega)$ defined by $f(Lk) = \{\omega : \omega = \ell k\omega_0$ for some $\ell \in L\}$. i) We must first show that f is well-defined. Suppose $Lk_1 = Lk_2$ for $k_1, k_2 \in K$. Then

$$f(Lk_1) = \{\omega : \omega = \ell k_1\omega_0 \text{ for some } \ell \in L\}$$

$$= \{\omega : \omega = \ell k_1 k_2^{-1} k_2\omega_0 \text{ for some } \ell \in L\}$$

$$= f(Lk_2) ,$$

since $Lk_1 = Lk_2$ implies that $Lk_1k_2^{-1} = L$ so that $\ell k_1k_2^{-1} \varepsilon L$. ii) We must now show that f is an onto mapping. For fixed $\omega_1 \varepsilon \Omega$, consider $\{\omega : \omega = \ell\omega_1$ for some $\ell \varepsilon L\}$. Since K is isomorphic to Ω, there exists an element $k_1 \varepsilon K$ such that $\omega_1 = k_1\omega_0$. Consider

$$f(Lk_1) = \{\omega : \omega = \ell k_1\omega_0 \text{ for some } \ell \varepsilon L\}$$

$$= \{\omega : \omega = \ell\omega_1 \text{ for some } \ell \varepsilon L\} .$$

iii) Finally, we must show that f is one-to-one. Suppose $f(Lk_1) = f(Lk_2)$. Then

$$\{\omega : \omega = \ell k_1\omega_0 \text{ for some } \ell \varepsilon L\} =$$

$$= \{\omega : \omega = \ell k_2\omega_0 \text{ for some } \ell \varepsilon L\} ,$$

and thus $\ell k_1 = \ell' k_2$ for some $\ell' \varepsilon L$. Therefore, $Lk_1 = Lk_2$.

The above lemmas can be used as follows. Suppose a problem involves determining in which orbit of a group K an unknown ω lies. Lemma 2.8 implies the problem is well defined. If the group K is transitive on Ω the problem is degenerate since all ω lie in a single orbit (Lemma 2.9). Suppose the group K is a proper subgroup of a group H isomorphic on Ω; then by Lemma 2.10 the orbits of K on Ω and the cosets of K in H are in one-to-one correspondence. By identifying orbits and cosets the problem of determining in which orbit ω lies is equivalent to the problem of choosing a coset.

Suppose the group K is transitive on the space Ω but not isomorphic to Ω. Then the problem of determining in which orbit ω lies can be transformed to the problem of choosing a coset by the above results and the following lemma.

Definition 2.16. If K is transitive on Ω, let K_{ω_0} be the isotropy group of ω_0 in Ω where $k \in K_{\omega_0}$ implies that $k\omega_0 = \omega_0$.

Lemma 2.11. If K is transitive on Ω, then there is a one-to-one correspondence between $C_L(K_{\omega_0}:K)$ and elements in Ω.

Proof. Define a function g by $g\{kK_{\omega_0}\} = k\omega_0$. If $k_1 K_{\omega_0} = k_2 K_{\omega_0}$, then

$$g\{k_1K_{\omega_0}\} = k_1\omega_0 = k_2k_2^{-1}k_1\omega_0$$

$$= k_2\omega_0 = g\{k_2K_{\omega_0}\} ,$$

so g is well-defined. Since K is transitive, if $\omega \in \Omega$, then $\omega = k^*\omega_0$ for some $k^* \in K$ and thus $\omega = g\{k^*K_{\omega_0}\}$, so that g is onto. Finally, g is one-to-one since if $g\{k_1K_{\omega_0}\} = g\{k_2K_{\omega_0}\}$, then $k_1\omega_0 = k_2\omega_0$ implies that $k_2^{-1}k_1 \in K_{\omega_0}$, which in turn implies that $k_1K_{\omega_0} = k_2K_{\omega_0}$.

Finally, we introduce the notion of decomposing a group into the product of subgroups.

Definition 2.17. An Abelian group K is the direct product of two subgroups K_1 and K_2 if every $k \in K$ can be written as $k = k_1k_2 = k_2k_1$ for one and only one $k_1 \in K_1$ and $k_2 \in K_2$.

Thus if a group K of transformations on a space Ω is the direct product of K_1 and K_2, then if $k \in K$ the action of k on a space Ω can be expressed as the composite action of an element k_1 in K_1 and an element k_2 in K_2. If K is not Abelian the notion of a direct product can be replaced by the notion of a semi-direct product:

Definition 2.18. A group K is the (left) semi-direct product of two subgroups K_1 and K_2 if every $k \in K$ can be written uniquely as $k = k_2 k_1$ for one and only one $k_1 \in K_1$ and $k_2 \in K_2$. The decomposition of k into $k_2 k_1$ will be called the left decomposition. A right decomposition can be defined similarly.

CHAPTER III

THE STRUCTURE OF THE ESTIMATION PROBLEM

Section 3.1. Definition of an Equivariant Estimation Problem

The estimation problems studied in this dissertation involve six concepts: $\mathcal{X}_0$, $\mathcal{A}$, P_θ, Θ, $L(\cdot,\cdot)$, and $\mathcal{D}$. We denote by $\mathcal{X}_0$ a sample space and by $\mathcal{A}$ a σ-algebra of subsets of $\mathcal{X}_0$. P_θ is a probability measure on $\mathcal{A}$, the index θ belonging to Θ, the parameter space. $L(\cdot,\cdot)$ is a function defined from $\Theta \times \Theta$ to R^+, the positive real line, and is called the loss function. $\mathcal{D}$ denotes the space of all mappings from $\mathcal{X}_0$ to Θ; these mappings are called estimators.

The problem is to estimate the value of θ. Given an estimator $d(\cdot) \in \mathcal{D}$ the statistician observes a value $z \in \mathcal{X}_0$ according to the measure P_θ, estimates θ to be $\hat{\theta} = d(z)$, and suffers loss $L(\theta,\hat{\theta})$. The goal is to find an estimator $d(\cdot)$ which minimizes the expected value of $L(\theta,\hat{\theta})$ for all values of θ. The search for such an estimator is usually fruitless unless attention is restricted to some subclass of the class of estimators (for example, unbiased or sufficient estimators), or alternatively, unless the parameter space is reduced in some manner. Although the most common method for reducing the class of available estimators is probably by using the criterion of unbiasedness, a second method of reduction is by equivariance, as was defined in Chapter I. In applying this latter criterion, the statistician requires that his estimator be equivariant under some group of transformations.

In this dissertation an estimation problem is called <u>equivariant</u>

if a group H^* of transformations on X_0 is defined which affects the problem in the following manner. i) H^* induces a group $\bar{H}$ of transformations on the parameter space Θ. If z has a distribution with parameter value θ, then $h^* z$ has a distribution with parameter value $\bar{h}\theta$. ii) $\bar{H}$ is isomorphic to Θ, and iii) the loss function $L(\cdot,\cdot)$ satisfies the invariance condition

$$L(\theta,\hat{\theta}) = L(\bar{h}\theta, \bar{h}\hat{\theta}) .$$

iv) In an equivariant estimation problem, attention is restricted to estimators which satisfy the property

$$\bar{h}[d(z)] = d[h^*(z)] .$$

As an illustration of the requirements of an equivariant decision problem, consider the problem of estimating the true mean of a set of measurements. Suppose the true mean is two (feet); transforming the measurements to inches makes the true mean 24 (inches). If the problem is equivariant the estimator of the mean when the measurements are taken in inches should be 12 times as great as the estimator obtained if the measurements are taken in feet. Moreover, regardless of whether the measurements are taken in feet or inches, the loss incurred in estimation should not be affected.

The notion of an equivariant estimation problem can be made more precise by introducing the following definition:

<u>Definition 3.1.1.</u> i) A function f defined on a space Ω is said to be <u>invariant</u> under a group K of transformations on Ω if

$$f(k\omega) = f(\omega)$$

for $\omega \in \Omega$ and $k \in K$. ii) The function f is said to be <u>equivariant</u> under K if

$$f(k\omega) = [\tilde{k}\, f]\ (\omega)$$

for $\omega \in \Omega$, $k \in K$ and $\tilde{k} \in \tilde{K}$, where $\tilde{K}$ is the group of transformations induced by K on the space of functions defined on Ω.

We may say that a function is equivariant if it commutes with the group operator, and invariant if it is constant on the orbits of K in Ω.

An equivariant estimation problem is thus one in which i) the distribution of the random variable is equivariant, ii) the loss function is invariant, and iii) attention is restricted to decision rules which are equivariant. If attention is restricted to the class of equivariant estimators the statistician may be able to obtain the minimum risk equivariant estimator. Under suitable conditions, the minimum risk equivariant estimator is also minimum risk (or at least admissible) in the space of all estimators (Ferguson (1967)).

Section 3.2. Additional Structure of the Proposed Problems

In this dissertation we consider estimation problems which have a common basic structure which gives rise to both an equivariant and an invariant estimation problem. It is assumed that any problem under consideration yields a sufficient statistic x taking values in a space $\mathcal{X}$. A sub-isomorphic group G_1^* of transformations on $\mathcal{X}_0$ is given which induces sub-isomorphic groups G_1 and $\bar{G}_1$ of transformations on $\mathcal{X}$ and Θ,

respectively. The problem is to estimate the orbit of Θ (with respect to $\bar{G}_1$) in which θ lies.

It is assumed that a group G_2^* of transformations on X_0 can also be defined such that if

$$G^* = \{[g_2, g_1] : g_1 \in G_1^* \text{ and } g_2 \in G_2^*\} ,$$

then an operation can be defined on G^* which makes G^* a group which is the semi-direct product of isomorphic images of G_2^* and G_1^*. Furthermore, G^* induces groups G and $\bar{G}$ of transformations on X and Θ, respectively, to which G and $\bar{G}$ are isomorphic. Finally, it is assumed that since $\bar{G}$ is isomorphic to the space Θ, $\bar{G}$ gives rise to an equivariant decision problem.

Suppose reference points x_0 and θ_0 are chosen in X and Θ and identified with the identity elements in G and $\bar{G}$, respectively. Then since G and $\bar{G}$ are isomorphic to X and Θ, any point $x \in X$ can be written

$$x = g_x x_0$$

for a unique g_x in G; similarly, any point $\theta \in \Theta$ can be written

$$\theta = \bar{g}_\theta \theta_0$$

for a unique $\bar{g}_\theta$ in $\bar{G}$. Therefore, once x_0 and θ_0 are chosen, points in X and Θ can be identified with operators in G and $\bar{G}$ in a natural way.

Section 3.3. An Example of the Estimation Structure Under Consideration

As an example of the structure of the estimation problems considered in this dissertation, consider the problem of estimating the mean μ and variance σ^2 of a normal distribution, given a sample $\underline{z} = (z_1,\ldots,z_n)$ of n observations. The sample space is $\mathcal{X}_0 = \{\underline{z}\}$ and the parameter space is

$$\Theta = \{\theta = (\mu,\sigma^2) : \mu \in R \text{ and } \sigma^2 \in R^+\} .$$

Consider the class H^* of location and scale transformations on $\mathcal{X}_0$

$$H^* = \{[c,\lambda^2] : c \in R,\ \lambda^2 \in R^+\} ,$$

where the operation is defined on H^* by

$$[c,\lambda^2]\ [c_1,\lambda_1^2] = [\lambda c_1 + c,\ \lambda^2\lambda_1^2] .$$

Lemma 3.3.1. H^* is a group.

Proof. i) Since $\lambda c_1 + c \in R$ and $\lambda^2\lambda_1^2 \in R^+$, H^* is closed. ii) For associativity, we have

$$[c,\lambda^2]\ ([c_1,\lambda_1^2]\ [c_2,\lambda_2^2]) = [c,\lambda^2]\ [\lambda_1 c_2 + c_1,\ \lambda_1^2\lambda_2^2]$$

$$= [\lambda\lambda_1 c_2 + \lambda c_1 + c,\ \lambda^2\lambda_1^2\lambda_2^2]$$

and

$$([c,\lambda^2]\ [c_1,\lambda_1^2])\ [c_2,\lambda_2^2] = [\lambda c_1 + c,\ \lambda^2\lambda_1^2]\ [c_2,\lambda_2^2]$$

$$= [\lambda\lambda_1 c_2 + \lambda c_1 + c,\ \lambda^2\lambda_1^2\lambda_2^2] .$$

iii) The identity element is $[0,1]$ since

$$[0,1]\ [c,\lambda^2] = [1c + 0,\ 1\lambda^2] = [c,\lambda^2]$$

and

$$[c,\lambda^2]\ [0,1] = [\lambda 0 + c,\ \lambda^2 1] = [c,\lambda^2]\ .$$

iv) The inverse of $[c,\lambda^2]$ is $[-\ c\lambda^{-1},\ \lambda^{-2}]$ since

$$[c,\lambda^2]\ [-\ c\lambda^{-1},\ \lambda^{-2}] = [\lambda(-\ c\lambda^{-1}) + c,\ \lambda^2\lambda^{-2}]$$

$$= [0,1]$$

and

$$[-\ c\lambda^{-1},\ \lambda^{-2}]\ [c,\lambda] = [\lambda^{-1}c - c\lambda^{-1},\ \lambda^{-2}\lambda^2]$$

$$= [0,1]\ .$$

In order to make H^* a group of transformations on X_0 we now define how an element in H^* acts on an element in X_0. Let

$$[c,\lambda^2]\ (z_1,\ldots,z_n) = (\lambda z_1 + c,\ldots,\ \lambda z_n + c)\ .$$

(Note that square brackets are used to denote transformations in a group, and round brackets to denote elements in a space.)

The group of transformations induced by H^* on Θ is

$$\bar{H} = \{[c,\lambda^2]\ :\ c \in R,\ \lambda^2 \in R^+\}\ ,$$

where the operation in $\bar{H}$ is defined in exactly the same way as in H^*. $\bar{H}$ acts on Θ as

$$[c,\lambda^2]\ (\mu,\sigma^2) = (\lambda\mu + c,\ \lambda^2\sigma^2)\ .$$

Lemma 3.3.2. $\bar{H}$ is isomorphic to Θ.

Proof. Given θ_1 and $\theta_2 \in \Theta$, where

$$\theta_1 = (\mu_1,\sigma_1^2)$$

$$\theta_2 = (\mu_2,\sigma_2^2)\ ,$$

$$(\mu_1,\sigma_1^2) = [a,b^2]\ (\mu_2,\sigma_2^2)$$

where

$$b^2 = \sigma_1^2/\sigma_2^2$$

and

$$a = \mu_1 - (\sigma_1/\sigma_2)\ \mu_2\ ,$$

and this representation is unique. The isomorphism between $\bar{H}$ and Θ is represented by choosing the reference point

$$\theta_0 = (0,1)$$

in Θ so that if $\theta = (\mu,\sigma^2) \in \Theta$, then

$$\theta = \bar{h}_\theta\theta_0 = [\mu,\sigma^2]\ (0,1) = (\mu,\sigma^2)\ ,$$

and thus $[\mu,\sigma^2]$ and (μ,σ^2) are identified with each other.

The problem gives rise to the two-dimensional sufficient statistic $(\bar{x},s^2)$ and thus the sufficient statistic space $X = R \times R^+$. The induced group H on X contains elements $h = [c,\lambda^2]$, where

$$[c,\lambda^2]\ (\bar{x},s^2) = (\lambda\bar{x} + c,\ \lambda^2 s^2)\ ,$$

and where the group operation is defined as before. Clearly, H is isomorphic to $\mathcal{X}$, and H and $\bar{H}$ are isomorphic groups.

By defining an invariant loss function such as

$$L(\theta,\hat{\theta}) = (\mu - d_1(x))^2 + (1/\sigma^4)\ (\sigma^2 - d_2(x))^2$$

where

$$d(x) = (d_1(x),\ d_2(x))$$

is equivariant under $\bar{H}$, we have an estimation problem equivariant under the group H.

However, suppose the problem is not to estimate (μ,σ^2), but to estimate a function of (μ,σ^2) which is invariant under scale changes. Let G_1^* be the class of scale changes on $(z_1,\ldots,z_n)$ where

$$G_1^* = \{[\lambda]_1 : \lambda \in R^+\}$$

with operation

$$[\lambda_1]_1\ [\lambda_2]_1 = [\lambda_1\lambda_2]_1$$

and where

$$[\lambda]_1\ (z_1,\ldots,z_n) = (\lambda z_1,\ldots,\lambda z_n)\ .$$

<u>Lemma 3.3.3</u>. G_1^* is a group of transformations on $\mathcal{X}_0$.

<u>Proof</u>. Since $\lambda_1\lambda_2 \in R^+$, G_1^* is closed. Associativity is immediate. The identity element is $[1]_1$, and the inverse of $[\lambda]_1$ is $[\lambda^{-1}]_1$.

G_1^* induces a group of transformations $\bar{G}_1$ on Θ, where the operation

in $\bar{G}_1$ is identical to the operation in G_1^*, and if

$$\bar{g} = [\lambda]_1 \in \bar{G}_1 \text{ for } \lambda \in R^+ ,$$

then

$$\bar{g}\theta = [\lambda]_1 (\mu,\sigma^2) = (\lambda\mu, \lambda^2\sigma^2) .$$

Lemma 3.3.4. $\bar{G}_1$ is not transitive on Θ and thus not isomorphic to Θ.

Proof. The proof is immediate since if $[\mu_1,\sigma_1^2] = [\lambda]_1 (\mu_2,\sigma_2^2)$ then $\mu_1/\sigma_1 = \mu_2/\sigma_2$ and thus for some $\theta_1,\theta_2 \in \Theta$ there does not exist a $\bar{g} \in \bar{G}_1$ such that $\theta_1 = \bar{g}\theta_2$.

Since $\bar{G}_1$ is not transitive on Θ, $S_{\bar{G}_1}(\theta)$ contains more than one orbit and it makes sense to estimate orbits.

In order to complete the necessary structure, let G_2^* be the class of location changes on X_0. We define

$$G_2^* = \{[c]_2 : c \in R\}$$

with the operation

$$[c_1]_2 [c_2]_2 = [c_1 + c_2]_2 .$$

Lemma 3.3.5. G_2^* is a group.

Proof. G_2^* is closed since $c_1 + c_2 \in R$, and associativity is immediate. The identity is $[0]_2$, and the inverse of $[c]_2$ is $[-c]_2$.

Define the operation of G_2^* on X_0 as

$$[c]_2\ (z_1,\ldots,z_n) = (z_1 + c,\ldots,z_n + c)\ .$$

Definition 3.3.1. Let G^* be the set of ordered pairs $[g_2,g_1^2]$ for $g_2 \in G_2^*$ and $g_1 \in G_1^*$, where $([\lambda]_2)^2 = [\lambda^2]_2$, and let the following operation be defined on G^*:

$$[g_2,g_1^2]\ [g_2',g_1'^2] = [g_1g_2' + g_2,\ g_1^2g_2'^2]\ .$$

Thus G^* is a class of location and scale changes on X_0.

Lemma 3.3.6. G^* is a group of transformations on X_0.

Proof. i) Since $g_1g_2' + g_2 \in R$ and $g_1^2g_1'^2 \in R^+$, G^* is closed.

ii) To demonstrate associativity,

$$[g_2,g_1^2]\ ([g_2',g_1'^2]\ [g_2'',g_1''^2]) = [g_2,g_1^2]\ [g_1'g_2'' + g_2',\ g_1'^2g_1''^2]$$

$$= [g_1g_1'g_2'' + g_1g_2' + g_2,\ g_1^2g_1'^2g_1''^2]$$

and

$$([g_2,g_1^2]\ [g_2',g_1'^2])\ [g_2'',g_1''^2] = [g_1g_2' + g_2,\ g_1^2g_1'^2]\ [g''_2,g_1''^2]$$

$$= [g_1g_1'g_2'' + g_1g_2' + g_2,\ g_1^2g_1'^2g_1''^2]\ .$$

iii) The identity is $[0,1]$, since

$$[0,1]\ [g_2,g_1^2] = [g_2,g_1^2]$$

and

$$[g_2,g_1^2]\ [0,1] = [g_2,g_1^2]\ .$$

iv) The inverse of $[g_2,g_1^2]$ is $[-g_2g_1^{-1},\ g_1^{-2}]$ since

$$[g_2, g_1^2]\ [-g_2 g_1^{-1},\ g_1^2] = [g_1(-g_2 g_1^{-1}) + g_2,\ g_1^2 g_1^{-2}]$$

$$= [0,1]$$

and

$$[-g_2 g_1^{-1},\ g_1^{-2}]\ [g_2, g_1^2] = [g_1^{-1} g_2 - g_2 g_1^{-1},\ g_1^{-2} g_1^2]$$

$$= [0,1]\ .$$

Let $\bar{G}$ be the group of transformations induced on Θ by G^*.

<u>Lemma 3.3.7</u>. $\bar{G}$ is a group of transformations isomorphic to Θ.

<u>Proof</u>. The proof is immediate since it is apparent that $\bar{G}$ is identical to $\bar{H}$ and Lemma 3.3.2 shows that $\bar{H}$ is isomorphic to Θ.

<u>Lemma 3.3.8</u>. G^* is not Abelian.

<u>Proof</u>. $[g_2, g_1^2]\ [g_2', g_1'^2] = [g_1 g_2' + g_2,\ g_1^2 g_2'^2]$ but

$$[g_2', g_1'^2]\ [g_2, g_1^2] = [g_1' g_2 + g_2',\ g_1'^2 g_1^2]\ .$$

<u>Lemma 3.3.9</u>. G_1^* and G_2^* are isomorphic to the subgroups H_1^* and H_2^* of H^*, respectively, where if $h_1 \in H_1^*$ then

$$h_1 = [0, \lambda^2]$$

and if $h_2 \in H_2^*$, then

$$h_2 = [c, 1]\ .$$

Thus $\bar{G}_1$ is a sub-isomorphic group with respect to Θ.

<u>Proof</u>. The proof is immediate if one identifies $[\lambda]_1$ in G_1^* with

$[0,\lambda^2]$ in H_1^* and $[c]_2$ in G_2^* with $[c,1]$ in H_2^*.

Lemma 3.3.10. H^* is not the direct product of H_1^* and H_2^*.

Proof. The proof is immediate since H^* is not Abelian, and direct products must be Abelian.

Lemma 3.3.11. H^* is the semi-direct product of H_1^* and H_2^*.

Proof. The proof is immediate since any $[a,b^2] \in H^*$ can be written $[a,1]\ [0,b^2]$.

The structure is complete since $\bar{H}$ (or $\bar{G}$) is isomorphic to Θ, $\bar{H}$ is the semi-direct product of $\bar{H}_1$ and $\bar{H}_2$, and the problem is to estimate the orbit of $\bar{H}_1$ (or $\bar{G}_1$) in which θ lies.

CHAPTER IV

INVARIANT ESTIMATION OF ORBITS OF THE PARAMETER SPACE

Section 4.1. A Group Theoretic Function Which Indexes Orbits

In this chapter we will derive an estimator for the orbit of Θ (with respect to a sub-isomorphic group L) which is a natural function of the sufficient statistic. We begin by showing that if K is a group of transformations isomorphic to Θ and L is a subgroup of K, then estimating orbits of Θ is equivalent to estimating the cosets of L in K. Thus if L is a normal subgroup, estimating the orbit of Θ is equivalent to estimating elements of K/L. Our estimator will be obtained from group theoretic considerations and not by trying to minimize the appropriate invariant risk function. However, in Chapter VII we will relate our estimator to the minimum risk equivariant estimator.

The problem of estimating the orbits of a sub-isomorphic group L in Θ is facilitated by introducing the notion of a function which indexes orbits.

Definition 4.1.1. Consider a group L of transformations sub-isomorphic to a space Ω. A function f defined on the space Ω will be said to index the orbits of L in Ω provided $f(\omega_1) = f(\omega_2)$ for ω_1, $\omega_2 \in \Omega$ only if ω_1 and ω_2 belong to the same orbit of L in Ω. That is, $f(\omega_1) = f(\omega_2)$ if and only if $\omega_1 = \ell\omega_2$ for $\ell \in L$.

Thus a function indexes orbits if it is constant on an orbit and if it takes different values on different orbits.

If $\Omega = X_0$, then a function which indexes the orbits of L in Ω is

usually called a maximal invariant (see, for example, Wijsman (1957)).

Lemma 4.1.1. If f indexes the orbits of L (a group of transformations) in a space Ω, then f is invariant with respect to L, and any other function invariant with respect to L is a function of f.

Proof. If f indexes the orbits of L in Ω then

$$f(\omega_1) = f(\ell\omega_1) \text{ if } \ell \in L \text{ for } \omega_1 \in \Omega$$

and thus f is invariant with respect to L. If f^* is invariant with respect to L then

$$f^*(\omega_1) = f^*(\ell\omega_1) \text{ if } \ell \in L \text{ and } \omega_1 \in \Omega$$

and thus f^* can be written as a function of f.

Definition 4.1.2. If f is a function which is invariant and such that any other invariant function is a function of f, then f is called a maximal invariant function.

Thus, any function which indexes the orbits of L in Ω is a maximal invariant function under L.

Continuing the example of estimating $\theta = (\mu,\sigma^2)$ from a normal distribution, consider estimating the orbits of the group $\bar{H}_2$ of location changes in Θ. The function which indexes the orbits of $\bar{H}_2$ in Θ must be invariant under transformations of the form

$$[c,1]\ (\mu,\sigma^2) = (\mu + c,\ \sigma^2)\ .$$

Thus the problem of estimating the orbits of $\bar{H}_2$ is the problem of

estimating the element of $\bar{H}_1$ associated with (μ,σ^2), or equivalently, estimating σ^2. Since μ is a nuisance parameter and the problem is to estimate the coset of $\bar{H}_2$, $\bar{H}_2$ might be called a <u>nuisance group</u>.

Now consider the problem of estimating the orbits of the group $\bar{H}_1$ in the same problem. Since

$$[0,\lambda^2]\ (\mu,\sigma^2) = (\lambda\mu,\ \lambda^2\sigma^2)\ ,$$

neither μ nor σ^2 indexes the orbits of $\bar{H}_1$ in Θ, so the problem in this case does not reduce to the problem of estimating a single parameter in the presence of a nuisance parameter.

The problem of estimation in the presence of nuisance parameters has been explored elsewhere (Zacks, 1971); we will concentrate on problems of the second type, in which orbits of the nuisance group are not indexed by the parameters themselves. We will first show how to obtain a function which indexes the orbits of $\bar{H}_1$ in H, provided that $\bar{H}_2$ is a normal subgroup of a group $\bar{H}$, where $\bar{H}$ is the semi-direct product of $\bar{H}_1$ and $\bar{H}_2$.

<u>Lemma 4.1.2</u>. Suppose L_1 and L_2 are subgroups of a group L, where L_2 is normal in L. If

$$\ell = \ell_1'\ \ell_2\ \ell_1$$

for $\ell \in L$, ℓ_1', $\ell_1 \in L_1$ and $\ell_2 \in L_2$, then for some $\ell_2' \in L_2$

$$\text{i)}\ \ell = \ell_2'\ \ell_1'\ \ell_1 \tag{4.1.1}$$

and

$$\text{ii)}\ (\ell_1')^{-1}\ \ell_2'\ \ell_1' = \ell_2\ . \tag{4.1.2}$$

Proof. We first note that since L_2 is normal in L we can define $\ell_2' \in L_2$ by

$$\ell_1' \ell_2 (\ell_1')^{-1} = \ell_2'$$

so that $\ell_1' \ell_2 = \ell_2' \ell_1'$.

i) If $\ell = \ell_1' \ell_2 \ell_1 = (\ell_1' \ell_2) \ell_1$, then

$$\ell = \ell_2' \ell_1' \ell_1 .$$

ii) $(\ell_1')^{-1} \ell_2' \ell_1' = (\ell_1')^{-1} (\ell_2' \ell_1') = (\ell_1')^{-1} \ell_1' \ell_2 = \ell_2$.

Corollary to Lemma 4.1.2. If L is the semi-direct product of L_2 and L_1 and L_2 is normal in L, then for every $\ell \in L$ and $\ell_1 \in L_1$, there exist ℓ_2', $\ell_2'' \in L_2$ and $\ell_1' \in L_1$ such that

$$\text{i) } \ell_1 \ell = \ell_2' \ell_1'$$

and

$$\text{ii) } \ell = \ell_2'' \ell_1'' \text{ where } \ell_1'' = \ell_1^{-1} \ell_1'$$

Proof. The proof of i) follows immediately from the fact that L is the semi-direct product of L_2 and L_1. Part ii) follows from Theorem 4.1.2.

For any $\omega \in \Omega$, let ℓ_ω be the element in L such that

$$\omega = \ell_\omega \omega_0$$

where ω_0 is the reference point chosen in Ω. Recalling that since L is

the semi-direct product of L_2 and L_1, any $\ell \in L$ can be written

$$\ell = \ell_2 \, \ell_1$$

for some $\ell_1 \in L_1$ and $\ell_2 \in L_2$, the following theorem shows how to use the group structure to construct a function on a space Ω which indexes the orbits of L_1 in Ω, or equivalently, is a maximal invariant function under L_1.

Theorem 4.1.1. Let L_1 be a sub-isomorphic group of transformations on a space Ω. Suppose a group L_2 of transformations on Ω exists, such that if L is the semi-direct product of L_2 and L_1, then L is isomorphic to Ω and L_2 is a normal subgroup of L. Suppose

$$\omega = \ell_\omega \, \omega_0 = \ell_{2\omega} \, \ell_{1\omega} \, \omega_0$$

where $\ell_{2\omega} \in L_2$ and $\ell_{1\omega} \in L_1$. Then

$$\psi(\omega) = \ell_{1\omega}^{-1} \, \omega = \ell_{1\omega}^{-1} \, \ell_{2\omega} \, \ell_{1\omega} \, \omega_0$$

indexes the orbits of L_1 in Ω.

Proof. Let $\omega' = \ell_1' \, \omega$ for some $\ell_1 \in L_1$. Then

$$\omega' = \ell_1' \, \ell_{2\omega} \, \ell_{1\omega} \, \omega_0 = \ell_{2\omega}' \, \ell_{1\omega}'' \, \omega_0$$

for some $\ell_{2\omega}' \in L_2$ and some $\ell_{1\omega}'' \in L_1$. Then by definition,

$$\psi(\omega') = (\ell_{1\omega}'')^{-1} \, \ell_{2\omega}' \, \ell_{1\omega}'' \, \omega_0$$

$$= (\ell_1' \, \ell_{1\omega})^{-1} \, \ell_{2\omega}' \, (\ell_1' \, \ell_{1\omega}) \, \omega_0 \, ,$$

by part ii) of the Corollary to Lemma 4.1.2. Then

$$\psi(\omega') = \ell_{1\omega}^{-1} (\ell_1')^{-1} \ell_{2\omega}' \ell_1' \ell_{1\omega} \omega_0$$

$$= \ell_{1\omega}^{-1} \ell_{2\omega} \ell_{1\omega} \omega_0 ,$$

by part ii) of Lemma 4.1.2. Thus we have that

$$\psi(\omega') = \ell_{1\omega}^{-1} \ell_{2\omega} \ell_{1\omega} \omega_0 = \psi(\omega) ,$$

and ψ is constant on orbits.

Now let $\phi(\omega)$ be some other invariant function. Since $\phi(\omega)$ is invariant, $\phi(\omega) = \phi(\omega')$ where $\omega' = \ell_1\omega$ for some $\ell_1 \in L_1$. Since $(\ell_{1\omega})^{-1} \in L_1$,

$$\phi(\omega) = \phi[(\ell_{1\omega})^{-1} \omega] = \phi[\psi(\omega)] ,$$

so that any other function constant on the orbits of Ω is a function of ψ. Thus ψ indexes the orbits of Ω.

We will now apply Theorem 4.1.1 to the estimation problem defined in Chapter III.

Section 4.2. A Maximal Invariant Parametric Function

Applying the results of Theorem 4.1.1 to the estimation problem defined in Chapter III, we have immediately a corollary to the theorem.

Corollary to Theorem 4.1.1. Let $\bar{H}_1$ be a sub-isomorphic group of transformations on the parameter space Θ. Suppose a group $\bar{H}_2$ of transformations on Θ exists, such that if $\bar{H}$ is the semi-direct product of $\bar{H}_2$ and $\bar{H}_1$, then $\bar{H}$ is a group isomorphic to Θ and $\bar{H}_2$ is a normal subgroup of $\bar{H}$. Then

$$\psi(\theta) = \bar{h}_{1\theta}^{-1}\ \theta = \bar{h}_{1\theta}^{-1}\ \bar{h}_{2\theta}\ \bar{h}_{1\theta}\ \theta_0$$

indexes the orbits of $\bar{H}_1$ in Θ. We call $\psi(\theta)$ a <u>maximal invariant parametric function</u>.

Let us return to the example of estimating (μ,σ^2) from a normal distribution to show how Theorem 4.1.1 is applied.

<u>Lemma 4.2.1</u>. The group $\bar{H}_2$ of location changes is a normal subgroup of the group $\bar{H}$ of location and scale changes.

<u>Proof</u>. $\bar{h}\ \bar{h}_2\ \bar{h}^{-1} = [c,\lambda^2]\ [c',1]\ [c,\lambda^2]^{-1}$

$$= [c,\lambda^2]\ [c',1]\ [-c\lambda^{-1},\ \lambda^{-2}]$$

$$= [c,\lambda^2]\ [-c\lambda^{-1} + c',\ \lambda^{-2}]$$

$$= [\lambda(-c\lambda^{-1} + c) + c',\ \lambda^2\lambda^{-2}]$$

$$= [c',1] \in \bar{H}_2 .$$

We can apply the Corollary to Theorem 4.1.1 to obtain an index of the orbits of $\bar{H}_1$ (the group of scale changes in Θ. Since

$$\theta = (\mu,\sigma^2) = [\mu,\sigma^2]\ (0,1)$$

$$= [\mu,1]\ [0,\sigma^2]\ (0,1)\ ,$$

we have

$$\psi(\theta) = [0,\sigma^2]^{-1}\ [\mu,1]\ [0,\sigma^2]\ (0,1)$$

$$= [0,\sigma^{-2}]\ (\mu,\sigma^2)$$

$$= (\sigma^{-1}\mu,\ 1)$$

$$= (\mu/\sigma, 1)\ ,$$

and thus $\psi(\theta)$ indexes the orbits of $\bar{H}_1$ in Θ. Now the problem of estimating the orbits is equivalent to the problem of estimating μ/σ. Note, however, that Theorem 4.1.1 does not apply to the problem of estimating the orbits of $\bar{H}_2$ since

<u>Lemma 4.2.2.</u> $\bar{H}_1$ is not a normal subgroup of $\bar{H}$.

<u>Proof.</u> $\bar{h}\ \bar{h}_1\ \bar{h}^{-1} = [c,\lambda^2]\ [0,\lambda_1^2]\ [c,\lambda^2]^{-1}$

$$= [c,\lambda^2]\ [0,\lambda_1^2]\ [-c\lambda^{-1},\ \lambda^{-2}]$$

$$= [c,\lambda^2]\ [-c\lambda_1\lambda^{-1},\ \lambda_1^2\lambda^{-2}]$$

$$= [\lambda(-c\lambda_1\lambda^{-1}) + c,\ \lambda^2\lambda_1^2\lambda^{-2}]$$

$$= [-c\lambda_1 + c,\ \lambda_1^2] \notin \bar{H}_1$$

Since the orbits of $\bar{H}_2$ in Θ are indexed by σ^2, as was seen to be the case in Section 4.1, the proposed procedure only applies to certain problems - those in which a particular subgroup ($\bar{H}_2$) is normal.

<u>Section 4.3. An Invariant Sufficient Estimator of the Index of Orbits</u>

The problem of estimating the orbit of $\bar{H}_1$ in Θ in which θ lies is now reduced to the problem of estimating the index of the orbit, a maximal invariant parametric function. In the estimation problem defined in Chapter III, the group G is isomorphic to the sufficient statistic

space, X, so we can obtain an invariant sufficient estimator of $\psi(\theta)$ by forming a function on X in the same way as $\psi(\theta)$ was formed on Θ.

The Stein theorem, as presented by Hall, Wijsman and Ghosh (1965), stated below without proof, assures us that this will be the same estimator we would obtain if we first reduced the sample space by invariance and then further reduced it by sufficiency. We will state only the portions of the Stein theorem relevant to the problems under consideration.

First, we must present some preliminaries relevant to the Stein theorem. Denoting the sample space as

$$(X_0, A, P)$$

where X_0 is the space of observations z with probability distribution P and σ-algebra A, the sufficient statistic space can be represented

$$(X, A_s, P_s)$$

where A_s and P_s are the σ-algebra and probability measure induced on X by the sufficient statistic x. The reduction of the sample space by invariance yields

$$(X_0/H_1^*, A(H_1^*), P(H_1^*))$$

where X_0/H_1^* is the space of orbits of H_1^* in X_0 and $A(H_1^*)$ and $P(H_1^*)$ are the induced σ-algebra and probability measure.

If we reduce the sample space by sufficiency and then reduce the sufficient space by invariance, then we obtain the space of <u>invariant sufficient</u> statistics

$$(X/H_1,\ A_s(H_1),\ P_s(H_1))\ .$$

If we reduce the sample space first by invariance and then by sufficiency, we obtain the space of <u>sufficient</u> <u>invariant</u> statistics

$$((X_0/H_1^*)_s,\ (A(H_1^*))_s,\ (P(H_1^*))_s)\ .$$

<u>Lemma 4.3.1.</u> (Stein) Suppose $U(x)$ is a measurable function which indexes orbits of X/H_1 and $V(x)$ is a measurable function which indexes orbits of $(X_0/H_1^*)_s$. Then $U(x) = V(x)$ if $h(A_s) = A_s$ for all $h \in H_1$ (that is, if the group preserves the measurability of functions).

The following Lemma is well-known.

<u>Lemma 4.3.2.</u> If a space (Y, B, P) is isomorphic to a subspace of R^n with the Borel σ-algebra, then the σ-algebra is preserved by nonsingular transformations and location changes.

The above lemma insures us that Lemma 4.3.1 will apply to our problems.

The importance of Lemma 4.3.1 is in that it is easier to attack a problem by first applying sufficiency and then invariance than the other way around; however, the more natural approach is to reduce by invariance and then by sufficiency. Lemma 4.3.1 implies that both approaches yield the same result, so we will take the former approach for the sake of simplicity.

<u>Theorem 4.3.1.</u> If we consider the estimation problem defined in

Chapter III with the group of location and scale changes isomorphic to the sufficient statistic space, then an invariant sufficient estimator is a sufficient invariant estimator.

Let x_0 be the reference point chosen in $\mathcal{X}$, and let $h_x \in H$ be such that

$$x = h_x x_0 .$$

Recall that $\bar{H}_1$ and $\bar{H}_2$ induce groups H_1 and H_2 on $\mathcal{X}$, so that any $x \in \mathcal{X}$ can be represented as

$$x = h_x x_0 = h_{2x} h_{1x} x_0 .$$

<u>Theorem 4.3.2.</u> If H_2 is a normal subgroup of H, then

$$T(x) = h_{1x}^{-1} x = h_{1x}^{-1} h_{2x} h_{1x} x_0$$

is an invariant sufficient estimator of $\psi(\theta)$; equivalently, $T(x)$ is an invariant sufficient estimator of the orbits of $\bar{H}_1$ in Θ.

<u>Proof.</u> Since x is sufficient, $T(x)$ is a function of the sufficient statistic. That $T(x)$ is invariant can be demonstrated in the same way as $\psi(\theta)$ was shown to be invariant in Theorem 4.1.1, using results analagous to those of Lemma 4.1.2. Thus $T(x)$ can be said to index the orbits of H_1 in $\mathcal{X}$, and then due to the isomorphism between H and $\bar{H}$, $T(x)$ can be used as an estimator of $\psi(\theta)$.

Again referring to the example of estimating (μ,σ^2) from a normal distribution, the sufficient statistic $(\bar{x},s^2)$ can be written

$$(\bar{x}, s^2) = [\bar{x}, s^2]\ (0,1)$$

$$= [\bar{x}, 1]\ [0, s^2]\ (0,1)\ .$$

An invariant sufficient estimator of the orbits of $\bar{H}_1$ is then

$$T(x) = [0, s^2]^{-1}\ [\bar{x}, 1]\ [0, s^2]\ (0,1)$$

$$= [0, s^{-2}]\ (\bar{x}, s^2)$$

$$= (s^{-1}\bar{x},\ 1)$$

$$= (\bar{x}/s,\ 1)\ .$$

Since $T(x) = (\bar{x}/s)$ estimates the orbits of $\bar{H}_1$ in Θ, we can say that the invariant sufficient statistic $\bar{x}/s$ estimates the index of orbits, μ/σ.

Note that $(\bar{x}/s)^{-1} = s/\bar{x}$ is the <u>coefficient of variation</u> found in elementary statistics texts. Therefore $s/\bar{x}$ is an invariant sufficient estimator of the orbits of the scale transformation group, and thus of σ/μ, the population coefficient of variation, which indexes the orbits.

CHAPTER V

APPLICATION TO THE LINEAR MODEL: NONSTOCHASTIC INPUT MATRIX

In this chapter the method of invariant estimation developed in Chapter IV is used to obtain invariant estimators of the parameters in a linear model.

Consider the model

$$\underline{y} = \underline{Z}\,\underline{\beta} + \underline{\varepsilon}\ ,$$

where $\underline{y}$ and $\underline{\varepsilon}$ are $n \times 1$ vectors of observable and nonobservable random variables, respectively; $\underline{Z} = [\underline{1}, \underline{z}_1, \ldots, \underline{z}_k]$ is an $n \times p$ $(p = k + 1)$ nonstochastic input matrix $(n > p)$ of full rank, and $\underline{\beta}$ is a $p \times 1$ vector $[\alpha, \beta_1, \ldots, \beta_k]'$ of unknown parameters to be estimated. We assume that the n values of each of $z_1, \ldots, z_k$ have been expressed in deviation form and make the usual assumptions that $E(\underline{\varepsilon}) = \underline{0}$ and $E(\underline{\varepsilon}\,\underline{\varepsilon}') = \sigma^2\,\underline{I}$.

Suppose that the k input variables $z_1, \ldots, z_k$ and the independent variable y have arbitrary units of measurement. (This assumption is quite common in many areas of research, especially in the social sciences when dealing with such measures as attitudinal or belief intensity scales.) The problem is to estimate $\underline{\beta}$ in a manner invariant under scale changes on $z_1, \ldots, z_k$ and y.

In order to apply the methods of Chapter IV, the appropriate group structure must be defined. Although in the nonstochastic case of the linear model, y alone is usually considered to be the random variable, to apply our procedure scale changes on $z_1, \ldots, z_k$ must be expressed as a group of operators on the sample space. Thus we define the estimation problem as follows.

Let $\mathcal{D}(\underline{Z}_0)$ be the space of all $n \times p$ matrices such that if $\underline{Z} \in \mathcal{D}(\underline{Z}_0)$, then

$$\underline{Z} = \underline{Z}_0 \, \underline{D}$$

for some positive definite diagonal matrix $\underline{D}$ and fixed $\underline{Z}_0$, where $\underline{Z}_0 = [\underline{1}, \underline{z}_{01}, \ldots, \underline{z}_{0k}]$ is such that

$$\underline{Z}_0' \, \underline{Z}_0 = \underline{I} \; .$$

Since $\underline{Z}$ is nonstochastic, we may treat it as a random variable with a one-point distribution in $\mathcal{D}(\underline{Z}_0)$. That is, we assume that there exists a $\underline{D}^*$ (diagonal) such that

$$P(\underline{Z} = \underline{Z}_0 \, \underline{D}^*) = 1 \; .$$

Note that the leading element of $\underline{D}^*$ must be a one. The sample space is then defined as

$$\mathcal{X}_0 = \mathcal{D}(\underline{Z}_0) \times R^n \; ,$$

where

$$z = (\underline{Z}, \underline{y})$$

is a point in $\mathcal{X}_0$.

Consider the class of transformations H^* defined on $\mathcal{X}_0$ and containing elements $[\underline{\Lambda}, \underline{c}, \gamma^2]$, where $\underline{\Lambda}$ is a $p \times p$ positive definite diagonal matrix $\underline{\Lambda} = \mathrm{diag}(1, \lambda_1, \ldots, \lambda_k)$, $\underline{c}$ is the $p \times 1$ vector $[c_1, \ldots, c_p]'$, and γ^2 is a positive scalar. The operation of H^* on $\mathcal{X}_0$ is defined by

$$[\underline{\Lambda}, \underline{c}, \gamma^2] \; (\underline{Z}, \underline{y}) = (\underline{Z}\,\underline{\Lambda}, \gamma\,\underline{y} + \underline{Z}\,\underline{\Lambda}\,\underline{c})$$

and the operation in H^* is defined by

$$[\underline{\Lambda}, \underline{c}, \gamma^2] \; [\underline{\Lambda}_1, \underline{c}_1, \gamma_1^2] = [\underline{\Lambda}_1\,\underline{\Lambda}, \underline{\Lambda}^{-1}\,\gamma\,\underline{c}_1 + \underline{c}, \gamma^2\,\gamma_1^2] \; .$$

<u>Lemma 5.1</u>. H^* is a group.

<u>Proof</u>. i) Since $\underline{\Lambda}_1\,\underline{\Lambda}$ is a $p \times p$ positive definite diagonal matrix with a one as its first component, $\underline{\Lambda}^{-1}\,\gamma\,\underline{c}_1 + \underline{c}$ is a $p \times 1$ vector, and $\gamma^2\,\gamma_1^2$ is a positive scalar, H^* is closed.

ii) $[\underline{\Lambda}, \underline{c}, \gamma^2] \; ([\underline{\Lambda}_1, \underline{c}_1, \gamma_1^2] \; [\underline{\Lambda}_2, \underline{c}_2, \gamma_2^2]) =$

$$= [\underline{\Lambda}, \underline{c}, \gamma^2] \; [\underline{\Lambda}_2\,\underline{\Lambda}_1, \underline{\Lambda}_1^{-1}\,\gamma_1\,\underline{c}_2 + \underline{c}_1, \gamma_1^2\,\gamma_2^2]$$

$$= [\underline{\Lambda}_2\,\underline{\Lambda}_1\,\underline{\Lambda}, \underline{\Lambda}^{-1}\,\gamma(\underline{\Lambda}_1^{-1}\,\gamma_1\,\underline{c}_2 + \underline{c}_1) + \underline{c}, \gamma^2\,\gamma_1^2\,\gamma_2^2]$$

$$= [\underline{\Lambda}_2\,\underline{\Lambda}_1\,\underline{\Lambda}, \underline{\Lambda}^{-1}\,\underline{\Lambda}_1^{-1}\,\gamma\,\gamma_1\,\underline{c}_2 + \underline{\Lambda}^{-1}\,\gamma\,\underline{c}_1 + \underline{c}, \gamma^2\,\gamma_1^2\,\gamma_2^2]$$

and

$$([\underline{\Lambda}, \underline{c}, \gamma^2] \; [\underline{\Lambda}_1, \underline{c}_1, \gamma_1^2]) \; [\underline{\Lambda}_2, \underline{c}_2, \gamma_2^2] =$$

$$= [\underline{\Lambda}_1\,\underline{\Lambda}, \underline{\Lambda}^{-1}\,\gamma\,\underline{c}_1 + \underline{c}, \gamma^2\,\gamma_1^2] \; [\underline{\Lambda}_2, \underline{c}_2, \gamma_2^2]$$

$$= [\underline{\Lambda}_2\,\underline{\Lambda}_1\,\underline{\Lambda}, (\underline{\Lambda}_1\,\underline{\Lambda})^{-1}\,\gamma\,\gamma_1\,\underline{c}_2 + \underline{\Lambda}^{-1}\,\gamma\,\underline{c}_1 + \underline{c}, \gamma^2\gamma_1^2\,\gamma_2^2]$$

$$= [\underline{\Lambda}_2\,\underline{\Lambda}_1\,\underline{\Lambda}, \underline{\Lambda}^{-1}\,\underline{\Lambda}_1^{-1}\,\gamma\,\gamma_1\,\underline{c}_2 + \underline{\Lambda}^{-1}\,\gamma\,\underline{c}_1 + \underline{c}, \gamma^2\gamma_1^2\,\gamma_2^2]$$

so the operation is associative.

iii) The identity element is $[\underline{I}, \underline{0}, 1]$ since

$$[\underline{I}, \underline{0}, 1] \; [\underline{\Lambda}, \underline{c}, \gamma^2] = [\underline{\Lambda}, \underline{c}, \gamma^2]$$

and

$$[\underline{\Lambda}, \underline{c}, \gamma^2]\ [\underline{I}, \underline{0}, 1] = [\underline{\Lambda}, \underline{c}, \gamma^2]\ .$$

iv) The inverse of $[\underline{\Lambda}, \underline{c}, \gamma^2]$ is $[\underline{\Lambda}^{-1}, -\gamma^{-1}\ \underline{\Lambda}\ \underline{c}, \gamma^{-2}]$

since

$$[\underline{\Lambda}, \underline{c}, \gamma^2]\ [\underline{\Lambda}^{-1}, -\gamma^{-1}\ \underline{\Lambda}\ \underline{c}, \gamma^{-2}] =$$

$$= [\underline{\Lambda}^{-1}\ \underline{\Lambda}, \underline{\Lambda}^{-1}\ \gamma(-\gamma^{-1}\ \underline{\Lambda}\ \underline{c}) + \underline{c}, \gamma^2\ \gamma^{-2}]$$

$$= [\underline{I}, \underline{0}, 1]$$

and

$$[\underline{\Lambda}^{-1}, -\gamma^{-1}\ \underline{\Lambda}\ \underline{c}, \gamma^{-2}]\ [\underline{\Lambda}, \underline{c}, \gamma^2] =$$

$$= [\underline{\Lambda}\ \underline{\Lambda}^{-1}, \underline{\Lambda}\ \gamma^{-1}\ \underline{c} - \gamma^{-1}\ \underline{\Lambda}\ \underline{c}, \gamma^{-2}\ \gamma^2]$$

$$= [\underline{I}, \underline{0}, 1]\ .$$

Thus H^* is a group.

The parameter space for the linear models problem is

$$\Theta = \Delta_p \times R^p \times R^+\ ,$$

where Δ_p is the space of all $p \times p$ positive definite diagonal matrices and $\theta \in \Theta$ is of the form

$$\theta = (\underline{D}_n^*, \underline{\beta}, \sigma^2)$$

where

$$\underline{D}_n^* = (n-1)^{-1/2}\ \underline{D}^* .$$

The induced group $\bar{H}$ on Θ contains elements $[\underline{\Lambda}, \underline{c}, \gamma^2]$ which act on Θ as follows:

$$[\underline{\Lambda}, \underline{c}, \gamma^2]\ (\underline{D}_n^*, \underline{\beta}, \sigma^2) = (\underline{D}_n^*\ \underline{\Lambda}, \underline{\Lambda}^{-1}\ \gamma\ \underline{\beta} + \underline{c}, \gamma^2\ \sigma^2) .$$

To see that H^* on X_0 induces the above operation on Θ, consider

$$\underline{y} = \underline{Z}\ \underline{\beta} + \underline{\varepsilon}$$

$$\gamma\ \underline{y} = \gamma\ \underline{Z}\ \underline{\beta} + \gamma\ \underline{\varepsilon}$$

$$= \gamma\ \underline{Z}\ \underline{\Lambda}\ \underline{\Lambda}^{-1}\ \underline{\beta} + \gamma\ \underline{\varepsilon}$$

$$\gamma\ \underline{y} + \underline{Z}\ \underline{\Lambda}\ \underline{c} = \gamma\ \underline{Z}\ \underline{\Lambda}\ \underline{\Lambda}^{-1}\ \underline{\beta} + \underline{Z}\ \underline{\Lambda}\ \underline{c} + \gamma\ \underline{\varepsilon}$$

$$= \underline{Z}\ \underline{\Lambda}(\gamma\ \underline{\Lambda}^{-1}\ \underline{\beta} + \underline{c}) + \gamma\ \underline{\varepsilon} ,$$

so that if $\underline{y}$ is transformed to $\gamma\ \underline{y} + \underline{Z}\ \underline{\Lambda}\ \underline{c}$, then $\underline{Z}$ is transformed to $\underline{Z}\ \underline{\Lambda}$, and since

$$\underline{Z} = \underline{Z}_0\ \underline{D}^* ,$$

$$\underline{Z}\ \underline{\Lambda} = \underline{Z}_0\ \underline{D}^*\ \underline{\Lambda}$$

so $\underline{D}^*$ is transformed to $\underline{D}^*\ \underline{\Lambda}$. Also, σ^2 is transformed to $\gamma^2\ \sigma^2$ and $\underline{\beta}$ is transformed to $\gamma\ \underline{\Lambda}^{-1}\ \underline{\beta} + \underline{c}$.

<u>Lemma 5.2</u>. If the operation on $\bar{H}$ is defined by

$$[\underline{\Lambda}, \underline{c}, \gamma^2]\ [\underline{\Lambda}_1, \underline{c}_1, \gamma_1^2] = [\underline{\Lambda}_1\ \underline{\Lambda}, \gamma\ \underline{\Lambda}^{-1}\ \underline{c}_1 + \underline{c}, \gamma^2\ \gamma_1^2] ,$$

then $\bar{H}$ is a group.

Proof. The proof is exactly the same as for Lemma 5.1.

Lemma 5.3. If we assume that $\underline{\varepsilon}$ is normally distributed, then the sufficient statistic in the linear model problem is

$$(D_n(\underline{Z}),\ \hat{\underline{\beta}},\ s^2)\ ,$$

where

$$D_n(\underline{Z}) = \mathrm{diag}((n-1)^{-1/2},\ d_1^{1/2},\ldots,d_k^{1/2})$$

and d_i is the (i+1)st diagonal element of $(n-1)^{-1}\ \underline{Z}'\underline{Z}$ (note that since values of $z_1,\ldots,z_k$ are in deviation form, $D_n(\underline{Z})$ contains standard deviations of the k input variables); $\hat{\underline{\beta}}$ is the usual least squares estimator

$$\hat{\underline{\beta}} = (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\underline{y}\ ;$$

and $s^2 = s^2_{y\cdot\underline{Z}} = s^2_{\varepsilon}$ is the residual or error mean square,

$$s^2 = (\underline{y} - \underline{Z}\ \hat{\underline{\beta}})'\ (\underline{y} - \underline{Z}\ \hat{\underline{\beta}})/(n-p)\ .$$

Proof. It is well known that for the case in which $\underline{Z}$ is a fixed input matrix, $(\hat{\underline{\beta}},\ s^2)$ is jointly sufficient for $(\underline{\beta},\ \sigma^2)$ (see Graybill (1961), for example). Thus the conditional frequency distribution

$$g(z_1,\ldots,z_k,\ y|\hat{\underline{\beta}},\ s^2)$$

is independent of $(\underline{\beta},\ \sigma^2)$. Since $\underline{Z}$ and thus $D_n(\underline{Z})$ has a one point distribution, it is immediate that the conditional density

$$g[(z_1,\ldots,z_k,\ y|\hat{\underline{\beta}},\ s^2)|D_n(\underline{Z})] = g(z_1,\ldots,z_k,\ y|\hat{\underline{\beta}},\ s^2,\ D_n(Z))$$

is independent of $(\underline{D}_n^*,\ \underline{\beta},\ \sigma^2)$, so that $(D_n(\underline{Z}),\ \hat{\underline{\beta}},\ s^2)$ is sufficient for

$(\underline{D}_n^*, \underline{\beta}, \sigma^2)$.

The class of operators induced on X by H^* and $\bar{H}$ is

$$H = \{[\underline{\Lambda}, \underline{c}, \sigma^2]\}$$

where the operation in H is defined as were the operations in H^* and $\bar{H}$, and where

$$[\underline{\Lambda}, \underline{c}, \gamma^2]\ (D_n(\underline{Z}), \hat{\underline{\beta}}, s^2) = (D_n(\underline{Z})\ \underline{\Lambda}, \gamma\ \underline{\Lambda}^{-1}\ \hat{\underline{\beta}} + \underline{c}, \gamma^2\ s^2)\ .$$

It is immediate that H is a group of transformations on X.

Although the group structure we have introduced appears quite cumbersome, it is motivated by the following lemma.

<u>Lemma 5.4.</u> H is isomorphic to X, $\bar{H}$ is isomorphic to Θ, and H and $\bar{H}$ are isomorphic.

<u>Proof.</u> The isomorphism between H and $\bar{H}$ is immediate since the elements and operations in the two groups are identical. To show that $\bar{H}$ is isomorphic to Θ and that H is isomorphic to X, $(\underline{I}, \underline{0}, 1)$ is chosen as a reference point in both Θ and X. Then any point $(\underline{D}_n^*, \underline{\beta}, \sigma^2) \in \Theta$ can be written uniquely as

$$[\underline{D}_n^*, \underline{\beta}, \sigma^2]\ (\underline{I}, \underline{0}, 1) = (\underline{D}_n^*, \underline{\beta}, \sigma^2)$$

for $[\underline{D}_n^*, \underline{\beta}, \sigma^2] \in \bar{H}$; and any point $(D_n(\underline{Z}), \hat{\underline{\beta}}, s^2) \in X$ can be written uniquely as

$$[D_n(\underline{Z}), \hat{\underline{\beta}}, s^2]\ (\underline{I}, \underline{0}, 1) = (D_n(\underline{Z}), \hat{\underline{\beta}}, s^2)$$

for $[D_n(\underline{Z}), \hat{\underline{\beta}}, s^2] \in H$. Thus $\bar{H}$ is isomorphic to Θ and H is isomorphic

to X.

We now consider the class G_1^* of scale changes on $X_0 = \{(\underline{Z}, \underline{y})\}$, containing elements $[\underline{\Lambda}, \gamma^2]_1$, where

$$[\underline{\Lambda}, \gamma^2]_1 \ (\underline{Z}, \underline{y}) = (\underline{Z}\ \underline{\Lambda}, \gamma\ \underline{y}) \ .$$

<u>Lemma 5.5</u>. G_1^* is a group under the operation defined by

$$[\underline{\Lambda}, \gamma^2]_1 \ [\underline{\Lambda}_1, \gamma_1^2]_1 = [\underline{\Lambda}_1\ \underline{\Lambda}, \gamma^2\ \gamma_1^2]_1 \ .$$

<u>Proof</u>. Closure and associativity are immediate. The identity is $[\underline{I}, 1]_1$, and $[\underline{\Lambda}, \gamma^2]_1^{-1} = [\underline{\Lambda}^{-1}, \gamma^{-2}]_1$.

Another class G_2^* of transformations on X_0 exists, where G_2^* is the class of location changes containing elements $[\underline{c}]_2$, where

$$[\underline{c}]_2 \ (\underline{Z}, \underline{y}) = (\underline{Z}, \underline{y} + \underline{c}) \ .$$

<u>Lemma 5.6</u>. If the operation in G_2^* is defined by

$$[\underline{c}]_2 \ [\underline{c}']_2 = [\underline{c} + \underline{c}']_2 \ ,$$

then G_2^* is a group.

<u>Proof</u>. G_2^* is closed and associative with identity element $[\underline{0}]_2$ and inverse $[-\underline{c}]_2$.

<u>Lemma 5.7</u>. The set G^* of ordered pairs

$$[[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]$$

with the operation

$$[[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]\ [[\underline{c}']_2, [\underline{\Lambda}', \gamma'^2]_1] =$$

$$= [[\underline{\Lambda}^{-1}\ \gamma\ \underline{c}' + \underline{c}]_2, [\underline{\Lambda}'\ \underline{\Lambda}, \gamma^2\ \gamma'^2]_1]$$

is a group of transformations isomorphic to $\bar{H}$, and thus also isomorphic to Θ.

<u>Proof</u>. i) Closure is apparent upon inspection.

ii) $[[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]([[\underline{c}']_2, [\underline{\Lambda}', \gamma'^2]_1][[c'']_2, [\underline{\Lambda}'', \gamma''^2]_1])$

$$= [[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]\ [[\underline{\Lambda}'^{-1}\ \gamma'\ \underline{c}'' + \underline{c}']_2, [\underline{\Lambda}''\underline{\Lambda}', \gamma'^2\gamma''^2]_1]$$

$$= [[\underline{\Lambda}^{-1}\underline{\Lambda}'^{-1}\gamma\ \gamma'\ \underline{c}'' + \underline{\Lambda}^{-1}\ \gamma\ \underline{c}' + \underline{c}]_2, [\underline{\Lambda}''\underline{\Lambda}'\underline{\Lambda}, \gamma^2\gamma'^2\gamma''^2]_1]$$

and

$$([[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]\ [[\underline{c}']_2, [\underline{\Lambda}', \gamma'^2]_1])[[\underline{c}'']_2, [\underline{\Lambda}'', \gamma''^2]_1]$$

$$= [[\underline{\Lambda}^{-1}\ \gamma\ \underline{c}' + \underline{c}]_2, [\underline{\Lambda}'\underline{\Lambda}, \gamma^2\gamma'^2]_1][[\underline{c}'']_2, [\underline{\Lambda}'', \gamma''^2]_1]$$

$$= [[\underline{\Lambda}^{-1}\underline{\Lambda}'^{-1}\ \gamma\ \gamma'\ \underline{c}'' + \underline{\Lambda}^{-1}\ \gamma\ \underline{c}' + \underline{c}]_2,$$

$$[\underline{\Lambda}''\ \underline{\Lambda}'\ \underline{\Lambda}, \gamma^2\ \gamma'^2\ \gamma''^2]_1]$$

so the operation is associative.

iii) The identity element is $[[\underline{0}]_2, [\underline{I}, 1]_1]$ since

$$[[\underline{0}]_2, [\underline{I}, 1]_1]\ [[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1] = [[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]$$

and

$$[[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]\ [[\underline{0}]_2, [\underline{I}, 1]_1] = [[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]\ .$$

iv) The inverse of $[[\underline{c}]_2, [\underline{\Lambda}, \gamma^2]_1]$ is

$$[[-\gamma^{-1}\,\underline{\Lambda}\,\underline{c}]_2,\ [\underline{\Lambda}^{-1},\ \gamma^{-2}]_1]$$

since

$$[[-\gamma^{-1}\,\underline{\Lambda}\,\underline{c}]_2,\ [\underline{\Lambda}^{-1},\ \gamma^{-2}]_1]\ [[\underline{c}]_2,\ [\underline{\Lambda},\ \gamma^2]_1] =$$

$$= [[\underline{0}]_2,\ [\underline{I},\ 1]_1]$$

and

$$[[\underline{c}]_2,\ [\underline{\Lambda},\ \gamma^2]_1]\ [[-\gamma^{-1}\,\underline{\Lambda}\,\underline{c}]_2,\ [\underline{\Lambda}^{-1},\ \gamma^{-2}]_1] =$$

$$= [[\underline{0}]_2,\ [\underline{I},\ 1]_1]\ .$$

Thus G^* is a group. To show that G^* is isomorphic to $\bar{H}$, we need only identify the element $[[\underline{c}]_2,\ [\underline{\Lambda},\ \gamma^2]_1] \in G^*$ with the element $[\underline{\Lambda},\ \underline{c},\ \gamma^2] \in \bar{H}$.

<u>Lemma 5.8</u>. If $\bar{H}_1$ contains elements of $\bar{H}$ of the form $[\underline{\Lambda},\ \underline{0},\ \gamma^2]$ and $\bar{H}_2$ contains elements of $\bar{H}$ of the form $[\underline{I},\ \underline{c},\ 1]$, then G_1^* and $\bar{H}_1$ are isomorphic, G_2^* and $\bar{H}_2$ are isomorphic, and $\bar{H}_1$ and $\bar{H}_2$ are subgroups of $\bar{H}$.

<u>Proof</u>. i) If we identify $[\underline{\Lambda},\ \gamma^2]_1 \in G_1^*$ with $[\underline{\Lambda},\ \underline{0},\ \gamma^2] \in \bar{H}_1$ and $[\underline{c}]_2 \in G_2^*$ with $[\underline{I},\ \underline{c},\ 1] \in \bar{H}_2$, then it is apparent that G_1^* is isomorphic to $\bar{H}_1$ and G_2^* is isomorphic to $\bar{H}_2$. ii) First consider $\bar{H}_1$. For $[\underline{\Lambda},\ \underline{0},\ \gamma^2]$ and $[\underline{\Lambda}_1,\ \underline{0},\ \gamma_1^2] \in \bar{H}_1$,

$$[\underline{\Lambda},\ \underline{0},\ \gamma^2]\ [\underline{\Lambda}_1,\ \underline{0},\ \gamma_1^2] = [\underline{\Lambda}_1\,\underline{\Lambda},\ \underline{0},\ \gamma^2\,\gamma_1^2] \in \bar{H}_1$$

so $\bar{H}_1$ is closed. Also,

$$[\underline{\Lambda},\ \underline{0},\ \gamma^2]^{-1} = [\underline{\Lambda}^{-1},\ \underline{0},\ \gamma^{-2}] \in \bar{H}_1$$

so that $\bar{H}_1$ contains identities and by Lemma 2.1, $\bar{H}_1$ is a subgroup of $\bar{H}$.

iii) For $[\underline{I}, \underline{c}, 1]$ and $[\underline{I}, \underline{c}_1, 1] \in \bar{H}_2$,

$$[\underline{I}, \underline{c}, 1]\ [\underline{I}, \underline{c}_1, 1] = [\underline{I}, \underline{c}_1 + \underline{c}, 1] \in \bar{H}_2$$

so $\bar{H}_2$ is closed. Also,

$$[\underline{I}, \underline{c}, 1]^{-1} = [\underline{I}, -\underline{c}, 1] \in \bar{H}_2$$

so $\bar{H}_2$ contains inverses and is therefore a subgroup of $\bar{H}$.

The problem is to estimate orbits of $\bar{H}_1$ in Θ.

<u>Lemma 5.9</u>. $\bar{H}_2$ is a normal subgroup of $\bar{H}$.

<u>Proof</u>. $\bar{h}\ \bar{h}_2\ \bar{h}^{-1} = [\underline{\Lambda}, \underline{c}, \gamma^2]\ [\underline{I}, \underline{c}', 1]\ [\underline{\Lambda}, \underline{c}, \gamma^2]^{-1}$

$$= [\underline{\Lambda}, \underline{c}, \gamma^2]\ [\underline{I}, \underline{c}', 1]\ [\underline{\Lambda}^{-1}, -\gamma^{-1}\ \underline{\Lambda}\ \underline{c}, \gamma^{-2}]$$

$$= [\underline{\Lambda}, \underline{c}, \gamma^2]\ [\underline{\Lambda}^{-1}, -\gamma^{-1}\ \underline{\Lambda}\ \underline{c} + \underline{c}', \gamma^{-2}]$$

$$= [\underline{I}, \underline{\Lambda}^{-1}\ \gamma\ \underline{c}', 1] \in \bar{H}_2$$

and $\bar{H}_2$ is a normal subgroup of $\bar{H}$.

Lemma 5.9 allows us to obtain an index of the orbits of $\bar{H}_1$ in Θ.

<u>Theorem 5.1</u>. In the linear model estimation problem,

$$\Psi(\theta) = \bar{h}_{1\theta}^{-1}\ \bar{h}_{2\theta}\ \bar{h}_{1\theta}\ \theta_0 = [\underline{I}, \underline{D}_n^*\ \underline{\beta}/\sigma, 1]$$

is a maximal invariant parametric function.

<u>Proof</u>. Using the result of the Corollary to Theorem 4.1.1,

$$\Psi(\theta) = \bar{h}_{1\theta}^{-1}\ \bar{h}_{2\theta}\ \bar{h}_{1\theta}\ \theta_0 =$$

$$= [\underline{D}_n^*,\ \underline{0},\ \sigma^2]^{-1}\ [\underline{I},\ \underline{\beta},\ 1]\ [\underline{D}_n^*,\ \underline{0},\ \sigma^2]\ (\underline{I},\ \underline{0},\ 1)$$

$$= [\underline{D}_n^{*-1},\ \underline{0},\ \sigma^{-2}]\ (\underline{D}_n^*,\ \underline{\beta},\ \sigma^2)$$

$$= (\underline{I},\ \underline{D}_n^*\ \underline{\beta}/\sigma,\ 1)\ .$$

Thus the parametric function

$$\underline{B} = \underline{D}_n^*\ \underline{\beta}/\sigma$$

indexes the cosets of $\bar{H}_1$ in Θ, and the problem of estimating $(\underline{D}_n^*,\ \underline{\beta},\ \sigma^2)$ invariantly under scale changes is equivalent to the problem of estimating $\underline{B}$.

Note that $\underline{D}_n^*\ \underline{\beta}/\sigma$ is a parametric counterpart of the "standardized regression coefficient" or "beta coefficient" widely used in social science research.

To obtain an invariant sufficient estimator of $\underline{B}$, we compute $T(x)$ as indicated by Theorem 4.3.2. The estimator we will obtain is not, however, the estimator commonly used by social scientists.

<u>Theorem 5.2</u>. In the linear model problem,

$$T(x) = (\underline{I},\ D_n(\underline{Z})\ \hat{\underline{\beta}}/s,\ 1)\ .$$

<u>Proof</u>. $T(x) = [D_n(\underline{Z}),\ \underline{0},\ s^2]^{-1}[\underline{I},\ \hat{\underline{\beta}},\ 1][D_n(\underline{Z}),\ \underline{0},\ s^2]\ (\underline{I},\ \underline{0},\ 1)$

$$= [D_n^{-1}(\underline{Z}),\ \underline{0},\ s^{-2}]\ (D_n(\underline{Z}),\ \hat{\underline{\beta}},\ s^2)$$

$$= (\underline{I},\ D_n(\underline{Z})\ \hat{\underline{\beta}}/s,\ 1)\ .$$

Thus, by Theorem 4.3.2,

$$\underline{\tilde{B}} = D_n(\underline{Z})\ \underline{\hat{\beta}}/s$$

is an invariant sufficient estimator of $\underline{B}$.

Note that the usual social science estimator of $\underline{B}$ uses s_y, the marginal standard deviation of y, in the denominator, rather than the residual standard deviation. Chapter VI will demonstrate that $\underline{\tilde{B}}$ is a reasonably good estimator of $\underline{B}$ and is in fact superior to the usual social science estimator according to several criteria.

CHAPTER VI

COMPARISON OF THE ESTIMATORS OF THE MAXIMAL INVARIANT PARAMETRIC FUNCTION

In this chapter we consider the properties of the proposed function

$$\underline{\tilde{B}} = D_n(\underline{Z})\ \underline{\hat{\beta}}/s$$

as an estimator of the maximal invariant parametric function

$$\underline{B} = \underline{D}_n^*\ \underline{\beta}/\sigma$$

and will compare $\underline{\tilde{B}}$ to the usual social science estimator of $\underline{B}$. We will show that the invariant sufficient estimator $\underline{\tilde{B}}$ of the maximal invariant parametric function $\underline{B}$ is consistent, has small bias, and has a relatively simple distribution. We will also demonstrate that the social science estimator possesses none of these properties. Throughout this chapter we assume that the error component $\underline{\varepsilon}$ follows a normal distribution.

Section 6.1. Properties of $\underline{\tilde{B}}$

Definition 6.1.1. If θ_n is a parameter which depends on n and θ_n tends to a constant α as $n \to \infty$, then T_n will be called a consistent estimator of θ_n if T_n converges to α (in probability).

Theorem 6.1.1. $\underline{\tilde{B}}$ is a consistent estimator of $\underline{B}$.

Proof. Assuming $\underline{D}_n^*$ is chosen so that it converges to some fixed diagonal matrix $\underline{\Delta}$, we must show that $\underline{\tilde{B}}$ converges to $\underline{\Delta}\ \underline{\beta}/\sigma$ (in probability). If $\underline{D}_n^*$ converges to $\underline{\Delta}$, then $D_n(\underline{Z})$ also converges to $\underline{\Delta}$ (in probability). Since s^2 converges to σ^2, Slutsky's theorem (Cramér,

1946) assures us that s converges to σ; and since $\hat{\underline{\beta}}$ converges to $\underline{\beta}$, we have (again by Slutsky's theorem), that $\tilde{\underline{B}}$ converges to $\underline{\Delta}\ \underline{\beta}/\sigma$ (in probability).

Lemma 6.1.1. In the rank two (k=1) linear model

$$y = \alpha + \beta X + \varepsilon$$

the random variable

$$(n-1)^{1/2}\ \tilde{B} = (n-1)^{1/2}\ s_Z\ \hat{\beta}/s\ ,$$

where

$$s_Z^2 = \sum_{i=1}^{n} (z_i - \bar{z})^2/(n-1)$$

and

$$s^2 = \sum_{i=1}^{n} (y_i - \hat{\alpha} - \hat{\beta}\ z_i)^2/(n-2)$$

is distributed as a noncentral t random variable with n-2 degrees of freedom and noncentrality parameter

$$\delta = (n-1)^{1/2}\ s_Z\ \beta/\sigma\ .$$

Proof. Under normal theory assumptions,

$$\hat{\beta} \sim N(\beta,\ \sigma^2/(n-1)\ s_Z^2)\ ,$$

and thus

$$T_1 = \frac{(n-1)^{1/2}\ s_Z\ \hat{\beta}}{\sigma} \sim N\left(\frac{(n-1)^{1/2}\ s_Z\ \beta}{\sigma}\ ,\ 1\right)$$

Similarly

$$T_2 = (n-2)\ s^2/\sigma^2 \sim \chi^2_{n-2}\ .$$

Since T_1 and T_2 are independent (Graybill (1961))

$$\frac{(n-1)^{1/2}\ s_Z\ \hat{\beta}/\sigma}{\left[(n-2)\ s^2/(n-2)\ \sigma^2\right]^{1/2}} = \frac{(n-1)^{1/2}\ s_Z\ \hat{\beta}}{s}$$

is distributed as a noncentral t random variable with n-2 degrees of freedom and noncentrality parameter $\delta = (n-1)^{1/2}\ s_Z\ \beta/\sigma$.

<u>Corollary to Lemma 6.1.1</u>. For the rank two case, the random variable $\tilde{B}$ is distributed as $(n-1)^{-1/2}$ times a noncentral t random variable with n-2 degrees of freedom and noncentrality parameter $\delta = (n-1)^{1/2}\ s_Z\ \beta/\sigma$.

<u>Proof</u>. Let $t = (n-1)^{1/2}\ \tilde{B}$. By Lemma 6.1.1, t follows a t distribution with n-2 degrees of freedom and noncentrality parameter δ. Since $\tilde{B} = (n-1)^{-1/2}\ t$, the proof is immediate.

<u>Lemma 6.1.2</u>. For the $k > 1$ case,

$$T_0 = (n-1)^{1/2}\ \underline{D}_0\ \underline{\hat{\beta}}/s$$

$$= [(n-1)^{1/2}\ s_{00}^{1/2}\ \hat{\alpha}/s,\ (n-1)^{1/2}\ s_{11}^{1/2}\ \hat{\beta}_1/s, \ldots, (n-1)^{1/2}\ s_{kk}^{1/2}\ \hat{\beta}_k/s]'$$

follows a multivariate noncentral t distribution with n-p degrees of freedom and noncentrality parameter $\underline{\delta}$, where

$$\underline{D}_0 = \mathrm{diag}(s_{00}^{1/2},\ s_{11}^{1/2}, \ldots, s_{kk}^{1/2})$$

and $(1/s_{ij})$ is the (i, j) element of $(n-1)\ (\underline{Z}'\underline{Z})^{-1}$ for i, j = 0,1,...,k; and

$$\underline{\delta} = [(n-1)^{1/2}\ s_{00}^{1/2}\ \alpha/\sigma,\ (n-1)^{1/2}\ s_{11}^{1/2}\ \beta_1/\sigma,\ldots,(n-1)^{1/2}\ s_{kk}^{1/2}\ \beta_k/\sigma]'\ .$$

Proof. The proof follows from the fact that (Graybill, 1961)

$$u_i = (n-1)^{1/2}\ s_{ii}^{1/2}\ \hat{\beta}_i/s$$

(calling $\hat{\beta}_0 = \hat{\alpha}$) has a marginal noncentral t distribution for $i = 1,\ldots,p$.

Theorem 6.1.2. $\tilde{\underline{B}} = D_n(\underline{Z})\ \hat{\underline{\beta}}/s$ is distributed as a multivariate noncentral t random variable with n-p degrees of freedom and noncentrality parameter $\underline{\delta}$ (where $\underline{\delta}$ is defined as in Lemma 6.1.2), up to the scale factor $(n-1)^{1/2}\ \underline{D}_0\ D_n^{-1}\ (\underline{Z})$.

Proof. The proof follows immediately from Lemma 6.1.2.

To obtain an expression for the density of T_0 consider a nonsingular matrix $\underline{A}$ such that

$$\underline{A}'^{-1}\underline{Z}'\underline{Z}\ \underline{A}^{-1} = \underline{I}\ .$$

Then the model

$$\underline{y} = \underline{Z}\ \underline{\beta} + \underline{\varepsilon}$$

can be written

$$\underline{y} = \underline{Z}\ \underline{A}^{-1}\ \underline{A}\ \underline{\beta} + \underline{\varepsilon}$$

$$= \underline{W}\ \underline{\eta} + \underline{\varepsilon}$$

where

$$\underline{W} = \underline{Z}\ \underline{A}^{-1}$$

and

$$\underline{\eta} = \underline{A}\ \underline{\beta}\ .$$

Then

$$\underline{W}'\ \underline{W} = \underline{A}^{-1'}\ \underline{Z}'\ \underline{Z}\ \underline{A}^{-1} = \underline{I}$$

and

$$\hat{\underline{\eta}} = \underline{A}\ \hat{\underline{\beta}}\ ,$$

where

$$\mathrm{Var}(\hat{\underline{\eta}}) = \sigma^2\ \underline{I}\ ,$$

so that the $\hat{\eta}_i$ are independent, identically distributed normal random variables with means η_i and variances σ^2. Thus

$$\hat{\eta}_i/s \text{ for } i = 1,\ldots,p$$

are independent, identically distributed noncentral t random variables, and the joint density of

$$[\hat{\eta}_1/s,\ldots,\hat{\eta}_p/s]'$$

is the product of p univariate t densities:

$$\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\pi^{p/2}}\left(1+\sum_{i=1}^{p} t_i^2\right)^{-\frac{p+1}{2}} \prod_{i=1}^{p} (dt_i/\sqrt{n}) .$$

Since $\hat{\underline{\beta}} = \underline{A}\,\hat{\underline{\eta}}$, the joint density of $T_0 = (n-1)^{1/2}\,\underline{D}_0\,\hat{\underline{\beta}}/s = \underline{w}$ is

$$\frac{\Gamma\left(\frac{p+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\pi^{p/2}}\left(1+\underline{w}'\,(\underline{Z}'\underline{Z})\,\underline{w}\right)^{-\frac{p+1}{2}} |\underline{Z}'\underline{Z}|^{1/2} \prod_{i=1}^{p} (dw_i/\sqrt{n}) .$$

The following two lemmas will be used in Theorem 6.1.3 to show that $\tilde{\underline{B}}$ is asymptotically unbiased.

<u>Lemma 6.1.3</u>. If the random variable u is distributed as a chi-square random variable with r degrees of freedom, then

$$E(u^{1/2}) = \frac{\Gamma\left(\frac{r-1}{2}\right)}{2^{1/2}\,\Gamma\left(\frac{r}{2}\right)} .$$

<u>Proof</u>. If $u \sim \chi_r^2$, then for $u > 0$,

$$f(u) = \frac{1}{2^{r/2}\,\Gamma\left(\frac{r}{2}\right)} u^{\frac{r}{2}-1} e^{-\frac{u}{2}}$$

and

$$E(u^{1/2}) = \frac{1}{2^{r/2}\,\Gamma\left(\frac{r}{2}\right)} \int_0^\infty u^{\frac{r-1}{2}} e^{-\frac{u}{2}}\, du$$

$$= \frac{2^{\frac{r-1}{2}}\,\Gamma\left(\frac{r-1}{2}\right)}{2^{r/2}\,\Gamma\left(\frac{r}{2}\right)} \int_0^\infty \frac{u^{\frac{r-1}{2}} e^{-\frac{u}{2}}}{2^{\frac{r-1}{2}}\,\Gamma\left(\frac{r-1}{2}\right)}\, du$$

$$= \frac{\Gamma\left(\frac{r-1}{2}\right)}{2^{1/2}\,\Gamma\left(\frac{r}{2}\right)} .$$

<u>Lemma 6.1.4.</u> If v is distributed as a noncentral t random variable with r degrees of freedom and noncentrality parameter δ, then

$$E(v) = \delta\left(\frac{r}{2}\right)^{1/2} \frac{\Gamma\left(\frac{r-1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} .$$

<u>Proof.</u> The first moment of the noncentral t is found by noting that if $v \sim t_r(\delta)$, then

$$v = u_1/(u_2/n)^{1/2}$$

where $u_1 \sim N(\delta, 1)$ and $u_2 \sim \chi^2_r$, and u_1 and u_2 are independent. Then

$$v = \frac{u_1 - \delta}{\left(\frac{u_2}{r}\right)^{1/2}} + \frac{\delta}{\left(\frac{u_2}{r}\right)^{1/2}} = w + \frac{\delta\sqrt{r}}{\sqrt{u_2}} \quad .$$

Since $w \sim N(0, 1)$, $E(w) = 0$. Since $u_2 \sim \chi^2_r$,

$$E(u_2^{-1/2}) = \Gamma\left(\frac{r-1}{2}\right)/2^{1/2}\ \Gamma\left(\frac{r}{2}\right)$$

by Lemma 6.1.3. Thus,

$$E(v) = E\left(\frac{\delta\sqrt{r}}{\sqrt{u_2}}\right) = \delta\left(\frac{r}{2}\right)^{1/2} \frac{\Gamma\left(\frac{r-1}{2}\right)}{\Gamma\left(\frac{r}{2}\right)} \quad .$$

<u>Theorem 6.1.3</u>. The bias in $\underline{\tilde{B}}$ is asymptotically zero.

<u>Proof</u>. Consider the case $k = 1$. Since by Lemma 6.1.1, $(n-1)^{1/2}\ \tilde{B}$ has a noncentral t distribution with n-2 degrees of freedom and noncentrality parameter $(n-1)^{1/2}\ s_Z\ \beta/\sigma$, we have from Lemma 6.1.4 that

$$E[(n-1)^{1/2}\ \tilde{B}] = \frac{(n-1)^{1/2}\ s_Z\ \beta}{\sigma} \left(\frac{n-2}{2}\right)^{1/2} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \quad .$$

Thus,

$$E(\tilde{B}) = \frac{s_Z\ \beta}{\sigma} \left(\frac{n-2}{2}\right)^{1/2} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \quad .$$

Stirling's formula gives the following approximation to the gamma function:

$$\Gamma(x) \approx e^{-x}\, x^{x-1/2}\, \sqrt{2\pi}\ .$$

Thus

$$\Gamma\left(\frac{n-3}{2}\right) \approx e^{-\frac{n-3}{2}} \left(\frac{n-3}{2}\right)^{\frac{n-4}{2}} (2\pi)^{1/2}$$

and

$$\Gamma\left(\frac{n-2}{2}\right) \approx e^{-\frac{n-2}{2}} \left(\frac{n-2}{2}\right)^{\frac{n-3}{2}} (2\pi)^{1/2}\ .$$

Now

$$\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \approx \frac{e^{-\frac{n-3}{2}} \left(\frac{n-3}{2}\right)^{\frac{n-4}{2}} (2\pi)^{1/2}}{e^{-\frac{n-2}{2}} \left(\frac{n-2}{2}\right)^{\frac{n-3}{2}} (2\pi)^{1/2}}$$

$$= \frac{(2e)^{1/2}}{(n-2)^{1/2}} \left(\frac{n-3}{n-2}\right)^{\frac{n-4}{2}}$$

$$= \left(\frac{2e}{n-2}\right)^{1/2} \left(1 - \frac{1}{n-2}\right)^{\frac{n-4}{2}}$$

$$= \left(\frac{2e}{n-2}\right)^{1/2} \left[\left(1 - \frac{1}{n-2}\right)^{n-2}\right]^{1/2} \left(1 - \frac{1}{n-2}\right)^{-1} .$$

For large n,

$$\left(1 - \frac{1}{n-2}\right)^{n-2} \to e^{-1}$$

and

$$\left(1 - \frac{1}{n-2}\right) \to 1 ,$$

so that

$$\frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \to \left(\frac{2e}{n-2}\right)^{1/2} e^{-1/2} = \left(\frac{2}{n-2}\right)^{1/2}$$

for large n. Thus for large n,

$$E(\tilde{B}) \approx \frac{s_Z \beta}{\sigma} \left(\frac{n-2}{2}\right)^{1/2} \left(\frac{2}{n-2}\right)^{1/2} = \frac{s_Z \beta}{\sigma} = B .$$

<u>Corollary to Theorem 6.1.3</u>. The unbiased estimator of B, in the rank two case, is

$$B^* = \left(\frac{2}{n-2}\right)^{1/2} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-3}{2}\right)} \tilde{B} .$$

Proof. The proof is immediate since by Theorem 6.1.3,

$$E(\tilde{B}) = \left(\frac{n-2}{2}\right)^{1/2} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} B \ .$$

Theorem 6.1.4. For $k > 1$, the unbiased estimator of $\underline{B}$ is

$$\underline{B}^* = \left(\frac{2}{n-2}\right)^{1/2} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-3}{2}\right)} \tilde{\underline{B}} \ .$$

Proof. If the model $\underline{y} = \underline{Z}\ \underline{\beta} + \underline{\varepsilon}$ is reparameterized as

$$\underline{y} = \underline{W}\ \underline{\eta} + \underline{\varepsilon}$$

where $\underline{W} = \underline{Z}\ \underline{A}^{-1}$ and $\underline{\eta} = \underline{A}\ \underline{\beta}$ for some nonsingular matrix $\underline{A}$ such that

$$\underline{A}^{-1\prime}\ \underline{Z}'\ \underline{Z}\ \underline{A}^{-1} = \underline{W}'\ \underline{W} = \underline{I} \ ,$$

then

$$\tilde{\underline{B}}_A = D_n(\underline{W})\ \hat{\underline{\eta}}/s = (\tilde{b}_1, \ldots, \tilde{b}_p)$$

and

$$\underline{B}_A = D_n(W)\ \eta/\sigma = (b_1, \ldots, b_p) \ .$$

Now,

$$E(\tilde{b}_i) = g\ b_i$$

where

$$g = \left(\frac{n-2}{2}\right)^{1/2} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} ,$$

so

$$E(\underline{\tilde{B}}_A) = g\ \underline{B}_A$$

where

$$\underline{B}_A = D_n(\underline{W})\ \underline{n}/\sigma$$

$$= D_n(\underline{W})\ \underline{A}\ \underline{\beta}/\sigma\ .$$

But

$$\underline{\tilde{B}} = D_n(\underline{Z})\ \underline{\hat{\beta}}/s$$

$$= D_n(\underline{Z})\ \underline{A}^{-1}\ \underline{\hat{n}}/s$$

$$= D_n(\underline{Z})\,\underline{A}^{-1}\ \underline{\tilde{B}}_A$$

since $\underline{W}'\underline{W} = \underline{I}$.

Then

$$E(\underline{\tilde{B}}) = D_n(\underline{Z})\ \underline{A}^{-1}\ E(\underline{\tilde{B}}_A)$$

$$= g\ D_n(\underline{Z})\ \underline{A}^{-1}\ D_n(\underline{W})\ \underline{A}\ \underline{\beta}/\sigma$$

$$= g\ D_n(\underline{Z})\ \beta/\sigma$$

$$= \left(\frac{n-2}{2}\right)^{1/2} \frac{\Gamma\left(\frac{n-3}{2}\right)}{\Gamma\left(\frac{n-2}{2}\right)} \underline{B} .$$

Note that $[2/(n-2)]^{1/2}\ \Gamma\left(\frac{n-2}{2}\right)/\Gamma\left(\frac{n-3}{2}\right)$ is less than one. Thus, $\underline{B}^*$ will tend to be a bit smaller than $\tilde{\underline{B}}$, in absolute value.

We now have that the natural invariant sufficient estimator $\tilde{\underline{B}}$ of the maximal invariant parametric function $\underline{B}$ is consistent, has very small bias for large n, and has a relatively simple distribution. Theorem 6.1.5 will show, however, that the unbiased form, $\underline{B}^*$, is not efficient in the Cramér-Rao sense. We refer first to a theorem presented in Zacks (1971), presenting only the essentials and omitting the proof.

<u>Lemma 6.1.5</u>. (The generalized Cramér-Rao inequality.) Let $\{f(x;\ \theta)\ :\ \theta \in \Theta\}$ be a family of density functions satisfying certain general regularity conditions (all of which hold for this situation). Let $g(\underline{\theta})$ be a vector-valued function on Θ such that the matrix of partial derivatives with respect to $\underline{\theta}$ exists for all $\theta \in \Theta$. Let $\hat{g}(x)$ be an unbiased estimator of $g(\underline{\theta})$ such that

$$\int \left|\hat{g}(x)\ \frac{\partial}{\partial\theta_i}\ f(x;\ \theta_1,\ldots,\theta_r)\right| \mu dx < \infty$$

for all $i = 1,\ldots,r$ and all θ. If the matrix of partial derivatives is nonsingular, then $\hat{g}(x)$ attains the minimum possible generalized variance if and only if $\hat{g}(x)$ is a function of the minimal sufficient statistic for $\underline{\theta}$ and for each $i = 1,\ldots,r$,

$$\frac{\partial}{\partial\theta_i} \log f(x;\ \theta_1,\ldots,\theta_r) = c_i(\underline{\theta})[\hat{g}(X) - g(\underline{\theta})]$$

where $c_i(\underline{\theta})$ may depend on θ but not on x.

<u>Theorem 6.1.5.</u> $\underline{B}^*$ is not an efficient estimator of $\underline{B}$ (in the Cramér-Rao sense).

<u>Proof.</u> Since $\hat{\underline{\beta}} \sim N(\underline{\beta},\ \sigma^2(\underline{Z}'\underline{Z})^{-1})$ and $s^2 \sim \sigma^2\ \chi^2_{n-p}/(n-p)$,

$$f(\hat{\underline{\beta}},\ s^2;\ \underline{\theta}) = \frac{\sigma^2\ u^{\frac{1}{2}(n-p-2)}}{2^{n/2}\ \pi^{p/2}\ |(\underline{Z}'\underline{Z})^{-1}|^{1/2}\ \Gamma\left(\frac{n-p}{2}\right)} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2\sigma^2}\ [(\hat{\underline{\beta}}-\underline{\beta})'\ \underline{Z}'\underline{Z}\ (\hat{\underline{\beta}}-\underline{\beta}) + \sigma^2\ u]\right\}$$

and

$$\log f = 2 \log \sigma + \frac{n-p-2}{2} \log u - \frac{n}{2} \log 2 - \frac{p}{2} \log \pi$$

$$- \log|(\underline{Z}'\underline{Z})^{-1}| - \log \Gamma\left(\frac{n-p}{2}\right) - \frac{1}{2\sigma^2}\ [(\hat{\underline{\beta}}-\underline{\beta})'\ \underline{Z}'\underline{Z}'\ (\hat{\underline{\beta}}-\underline{\beta}) + \sigma^2\ u]$$

and

$$\frac{\partial}{\partial\sigma^2} \log f = \frac{\partial}{\partial\sigma^2} (\log \sigma^2) - \frac{\partial}{\partial\sigma^2} \left[\frac{1}{2\sigma^2}\ (\hat{\underline{\beta}}-\underline{\beta})'\ \underline{Z}'\underline{Z}\ (\hat{\underline{\beta}}-\underline{\beta})\right]$$

$$= \frac{\partial}{\partial\sigma^2} \log \sigma^2 + \frac{1}{2\sigma^4}\ (\hat{\underline{\beta}}-\underline{\beta})'\ \underline{Z}'\underline{Z}\ (\hat{\underline{\beta}}-\underline{\beta})\ ,$$

which is not a function of $c(\underline{B})$ $(\underline{B}^* - \underline{B})$ so that $\underline{B}^*$ is not efficient in that it does not attain the Cramér-Rao lower bound.

Even though $\underline{B}^*$ does not attain the Cramér-Rao lower bound, $\underline{B}^*$ may be considered to be efficient in that it is the uniformly minimum variance unbiased estimator of $\underline{B}$. This is so because $\underline{B}^*$ is unbiased for $\underline{B}$ and is also a function of the complete sufficient statistic $(\underline{\beta}, \sigma^2)$ (see Graybill, 1961).

$\underline{B}^*$ also has smaller mean squared error than $\underline{\tilde{B}}$. For simplicity, consider $k = 1$ and let $\tilde{B} = \eta B^*$, where

$$\eta = [(n-2)/2]^{1/2} \Gamma\left(\frac{n-3}{2}\right)/\Gamma\left(\frac{n-2}{2}\right) > 1 .$$

Now, $E(B^* - B)^2 = \mathrm{Var}(B^*) = v \geq 0$ and

$$\begin{aligned} E(\tilde{B}-B)^2 &= E(\tilde{B}-\eta B)^2 + E(\eta B-B)^2 \\ &= E(\eta B^*-\eta B)^2 + E(\eta B-B)^2 \\ &= \eta^2 E(B^*-B)^2 + (\eta-1)^2 B^2 \\ &= \eta^2 v + (\eta-1)^2 B^2 . \end{aligned}$$

Suppose

$$v > \eta^2 v + (\eta-1)^2 B^2 .$$

Then

$$v - \eta^2 v > (\eta-1)^2 B^2$$

or

$$(1-\eta^2)\ v > (\eta-1)^2\ B^2$$

and

$$v < \frac{(\eta-1)^2}{1-\eta^2}\ B^2$$

since if $\eta > 1$, $1 - \eta^2 < 0$. Then

$$v < \frac{(1-\eta)^2}{(1-\eta)(1+\eta)}\ B^2 = \frac{1-\eta}{1+\eta}\ B^2 .$$

Since $\eta > 1$, $1 - \eta < 0$ and

$$\frac{1-\eta}{1+\eta}\ B^2 < 0 ,$$

which implies that $v < 0$, which cannot be. Thus, v must be less than or equal to $\eta^2\ v + (\eta-1)^2\ B^2$, or

$$E(B^*-B)^2 \leq E(\tilde{B}-B)^2 .$$

Since $\underline{B}^*$ is uniformly minimum variance unbiased and has smaller mean squared error than $\underline{\tilde{B}}$, it might appear that $\underline{B}^*$ would be preferred to $\underline{\tilde{B}}$ as an estimator of $\underline{B}$. However, we feel that in practice, since the bias in $\underline{\tilde{B}}$ is quite small and since computation of the quantity

$$[2/(n-2)]^{1/2}\ \Gamma\left(\frac{n-2}{2}\right)/\Gamma\left(\frac{n-3}{2}\right)$$

is quite laborious, $\underline{\tilde{B}}$ is the natural estimator to use unless the problem specifically requires use of the unbiased form. Our intent is not,

furthermore, to compare the relative merits of the alternative forms of $\underset{\sim}{B}$, but rather to compare $\underset{\sim}{B}$ to the commonly used estimator.

Section 6.2. Properties of the Usual Social Science Estimator

The estimator of $\underline{B}$ commonly used in the social sciences is

$$\hat{\underline{B}} = D_n(\underline{Z})\ \hat{\underline{\beta}}/s_y\ ,$$

where

$$s_y^2 = \sum_{i=1}^{n} (y_i - \bar{y})^2/(n-1)\ .$$

We will show that $\hat{\underline{B}}$ does not possess desirable properties: although $\hat{\underline{B}}$ is invariant and is a function of the sufficient statistic for $(\underline{D}_n^*, \underline{\beta}, \sigma^2)$, $\hat{\underline{B}}$ is not a consistent estimator of $\underline{B}$, it is not unbiased, and it has an unwieldly distribution.

Lemma 6.2.1. $\hat{\underline{B}}$ is not a consistent estimator of $\underline{B}$.

Proof. Recall that in order for $\hat{\underline{B}}$ to be a consistent estimator of $\underline{B}$, $\hat{\underline{B}}$ must converge to $\underline{\Delta}\ \underline{\beta}/\sigma$ in probability, assuming that $\underline{D}_n^*$ is chosen to converge to a diagonal matrix $\underline{\Delta}$. We note that

$$\underline{y}'\underline{y}' = \underline{y}'\ \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}'\ \underline{y} + \underline{y}'\ (\underline{I} - \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}')\ \underline{y}$$

$$= \hat{\underline{\beta}}'\ \underline{Z}'\ \underline{Z}\ \hat{\underline{\beta}} + \underline{y}'\ (\underline{I} - \underline{Z}\ (\underline{Z}'\underline{Z})^{-1}\ \underline{Z}')\ \underline{y}\ ,$$

where

$$\hat{\underline{\beta}} = [\hat{\alpha}, \hat{\beta}_1, \ldots, \hat{\beta}_k]' = [\hat{\alpha}, \hat{\underline{\beta}}_1]'$$

and

$$\underline{Z}'\underline{Z} = \begin{bmatrix} n & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & \underline{Z}_1'\underline{Z}_1 & \\ 0 & & & \end{bmatrix}$$

for $\underline{Z}_1'\underline{Z}_1$ a $k \times k$ matrix. Then

$$\underline{y}'\underline{y} = \sum_{i=1}^{n} y_i^2 = \left(\sum_{i=1}^{n} y_i\right)^2 /n + \hat{\underline{\beta}}_1' \underline{Z}_1' \underline{Z}_1 \hat{\underline{\beta}}_1 + SS_E$$

where

$$SS_E = \underline{y}' \left(\underline{I} - \underline{Z} (\underline{Z}'\underline{Z})^{-1} \underline{Z}'\right) \underline{y}$$

is the error or residual sum of squares. Thus

$$\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2 /n = \hat{\underline{\beta}}_1 \underline{Z}_1' \underline{Z}_1 \hat{\underline{\beta}}_1 + SS_E$$

and

$$(n-1)^{-1} \left[\sum_{i=1}^{n} y_i^2 - \left(\sum_{i=1}^{n} y_i\right)^2 /n\right] = s_y^2 .$$

Now, SS_E converges to $(n-p)\,\sigma^2$ and $\underline{Z}_1'\underline{Z}_1/(n-1)$ will converge to some positive definite matrix $\underline{K}$. Thus, s_y^2 converges to $(\underline{\beta}_1' \underline{K} \underline{\beta}_1 + \sigma^2)^{1/2} \neq \sigma$, and $\hat{\underline{B}}$ does not converge to $\underline{B}$ (in probability).

<u>Lemma 6.2.2</u>. $\hat{\underline{B}}$ is not an unbiased estimator of $\underline{B}$.

<u>Proof</u>. If $\underline{D}_n^*$ is chosen to converge to some diagonal matrix $\underline{\Delta}$, then $\underline{B} = \underline{D}_n^* \beta/\sigma$ converges to $\underline{\Delta}\, \underline{\beta}/\sigma$. By Lemma 6.2.1, $\hat{\underline{B}}$ converges to

$\underline{\Delta}\ \underline{\beta}/(\underline{\beta}_1'\ \underline{K}\ \underline{\beta}_1 + \sigma^2)^{1/2} \neq \underline{\Delta}\ \underline{\beta}/\sigma$ unless $\underline{\beta} = \underline{0}$ or $\underline{K} = \underline{0}$. Since the parameter and the estimator do not converge to the same value, $\hat{\underline{B}}$ cannot be unbiased for $\underline{B}$.

We also note that since $\underline{K}$ is positive definite, $\underline{\beta}_1'\ \underline{K}\ \underline{\beta}_1 > 0$, so that $\underline{\Delta}\ \underline{\beta}/(\underline{\beta}_1'\ \underline{K}\ \underline{\beta}_1 + \sigma^2)^{1/2} < \underline{B}$, and $\hat{\underline{B}}$ tends to underestimate $\underline{B}$ in absolute value. The effects of this underestimation will be illustrated in Chapter X.

Lemma 6.2.3. For $k = 1$, $\hat{B}$ can be written

$$\hat{B} = \frac{u^{1/2}}{(u+r)^{1/2}}$$

where u is distributed as a noncentral F random variable with 1 and $r = n - 2$ degrees of freedom and noncentrality parameter $\lambda = (n-1)\ s_Z^2\ \beta^2/2\sigma^2$.

Proof. We note that for $k = 1$, s_y^2 can be written

$$s_y^2 = \hat{\beta}^2\ s_Z^2 + \frac{n-2}{n-1}\ s^2 \ .$$

Thus

$$\hat{B} = \frac{s_Z\ \hat{\beta}}{s_y} = \frac{s}{s_y}\ \frac{s_Z\ \hat{\beta}}{s}$$

$$= \frac{s}{\left[\hat{\beta}\ s_Z^2 + \frac{n-2}{n-1}\ s^2\right]^{1/2}}\ \frac{s_Z\ \hat{\beta}}{s}$$

$$= \frac{1}{\left[\frac{\hat{\beta}^2 \; s_Z^2}{s^2} + \frac{n-2}{n-1}\right]^{1/2}} \frac{s_Z \; \hat{\beta}}{s} .$$

Let $u = (n-1) \; s_Z^2 \; \hat{\beta}^2/s^2$. Then since

$$\hat{\beta} \sim N(\beta, \; \sigma^2/(n-1) \; s_Z^2) \; ,$$

$$(n-1)^{1/2} \; s_Z \; \hat{\beta}/\sigma \sim N(n-1)^{1/2} \; s_Z \; \beta/\sigma, \; 1)$$

and

$$(n-1) \; s_Z^2 \; \hat{\beta}^2/\sigma^2 \sim \chi_1^2 \; (\lambda)$$

where

$$\lambda = (n-1) \; s_Z^2 \; \beta^2/\sigma^2 \; .$$

Also,

$$(n-2) \; s^2/\sigma^2 \sim \chi_{n-2}^2$$

so that

$$s^2/\sigma^2 \sim \chi_{n-2}^2/(n-2)$$

and

$$(n-1) \; s_Z^2 \; \hat{\beta}^2/s^2 \sim F_{1, \; n-2} \; (\lambda) \; .$$

Then $u^{1/2} = (n-1)^{1/2} \; s_Z \; \hat{\beta}/s$ and

$$\hat{B} = \frac{1}{\left[\frac{u}{n-1} + \frac{n-2}{n-1}\right]^{1/2}} \frac{u^{1/2}}{(n-1)^{1/2}}$$

$$= \frac{1}{\left[\frac{u + (n-2)}{n-1}\right]^{1/2}} \frac{u^{1/2}}{(n-1)^{1/2}}$$

$$= \frac{1}{[u + r]^{1/2}} u^{1/2}$$

where $r = n - 2$.

Lemma 6.2.4. The distribution of $\hat{B}$, in the rank two case, is

$$f(w) = \frac{2e^{-\lambda} (1-w^2)^{\frac{r-2}{2}}}{\Gamma\left(\frac{r}{2}\right)} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma\left(\frac{2i + 1 + r}{2}\right)}{\Gamma\left(\frac{2i + 1}{2}\right)} w^{2i} .$$

Proof. From Lemma 6.2.3 we have that $\hat{B} = u^{1/2}/(r + u)^{1/2}$, where $u \sim F_{1,r}(\lambda)$. That is, for $u \geq 0$,

$$f(u) = \sum_{i=0}^{\infty} \frac{e^{-\lambda} \lambda^i}{i!} \frac{r^{r/2} \Gamma\left(\frac{1 + r + 2i}{2}\right)}{\Gamma\left(\frac{2i + 1}{2}\right) \Gamma\left(\frac{r}{2}\right)} \frac{u^{\frac{2i - 1}{2}}}{(r + u)^{\frac{2i + 1 + r}{2}}} .$$

Let

$$w = u^{1/2}/(r + u)^{1/2} .$$

Then

$$w^2 = u/r + u)$$

$$w^2 r + w^2 u = u$$

$$w^2 r = u(1 - w^2)$$

$$w^2 r (1 - w^2)^{-1} = u$$

and

$$\frac{du}{dw} = 2rw (1 - w^2)^{-1} + w^2 r (-1) (1 - w^2)^{-2} (-2w)$$

$$= \frac{2rw}{1 - w^2} + \frac{2w^3 r}{(1 - w^2)^2}$$

$$= \frac{2rw - 2rw^3 + 2rw^3}{(1 - w^2)^2}$$

$$= 2rw/(1 - w^2)^2 .$$

Also,

$$r + u = r + \frac{rw^2}{1-w^2}$$

$$= \frac{r - rw^2 + rw^2}{1 - w^2}$$

$$= r/(1 - w^2) .$$

Then

$$f(w) = K \sum_{i=0}^{\infty} G_i \frac{[rw^2/(1 - w^2)]^{\frac{2i - 1}{2}}}{[r/(1 - w^2)]^{\frac{2i + 1 + r}{2}}} \frac{2rw}{(1 - w^2)^2} ,$$

where

$$K = e^{-\lambda} r^{r/2}/\Gamma\left(\frac{r}{2}\right)$$

and

$$G_i = \lambda^i \Gamma\left(\frac{2i + 1 + r}{2}\right)/i! \Gamma\left(\frac{2i + 1}{2}\right) .$$

Simplifying the expression for f(w),

$$f(w) = K \sum_{i=0}^{\infty} G_i \frac{(rw^2)^{\frac{2i - 1}{2}} (1 - w^2)^{\frac{2i + 1 + r}{2}} 2rw}{(1 - w^2)^{\frac{2i - 1}{2}} r^{\frac{2i + 1 + r}{2}} (1 - w^2)^2}$$

$$= 2K \sum_{i=0}^{\infty} G_i \frac{r^{\frac{2i + 1}{2}} w^{2i} (1 - w^2)^{\frac{2i + 1 + r}{2}}}{(1 - w^2)^{\frac{2i - 3}{2}} r^{\frac{2i + 1 + r}{2}}}$$

$$= \frac{2K}{r^{r/2}} \sum_{i=0}^{\infty} G_i w^{2i} (1 - w^2)^{\frac{r-2}{2}}$$

$$= \frac{2K (1 - w^2)^{\frac{r-2}{2}}}{r^{r/2}} \sum_{i=0}^{\infty} G_i w^{2i} .$$

Thus

$$f(w) = \frac{2(1 - w^2)^{\frac{r-2}{2}}}{r^{r/2}} \frac{e^{-\lambda} r^{r/2}}{\Gamma\left(\frac{r}{2}\right)} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma\left(\frac{2i + 1 + r}{2}\right)}{\Gamma\left(\frac{2i + 1}{2}\right)} w^{2i}$$

$$= \frac{2e^{-\lambda} (1 - w^2)^{\frac{r-2}{2}}}{\Gamma\left(\frac{r}{2}\right)} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma\left(\frac{2i + 1 + r}{2}\right)}{\Gamma\left(\frac{2i + 1}{2}\right)} w^{2i} .$$

Clearly, $\hat{\underline{B}}$ does not have a simple distribution, even for the k = 1 case.

We note in closing that $\tilde{\underline{B}}$, but not $\underline{B}^*$ or $\hat{\underline{B}}$, is a maximum likelihood estimator of $\underline{B}$ since by the invariant property of maximum likelihood statistics (Hogg and Craig, 1965) they are single-valued functions of the maximum likelihood estimator of $(\underline{\beta}, \sigma^2)$.

The results presented thus far indicate that the estimator $\tilde{\underline{B}}$ introduced in this dissertation is superior to the usual social science estimator $\hat{\underline{B}}$ as an estimator of $\underline{B}$, the "beta coefficient." $\tilde{\underline{B}}$ is consistent, has a relatively simple distribution, and an unbiased form of $\tilde{\underline{B}}$ can be constructed which has uniformly minimum variance; none of these properties are true for $\hat{\underline{B}}$. The ramifications of the use of $\tilde{\underline{B}}$ instead of $\hat{\underline{B}}$ will be illustrated in Chapter X in numerical examples employing actual social science and engineering research data.

CHAPTER VII

RELATIONSHIP OF PROPOSED ESTIMATOR TO THE MINIMUM RISK EQUIVARIANT AND FIDUCIAL ESTIMATORS

Section 7.1. Relationship to the Minimum Risk Equivariant Estimator

In the linear model problem under consideration, the group of transformations $\bar{H} = \{[\underline{\Lambda}, \underline{c}, \gamma^2]\}$ is isomorphic to the parameter space Θ, and thus if an invariant loss function is defined, an equivariant decision problem is obtained. Since $\bar{H}$ is isomorphic to Θ, the maximal invariant function with respect to $\bar{H}$ is a constant and the problem of finding a minimum risk equivariant estimator of $\theta = (\underline{D}_n^*, \underline{\beta}, \sigma^2)$ which is a function of the sufficient statistic $x = (D_n(\underline{Z}), \hat{\underline{\beta}}, s^2)$ can be treated in the usual manner. We also note that if $\bar{H}$ were transitive on, but not isomorphic to Θ, an equivariant decision problem could still be defined (Kudō, 1955).

Consider an equivariant estimator

$$e = (\underline{e}_1, \underline{e}_2, e_3) ,$$

where $\underline{e}_1 \in \mathcal{D}(\underline{Z}_0)$, $\underline{e}_2 \in R^n$, $e_e \in R^+$, and

$$[\underline{\Lambda}, \underline{c}, \gamma^2](\underline{e}_1, \underline{e}_2, e_3) = (\underline{e}_1 \underline{\Lambda}, \gamma \underline{\Lambda}^{-1} \underline{e}_2 + \underline{c}, \gamma^2 e_3) .$$

Lemma 7.1.1. The loss function

$$L(\theta, e) = \text{trace}(\underline{D}_n^{*-1})' (\underline{e}_1 - \underline{D}_n^*)' (\underline{e}_1 - \underline{D}_n^*) \underline{D}_n^{*-1}$$

$$+ \sigma^{-2} (\underline{e}_2 - \underline{\beta})' \underline{D}_n^{*\prime} \underline{D}_n^* (\underline{e}_2 - \underline{\beta})$$

$$+ \sigma^{-4} (e_3 - \sigma^2)^2$$

is invariant under H^*.

Proof.

$$L(\bar{h}\theta, he) = \text{trace}[(\underline{D}_n^* \underline{\Lambda})^{-1}]' (\underline{e}_1 \underline{\Lambda} - \underline{D}_n^* \underline{\Lambda})' (\underline{e}_1 \underline{\Lambda} - \underline{D}_n^* \underline{\Lambda}) (\underline{D}_n^* \underline{\Lambda})^{-1}$$

$$+ (\gamma\sigma)^{-2}[\gamma\underline{\Lambda}^{-1}\underline{e}_2 + \underline{c} - (\gamma\underline{\Lambda}^{-1}\underline{\beta}+\underline{c})]'(\underline{D}_n^*\underline{\Lambda})'(\underline{D}_n^*\underline{\Lambda})[\gamma\underline{\Lambda}^{-1}\underline{e}_2+\underline{c}-(\gamma\underline{\Lambda}^{-1}\underline{\beta}+\underline{c})]$$

$$+ (\gamma\sigma)^{-4} (\gamma^2 e_3^2 - \gamma^2 \sigma^2)^2$$

$$= \text{trace}(\underline{\Lambda}^{-1} \underline{D}_n^{*-1})' [(\underline{e}_1 - \underline{D}_n^*) \underline{\Lambda}]' [\underline{e}_1 - \underline{D}_n^*) \underline{\Lambda}] \underline{\Lambda}^{-1} \underline{D}_n^{*-1}$$

$$+ (\gamma\sigma)^{-2} [\gamma\underline{\Lambda}^{-1} (\underline{e}_2 - \underline{\beta})]' \underline{\Lambda}' (\underline{D}_n^*)' \underline{D}_n^* \Lambda[\gamma\underline{\Lambda}^{-1} (\underline{e}_2 - \underline{\beta})]$$

$$+ (\gamma\sigma)^{-4} [\gamma^2(e_3 - \sigma^2)]^2$$

$$= \text{trace}(\underline{D}_n^{*-1})' (\underline{\Lambda}^{-1})' \underline{\Lambda}' (\underline{e}_1 - \underline{D}_n^*)' (\underline{e}_1 - \underline{D}_n^*) \underline{\Lambda} \underline{\Lambda}^{-1} \underline{D}_n^{*-1}$$

$$+ (\gamma\sigma)^{-2} \gamma^2[(\underline{e}_2 - \underline{\beta})' (\underline{\Lambda}^{-1})' \underline{\Lambda}' (\underline{D}_n^*)' \underline{D}_n^* \underline{\Lambda} \underline{\Lambda}^{-1} (\underline{e}_2 - \underline{\beta})]$$

$$+ (\gamma\sigma)^{-4} \gamma^4 (e_3 - \sigma^2)^2$$

$$= \text{trace}(\underline{D}_n^{*-1})' (\underline{e}_1 - \underline{D}_n^*)' (\underline{e}_1 - \underline{D}_n^*) \underline{D}_n^{*-1}$$

$$+ \sigma^{-2} (\underline{e}_2 - \underline{\beta})' (\underline{D}_n^*)' \underline{D}_n^* (\underline{e}_2 - \underline{\beta}) + \sigma^{-4} (e_3 - \sigma^2)^2$$

$$= L(\theta, e) .$$

Recall that risk is defined as the expected value of the loss.

Theorem 7.1.1. Using the loss function defined in Lemma 7.1.1,

$$e = (\underline{e}_1, \underline{e}_2, e_3) = (D_n(\underline{Z}), \hat{\underline{\beta}}, s^2)$$

is a minimum risk equivariant estimator of $\theta = (\underline{D}_n^*, \underline{\beta}, \sigma^2)$.

Proof. Consider the first term of the loss function. Its expected value,

$$E[\text{trace}(\underline{D}_n^{*-1})' (\underline{e}_1 - \underline{D}_n^*)' (\underline{e}_1 - \underline{D}_n^*) \underline{D}_n^{*-1}] ,$$

can be written

$$E[\text{trace}(\underline{e}_1 - \underline{D}_n^*) \underline{D}_n^{*-2} (\underline{e}_1 - \underline{D}_n^*)']$$

since traces are invariant under such changes. Since this term is non-negative and $D_n(\underline{Z}) = \underline{D}_n^*$ with probability one, this term will be zero if $\underline{e}_1 = D_n(\underline{Z})$, and thus minimized.

The second term of the risk function is

$$E[\sigma^{-2} (\underline{e}_2 - \underline{\beta})' \underline{D}_n^{*\prime} \underline{D}_n^* (\underline{e}_2 - \beta)] .$$

Since $\hat{\underline{\beta}}$ is unbiased for $\underline{\beta}$, this term can be expanded as

$$E[\sigma^{-2} (\underline{e}_2 - \hat{\underline{\beta}})' \underline{D}_n^{*\prime} \underline{D}_n^* (\underline{e}_2 - \hat{\underline{\beta}})] + E[\sigma^{-2} (\hat{\underline{\beta}} - \underline{\beta})' \underline{D}_n^{*\prime} \underline{D}_n^* (\hat{\underline{\beta}} - \underline{\beta}] ,$$

where the second term does not depend on $\underline{e}_2$ and the first term is minimized by letting $\underline{e}_2 = \hat{\underline{\beta}}$.

Similarly, the third term in the risk function can be expanded into

$$E[\sigma^{-4} (e_3 - s^2)^2] + E[\sigma^{-4} (s^2 - \sigma^2)^2] ,$$

and this term is minimized by taking $e_3 = s^2$.

Theorem 7.1.2. $T(x) = (\underline{I}, \tilde{\underline{B}}, 1)$ is a maximal invariant function of

the minimum risk equivariant estimator.

Proof. $T(x) = (\underline{I}, \tilde{\underline{B}}, 1) = h_{1x}^{-1} (\underline{D}_n(\underline{Z}), \hat{\underline{\beta}}, s^2)$ is a function of the minimum risk equivariant estimator $(D_n(\underline{Z}), \hat{\underline{\beta}}, s^2)$ and $T(x)$ indexes the orbits of H_1 in X.

Section 7.2. Relationship to the Fiducial Estimator

We will now show that $\tilde{\underline{B}}$ is also a maximal invariant function of the fiducial estimator. Although a theorem of Hora and Buehler (1966) can be used which immediately shows that the minimum risk equivariant estimator is a minimum risk fiducial estimator, we will show this result directly in order to demonstrate the similarity between the structure of the problem studied in this dissertation and the corresponding fiducial problem. We do not intend, however, to pursue the philosophical implications of the fiducial argument.

Our approach to fiducial estimation will follow that of Fraser (1961 b); that is, we shall relate our proposed estimator to the fiducial estimator as defined by Fraser, which is not necessarily the same as the one proposed by Fisher (1934).

Definition 7.2.1. A pivotal quantity is a function of the sufficient statistic and the parameter which has a fixed distribution regardless of the value of the parameter.

Fraser developed a pivotal quantity as follows. By the isomorphisms of H to X and $\bar{H}$ to Θ and the isomorphism between H and $\bar{H}$, each point $x \in X$ corresponds to a point $\theta \in \Theta$, where the distribution of x has index value θ. Since $\theta = \bar{h}_\theta \, \theta_0$,

$$\bar{h}_\theta^{-1} x = \bar{h}_\theta^{-1} h_x \, x_0$$

is a random variable with a θ_0-distribution. Thus

$$p = \bar{h}_\theta^{-1} \, h_x$$

is a function of θ and x having a fixed distribution regardless of the value of θ; p is a pivotal quantity. (Note that the fiducial argument relies heavily on the permissibility of applying transformations in $\bar{H}$ to elements of X. This procedure is apparently philosophically justified by the relationships between X, H, $\bar{H}$, and Θ.)

<u>Lemma 7.2.1.</u> In the linear model estimation problem defined in Chapter V, the pivotal quantity (with respect to $\bar{H}$) is

$$p = [D_n(\underline{Z}) \, \underline{D}_n^{*-1}, \, \underline{D}_n^* \, (\hat{\underline{\beta}} - \underline{\beta})/\sigma, \, s^2/\sigma^2]$$

$$= [\underline{U}_1, \, \underline{U}_2, \, U_3^2] \; .$$

<u>Proof.</u> In the linear model problem,

$$\bar{h}_\theta = [\underline{D}_n^*, \, \underline{\beta}, \, \sigma^2] \; ,$$

which implies that

$$\bar{h}_\theta^{-1} \, h_x \, x_0 = [\underline{D}_n^*, \, \underline{\beta}, \, \sigma^2]^{-1} \, [D_n(\underline{Z}), \, \hat{\underline{\beta}}, \, s^2] \, (\underline{I}, \, \underline{0}, \, 1)$$

$$= [\underline{D}_n^*, \, \underline{\beta}, \, \sigma^2]^{-1} \, (D_n(\underline{Z}), \, \hat{\underline{\beta}}, \, s^2)$$

$$= [\underline{D}_n^{*-1}, \, -\sigma^{-1} \, \underline{D}_n^* \, \underline{\beta}, \, \sigma^{-2}] \, (D_n(\underline{Z}), \, \hat{\underline{\beta}}, \, s^2)$$

$$= (D_n(\underline{Z}) \, \underline{D}_n^{*-1}, \, \sigma^{-1} \, \underline{D}_n^* \, \hat{\underline{\beta}} - \sigma^{-1} \, \underline{D}_n^* \, \underline{\beta}, \, \sigma^{-2} \, s^2)$$

$$= (D_n(\underline{Z})\ \underline{D}_n^{*-1},\ \underline{D}_n^{*}\ (\hat{\underline{\beta}} - \underline{\beta})/\sigma,\ s^2/\sigma^2)\ .$$

To verify that p is indeed a pivotal quantity, we first note that p is a function of the sufficient statistic and the parameter. Now,

$$P(\underline{U}_1 = \underline{I}) = 1\ ,$$

$\underline{U}_2$ is distributed as $N(\underline{0}, \underline{I})$, since

$$\underline{Z}_0'\ \underline{Z}_0 = \underline{I}\ ,$$

and U_3^2 is distributed as $\chi^2/(n-p)$, so that p has a fixed distribution and is a pivotal quantity.

As a function of x and θ, p is invariant under transformations in H. Following Fraser's proof, if h is a transformation in H, then

$$x = h_x\ x_0$$

is transformed into

$$hx = h\ h_x\ x_0$$

and a distribution

$$\theta = \bar{h}_\theta\ \theta_0$$

for x becomes the distribution

$$h\theta = h\ \bar{h}_\theta\ \theta_0$$

for hx. Hence

$$p' = (h\ \bar{h}_\theta)^{-1}\ (h\ h_x)$$

$$= \bar{h}_\theta^{-1}\ h^{-1}\ h\ h_x$$

$$= \bar{h}_\theta^{-1}\ h_x = p\ .$$

Specifically for the linear model,

$$h\ h_x = [\underline{\Lambda},\ \underline{c},\ \gamma^2]\ [D_n(\underline{Z}),\ \hat{\underline{\beta}},\ s^2]$$

$$= [D_n(\underline{Z})\ \underline{\Lambda},\ \gamma\ \underline{\Lambda}^{-1}\ \hat{\underline{\beta}} + \underline{c},\ \gamma^2\ s^2]$$

and

$$(h\ \bar{h}_\theta)^{-1} = \bar{h}_\theta^{-1}\ h^{-1} = [\underline{D}_n^*,\ \underline{\beta},\ \sigma^2]^{-1}\ [\underline{\Lambda},\ \underline{c},\ \gamma^2]^{-1}$$

$$= [\underline{D}_n^{*-1},\ -\sigma^{-1}\ \underline{D}_n^*\ \underline{\beta},\ \sigma^{-2}]\ [\underline{\Lambda}^{-1},\ -\gamma^{-1}\ \underline{\Lambda}\ \underline{c},\ \gamma^{-2}]$$

$$= [\underline{\Lambda}^{-1}\ \underline{D}_n^{*-1},\ -\sigma^{-1}\ \gamma^{-1}\ \underline{D}_n^*\ \underline{\Lambda}\ \underline{c} - \sigma^{-1}\ \underline{D}_n^*\ \underline{\beta},\ \sigma^{-2}\ \gamma^{-2}]$$

$$= [\underline{\Lambda}^{-1}\ \underline{D}_n^{*-1},\ -\sigma^{-1}\ \underline{D}_n^*\ (\gamma^{-1}\ \underline{\Lambda}\ \underline{c} + \underline{\beta}),\ \sigma^{-2}\ \gamma^{-2}]\ .$$

Then $(h\ \bar{h}_\theta)^{-1}\ h\ h_x =$

$$[\underline{\Lambda}^{-1}\ \underline{D}_n^{*-1},\ -\sigma^{-1}\ \underline{D}_n^*(\gamma^{-1}\ \underline{\Lambda}\ \underline{c} + \underline{\beta}),\ \sigma^{-2}\ \gamma^{-2}][D_n(\underline{Z})\ \underline{\Lambda},\ \gamma\ \underline{\Lambda}^{-1}\ \hat{\underline{\beta}} + \underline{c},\ \gamma^2\ s^2]$$

$$= [D_n(\underline{Z})\ \underline{\Lambda}\ \underline{\Lambda}^{-1}\ \underline{D}_n^{*-1},\ \sigma^{-1}\ \gamma^{-1}\ \underline{D}_n^*\ \underline{\Lambda}\ (\gamma\ \underline{\Lambda}^{-1}\ \hat{\underline{\beta}} + \underline{c}) - \sigma^{-1}\ \underline{D}_n^*(\gamma^{-1}\ \underline{\Lambda}\ \underline{c} + \underline{\beta}),$$

$$\sigma^{-2}\ \gamma^{-2}\ \gamma^2\ s^2]$$

$$= [D_n(\underline{Z})\ \underline{D}_n^{*-1},\ \sigma^{-1}\ \underline{D}_n^*\ \hat{\underline{\beta}} - \sigma^{-1}\ \underline{D}_n^*\ \underline{\beta},\ \sigma^{-2}\ s^2]$$

$$= [D_n(\underline{Z})\ \underline{D}_n^{*-1},\ \underline{D}_n^*\ (\hat{\underline{\beta}} - \underline{\beta})/\sigma,\ s^2/\sigma^2]$$

$$= p\ .$$

Any function of x and θ which is invariant under H is a function of the pivotal quantity $p = \bar{h}_\theta^{-1} h_x$. The proof of this statement is again Fraser's: Let $F(x; \theta)$ be a function of x and θ which is invariant under H. Then

$$F(hx; \bar{h}\theta) = F(x; \theta) \text{ for } h \in H \text{ and } \bar{h} \in \bar{H} .$$

Then

$$\begin{aligned} F(x; \theta) &= F(h_x x_0; \bar{h}_\theta \theta_0) \\ &= F(\bar{h}_\theta^{-1} h_x x_0; \bar{h}_\theta^{-1} \bar{h}_\theta \theta_0) \\ &= F(\bar{h}_\theta^{-1} h_x x_0; \theta_0) , \end{aligned}$$

so that F has been expressed as a function of $\bar{h}_\theta^{-1} h_x = p$. Any invariant pivotal quantity is a function of the pivotal quantity p, so that in a sense p is maximal.

The fiducial distribution for θ can be obtained as follows. Let p be a variable with the fixed pivotal quantity distribution, let x be the observed value of the sufficient statistic, and let $\hat{\theta}$ be a variable designating possible values of θ given the frequency and observational information. The pivotal equation can be solved for θ in terms of p and x:

$$p = \bar{h}_\theta^{-1} h_x$$

$$\bar{h}_\theta p = h_x$$

$$\bar{h}_\theta = h_x p^{-1} ,$$

which implies that

$$\hat{\theta} = \bar{h}_\theta \, \theta_0 = h_x \, p^{-1} \, \theta_0 \, .$$

<u>Theorem 7.2.1</u>. In the linear model problem, the fiducial distribution is

$$\hat{\theta} = (D_n(\underline{Z}), \, \hat{\underline{\beta}} - s \, D_n^{-1} \, (\underline{Z}) \, t_{n-p}, \, (n-p) \, s^2/\chi^2) \, .$$

<u>Proof</u>. From Lemma 7.2.1, the pivotal quantity in the linear model is

$$p = [\underline{u}_1, \, \underline{u}_2, \, u_3^2]$$

and

$$p^{-1} = [\underline{u}_1^{-1}, \, - u_3^{-1} \, \underline{u}_1 \, \underline{u}_2, \, u_3^{-2}]$$

$$= [\underline{I}, \, - u_3^{-1} \, \underline{u}_2, \, u_3^{-2}] \, ,$$

with probability one. Now,

$$p^{-1} \, \theta_0 = (\underline{I}, \, - u_3^{-1} \, \underline{u}_2, \, u_3^{-2})$$

and thus

$$\hat{\theta} = h_x \, p^{-1} \, \theta_0 = [D_n(\underline{Z}), \, \hat{\underline{\beta}}, \, s^2] \, (\underline{I}, \, - u_3^{-1} \, \underline{u}_2, \, u_3^{-2})$$

$$= (D_n(\underline{Z}), \, s \, D_n^{-1}(\underline{Z}) \, (- u_3^{-1} \, \underline{u}_2) + \hat{\underline{\beta}}, \, s^2 \, u_3^{-2})$$

$$= (D_n(\underline{Z}), \, \hat{\underline{\beta}} - s \, D_n^{-1}(\underline{Z}) \, \underline{u}_2/u_3, \, s^2/u_3^2) \, .$$

Recall that $u_3^2 \sim \chi^2/(n-p)$. Thus

$$\frac{s^2}{U_3^2} \sim \frac{s^2}{\chi^2/(n-p)} = \frac{(n-p)\ s^2}{\chi^2} .$$

Also, since $\underline{U}_2 \sim N(\underline{0}, \underline{I})$,

$$\frac{s\ D_n^{-1}\ (\underline{Z})\ \underline{U}_2}{U_3} \sim s\ D_n^{-1}\ (\underline{Z})\ t_{n-p}$$

where t_{n-p} is a t random variable with n-p degrees of freedom. The fiducial variables are therefore

$$\underline{D}_n^* = D_n(\underline{Z})$$

$$\sigma^2 = (n-p)\ s^2/\chi^2$$

and

$$\underline{\beta} = \hat{\underline{\beta}}\ s\ D_n^{-1}(\underline{Z})\ t_{n-p} .$$

That is,

$$\hat{\theta} = (D_n(\underline{Z}),\ \hat{\underline{\beta}} - s\ D_n^{-1}(\underline{Z})\ t_{n-p},\ (n-p)\ s^2/\chi^2)$$

gives the fiducial distribution.

Under squared error loss the fiducial estimator is the mean of the fiducial distribution. We will now show that the fiducial estimator of θ is the minimum risk equivariant estimator, $(D_n(\underline{Z}),\ \hat{\underline{\beta}},\ s^2)$. We may restrict our attention to the pivotal quantity

$$(\underline{T}_1,\ \underline{T}_2) = (\underline{D}_n^*\ (\hat{\underline{\beta}} - \underline{\beta})/\sigma,\ s^2/\sigma^2) ,$$

and for simplicity we will consider the case for k = 1. Then the pivotal quantity becomes

$$(T_1, T_2) = \left(\frac{\hat{\beta} - \beta}{\sigma/s_z}, s^2/\sigma^2\right).$$

The joint probability element of (T_1, T_2) is

$$c e^{-(\hat{\beta} - \beta)^2/(2\sigma^2/s_Z^2)} (s^2/\sigma^2)^{\frac{r}{2} - 1} e^{-\frac{s^2}{2\sigma^2}} d\left(\frac{\hat{\beta} - \beta}{\sigma/s_z}\right) d(s^2/\sigma^2)$$

where $r = n - 2$ and

$$c = \frac{1}{\pi^{1/2} 2^{\frac{r+1}{2}} \sigma r \Gamma\left(\frac{r}{2}\right)}.$$

In order to obtain the fiducial distribution we need to find the probability element $d\beta\, d\sigma^2$. It can be shown (Nachbin, 1965) that the left invariant Haar measure generated by (λ, γ) is of the form

$$d\left(\frac{\hat{\beta} - \beta}{\sigma/s_Z}\right) = \frac{\sigma^2}{s_Z^2} d\beta$$

and

$$d\left(\frac{s^2}{\sigma^2}\right) = \frac{s^2}{\sigma^2} d\sigma^2$$

where $(\sigma/s_Z)^2$ and $(s/\sigma)^2$ are the appropriate modular functions.

The fiducial probability element can be written

$$ce^{-(\hat{\beta}-\beta)^2/(2\sigma^2/s_Z^2)}\,(s^2/\sigma^2)^{\frac{r}{2}-1}\,e^{-\frac{s^2}{2\sigma^2}}\left(\frac{\sigma}{s_Z}\right)^2\left(\frac{s^2}{\sigma^2}\right)d\beta\,d\sigma^2\ .$$

Since the loss is squared error and the fiducial estimation problem is analagous to a Bayesian estimation problem with a Haar invariant prior (Good, 1972; Fraser, 1961 b), the fiducial estimator is the mean of the fiducial density (the posterior). These means are obtained (similar to Fraser, 1964) by integrating the marginal fiducial densities of β and σ^2. These marginals can be represented as

$$\beta \sim \hat{\beta} - \frac{s}{s_Z}\,T_3$$

and

$$\sigma^2 \sim s^2/T_4$$

where T_3 is a t random variable with $r = n - 2$ degrees of freedom and T_4 is r^{-1} times a chi-square variable with r degrees of freedom.

Then in order to find the means we evaluate

$$\int \beta ce^{-(\hat{\beta}-\beta)^2/(2\sigma^2/s^2)}\left(\frac{s^2}{\sigma^2}\right)^{\frac{r}{2}-1} e^{-s^2/2\sigma^2}\left(\frac{\sigma}{s_Z}\right)^2\left(\frac{s^2}{\sigma^2}\right)d\beta\,d\sigma^2$$

and

$$\int \sigma^2\,ce^{-(\hat{\beta}-\beta)^2/(2\sigma^2/s^2)}\left(\frac{s^2}{\sigma^2}\right)^{\frac{r}{2}-1} e^{-s^2/2\sigma^2}\left(\frac{\sigma}{s_Z}\right)^2\left(\frac{s^2}{\sigma^2}\right)d\beta\,d\sigma^2\ .$$

Since these integrals are Haar invariant, they are the same as

$$\int \frac{\hat{\beta} - \beta}{\sigma/s_Z} \, c e^{-(\hat{\beta}-\beta)^2/(2\sigma^2/s^2)} \left(\frac{s^2}{\sigma^2}\right)^{\frac{r}{2}-1} e^{-s^2/2\sigma^2} \, d\left(\frac{\hat{\beta} - \beta}{\sigma/s_Z}\right) d\left(\frac{s^2}{\sigma^2}\right)$$

and

$$\int \frac{s^2}{\sigma^2} \, c e^{-(\hat{\beta}-\beta)^2/(2\sigma^2/s_Z^2)} \left(\frac{s^2}{\sigma^2}\right)^{\frac{r}{2}-1} e^{-s^2/2\sigma^2} \, d\left(\frac{\hat{\beta} - \beta}{\sigma/s_Z}\right) d\left(\frac{s^2}{\sigma^2}\right) .$$

Thus,

$$E\left(\frac{\hat{\beta} - \beta}{\sigma/s_Z}\right) = E_f \; (\beta) = \hat{\beta}$$

and

$$E(s^2/\sigma^2) = E_f \; (\sigma^2) = s^2$$

and the fiducial estimates are also the minimum risk equivariant estimators. Thus $\underline{\tilde{B}}$ is a maximal invariant function of the fiducial estimator, as well as a maximal invariant function of the minimum risk equivariant estimator.

CHAPTER VIII

APPLICATION TO THE LINEAR MODEL: STOCHASTIC INPUT MATRIX

Section 8.1. Alternative Parametric Formulations

It is often the case that the input matrix $\underline{Z}$ in the linear model

$$\underline{y} = \underline{Z}\,\underline{\beta} + \underline{\varepsilon}$$

cannot be fixed by the experimenter. For example in social science research, it may be the case that the levels of the independent variables $z_1, \ldots, z_k$ cannot be controlled by the experimenter, but are rather random observations on the k variables. This is usually the case in areas such as survey research. We will show that, depending on what the experimenter wishes to measure, there are two maximal invariant parametric functions which arise naturally. In one case, the natural invariant estimator will be $\underline{\tilde{B}}$, and in the other case the natural invariant estimator will be the social science estimator $\underline{\hat{B}}$. We will consider only the rank two model and will assume that z and y are jointly distributed as bivariate normal random variables.

The group of transformations on the sample space is the same as in Chapter V:

$$G^* = \{[\lambda, c, \gamma]\}$$

where

$$[\lambda, c, \gamma]\ (z,y) = (z\lambda, \gamma y + z\lambda c)\ .$$

Depending upon the aims of the researcher, the joint distribution

of z and y can be parameterized in one of two alternative manners. One parameterization, to which we shall refer as Formulation One, defines the parameter space as

$$\Theta = \{(\mu_z, \mu_y, \sigma_z^2, \sigma_y^2, \beta)\} ,$$

where μ_z and μ_y are the means of z and y, respectively, and σ_z^2 and σ_y^2 are the marginal variances of the two variables. Formulation Two defines the parameter space as

$$\Theta = \{(\mu_z, \mu_y, \sigma_z^2, \sigma^2, \beta)\} ,$$

where σ^2 is the conditional variance of y, given z. We will show that the investigator chooses Formulation One or Formulation Two as he wishes to estimate the standardized change in y relative to the marginal or conditional standardized change in z, respectively.

For example, suppose the researcher is interested in predicting the rate of drug use in a community as a function of the community's attitude toward drug use. He might then decide to predict the standardized change in the rate of drug use for a one-standard deviation change in attitude in the community as compared to all communities. In this case, he should choose Formulation One. However, if he wished to predict the standardized change in the rate of drug use compared only to communities which have <u>the same</u> attitude toward drug use as the one under study, he should choose Formulation Two. (The reasons for the choice of formulation will become apparent when the parametric functions are obtained.)

The group of transformations induced on Θ is, as before,

$$\bar{G} = \{[\lambda, c, \gamma]\} ,$$

where, under Formulation One,

$$[\lambda, c, \gamma]\ (\mu_z, \mu_y, \sigma_z^2, \sigma_y^2, \beta) =$$

$$= (\lambda\mu_z, \gamma\mu_y, \lambda^2\sigma_z^2, \gamma^2\sigma_y^2, \gamma\lambda^{-1}\beta + c) ,$$

and under Formulation Two,

$$[\lambda, c, \gamma]\ (\mu_z, \mu_y, \sigma_z^2, \sigma^2, \beta) =$$

$$= (\lambda\mu_z, \gamma\mu_y, \lambda^2\sigma_z^2, \gamma^2\sigma^2, \gamma\lambda^{-1}\beta + c) .$$

The jointly sufficient statistic for $(\mu_z, \mu_y, \sigma_z^2, \sigma_y^2, \beta)$ is usually written

$$(\bar{z}, \bar{y}, s_z^2, s_y^2, \hat{\beta})$$

and the sufficient statistic for $(\mu_z, \mu_y, \sigma_z^2, \sigma^2, \beta)$ is usually written

$$(\bar{z}, \bar{y}, s_z^2, s^2, \hat{\beta}) ;$$

and the induced group G on the sufficient statistic space acts exactly like $\bar{G}$ acts on Θ.

The groups $\bar{H}$, $\bar{H}_1$, $\bar{H}_2$ and the induced groups H, H_1 and H_2 are defined as before, leading to the maximal invariant parametric functions

$$B_1 = \sigma_z\beta/\sigma_y \text{ for Formulation One}$$

and

$$B_2 = \sigma_z\beta/\sigma \text{ for Formulation Two} .$$

Note that both B_1 and B_2 are standardized regression coefficients; the

standardizations, however, are different.

Referring back to the example in which the researcher wishes to predict change in drug use from change in attitude towards drugs, B_1 measures standardized change in drug use rate for a one-standard deviation change in attitude, as compared to all communities, regardless of their attitudes. B_2 measures the standardized change in use rate as compared to other communities of the same attitude.

The invariant sufficient estimator of B_1 produced by the theory introduced in this dissertation is

$$\hat{B} = s_z \hat{\beta} / s_y$$

and the invariant sufficient estimator of B_2 is

$$\tilde{B} = s_z \hat{\beta} / s \ .$$

In the stochastic input case, both $\tilde{B}$ and $\hat{B}$ are reasonable estimators of their respective parameters. In the next subsection we will consider the consistency and distributions of the estimators.

Section 8.2. Properties of the Estimator in Formulation One

Let us first consider $\hat{B}$ as an estimator of B_1. We note that $\hat{B}$ is equivalent to the sample correlation coefficient $\hat{\rho}$ (recalling that $k = 1$). To show this, let

$$S_{zz} = \sum_{i=1}^{n} (z_i - \bar{z})^2 = (n-1)\ s_z^2 \ ,$$

$$S_{yy} = \sum_{i=1}^{n} (y_i - \bar{y})^2 = (n-1)\ s_y^2 \ ,$$

and

$$S_{zy} = \sum_{i=1}^{n} (z_i - \bar{z})\ (y_i - \bar{y})\ .$$

Then

$$\hat{B} = \frac{s_z \hat{\beta}}{s_y} = \frac{(n-1)^{1/2}\ s_z}{(n-1)^{1/2}\ s_y}\ \hat{\beta}$$

$$= \left(\frac{S_{zz}}{S_{yy}}\right)^{1/2} \hat{\beta}$$

$$= \left(\frac{S_{zz}}{S_{yy}}\right)^{1/2} \frac{S_{zy}}{S_{zz}}$$

$$= \frac{S_{zy}}{(S_{zz}\ S_{yy})^{1/2}} = \hat{\rho}\ .$$

To show that $\hat{B}$ is a consistent estimator of B_1, we note that, by Slutsky's theorem,

$$\hat{B} = \frac{s_z \hat{\beta}}{s_y} \quad \text{converges to} \quad \frac{\sigma_z \beta}{\sigma_y} = B_1\ .$$

For the distribution of $\hat{B}$, recall that in Lemma 6.2.2, when z was nonstochastic we found its distribution was

$$f(w) = \frac{2e^{-\lambda}\,(1-w^2)^{\frac{r-2}{2}}}{\Gamma\left(\frac{r}{2}\right)} \sum_{i=0}^{\infty} \frac{\lambda^i}{i!} \frac{\Gamma\left(\frac{2i+1+r}{2}\right)}{\Gamma\left(\frac{2i+1}{2}\right)} w^{2i}$$

where $r = n - 2$ and $\lambda = (n-1)\, s_z^{\,2}\, \beta^2/2\sigma^2$. That is,

$$f(w) = \frac{2e^{-(n-1)\, s_z^{\,2}\, \beta^2/2\sigma^2}\,(1-w^2)^{\frac{r-2}{2}}}{\Gamma\left(\frac{r}{2}\right)} \sum_{i=0}^{\infty} \left[\frac{(n-1)\, s_z^{\,2}\, \beta^2}{2\sigma^2}\right]^i \frac{\Gamma\left(\frac{2i+1+r}{2}\right)}{i!\;\Gamma\left(\frac{2i+1}{2}\right)} w^{2i} .$$

$f(w)$ can be shown to be the conditional density of $\hat{\beta}$ given z in the stochastic case. Considering the case in which z is normal, note that $f(w)$ depends on z only through $s_z^{\,2}$. Let $x = (n-1)\, s_z^{\,2}/\sigma_z^{\,2}$; then

$$x \sim \chi^2_{n-1}$$

and the marginal density of x is

$$g(v) = \frac{1}{2^{\frac{r+1}{2}}\,\Gamma\left(\frac{r+1}{2}\right)}\, v^{\frac{r-1}{2}}\, e^{-v/2} \text{ for } v > 0 ,$$

where $r = n - 2$.

Let

$$u = \frac{(n-1)\, s_z^{\,2}}{\sigma^2} = \frac{\sigma_z^{\,2}}{\sigma^2}\, \frac{(n-1)\, s_z^{\,2}}{\sigma_z^{\,2}} = \frac{\sigma_z^{\,2}}{\sigma^2}\, x .$$

Then

$$x = \frac{\sigma^2}{\sigma_z^2} u$$

and

$$dx = \frac{\sigma^2}{\sigma_z^2} du$$

and the density of u is

$$g(t) = \frac{1}{2^{\frac{r+1}{2}} \Gamma\left(\frac{r+1}{2}\right)} \left(\frac{\sigma^2}{\sigma_z^2} t\right)^{\frac{r-1}{2}} e^{-\frac{\sigma^2}{\sigma_z^2} t/2} \left(\frac{\sigma^2}{\sigma_z^2}\right) \text{ for } t > 0 .$$

The conditional density $f_{\hat{B}|u}(w)$ may be written

$$f_{\hat{B}|u}(w) = \frac{2e^{-t\beta^2/2} (1-w^2)^{\frac{r-2}{2}}}{\Gamma\left(\frac{r}{2}\right)} \sum_{i=0}^{\infty} \left(\frac{t\beta^2}{2}\right)^i \frac{\Gamma\left(\frac{2i+1+r}{2}\right)}{i!\ \Gamma\left(\frac{2i+1}{2}\right)} w^{2i} .$$

Then the joint density of $\hat{\beta}$ and u can be written $f_{\hat{B},u}(w,t) = f_{\hat{B}|u}(w)g(t) =$

$$= \frac{e^{-\frac{t}{2}(\beta^2 + \sigma^2/\sigma_z^2)} (1-w^2)^{\frac{r-2}{2}} t^{\frac{r-1}{2}} (\sigma/\sigma_z)^{r+1}}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r}{2}\right) \Gamma\left(\frac{r+1}{2}\right)} \sum_{i=0}^{\infty} \left(\frac{t\beta^2}{2}\right)^i \frac{\Gamma\left(\frac{2i+1+r}{2}\right)}{i!\ \Gamma\left(\frac{2i+1}{2}\right)} w^{2i}$$

$$= \frac{(1-w^2)^{\frac{r-2}{2}} (\sigma/\sigma_z)^{r+1}}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r}{2}\right) \Gamma\left(\frac{r+1}{2}\right)} \sum_{i=0}^{\infty} \frac{(w\beta)^{2i}}{2^i} \frac{\Gamma\left(\frac{2i+1+r}{2}\right)}{i!\ \Gamma\left(\frac{2i+1}{2}\right)} t^{\frac{r+2i-1}{2}} e^{-\frac{t}{2}(\beta^2+\sigma^2/\sigma_z^2)} .$$

The marginal density of $\hat{B}$ is $f_{\hat{B}}(w) = \int_t f_{\hat{B},u}(w,t)\, dt$

$$= C \sum_{i=0}^{\infty} G_i \int_0^{\infty} t^{\frac{r+2i-1}{2}} e^{-\frac{t}{2}(\beta^2+\sigma^2/\sigma_z^2)} dt$$

where

$$C = \frac{(1-w^2)^{\frac{r-2}{2}} (\sigma/\sigma_z)^{r+1}}{2^{\frac{r-1}{2}} \Gamma\left(\frac{r}{2}\right) \Gamma\left(\frac{r+1}{2}\right)}$$

and

$$G_i = \frac{(w\beta)^{2i}}{2^i} \frac{\Gamma\left(\frac{2i+1+r}{2}\right)}{i!\ \Gamma\left(\frac{2i+1}{2}\right)}$$

and the integrand is in the form of a gamma density with parameters

$$a = (r+2i+1)/2$$

and

$$b = 2/(\beta^2+\sigma^2/\sigma_z^{\ 2}) \ .$$

Thus

$$f_{\hat{B}}(w) = C \sum_{i=0}^{\infty} G_i \ \Gamma\left(\frac{r+2i+1}{2}\right) \left(\frac{2}{\beta^2+\sigma^2/\sigma_z^{\ 2}}\right)^{\frac{r+2i+1}{2}}$$

$$= C \sum_{i=0}^{\infty} \frac{\beta^{2i} \ w^{2i} \ 2^{\frac{r+1}{2}} \ \Gamma^2[(2i+1+r)/2]}{i! \ \Gamma\left(\frac{2i+1}{2}\right) \ (\beta^2+\sigma^2/\sigma_z^{\ 2})^{\frac{r+2i+1}{2}}}$$

$$= \frac{2(1-w^2)^{\frac{r-2}{2}} \ (\sigma/\sigma_z)^{r+1}}{} \sum_{i=0}^{\infty} \frac{\beta^{2i} \ w^{2i} \ \Gamma^2[(2i+1+r)/2]}{i! \ \Gamma\left(\frac{2i+1}{2}\right) \ (\beta^2+\sigma^2/\sigma_z^{\ 2})^{\frac{r+2i+1}{2}}} \ ;$$

$f_{\hat{B}}(w)$ is the density of $\hat{B}$ when y and z are bivariate normal.

Note that if $\beta = 0$, all the terms in the sum will be zero except the first (for i=0). Then $f_{\hat{B}}(w)$ becomes

$$f_{\hat{B}}(w) = \frac{2(1-w^2)^{\frac{r-2}{2}} \ (\sigma/\sigma_z)^{r+1}}{\Gamma\left(\frac{r}{2}\right) \ \Gamma\left(\frac{r+1}{2}\right)} \ \frac{\Gamma^2\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \ (\sigma^2/\sigma_z^{\ 2})^{\frac{r+1}{2}}}$$

$$= \frac{2(1-w^2)^{\frac{r-2}{2}} \; \Gamma\left(\frac{r+1}{2}\right)}{\Gamma\left(\frac{r}{2}\right) \; \Gamma\left(\frac{1}{2}\right)} .$$

Graybill (1961) gives the density of the sample correlation coefficient $\hat{\rho}$, when $\rho = 0$, as

$$f(\hat{\rho};0) = \frac{\Gamma\left(\frac{n-1}{2}\right) \; (1-\hat{\rho}^2)^{\frac{n-4}{2}}}{\Gamma\left(\frac{n-2}{2}\right) \; \Gamma\left(\frac{1}{2}\right)} \text{ for } -1 \leq \hat{\rho} \leq 1 .$$

If we consider that $f_{\hat{B}}(w)$ was derived for $w > 0$ and recall that $r = n - 2$, we see that for the case in which $\beta = 0$, the central density for $\hat{B}$ reduces to the central density for $\hat{\rho}$. Note also that the density, $f_{\hat{B}|s}(w)$, of $\hat{B}$ in the nonstochastic input case (Lemma 6.2.2) reduces to the same density, although $\hat{B}$ cannot be interpreted as a correlation coefficient if z is nonrandom.

Section 8.3. Properties of the Estimator in Formulation Two

Directing our attention to $\tilde{B}$, it is apparent that $\tilde{B}$ is a consistent estimator of B_2 since by Slutsky's theorem,

$$\tilde{B} = \frac{s_z \hat{\beta}}{s} \quad \text{converges to} \quad \frac{\sigma_z \beta}{\sigma} = B_2 .$$

When z was nonstochastic, we found that $\tilde{B}$ was distributed as the constant $(n-1)^{1/2}$ times a noncentral t distirbution with $r = n - 2$

degrees of freedom and noncentrality parameter $\delta = (n-1)^{1/2} s_z \beta/\sigma$ (Theorem 6.1.2). That is,

$$f(u) = \frac{(n-1)^{1/2} r^{r/2}}{\pi^{1/2} \Gamma\left(\frac{r}{2}\right)} \frac{e^{-\delta^2/2}}{(r+u^2)^{\frac{r+1}{2}}} \sum_{i=0}^{\infty} \frac{\Gamma[(r+i+1)/2]\, \delta^i\, 2^{i/2}\, u^i}{i!\, (r+u^2)^{i/2}} ,$$

or since $n-1 = r+1$ and $\delta = (r+1)^{1/2} s_z \beta/\sigma$,

$$f(u) = \frac{(r+1)^{1/2} r^{r/2} e^{-\frac{(r+1)s_z^2 \beta^2}{2\sigma^2}}}{\pi^{1/2} \Gamma\left(\frac{r}{2}\right) (r+u^2)^{\frac{r+1}{2}}} \sum_{i=0}^{\infty} \frac{\Gamma\left(\frac{r+i+1}{2}\right) \left[\frac{(r+1)s_z^2 \beta^2}{\sigma^2}\right]^{i/2} 2^{i/2} u^i}{i!\, (r+u^2)^{i/2}} .$$

In the stochastic input case $f(u)$ is the conditional density of $\hat{B}$ given z. Note that $f(u)$ depends on z only through s_z^2. Let $x = (r+1)\, s_z^2/\sigma_z^2$; then x is distributed as a chi-square variable with $n-1 = r+1$ degrees of freedom and thus has the density

$$g(w) = \frac{1}{2^{\frac{r+1}{2}} \Gamma\left(\frac{r+1}{2}\right)} w^{\frac{r-1}{2}} e^{-\frac{w}{2}} \text{ for } w > 0 .$$

Let

$$v = \frac{(r+1)\, s_z^2}{\sigma^2} = \frac{\sigma_z^2}{\sigma^2} \frac{(r+1)\, s_z^2}{\sigma_z^2} = \frac{\sigma_z^2}{\sigma^2} x .$$

Then $\sigma^2 v/\sigma_z^{\ 2} = x$ and $(\sigma^2/\sigma_z^{\ 2})\ dv = dx$ and the density of v is

$$g(t) = \frac{1}{2^{\frac{r+1}{2}}\ \Gamma\left(\frac{r+1}{2}\right)} \left(\frac{\sigma^2}{\sigma_z^{\ 2}}\ t\right)^{\frac{r-1}{2}} e^{-\frac{\sigma^2}{\sigma_z^{\ 2}}\ t/2} \left(\frac{\sigma^2}{\sigma_z^{\ 2}}\right) \text{ for } t > 0 .$$

Then the conditional density of $\tilde{B}$ given v may be written

$$f_{\tilde{B}|v}(u) = \frac{(r+1)^{1/2}\ r^{r/2}\ e^{-\frac{t\beta^2}{2}}}{\pi^{1/2}\ \Gamma\left(\frac{r}{2}\right)\ (r+u^2)^{\frac{r+1}{2}}} \sum_{i=0}^{\infty} \frac{\Gamma[(r+i+1)/2]\ (t\beta^2)^{i/2}\ 2^{i/2}\ u^i}{i!\ (r+u^2)^{i/2}} .$$

Then $f_{\tilde{B},v}(u,t) = f_{\tilde{B}|v}(u)\ g(t)$

$$= \frac{(r+1)^{1/2}\ r^{r/2}\ e^{-\frac{t}{2}(\beta^2+\sigma^2/\sigma_z^{\ 2})}\ (\sigma/\sigma_z)^{r+1}\ t^{\frac{r-1}{2}}}{2^{\frac{r+1}{2}}\ \Gamma\left(\frac{r+1}{2}\right)\ \pi^{1/2}\ \Gamma\left(\frac{r}{2}\right)\ (r+u^2)^{\frac{r+1}{2}}} \cdot$$

$$\cdot \sum_{i=0}^{\infty} \frac{\Gamma[(r+i+1)/2](t\beta^2)^{i/2}\ 2^{i/2}\ u^i}{i!\ (r+u^2)^{i/2}}$$

$$= \frac{(r+1)^{1/2}\ r^{r/2}\ (\sigma/\sigma_z)^{r+1}}{2^{\frac{r+1}{2}}\ \Gamma\left(\frac{r+1}{2}\right)\ \pi^{1/2}\ \Gamma\left(\frac{r}{2}\right)\ (r+u^2)^{\frac{r+1}{2}}} \cdot$$

$$\cdot \sum_{i=0}^{\infty} \frac{\Gamma[(r+i+1)/2]\ \beta^i\ 2^{i/2}\ u^i}{i!\ (r+u^2)^{i/2}}\ t^{\frac{r+i-1}{2}}\ e^{-\frac{t}{2}(\beta^2+\sigma^2/\sigma_z{}^2)} .$$

The marginal density of $\tilde{B}$ is obtained by integrating $f_{\tilde{B},V}(u,t)$ over t:

$$f_{\tilde{B}}(u) = k \sum_{i=0}^{\infty} g_i \int_0^{\infty} t^{\frac{r+i-1}{2}}\ e^{-\frac{t}{2}(\beta^2+\sigma^2/\sigma_z{}^2)}\ dt$$

where

$$k = \frac{(r+1)^{1/2}\ r^{r/2}\ (\sigma/\sigma_z)^{r+1}}{2^{\frac{r+1}{2}}\ \Gamma\left(\frac{r+1}{2}\right)\ \pi^{1/2}\ \Gamma\left(\frac{r}{2}\right)\ (r+u^2)^{\frac{r+1}{2}}} .$$

and

$$g_i = \frac{\Gamma[(r+i+1)/2]\ \beta^i\ 2^{i/2}\ u^i}{i!\ (r+u^2)^{i/2}} .$$

The integrand is in the form of a gamma with parameters

$$a=(r+i+1)/2$$

and

$$b=2/(\beta^2+\sigma^2/\sigma_z{}^2) \ .$$

Thus

$$f_{\tilde{B}}(u) = k \sum_{i=0}^{\infty} g_i \ \Gamma\left(\frac{r+i+1}{2}\right) \left(\frac{2}{\beta^2+\sigma^2/\sigma_z{}^2}\right)^{\frac{r+i+1}{2}}$$

$$= k \sum_{i=0}^{\infty} \frac{\Gamma^2\left(\frac{r+i+1}{2}\right) \beta^i \ u^i \ 2^{\frac{r+2i+1}{2}}}{i! \ (r+u^2)^{i/2} \ (\beta^2+\sigma^2/\sigma_z{}^2)^{\frac{r+i+1}{2}}}$$

$$= \frac{(r+1)^{1/2} \ r^{r/2} \ (\sigma/\sigma_z)^{r+1}}{\Gamma\left(\frac{r+1}{2}\right) \ \pi^{1/2} \ \Gamma\left(\frac{r}{2}\right) \ (r+u^2)^{\frac{r+1}{2}}} \cdot$$

$$\cdot \sum_{i=0}^{\infty} \frac{\Gamma^2\left(\frac{r+i+1}{2}\right) \ (2u\beta)^i}{i! \ (r+u^2)^{i/2} \ (\beta^2+\sigma^2/\sigma_z{}^2)^{\frac{r+i+1}{2}}} \ .$$

If $\beta = 0$, then $f_{\tilde{B}}(u)$ becomes

$$f_{\tilde{B}}(u) = \frac{(r+1)^{1/2}\ r^{r/2}\ (\sigma/\sigma_z)^{r+1}}{\Gamma\left(\frac{r+1}{2}\right)\ \pi^{1/2}\ \Gamma\left(\frac{r}{2}\right)\ (r+u^2)^{\frac{r+1}{2}}}\ \frac{\Gamma^2\left(\frac{r+1}{2}\right)}{(\sigma^2/\sigma_z^{\ 2})^{\frac{r+1}{2}}}$$

$$= \frac{(r+1)^{1/2}\ r^{r/2}\ \Gamma\left(\frac{r+1}{2}\right)}{\pi^{1/2}\ \Gamma\left(\frac{r}{2}\right)\ (r+u^2)^{\frac{r+1}{2}}},$$

which is $(r+1)^{1/2} = (n-1)^{1/2}$ times the central t distribution. In the central case, then, the distribution of $\tilde{B}$ is the same, whether z is stochastic or nonstochastic. This suggests a natural procedure for testing the hypothesis B = 0 in the stochastic input case.

CHAPTER IX

ESTIMATION IN THE LINEAR MODEL: INVARIANCE UNDER NONSINGULAR TRANSFORMATIONS

In this chapter the proposed method of estimating orbits is applied to the problem of estimating the parameters in the linear model in a manner invariant under nonsingular transformations. It was hoped that such a procedure might lead to an estimator "invariant" to the effects of multicolinearity.

Consider the problem of estimation in the model

$$\underline{y} = \underline{Z}\,\underline{\beta} + \underline{\varepsilon}$$

where $\underline{Z}$ is a non-stochastic input matrix such that $\underline{Z}'\,\underline{Z}$ is very nearly singular. (For simplicity of notation we do not assume $\underline{Z}$ has a column of ones and thus the model has no intercept.) That is, the eigenvalues λ_i of $\underline{Z}'\,\underline{Z}$ are such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$$

but

$$\lambda_p < \xi$$

where ξ is a pre-selected small positive number.

This situation arises when the input variables $z_1,\ldots,z_p$ are very highly correlated. The result is that the net regression coefficients $\hat{\beta}_1,\ldots,\hat{\beta}_p$ are very unreliable estimates of the effects of the independent variables, since the generalized variance of $\hat{\underline{\beta}}$,

$$V(\hat{\underline{\beta}}) = \sigma^2 \left| (\underline{Z}'\,\underline{Z})^{-1} \right| = \sigma^2 \prod_{i=1}^{p} \lambda_i^{-1} ,$$

tends to infinity as λ_p approaches zero.

It seems natural to try to obtain an estimator invariant under the group of orthogonal transformations on the input variables. However, application of an orthogonal transformation does not reduce the generalized variance. If we write the model as

$$\underline{y} = \underline{Z}\ \underline{P}\ \underline{P}'\ \underline{\beta} + \underline{\varepsilon}$$

$$= \underline{U}\ \underline{\delta} + \underline{\varepsilon}\ ,$$

where $\underline{P}$ is a $p \times p$ orthogonal matrix;

$$\underline{U} = \underline{Z}\ \underline{P}$$

and

$$\underline{\delta} = \underline{P}'\ \underline{\beta}\ .$$

Then

$$\underline{U}'\ \underline{U} = \underline{P}'\ \underline{Z}'\ \underline{Z}\ \underline{P} = \underline{Q}'\underline{D}\ \underline{Q}$$

where Q is orthogonal, and

$$\underline{D} = \mathrm{diag}(\lambda_1,\ldots,\lambda_p)\ .$$

Thus

$$\hat{\underline{\delta}} = \underline{P}'\ \hat{\underline{\beta}}\ ,$$

and

$$V(\hat{\underline{\delta}}) = \sigma^2|\underline{D}| = \sigma^2 \prod_{i=1}^{p} \lambda_i^{-1}\ ,$$

so that the generalized variance is not reduced.

Thus we consider a group of transformations

$$G^* = \{[\underline{N},\underline{c},\gamma]\} ,$$

where $\underline{N}$ is a $p \times p$ nonsingular matrix, $\underline{c}$ is an $n \times 1$ vector, and γ is a positive scalar, acting on the sample space

$$\mathcal{X}_0 = \{(\underline{Z},\underline{y})\} = \mathcal{A}(\underline{Z}_0) \times R^n$$

where

$$\underline{Z} \in \mathcal{A}(\underline{Z}_0)$$

if

$$\underline{Z} = \underline{Z}_0 \underline{A}$$

for a nonsingular $p \times p$ matrix $\underline{A}$ and a fixed matrix $\underline{Z}_0$ which satisfies the condition

$$\underline{Z}_0' \underline{Z}_0 = \underline{I} .$$

The action of G^* on $\mathcal{X}_0$ is defined by

$$[\underline{N},\underline{c},\gamma] \ (\underline{Z},\underline{y}) = (\underline{Z} \ \underline{N}, \gamma\underline{y} + \underline{Z} \ \underline{N} \ \underline{c}) .$$

We may write the model $\underline{y} = \underline{Z} \ \underline{\beta} + \underline{\varepsilon}$ as

$$\underline{y} = \underline{Z}_0 \ \underline{A} \ \underline{\beta} + \underline{\varepsilon}$$

$$= \underline{Z}_0 \ \underline{\tau} + \underline{\varepsilon} \qquad (9.1)$$

where

$$\underline{\tau} = \underline{A}\ \underline{\beta}\ .$$

The parameter space consists of elements of the form

$$\theta = (\underline{A},\underline{\beta},\sigma^2)$$

and the action of the induced group $\bar{G}$ on Θ is defined by

$$[\underline{N},\underline{c},\gamma]\ (\underline{A},\underline{\beta},\sigma^2) = (\underline{A}\ \underline{N},\ \gamma\underline{N}^{-1}\ \underline{\beta} + \underline{c},\ \gamma^2\sigma^2)\ .$$

The group of transformations $\bar{G}$ is isomorphic to Θ, where $(\underline{Z}_0,\underline{0},1)$ is chosen as a reference point in Θ and is also isomorphic to $\bar{H}$, the semi-direct product of the groups

$$\bar{H}_2 = \{[\underline{I},\underline{c},\gamma]\}$$

and

$$H_1 = \{[\underline{N},\underline{0},1]\}\ .$$

It is simple to show that $\bar{H}_2$ is a normal subgroup of $\bar{H}$. Thus the theory presented in this dissertation applies.

The maximal invariant parametric function is

$$\psi(\theta) = \bar{h}_{1\theta}^{-1}\ \theta = [\underline{A}^{-1},\underline{0},1]\ (\underline{A},\underline{\beta},\sigma^2)$$

$$= (\underline{I},\underline{A}\ \underline{\beta},\sigma^2)\ .$$

In order to estimate θ invariantly under the subgroup $\bar{H}_1$, one could use the maximal invariant function

$$T(x) = h_{1x}^{-1}\, x = [\underline{A}^{-1}, \underline{0}, 1]\ (\underline{A}, \hat{\underline{\beta}}, s^2)$$

$$= (\underline{I}, \underline{A}\ \hat{\underline{\beta}}, s^2)\ .$$

This estimator is nothing more than the estimator obtained under an orthonormalization of the problem. This is easily seen if we consider the model written in (9.1),

$$\underline{y} = \underline{Z}_0\ \underline{\tau} + \underline{\varepsilon}\ .$$

Since $\underline{Z}_0'\ \underline{Z}_0 = \underline{I}$, this model is orthonormalized:

$$\hat{\underline{\tau}} = (\underline{Z}_0'\ \underline{Z}_0)^{-1}\ \underline{Z}_0'\ \underline{y} = \underline{Z}_0'\ \underline{y}\ .$$

But $\underline{Z} = \underline{Z}_0\ \underline{A}$ implies that $\underline{Z}_0' = (\underline{Z}\ \underline{A}^{-1})' = \underline{A}^{-1'}\ \underline{Z}'$, so

$$\hat{\underline{\tau}} = \underline{A}^{-1'}\ \underline{Z}'\ \underline{y}$$

$$= \underline{A}^{-1'}\ (\underline{Z}'\ \underline{Z})\ (\underline{Z}'\ \underline{Z})^{-1}\ \underline{Z}'\ \underline{y}$$

$$= \underline{A}^{-1'}\ (\underline{A}'\ \underline{Z}_0'\ \underline{Z}_0\ \underline{A})\ \hat{\underline{\beta}}$$

$$= \underline{A}^{-1'}\ (\underline{A}'\ \underline{A})\ \hat{\underline{\beta}}$$

$$= \underline{A}\ \hat{\underline{\beta}}\ .$$

The generalized variance of this estimator is

$$V(\hat{\underline{\tau}}) = \sigma^2(\underline{Z}_0'\ \underline{Z}_0) = \sigma^2\underline{I}\ ,$$

so that the generalized variance is constant regardless of the value of $\underline{Z}'\ \underline{Z}$. Clearly, the estimator $\hat{\underline{\tau}}$ is unaffected by multicolinearity since the generalized variance of $\hat{\underline{\tau}}$ is $\sigma^2\underline{I}$ regardless of the generalized

variance of $\hat{\underline{\beta}}$.

Thus it is a simple matter to estimate $\underline{\beta}$ invariantly under the group of nonsingular transformations. Interpretation of $\hat{\underline{\tau}} = \underline{A}\,\hat{\underline{\beta}}$, however, is unclear. If the orthonormalized input variables have physical meaning in a given situation, the interpretation of $\hat{\underline{\tau}}$ presents no problem. Otherwise, this approach does not seem to be profitable in any actual research situations.

We note in closing that it might be possible to generate a group of transformations which affects only the smallest eigenvalues of $\underline{Z}'\,\underline{Z}$. If such a group can be formulated and the estimation problem carries the group naturally, it may lead to a quite useful tool with which to tackle the problem of multicolinearity.

CHAPTER X

EXAMPLES OF APPLICATION OF THE PROPOSED PROCEDURES

Section 10.1. Application to Sociological Data

As an application of the methods developed previously, we consider a sociological linear model provided by Dr. D. L. Klemmack, Department of Sociology, Virginia Polytechnic Institute and State University. Models such as the one presented are called "path models" in sociology. As indicated in Figure 10.1.1 the effect of any independent variable upon a dependent variable can be represented by an arrow, or "path" from the independent variable to the dependent variable. The strength of the relationship indicated by the arrow is denoted by a "path" or "beta" coefficient. This diagram leads to the following multiple regression equations

$$X_7 = \alpha_7 + \beta_{76} X_6 + \varepsilon_7$$

$$X_8 = \alpha_8 + \beta_{87} X_7 + \beta_{81} X_1 + \varepsilon_8$$

$$X_9 = \alpha_9 + \beta_{98} X_8 + \beta_{95} X_5 + \beta_{94} X_4 + \varepsilon_9$$

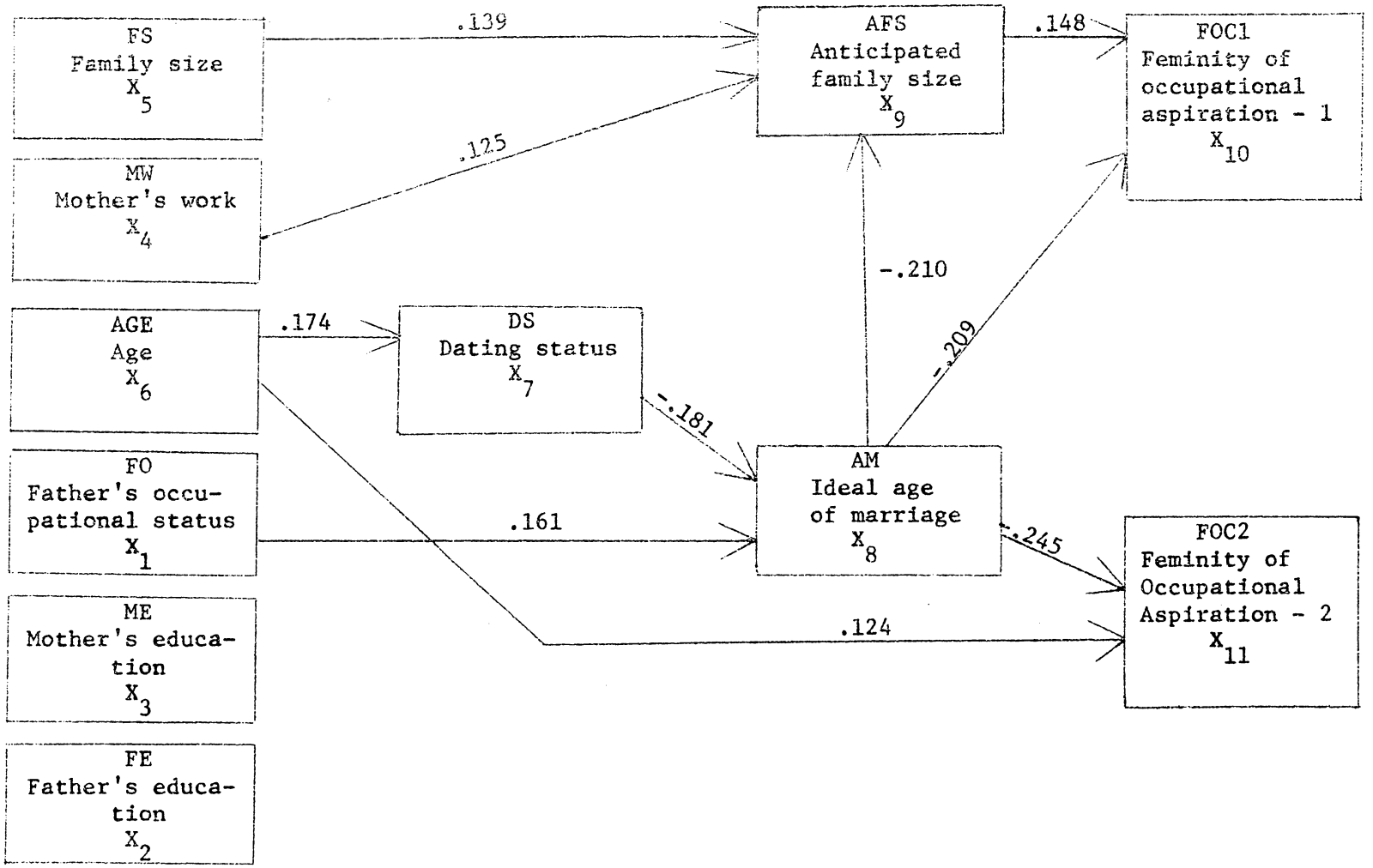
$$X_{10} = \alpha_{10} + \beta_{10,9} X_9 + \beta_{10,8} X_8 + \varepsilon_{10}$$

$$X_{11} = \alpha_{11} + \beta_{11,8} X_8 + \beta_{11,6} X_6 + \varepsilon_{11} .$$

We can also consider the equations

$$X_{10} = \alpha_{10} + \beta_{10,1} X_1 + \beta_{10,2} X_2 + \ldots + \beta_{10,9} X_9 + \varepsilon_{10}$$

Figure 10.1.1. Path diagram showing influence of family characteristics and age, marriage aspirations (from Klemmack).



$$X_{11} = \alpha_{11} + \beta_{11,1} X_1 + \beta_{11,2} X_2 + \ldots + \beta_{11,9} X_9 + \varepsilon_{10} .$$

We note from Figure 10.1.1 that the variables X_2 and X_3 appeared not to be significantly related to the intervening and/or dependent variables and were therefore deleted from the path diagram by Professor Klemmack.

Calculations of "path coefficients" were based on a sample of 300 observations (with some missing data) and were performed on the Virginia Polytechnic Institute and State University computer system with a modified version of the SPSS computer package. The results are presented in Tables 10.1.1 and 10.1.2.

Table 10.1.1 displays values of $\hat{B}$, $\tilde{B}$ and B^* for the five multiple regression equations indicated in the Path Diagram of Figure 10.1.1 and Table 10.1.2 displays $\hat{B}$, $\tilde{B}$ and B^* for the cases in which $X_1, X_2, \ldots, X_9$ are all included as independent variables to predict X_{10} and X_{11}.

To obtain $\tilde{B}$ from $\hat{B}$, it is necessary only to multiply $\hat{B}$ by s_y/s, since

$$\frac{s_y}{s} \hat{B} = \frac{s_y}{s} \frac{s_x \hat{\beta}}{s_y} = \frac{s_x \hat{\beta}}{s} = \tilde{B} .$$

For example, when considering the relationship

$$X_7 = \alpha_7 + \beta_{76} X_6 + \varepsilon_7$$

as in the first row of Table 10.1.1 $s_y = .5004$ and $s = .49377$, so that

$$\tilde{B} = \frac{.5004}{.49377} \hat{B} = (1.01343)\ (.17353) = .17586 .$$

To obtain B^* it was necessary to calculate the value of the constant

$$C = \left(\frac{2}{n-2}\right)^{1/2} \frac{\Gamma\left(\frac{n-2}{2}\right)}{\Gamma\left(\frac{n-3}{2}\right)} .$$

We used the following approximation to the gamma function, found in the C.R.C. Standard Mathematical Tables:

$$\Gamma(x) \approx x^x e^{-x} \left(\frac{2\pi}{x}\right)^{1/2} \left(1 + \frac{1}{12x}\right) .$$

For $n = 300$, $C \approx .99944$. Then to obtain B^*, one must simply multiply the value of $\tilde{B}$ by .99944. For example, in the first row of Table 1,

$$B^* = .99944\ \tilde{B} = .99944\ (.17586) = .17576 .$$

We note that there is very little difference in the values of $\hat{B}$, $\tilde{B}$ and B^* for any of the regressions considered. The reason that $\tilde{B}$ and B^* differ so little is, of course, due to the fact that C is close to one for a sample of 300. Comparison of s_y and s for each situation reveals that in no case is s_y very much larger than s; thus the ratio s_y/s will be close to one and $\tilde{B}$ will not differ much from $\hat{B}$.

That s and s_y are almost equal indicates very poor correlation between the dependent and independent variables. Inspection of the correlation matrix in Table 10.1.3 reveals that there is, indeed, very little correlation among any of the variables.

Further, we can look at the multiple coefficients of determination for each regression. These express the ratio of the variance in the

Table 10.1.1. Values of $\hat{B}$, $\tilde{B}$ and B^* for the regressions indicated in Figure 10.1.1.

Dependent Variable	Independent Variable(s)	Marginal Standard Deviation of Dependent Variable	Standard Error	Regression Coefficient	$\hat{B}$	$\tilde{B}$	B^*
X_7(DS)	X_6(AGE)	.5004	.49377	.08122	.17353	.17586	.17576
	X_1(FO)			.01257	.16149	.16572	.16563
X_8(AM)		1.8299	1.78320				
	X_7(DS)			-.66009	-.18052	-.18525	-.18515
	X_4(MW)			.10129	.12470	.12894	.12887
X_9(AFS)	X_5(FS)	1.1884	1.14936	.11198	.13902	.14374	.14366
	X_8(AM)			-.13669	-.21048	-.21763	-.21751
	X_8(AM)			-.05593	-.20874	-.21663	-.21651
X_{10}(FOC1)		.4903	.47245				
	X_9(AFS)			.06111	.14811	.15371	.15362
	X_6(AGE)			.04971	.12370	.12796	.12789
X_{11}(FOC2)		.4296	.41529				
	X_8(AM)			-.05762	-.24539	-.25385	-.25371

Table 10.1.2. Values of $\hat{B}$, $\tilde{B}$ and B^* for predicting X_{10}(FOC1) and X_{11}(FOC2) from the other nine variables.

Dependent Variable	Independent Variables	Marginal Standard Deviation of Dependent Variable	Standard Error	Regression Coefficient	$\hat{B}$	$\tilde{B}$	B^*
	X_1(FO)			.00071	.03388	.03488	.03486
	X_2(FE)			-.01508	-.05490	-.05651	-.05648
	X_3(ME)			.00876	.02442	.02514	.02513
	X_4(MW)			-.02230	-.06655	-.06850	-.06846
X_{10}(FOC1)	X_5(FS)	.4903	.47631	-.01210	-.03642	-.03749	-.03747
	X_6(AGE)			-.02006	-.04375	-.04504	.04502
	X_7(DS)			-.03958	-.04040	-.04159	-.04157
	X_8(AM)			-.05654	-.21101	-.21721	-.21709
	X_9(AFS)			.06794	.16465	.16949	.16940
	X_1(FO)			.00097	.05310	.05456	.05453
	X_2(FE)			.00197	.00818	.00840	.00840
	X_3(ME)			-.01764	-.05612	-.05766	-.05763
	X_4(MW)			-.00393	-.01338	-.01375	-.01374
X_{11}(FOC2)	X_5(FS)	.4296	.41809	-.00830	-.02850	-.02928	-.02926
	X_6(AGE)			.04447	.11067	.11372	.11366

Table 10.1.2 continued

Dependent Variable	Independent Variables	Marginal Standard Deviation of Dependent Variable	Standard Error	Regression Coefficient	$\hat{B}$	$\tilde{B}$	B^*
	X_7 (DS)			.02838	.03305	.03396	.03394
	X_8 (AM)			-.05337	-.22732	-.23358	-.23345
	X_9 (AFS)			.02876	.07955	.08174	.08169

Table 10.1.3. Correlation matrix for all eleven variables.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
X_1	1.00000	.50453	.25222	-.21324	-.02433	-.01155	.02036	.15781	-.09399	-.02168	.00266
X_2		1.00000	.44029	-.19749	.05317	-.14048	-.06958	.11578	-.09081	-.04628	-.04001
X_3			1.00000	.02456	.05223	-.08717	-.07356	.07645	-.00989	-.00572	-.07119
X_4				1.00000	-.12161	-.07787	-.05463	.00673	.10638	-.03619	-.02772
X_5					1.00000	.09305	-.00254	-.00798	.12553	-.01241	-.00865
X_6						1.00000	.17353	.04792	.06173	-.04372	.11194
X_7							1.00000	-.17724	.10460	.01308	.10631
X_8								1.00000	-.21075	-.23995	-.23947
X_9									1.00000	.19211	.12756
X_{10}										1.00000	.46052
X_{11}											1.00000

dependent variable accounted for, or explained by, the independent variable(s) to the total variance of the dependent variable. Table 10.1.4 shows that the coefficient of determination is quite small in every case, indicating that the relationships are weak ones. Much more dramatic results can be seen when $\tilde{B}$ and B^* are calculated for data in which the relationships are stronger, as our second example will show.

Before turning our attention to the second example, however, let us consider briefly significance tests on $(n-1)^{1/2}\ \underline{\tilde{B}}$. In order to convert $(n-1)^{1/2}\ \underline{\tilde{B}}$ to a vector having a multivariate noncentral t distribution, since we were given the standard errors of the regression coefficients each element of $(n-1)^{1/2}\ \underline{\tilde{B}}$ was multiplied by

$$\left[\frac{s_{x_i}}{s}\left(\hat{\sigma}_{\beta_i}\right)\right]^{-1} .$$

The results are presented in Table 10.1.5.

Since the elements of $(n-1)^{1/2}\ \underline{\tilde{B}}$ have marginal univariate t distributions (not independent), the components of $(n-1)^{1/2}\ \underline{\tilde{B}}$ are compared to critical values for testing whether they were significantly greater or less than zero. An asterisk next to a t-value in Table 10.1.5 indicates significance at the .05 level; two asterisks indicate significance at the .01 level. We note that the components of $(n-1)^{1/2}\ \underline{\tilde{B}}$ in each of the five regressions in Table 10.1.6 are found to be significantly greater or less than zero at at least the .05 level of significance.

Table 10.1.4. Multiple Coefficients of determination.

Dependent Variable	Independent Variable(s)	R^2	Dependent Variable	Independent Variables	R^2
X_7(DS)	X_6(AGE)	.03011		X_1(FO),X_2(FE),	
X_8(AM)	X_1(FO),X_7(DS)	.05748		X_3(ME), X_4(MW),	
X_9(AFS)	X_4(MW),X_5(FS),		X_{10}(FOC1)	X_5(FS),X_6(AGE),	.08817
	X_8(AM)	.08082		X_7(DS),X_8(AM),	
X_{10}(FOC1)	X_8(AM),X_9(AFS)	.07854		X_9(AFS)	
X_{11}(FOC2)	X_6(AGE),X_8(AM)	.07261			
				X_1(FO),X_2(FE),	
				X_3(ME),X_4(MW),	
			X_{11}(FOC2)	X_5(FS),X_6(AGE),	.08491
				X_7(DS),X_8(AM),	
				X_9(AFS)	

Table 10.1.5. Values of $(n-1)^{1/2}\ \tilde{\underline{B}} = \underline{t}$ for the regressions of FOC1 and FOC2 from $X_1,\ldots,X_9$.

Dependent Variable	Independent Variables	$(n-1)^{1/2}\ \tilde{\underline{B}} = \underline{t}$
	X_1(FO)	.48084
	X_2(FE)	- .72006
	X_3(ME)	.36419
	X_4(MW)	-1.05596
X_{10}(FOC1)	X_5(FS)	- .59659
	X_6(AGE)	- .70684
	X_7(DS)	- .65270
	X_8(AM)	-3.36134**
	X_9(AFS)	2.64745**
	X_1(FO)	.75234
	X_2(FE)	.10703
	X_3(ME)	- .83529
	X_4(MW)	- .21193
X_{11}(FOC2)	X_5(FS)	- .46595
	X_6(AGE)	1.78468*
	X_7(DS)	.53296
	X_8(AM)	-3.61579**
	X_9(AFS)	1.27699

Table 10.1.6. Values of $(n-1)^{1/2}\ \underline{\tilde{B}} = \underline{t}$ for the five regressions indicated in Figure 10.1.1.

Dependent Variable	Independent Variable(s)	$(n-1)^{1/2}\ \underline{\tilde{B}} = \underline{t}$
X_7(DS)	X_6(AGE)	2.847**
X_8(AM)	X_1(FO)	2.70961**
	X_7(DS)	-3.02647**
X_9(AFS)	X_4(MW)	2.09115*
	X_5(FS)	2.33117*
	X_8(AM)	-3.55605**
X_{10}(FOC1)	X_8(AM)	-3.4611**
	X_9(AFS)	2.45504**
X_{11}(FOC2)	X_6(AGE)	2.08846*
	X_8(AM)	-4.14179**

Section 10.2. Application to Engineering Research Data

We now consider data taken from routine operations records for a petroleum refining unit, as presented by Gorman and Toman (1966). There were 36 observations taken on 10 independent and one dependent variable.

Calculations were performed on the Virginia Polytechnic Institute and State University system using a modified version of the BMD03R.

The correlation matrix for these data is presented in Table 10.2.1. We note that in several instances the correlation coefficients are reasonably large, indicating strong relationships in certain pairs of variables. The multiple coefficient of determination is .9039, indicating that $X_1,\ldots,X_{10}$ are effective in explaining about 90% of the variation in y.

Table 10.2.2 gives the marginal standard deviations of all eleven variables, and Table 10.2.3 gives values of $\hat{\beta}$, $\tilde{B}$, $\hat{B}$ and B^* for the Gorman and Toman data. For n = 36,

$$C = .97742 ,$$

so that even for relatively small samples, the bias in $\tilde{B}$ is small. We note that the differences between values of $\hat{B}$ and $\tilde{B}$ are much greater for these data than in the previous example. In every case, $\tilde{B}$ and B^* are larger than $\hat{B}$, in absolute value, verifying that $\hat{B}$ does tend to underestimate B, in absolute value.

Values of $\underline{t} = (n-1)^{1/2}\,\underline{\tilde{B}}$ are presented in Table 10.2.4. As in the first example, components of $\underline{t}$ are treated as univariate t statistics and are indicated as being significantly less than or greater than zero

Table 10.2.1. Correlation matrix for data from Gorman and Toman.

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	y
X_1	1.000	-.04117	.5110	.1175	-.7082	-.8679	-.1158	-.0001792	-.1609	-.3153	-.1991
X_2		1.000	-.003293	-.1568	.06393	.09460	.1300	.005631	-.2415	-.3094	-.2078
X_3			1.000	.003326	-.5897	-.6484	-.09676	.3382	-.02926	-.3654	-.2252
X_4				1.000	-.06697	-.08928	-.02487	.07954	.05285	.05433	.1008
X_5					1.000	.8409	.2524	-.3652	-.07572	.3726	.1800
X_6						1.000	.1525	-.1977	.1114	.3967	.2562
X_7							1.000	-.1776	.5261	.4710	.4124
X_8								1.000	.2662	-.1168	.09065
X_9									1.000	.4357	.4511
X_{10}										1.000	.9144
y											1.000

Table 10.2.2. Marginal standard deviations of $X_1,\ldots,X_{10}$ and y for the Gorman and Toman data.

Variable	Marginal Standard Deviation
X_1	2.37164
X_2	.52699
X_3	.98786
X_4	.41831
X_5	.33922
X_6	.41350
X_7	.33858
X_8	2.10839
X_9	16.23543
X_{10}	9.86270
y	3.99514

$s = 1.46546$

Table 10.2.3. Values of $\hat{\beta}$, $\hat{B}$, $\tilde{B}$ and B^* for the Gorman and Toman data.

Variable	$\hat{\beta}$	$\hat{B}$	$\tilde{B}$	B^*
X_1	.12078	.071701	.195470	.19106
X_2	.95493	.125962	.343396	.33564
X_3	.08793	.021742	.059273	.05793
X_4	.37275	.039028	.106399	.10400
X_5	-1.69761	-.144142	-.392958	-.38409
X_6	.65778	.068080	.185600	.18141
X_7	-.26913	-.022808	-.062180	-.06078
X_8	.30607	.161523	.440344	.43040
X_9	-.00308	-.012523	-.034141	.03337
X_{10}	.42272	1.043558	2.844941	2.78070

Table 10.2.4. Values of $\underline{t} = (n-1)^{1/2}\ \underline{\tilde{B}}$ for the Gorman and Toman data.

Variable	$\underline{t} = (n-1)^{1/2}\ \underline{\tilde{B}}$
X_1	.52467
X_2	1.54150
X_3	.24986
X_4	.61067
X_5	- .98165
X_6	.35249
X_7	- .21484
X_8	1.97716*
X_9	- .11881
X_{10}	11.67196**

at the .05 and .01 levels of significance.

Section 10.3. Summary

The examples presented in the preceding sections illustrate the use of $\underline{\tilde{B}}$ instead of $\underline{\hat{B}}$ as a "standardized regression coefficient." Since the bias in $\underline{\tilde{B}}$ is quite small, even for small samples, since $\underline{\hat{B}}$ tends to underestimate the "population standardized regression coefficient" $\underline{B}$, and since the distribution of $\underline{\hat{B}}$ is unwieldly, we recommend the use of $\underline{\tilde{B}}$ instead of $\underline{\hat{B}}$ when the aim is to estimate the parameters in the linear model in an invariant manner.

We note that when the data indicate weak relationships among the dependent and independent variables, the numerical values of $\underline{\tilde{B}}$ and $\underline{\hat{B}}$ differ very little. Dramatic increases in the values for $\underline{\tilde{B}}$ over those for $\underline{\hat{B}}$ occur, however, if reasonably strong relationships are present. Examination of Tables 10.2.1 and 10.2.3 indicate that, in fact, correlations need be not much larger than about .15 in order for relatively large differences in the values of $\underline{\hat{B}}$ and $\underline{\tilde{B}}$ to be observed. Thus, in problems in which any reasonable relationships are present, use of $\underline{\tilde{B}}$ instead of $\underline{\hat{B}}$ is likely to lead to more satisfactory results.

REFERENCES

Brown, L. D. (1966). On the admissibility of invariant estimators of one or more location parameters. Ann. Math. Statist., 37, 1087-1136.

Cramér, H. (1946). Mathematical Methods of Statistics. Princeton University Press, Princeton, New Jersey.

Ferguson, T. S. (1967). Mathematical Statistics: A Decision Theoretic Approach. Academic Press, New York.

Fisher, R. A. (1934). Two new properties of mathematical likelihood. Proc. Roy. Statist. Soc., Ser. A, 14. 285-307.

Fraser, D. A. S. (1961a). On fiducial inference. Ann. Math. Statist., 32, 661-676.

Fraser, D. A. S. (1961b). The fiducial method and invariance. Biometrika, 48. 261-280.

Fraser, D. A. S. (1964). Fiducial inference for location and scale parameters. Biometrika, 51, 17-24.

Fraser, D. A. S. (1968). The Structure of Inference. John Wiley and Sons, New York.

Goldhaber, J. K. and Ehrlich, G. (1970). Algebra. Macmillan, London.

Good, I. J. (1972). Personal communication.

Gorman, J. W. and Toman, R. J. (1966). Selection of variables for fitting equations to data. Technometrics, 13, 27-51.

Graybill, F. A. (1961). An Introduction to Linear Statistical Models, Volume I. McGraw-Hill, New York.

Hall, W. J., Wijsman, R. A. and Ghosh, J. K. (1965). The relationship between sufficiency and invariance with applications in sequential analysis. Ann. Math. Statist., 36. 575-614.

Herstein, I. N. (1964). Topics in Algebra. Blaisdell, New York.

Hodgman, C. D., Editor. (1959). Chemical Rubber Company Standard Mathematical Tables. Chemical Rubber Publishing Company, Cleveland, Ohio.

Hogg, R. V. and Craig, A. T. (1965). Introduction to Mathematical Statistics. Macmillan, New York.

Hora, R. B. and Buehler, R. J. (1966). Fiducial theory and invariant estimation. Ann. Math. Statist., 37. 643-656.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group. Ann. Math. Statist., 25. 40-75.

James, A. T. (1964). Distribution of matrix variates and latent roots derived from normal samples. Ann. Math. Statist., 35. 475-501.

Kiefer, J. (1957). Invariance, minimax sequential estimation, and continuous time processes. Ann. Math. Statist., 28. 573-601.

Kudō, H. (1955). On minimax invariant estimates of the transformation parameter. Natural Science Report of the Ochanomizu University, 6. 31-73.

Nachbin, L. (1965). The Haar Integral. D. Van Nostrand, Princeton, New Jersey.

Pitman, E. J. G. (1939). The estimation of the location and scale parameters of a continuous population of any given form. Biometrika, 30. 391-421.

Stein, C. (1959). The admissibility of Pitman's estimator of a single location parameter. Ann. Math. Statist., 30. 970-979.

Wijsman, R. A. (1957). Random orthogonal transformations and their use in some classical distribution problems in multivariate analysis. Ann. Math. Statist., 28. 415-423.

Wijsman, R. A. (1965). Cross-sections of orbits and their application to densities of maximal invariants. Proc. 5th. Berk. Symp. on Math. Statist. and Probab., 1. 389-400.

Younger, M. S. (1969). The Principle of Invariance in Statistics. Master of Science Thesis, Virginia Polytechnic Institute and State University.

Zacks, S. (1971). The Theory of Statistical Inference. Wiley, New York.

The two page vita has been removed from the scanned document. Page 1 of 2

The two page vita has been removed from the scanned document. Page 2 of 2

INVARIANT ESTIMATION WITH

APPLICATION TO LINEAR MODELS

by

Mary Sue Younger

(ABSTRACT)

The method of invariant estimation proposed in this dissertation relies on defining a group of transformations on the sample space such that i) the group structure is isomorphic to the parameter space and "carries" the estimation problem in a natural manner (thus defining an equivariant estimation problem), and ii) the group structure generates orbits on the parameter space and the problem is to estimate the orbit in which the parameter lies (thus defining an invariant estimation problem). If the group of transformations can be expressed as the semi-direct product of two subgroups, one a "nuisance" group which is a normal subgroup, then an estimator of orbits under the nuisance group in the invariant estimation problem can be naturally obtained from the best estimator in the equivariant estimation problem.

The primary application is to the invariant estimation of the parameters in the general linear model under the (nuisance) group of scale changes on the dependent and independent variables. The invariant estimator of the regression coefficient is found to be a "standardized regression coefficient," but this standardized regression coefficient is not the same as the typical one ("beta coefficient") found in elementary statistics texts and social science research.

Comparison of the proposed estimator to the usual estimator, in the case in which the input matrix is nonstochastic, shows the proposed

estimator to be superior to the usual estimator in terms of such criteria as consistency, unbiasedness, and simplicity of distribution. In the case in which the input matrix is stochastic, some justification can be found for the use of the usual estimator.

Application of the proposed method of invariant estimation to the problem of obtaining estimators invariant under nonsingular transformations is straightforward, although the estimator obtained is difficult to interpret.