Multicollinearity and Imprecise Estimation

By S. D. Silvey

University of Glasgow

[Received December 1968]

Summary

For the standard linear model containing several explanatory variables, the precision of estimation of linear parametric functions is analysed in terms of latent roots and vectors of $\mathbf{X}'\mathbf{X}$, where $\mathbf{X}$ is the matrix of values of explanatory variables. This analysis provides a practical method for detecting multicollinearity, and it is demonstrated that it is also useful in solving problems of optimum choice of new values of explanatory variables.

1. Introduction

Multicollinearity is a term used in econometrics to denote the presence of linear relationships or "near linear relationships" among explanatory (independent, concomitant) variables in linear regression; see, for instance, Johnston (1963), Malinvaud (1966). The problems created by this state of affairs are well known and are illustrated clearly by a very simple example. Consider the single-variable linear regression model

$$y_i = \alpha+\beta x_i+\epsilon_i, \quad i = 1,2,\ldots,n, \tag{1.1}$$

with the usual assumptions—known x_i's, unknown α and β, and uncorrelated errors ϵ_i with common variance. There are two "explanatory" variables here, the constant "variable" which is identically 1 and the variable x. The values of these two explanatory variables occurring in the model (1.1) are linearly related if and only if $x_1 = x_2 = \ldots = x_n$, the common value being x_0, say. A scatter diagram makes it quite clear that when this is the case and $x_0 \neq 0$, there is no hope of estimating α and β separately. On the other hand, $\alpha+\beta x_0$ can be estimated perfectly well—by

$$\bar{y} = \frac{1}{n}(y_1+y_2+\ldots+y_n),$$

in fact.

Referring still to this simple example, it is clear that if the x_i's are not all equal but are very closely grouped around their mean $\bar{x}$, then α and β can be estimated separately, but estimates of both of them will be very imprecise (unless $\bar{x}$ is near zero, in which case α, but not β, may be fairly precisely estimated). It remains true, of course, that $\alpha+\beta\bar{x}$ can be estimated with relative precision.

We now consider the general linear model involving several explanatory variables,

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}, \tag{1.2}$$

where $\mathbf{y}$ is an n-vector of observations, $\mathbf{X}$ a known $n\times k$ matrix of explanatory variables, $\boldsymbol{\beta}$ an unknown k-vector of regression coefficients and $\boldsymbol{\epsilon}$ an error vector with zero mean and variance matrix $\sigma^2\mathbf{I}$. There is "extreme" multicollinearity here when there is at least one linear relationship among the columns of $\mathbf{X}$; that is, when the rank of $\mathbf{X}$ is less than k. Then $\boldsymbol{\beta}$ is not identifiable, and, as in the simple example, certain linear functions of $\boldsymbol{\beta}$ cannot be estimated at all, though others can.

It is not as immediately obvious what we mean by saying that there is "nearly" extreme multicollinearity in the model (1.2), though there are several obvious ways in which we might define fairly precisely the "degree of multicollinearity" present. However there is a sense in which such an exercise is pointless, since there is an infinite number of versions of any given linear model and while one version may exhibit certain features on which we concentrate attention in defining "degree of multicollinearity", another may not. For instance we may take the view, as has been done recently by Farrar and Glauber (1967), that multicollinearity is an opposite extreme from orthogonality, and base a definition of "degree of multicollinearity" on the departure from orthogonality of the columns of **X**. However, we can produce, by suitable reparametrization, an orthogonal version of (1.2). This is done by determining a non-singular matrix **T** such that the matrix $\mathbf{X}^* = \mathbf{XT}$ has orthogonal columns, which is always possible. Then if we set $\boldsymbol{\beta}^* = \mathbf{T}^{-1}\boldsymbol{\beta}$ we have

$$\mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{T}\mathbf{T}^{-1})\boldsymbol{\beta} = (\mathbf{XT})(\mathbf{T}^{-1}\boldsymbol{\beta}) = \mathbf{X}^*\boldsymbol{\beta}^*,$$

and the model

$$\mathbf{y} = \mathbf{X}^*\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$$

is a reparametrized version of (1.2) with orthogonal explanatory variables.

Rather than seek to define degree of multicollinearity, therefore, we might more profitably make a direct attack on the basic problem arising from multicollinearity, namely, that of imprecise estimation. So in this note we shall seek to answer the following two questions. Given the model (1.2), what can be estimated with relative precision and what cannot? Can we determine an efficient practical procedure for arriving at this distinction? Not surprisingly, it turns out that in answering these questions we obtain as a by-product an efficient practical method of analysing multicollinearity.

2. Imprecise Estimation

We start by considering, for the model (1.2), the extreme situation when certain real parameters cannot be estimated at all, in the absence of prior information about them. This arises when the rank of **X** is less than k.

As is well known, (see for instance Scheffé, 1959), when **X** does not have full rank, certain linear parametric functions, that is, functions of the form

$$\mathbf{c}'\boldsymbol{\beta} = c_1\beta_1 + c_2\beta_2 + \ldots + c_k\beta_k,$$

do not possess unbiased linear estimators. Indeed $\mathbf{c}'\boldsymbol{\beta}$ possesses an unbiased linear estimator if and only if $\mathbf{c}'$ can be expressed as a linear combination of the rows of **X**. When **X** has full rank, every $\mathbf{c}'$ can be expressed in this way, but otherwise some can not. An equivalent, and potentially much more useful condition for $\mathbf{c}'\boldsymbol{\beta}$ to be estimable, is that the vector **c** can be expressed as a linear combination of the columns (or rows) of $\mathbf{X}'\mathbf{X}$, a result proved by Rao (1965). This in turn is translated into an immediately useable form by the following result.

Theorem 1. The linear parametric function $\mathbf{c}'\boldsymbol{\beta}$ is estimable if and only if **c** is a linear combination of latent vectors of $\mathbf{X}'\mathbf{X}$ corresponding to non-zero latent roots of this matrix.

Proof. The range of $\mathbf{X}'\mathbf{X}$, that is, the set of vectors expressible as linear combinations of the columns of $\mathbf{X}'\mathbf{X}$ is spanned by a linearly independent set of latent vectors of $\mathbf{X}'\mathbf{X}$ corresponding to its non-zero roots. The result follows immediately from the result of Rao quoted above.

There is an illuminating way of looking at this result. As indicated in the introduction we obtain an orthogonal version of the model (1.2) by writing

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{XT})(\mathbf{T}^{-1}\boldsymbol{\beta}),$$

where $\mathbf{T}$ is a non-singular matrix chosen in such a way that the columns of $\mathbf{XT}$ are orthogonal. There are many possible choices of $\mathbf{T}$. One choice is $\mathbf{V}$, an orthogonal matrix whose columns are orthonormal latent vectors of $\mathbf{X}'\mathbf{X}$: since $\mathbf{VV}' = \mathbf{I}$, we may write

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{XV})(\mathbf{V}'\boldsymbol{\beta}) = \mathbf{Z}\boldsymbol{\gamma},$$

where $\mathbf{Z} = \mathbf{XV}$ and $\mathbf{V}'\boldsymbol{\beta} = \boldsymbol{\gamma}$. The columns of $\mathbf{XV}$ are orthogonal since

$$(\mathbf{XV})'\mathbf{XV} = \mathbf{V}'\mathbf{X}'\mathbf{XV} = \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_k\},$$

where the λ's are latent roots of $\mathbf{X}'\mathbf{X}$. Now suppose that $\mathbf{X}'\mathbf{X}$ has the latent root 0 of multiplicity j. Then j of the columns of $\mathbf{XV}$ are zero—say the last j. It follows that the last j components of $\boldsymbol{\gamma}$ are annihilated—they just do not appear in the model at all, essentially. So we cannot estimate $\gamma_{k-j+1}, \gamma_{k-j+2}, \ldots, \gamma_k$ from our observations. On the other hand, we can estimate $\gamma_1, \gamma_2, \ldots, \gamma_{k-j}$, or any linear combination of these. Therefore we can estimate $\mathbf{c}'\boldsymbol{\beta}$ iff $\mathbf{c}'\boldsymbol{\beta}$ transforms into $\alpha_1\gamma_1 + \alpha_2\gamma_2 + \ldots + \alpha_{k-j}\gamma_{k-j}$. Now $\mathbf{c}'\boldsymbol{\beta} = \mathbf{c}'\mathbf{VV}'\boldsymbol{\beta} = (\mathbf{V}'\mathbf{c})'\boldsymbol{\gamma}$. Hence we can estimate $\mathbf{c}'\boldsymbol{\beta}$ iff the last j components of $\mathbf{V}'\mathbf{c}$ are zero. This is what the theorem says if we write $\mathbf{V}'\mathbf{c} = \boldsymbol{\alpha}$ in the equivalent form $\mathbf{c} = \mathbf{V}\boldsymbol{\alpha}$.

This theorem is useful in practice because normally the order of the matrix $\mathbf{X}'\mathbf{X}$ is not excessive. (Its order is, of course, $k \times k$, where k is the number of explanatory variables). Moreover computer programs are available for calculating its latent roots in descending order of magnitude, $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_k$, say, and a corresponding orthonormal set, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ of latent vectors. Let $\mathbf{V}$ be the orthogonal matrix whose columns, in order, are those latent vectors, and suppose that the last j of the λ's are zero. Then to determine whether $\mathbf{c}'\boldsymbol{\beta}$ is estimable we need only calculate $\mathbf{V}'\mathbf{c}$; as above, $\mathbf{c}'\boldsymbol{\beta}$ is estimable if and only if the last j components of $\mathbf{V}'\mathbf{c}$ are all zero.

This analysis in terms of the latent roots and vectors of $\mathbf{X}'\mathbf{X}$ not only enables us to distinguish between linear parametric functions which can be estimated and those which cannot, but also provides a ready means of determining those which can be estimated precisely, and those which are estimable but can be estimated only imprecisely. To see this, let us suppose first that $\mathbf{v}$ is a unit latent vector of $\mathbf{X}'\mathbf{X}$ corresponding to a non-zero latent root λ. Then, by the above theorem, $\mathbf{v}'\boldsymbol{\beta}$ is estimable. Moreover the general version of the Gauss–Markov theorem, as proved for example by Scheffé (1959), shows that the minimum-variance unbiased linear estimator of $\mathbf{v}'\boldsymbol{\beta}$ is $\mathbf{v}'\hat{\boldsymbol{\beta}}$ where $\hat{\boldsymbol{\beta}}$ is *any* solution of the normal equations

$$\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}'\mathbf{y}.$$

Now

$$\mathbf{v}'\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{v}'\mathbf{X}'\mathbf{y},$$

and so

$$\lambda\mathbf{v}'\hat{\boldsymbol{\beta}} = \mathbf{v}'\mathbf{X}'\mathbf{y},$$

since $\mathbf{X'Xv} = \lambda\mathbf{v}$, or $\mathbf{v'X'X} = \lambda\mathbf{v}'$. It follows that

$$\begin{aligned}\lambda^2 \operatorname{var}(\mathbf{v}'\hat{\boldsymbol{\beta}}) &= \mathbf{v'X'}\operatorname{var}(\mathbf{y})\mathbf{Xv}\\ &= \sigma^2\mathbf{v'X'Xv}\\ &= \lambda\sigma^2\mathbf{v'v}\\ &= \lambda\sigma^2,\end{aligned}$$

since $\mathbf{v'v} = 1$.

Therefore

$$\operatorname{var}(\mathbf{v}'\hat{\boldsymbol{\beta}}) = \sigma^2/\lambda.$$

A similar argument shows that, if $\mathbf{v}_1$ and $\mathbf{v}_2$ are orthogonal latent vectors of $\mathbf{X'X}$ corresponding to non-zero roots, then $\operatorname{cov}(\mathbf{v}_1'\hat{\boldsymbol{\beta}}, \mathbf{v}_2'\hat{\boldsymbol{\beta}}) = 0$.

Now let $\mathbf{c}'\boldsymbol{\beta}$ be an estimable function. By the above theorem, $\mathbf{c}$ can be expressed in the form

$$\mathbf{c} = \alpha_1\mathbf{v}_1 + \alpha_2\mathbf{v}_2 + \ldots + \alpha_j\mathbf{v}_j, \tag{2.1}$$

where $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$ are orthonormal latent vectors of $\mathbf{X'X}$ corresponding to non-zero roots $\lambda_1, \lambda_2, \ldots, \lambda_j$. By the Gauss–Markov theorem, the minimum-variance unbiased linear estimator of $\mathbf{c}'\boldsymbol{\beta}$ is $\mathbf{c}'\hat{\boldsymbol{\beta}}$, and by the results just proved,

$$\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}})/\sigma^2 = \frac{\alpha_1^2}{\lambda_1} + \frac{\alpha_2^2}{\lambda_2} + \ldots + \frac{\alpha_j^2}{\lambda_j}. \tag{2.2}$$

The variance of $\mathbf{c}'\hat{\boldsymbol{\beta}}$ depends, amongst other things, on the length of the vector $\mathbf{c}$. So in order to "normalize" the discussion and concentrate attention on the *directions* which admit precise or imprecise estimation, we shall consider parametric functions $\mathbf{c}'\boldsymbol{\beta}$ for which $\mathbf{c'c} = 1$. When $\mathbf{c}$ is expressed in the form (2.1), this implies that $\sum\alpha_i^2 = 1$. The equation (2.2) then demonstrates quite clearly the distinction between precisely and imprecisely estimable functions. *Relatively precise estimation is possible in the directions of latent vectors of* $\mathbf{X'X}$ *corresponding to large latent roots*; *relatively imprecise estimation in these directions corresponding to small latent roots.* Knowledge of the latent roots of $\mathbf{X'X}$ provides a great deal of immediate information about the possibilities of precise estimation. For instance, from (2.2) we can easily deduce that

$$\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}})/\sigma^2 \leqslant 1/\min(\lambda_1, \lambda_2, \ldots, \lambda_k).$$

Thus if the minimum latent root of $\mathbf{X'X}$ is large, every linear parametric function can be estimated relatively precisely. On the other hand if $\mathbf{X'X}$ has at least one very small root, then there is the possibility of only imprecise estimation in at least one direction. If the latent roots of $\mathbf{X'X}$ are all equal, every $\mathbf{c}'\boldsymbol{\beta}$ such that $\mathbf{c'c} = 1$ can be estimated with equal precision, and so on.

It is perhaps worth remarking that the sum of the latent roots of $\mathbf{X'X}$ is its trace, and this is the sum of squares of the components of $\mathbf{X}$. Thus if n is appreciable and $\mathbf{X}$ does not have many zero or near zero components, the sum of the latent roots of $\mathbf{X'X}$ will be large. It follows that if the number of explanatory variables is reasonably small, $\mathbf{X'X}$ will have at least one fairly large latent root, so that relatively precise estimation will be possible in at least one direction. This is the analogue for the model (1.2) of the obvious result for model (1.1) that relatively precise estimation of at least one linear parametric function is always possible.

The very natural transformation introduced in this section is not, of course, new, and it will be recognized as that adopted in principal component analysis. Its use in connection with multicollinearity has already been proposed by Kendall (1957). Essentially what we have done is to change the emphasis in the analysis by asking about the precision of estimation of linear parametric functions instead of enquiring about which linear combinations of the explanatory variables should be omitted from the regression equation. As suggested by Farrar and Glauber (1967) it may be misleading to omit a linear combination of explanatory variables from the regression equation merely because we have insufficient evidence about its effect on the dependent variable. The change of emphasis which we have adopted has the merit of concentrating on the analysis of the information content of available data, and leaving other questions to be discussed separately.

3. MULTICOLLINEARITY

As indicated, at least implicitly, by Kendall (1957), calculation of the latent roots and an orthonormal set of latent vectors of $\mathbf{X}'\mathbf{X}$ provides a very clear cut analysis of multicollinearity.

Let us suppose that there is extreme multicollinearity in the model (1.2). Then $\mathbf{X}'\mathbf{X}$ has the latent root 0 of multiplicity $k-\text{rank}(\mathbf{X})$, since $\text{rank}(\mathbf{X}'\mathbf{X}) = \text{rank}(\mathbf{X})$. If $\text{rank}(\mathbf{X}) = k-1$, then there is essentially one unit vector $\mathbf{v}$ such that $\mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{0}$. (We say *essentially* one, because if $\mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{0}$ then also $\mathbf{X}'\mathbf{X}(-\mathbf{v}) = \mathbf{0}$). The fact that $\mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{0}$ implies, by the theorem, that $\mathbf{v}'\boldsymbol{\beta}$ is inestimable, as is any linear parametric function $\mathbf{c}'\boldsymbol{\beta}$, for which $\mathbf{c}$ has a non-zero component in the direction of $\mathbf{v}$.

Now $\mathbf{X}'\mathbf{X}\mathbf{v} = \mathbf{0}$ if and only if $\mathbf{X}\mathbf{v} = \mathbf{0}$. Consequently the multicollinearity is attributable to the fact that every row of $\mathbf{X}$ is orthogonal to $\mathbf{v}$; that is, the values of the explanatory variables at which observations have been made satisfy the linear relationship

$$v_1 x_1 + v_2 x_2 + \ldots + v_k x_k = 0,$$

where $(v_1, v_2, \ldots, v_k) = \mathbf{v}'$. From this we can deduce immediately how to choose a new set of values of the explanatory variables $x_1, x_2, \ldots, x_k$ in order to get rid of the multicollinearity. We simply choose a new set in a direction which is not orthogonal to $\mathbf{v}$, and the obvious way of doing this is to choose the new set in the direction of $\mathbf{v}$. This choice is not unique, of course.

If rank $\mathbf{X} = k-j$ and $j > 1$, then there are j orthogonal vectors, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$, say such that $\mathbf{X}\mathbf{v}_i = \mathbf{0}$, and correspondingly j independent linear relationships among the observed values of the explanatory variables. Then multicollinearity may be eradicated by taking a new observation at each of j new sets of values of the explanatory variables, one set in each of the directions $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$. Again this choice is not unique. However it is clear that the multicollinearity cannot be eradicated by taking less than j new observations.

A similar analysis applies when multicollinearity is present, but is not extreme. By this we now mean that $\mathbf{X}'\mathbf{X}$ has no zero latent roots, but has some small ones. Suppose first that $\mathbf{X}'\mathbf{X}$ has $k-1$ large roots and one small one, λ say. Estimation is then relatively imprecise in the direction of a unit latent vector corresponding to λ. Let this vector be $\mathbf{v}$ and suppose that we take an additional observation y_{n+1} at the values $\mathbf{x}'_{n+1} = (x_{1,n+1}, x_{2,n+1}, \ldots, x_{k,n+1})$ of the explanatory variables, where $\mathbf{x}'_{n+1} = l\mathbf{v}'$.

The model for the complete set of observations, including the new one is

$$\begin{pmatrix} \mathbf{y} \\ y_{n+1} \end{pmatrix} = \begin{pmatrix} \mathbf{X} \\ \mathbf{x}'_{n+1} \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \boldsymbol{\epsilon} \\ \epsilon_{n+1} \end{pmatrix}$$

or

$$\mathbf{y}_* = \mathbf{X}_* \boldsymbol{\beta} + \boldsymbol{\epsilon}_*, \quad \text{say.}$$

Now

$$\mathbf{X}'_* \mathbf{X}_* = \mathbf{X}'\mathbf{X} + \mathbf{x}_{n+1} \mathbf{x}'_{n+1} = \mathbf{X}'\mathbf{X} + l^2 \mathbf{v}\mathbf{v}'.$$

Further

$$\mathbf{X}'_* \mathbf{X}_* \mathbf{v} = \mathbf{X}'\mathbf{X}\mathbf{v} + l^2 \mathbf{v}\mathbf{v}'\mathbf{v} = \lambda \mathbf{v} + l^2 \mathbf{v} = (\lambda + l^2)\mathbf{v}.$$

Thus $\mathbf{v}$ is a latent vector of $\mathbf{X}'_* \mathbf{X}_*$ corresponding to the root $\lambda + l^2$.

A similar argument shows that any other latent vector of $\mathbf{X}'\mathbf{X}$, orthogonal to $\mathbf{v}$, is also a latent vector of $\mathbf{X}'_* \mathbf{X}_*$ and that the corresponding roots of the two matrices are equal. Hence by choosing $\mathbf{x}_{n+1}$ in the direction of $\mathbf{v}$, we improve the precision of estimation in the direction for which previously it was most imprecise.

The generalization is obvious. If $\mathbf{X}'\mathbf{X}$ has more than one small latent root, then there are several orthogonal directions in which estimation is relatively imprecise. Just as in the extreme case, several new observations have then to be taken in order to achieve the somewhat vague aim of improving the precision of estimation all round. The choice of new sets of values of the explanatory variables is governed by the directions of latent vectors corresponding to the small roots.

There is an interesting connection here with a particular result of Kiefer and Wolfowitz (1960), extended by Karlin and Studden (1966), concerning optimum experimental design. Suppose that we pose the following problem. We have already made observations on a dependent variable at certain values of explanatory variables and these conform to the linear model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, which we have been discussing. We wish to choose a new set of values of the explanatory variables, $\mathbf{x}_{n+1}$, subject to $\mathbf{x}'_{n+1} \mathbf{x}_{n+1} \leqslant l^2$, at which to take an additional observation on the dependent variable, and this choice has to be optimum in some sense. Two criteria of optimality have been considered in work on experimental design, and these, when applied to this particular problem, are:

(i) choose $\mathbf{x}_{n+1}$ in order to maximize the minimum latent root of the new information matrix $\mathbf{X}'_* \mathbf{X}_* = \mathbf{X}'\mathbf{X} + \mathbf{x}_{n+1} \mathbf{x}'_{n+1}$;

(ii) choose $\mathbf{x}'_{n+1}$ in order to maximize the determinant of $\mathbf{X}'_* \mathbf{X}_*$.

(For a general discussion of these criteria, and further references, see Kiefer, 1958).

The result of Kiefer and Wolfowitz (1960), referred to above, is that these two criteria, applied to a class of problems which apparently does not contain that under discussion, are equivalent in the sense that any design optimum relative to one is also optimum relative to the other.

It is interesting that this equivalence applies in the present instance also, and a constructive proof can be given of the fact that the choice of $\mathbf{x}_{n+1}$ in the direction of a latent vector of $\mathbf{X}'\mathbf{X}$ corresponding to its minimum latent root, and subject to $\mathbf{x}'_{n+1} \mathbf{x}_{n+1} = l^2$, is optimum relative to both criteria. We shall not go into details of this result here, since it is somewhat apart from our main line of discussion, and we shall content ourselves with the remark that the choice of $\mathbf{x}_{n+1}$ suggested above on an intuitive basis, can be justified on this more formal basis also.

4. Optimum Choice of New Values of the Explanatory Variables

As we have seen, choosing a new set of values of the explanatory variables in the direction of a latent vector of $\mathbf{X}'\mathbf{X}$ keeps everything tidy, since the augmented matrix $\mathbf{X}_*$ then has the property mentioned above: the latent vectors of $\mathbf{X}_*'\mathbf{X}_*$ coincide with those of $\mathbf{X}'\mathbf{X}$ and the latent roots of these matrices are simply related. However, in practice, complete freedom of choice of new values of the explanatory variable may not be available. Moreover, we might well have a more specific object than that of "improving estimation where it is most imprecise." Our discussion would not be complete therefore, without answers to the following related questions.

(i) How is precision of estimation affected by taking another observation on the dependent variable at the values $(x_{1,n+1}, x_{2,n+1}, \ldots, x_{k,n+1}) = \mathbf{x}_{n+1}'$ of the explanatory variables, when $\mathbf{x}_{n+1}$ is not necessarily in the direction of a latent vector of $\mathbf{X}'\mathbf{X}$?

(ii) How should $\mathbf{x}_{n+1}$ be chosen to improve as much as possible the estimation of a *specific* linear function $\mathbf{c}'\boldsymbol{\beta}$? This question arises when, for instance, we wish to improve the estimation of a particular regression coefficient or when we wish to predict y at the specific values $(c_1, c_2, \ldots, c_k) = \mathbf{c}'$ of the explanatory variables.

Although it is not strictly necessary we shall split the discussion of these questions into two parts:

(*a*) the case where $\mathbf{X}$ has full rank and every linear parametric function can be estimated before the new observation is taken;

(*b*) the case where $\mathbf{X}$ does not have full rank.

(*a*) *Rank* $(\mathbf{X}) = k$.

Let $\hat{\boldsymbol{\beta}}_n$ be the least-squares estimate of $\boldsymbol{\beta}$ from the original observations. Then

$$\operatorname{var} \hat{\boldsymbol{\beta}}_n = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

Let $\hat{\boldsymbol{\beta}}_{n+1}$ be the least-squares estimate of $\boldsymbol{\beta}$ obtained from the original set of observations augmented by a new observation on the dependent variable at the new values $\mathbf{x}_{n+1}$ of the explanatory variables. Then

$$\operatorname{var} \hat{\boldsymbol{\beta}}_{n+1} = \sigma^2(\mathbf{X}'\mathbf{X} + \mathbf{x}_{n+1}\mathbf{x}_{n+1}')^{-1},$$

if, as we shall assume, the error in the new observation is uncorrelated with previous errors and has the same variance as each of them.

The improvement in the estimation of $\mathbf{c}'\boldsymbol{\beta}$ given by the new observation at $\mathbf{x}_{n+1}$ is

$$\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_n) - \operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_{n+1}) = \sigma^2 \mathbf{c}'[(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\mathbf{X} + \mathbf{x}_{n+1}\mathbf{x}_{n+1}')^{-1}]\mathbf{c}. \tag{4.1}$$

This expression becomes much more manageable if we change the basis in R^k to an orthonormal set $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ of latent vectors of $\mathbf{X}'\mathbf{X}$, corresponding to its roots $\lambda_1, \lambda_2, \ldots, \lambda_k$, say. Let $\mathbf{V}$ be the orthogonal matrix whose columns are $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ and let

$$\mathbf{c} = \mathbf{V}\mathbf{a} \quad \text{and} \quad \mathbf{x}_{n+1} = \mathbf{V}\mathbf{z}.$$

Then we have

$$\mathbf{V'X'XV} = \operatorname{diag}\{\lambda_1, \lambda_2, \ldots, \lambda_k\} = \mathbf{D}, \quad \text{say;}$$

$$\begin{aligned}\frac{1}{\sigma^2}\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_n) &= \mathbf{c}'(\mathbf{X'X})^{-1}\mathbf{c}\\ &= \mathbf{a}'\mathbf{D}^{-1}\mathbf{a}\\ &= \textstyle\sum a_i^2/\lambda_i;\end{aligned}$$

and

$$\begin{aligned}\frac{1}{\sigma^2}\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_{n+1}) &= \mathbf{a}'\,[\mathbf{D}+\mathbf{zz}']^{-1}\mathbf{a}\\ &= \mathbf{a}'\left[\mathbf{D}^{-1}-\frac{\mathbf{D}^{-1}\mathbf{zz}'\mathbf{D}^{-1}}{1+\mathbf{z}'\mathbf{D}^{-1}\mathbf{z}}\right]\mathbf{a},\end{aligned}$$

by a standard, easily verified, result on matrix inversion.

So the improvement (4.1) in the estimation of $\mathbf{c}'\boldsymbol{\beta}$ brought about by a new observation at $\mathbf{x}_{n+1}$ is

$$\sigma^2\frac{\mathbf{a}'\mathbf{D}^{-1}\mathbf{zz}'\mathbf{D}^{-1}\mathbf{a}}{1+\mathbf{z}'\mathbf{D}^{-1}\mathbf{z}}$$

or, in non-matrix notation,

$$\sigma^2\left(\sum\frac{a_i z_i}{\lambda_i}\right)^2\bigg/\left(1+\sum\frac{z_i^2}{\lambda_i}\right). \tag{4.2}$$

The simplicity of this expression evidences again how well adapted to our present purposes is the orthogonal version of the original model obtained by writing

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{XV})(\mathbf{V}'\boldsymbol{\beta}).$$

Having calculated the latent roots and vectors of $\mathbf{X'X}$ we can very quickly calculate from (4.2) the improvement in the precision of estimation of any function $\mathbf{c}'\boldsymbol{\beta}$, achieved by taking a new observation at the values $\mathbf{x}_{n+1}$ of the explanatory variables.

We turn now to the question of how $\mathbf{x}_{n+1}$ should be chosen in order to improve the precision of estimation of a particular $\mathbf{c}'\boldsymbol{\beta}$. Naturally this choice depends on any restrictions which practical considerations impose. We shall consider the restriction that $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} \leqslant b^2$, noting that, from a practical point of view, it will seldom be the case that the choice is restricted in such a mathematically convenient way. For instance, in practice, one of the "explanatory" variables is nearly always the constant variable which takes the value 1, so that $\mathbf{x}_{n+1}$ is subject to the condition that a particular component takes the value 1 in addition to any other restrictions which may be imposed. Nevertheless some insight to the general problem is gained by considering the tractable condition that $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} \leqslant b^2$.

It is clear on intuitive grounds and easily proved from the expression (4.2), that for a given direction of $\mathbf{x}_{n+1}$ which gives some improvement in the precision of estimation of $\mathbf{c}'\boldsymbol{\beta}$, the greater the length of $\mathbf{x}_{n+1}$, the greater the improvement. So we might as well replace the condition $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} \leqslant b^2$ by $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} = b^2$ and we shall now show that, subject to the latter condition, the optimum direction for $\mathbf{x}_{n+1}$ is that of the vector

$$(\mathbf{I}+b^{-2}\mathbf{X'X})^{-1}\mathbf{c}.$$

Writing $\mathbf{c} = \mathbf{Va}$ and $\mathbf{x}_{n+1} = \mathbf{Vz}$, as above, the problem is to choose $z_1, z_2, \ldots, z_k$ to maximize (4.2) subject to the condition $\mathbf{z}'\mathbf{z} = b^2$, since $\mathbf{z}'\mathbf{z} = \mathbf{x}'_{n+1}\mathbf{x}_{n+1}$. This is a straightforward maximization problem. We introduce a Lagrange multiplier μ and differentiate, giving,

$$\frac{\dfrac{a_i}{\lambda_i}\sum\dfrac{a_i z_i}{\lambda_i}}{1+\sum\dfrac{z_i^2}{\lambda_i}} - \frac{\dfrac{z_i}{\lambda_i}\left(\sum\dfrac{a_i z_i}{\lambda_i}\right)^2}{\left(1+\sum\dfrac{z_i^2}{\lambda_i}\right)^2} + \mu z_i = 0, \quad i = 1, 2, \ldots, k$$

or

$$\frac{a_i}{\lambda_i}\frac{A}{B} - \frac{z_i}{\lambda_i}\frac{A^2}{B^2} + \mu z_i = 0, \quad i = 1, 2, \ldots, k, \tag{4.3}$$

where

$$A = \sum \frac{a_i z_i}{\lambda_i}$$

and

$$B = 1 + \sum \frac{z_i^2}{\lambda_i}.$$

For each i multiply the ith equation in (4.3) by z_i and add, giving

$$\frac{A^2}{B} - (B-1)\frac{A^2}{B^2} + \mu b^2 = 0,$$

that is,

$$\mu = -\frac{A^2}{b^2 B^2}$$

Therefore

$$z_i \frac{A^2}{B^2}\left(\frac{1}{b^2} + \frac{1}{\lambda_i}\right) = \frac{A}{B}\frac{a_i}{\lambda_i},$$

and so

$$z_i \propto \frac{a_i}{1 + \lambda_i/b^2}$$

or

$$\mathbf{z} \propto (\mathbf{I} + b^{-2}\mathbf{D})^{-1}\mathbf{a}.$$

This is equivalent to

$$\mathbf{x}_{n+1} \propto (\mathbf{I} + b^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{c},$$

as is easily verified.

We shall discuss this result in Section 5, after considering the case where rank $\mathbf{X} < k$.

(*b*) *Rank* $\mathbf{X} < k$.

There are two cases to be considered when rank $\mathbf{X} < k$:

(i) the new set of values $\mathbf{x}_{n+1}$ of the explanatory variables is such that $\mathbf{x}_{n+1}$ is a linear combination of latent vectors of $\mathbf{X}'\mathbf{X}$ corresponding to *non-zero* roots;

(ii) $\mathbf{x}_{n+1}$ does not satisfy this condition.

The first of these cases does not raise any new problems. Linear parametric functions estimable before the new observation has been taken of course remain estimable and no non-estimable function becomes estimable. Formula (4.2) remains valid for estimable $\mathbf{c}'\boldsymbol{\beta}$'s provided summation is restricted to positive λ_i's. And it remains true that, subject to $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} = b^2$, the optimum direction for $\mathbf{x}_{n+1}$ to improve the estimation of an originally estimable $\mathbf{c}'\boldsymbol{\beta}$ is that of $(\mathbf{I}+b^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$.

We shall not substantiate these remarks in detail. They follow fairly directly from a study of the orthogonal version of the model which is really the key to our present analysis. When we write $\mathbf{X}$ in the form

$$(\mathbf{XV})(\mathbf{V}'\boldsymbol{\beta}) = \mathbf{Z}\boldsymbol{\gamma}$$

(see the remarks following the proof of Theorem 1) and rank $\mathbf{X} = j < k$, then $\mathbf{Z}$ may be partitioned into

$$(\mathbf{Z}^{(1)}\ \mathbf{0})$$

where $\mathbf{Z}^{(1)}$ has order $n \times j$ and rank j: if $\boldsymbol{\gamma}$ is conformably partitioned into $\begin{pmatrix}\boldsymbol{\gamma}^{(1)}\\ \boldsymbol{\gamma}^{(2)}\end{pmatrix}$ and if we are interested only in originally estimable $\mathbf{c}'\boldsymbol{\beta}$'s, then $\boldsymbol{\gamma}^{(2)}$, which is annihilated by $\mathbf{Z}$, may be ignored completely and our model is essentially

$$\mathbf{y} = \mathbf{Z}^{(1)}\boldsymbol{\gamma}^{(1)} + \boldsymbol{\epsilon}.$$

All the parameters in this version of the model are estimable and it is not difficult to use the arguments of the previous section to substantiate the remarks made at the beginning of this one.

However it is certainly not immediately obvious that if we are interested in choosing a new set, $\mathbf{x}'_{n+1}$, of values of the explanatory variables in order to estimate, as well as possible, a *previously inestimable* $\mathbf{c}'\boldsymbol{\beta}$, the optimum direction of $\mathbf{x}_{n+1}$ is that of $(\mathbf{I}+b^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$. But this result is true, as we shall now demonstrate. We shall carry through the discussion entirely in terms of an orthogonal version of the model and then revert to the original version.

Suppose that rank $\mathbf{X} = j$. Then $\mathbf{X}'\mathbf{X}$ has j non-zero latent roots, $\lambda_1, \lambda_2, \ldots, \lambda_j$. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$ be an orthonormal set of vectors corresponding to these.

$\mathbf{X}'\mathbf{X}$ has the latent root zero of multiplicity $k-j$. There is considerable freedom of choice of $k-j$ orthonormal vectors corresponding to the root zero, when $k-j > 1$. It suits our present purposes to choose $\mathbf{v}_{j+1}$ in the subspace spanned by $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$ and $\mathbf{c}$. (Note that, since $\mathbf{c}'\boldsymbol{\beta}$ is inestimable $\mathbf{c}$ is linearly independent of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$.)

Specifically we take

$$\mathbf{v}_{j+1} \propto \mathbf{c} - (\mathbf{v}'_1\mathbf{c})\mathbf{v}_1 - (\mathbf{v}'_2\mathbf{c})\mathbf{v}_2 - \ldots - (\mathbf{v}'_j\mathbf{c})\mathbf{v}_j,$$

the choice of the coefficients of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$ ensuring that $\mathbf{v}_{j+1}$ is orthogonal to each of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_j$ and therefore that $\mathbf{X}'\mathbf{X}\mathbf{v}_{j+1} = \mathbf{0}$. Now let $\mathbf{v}_{j+2}, \mathbf{v}_{j+3}, \ldots, \mathbf{v}_k$ complete an orthonormal set of latent vectors of $\mathbf{X}'\mathbf{X}$, so that each $\mathbf{v}_{j+i}$ is a vector corresponding to the root zero. Then when we write

$$\mathbf{X}\boldsymbol{\beta} = (\mathbf{XV})(\mathbf{V}'\boldsymbol{\beta}) = \mathbf{Z}\boldsymbol{\gamma},$$

as before

$$\mathbf{Z} = (\mathbf{Z}^{(1)}\ \mathbf{0}),$$

where $\mathbf{Z}^{(1)}$ has order $n \times j$ and rank j: $\gamma_1, \gamma_2, \ldots, \gamma_j$ are initially estimable, but $\gamma_{j+1}, \gamma_{j+2}, \ldots, \gamma_k$ are not. Our choice of $\mathbf{v}_{j+1}$ ensures that

$$\mathbf{c}'\boldsymbol{\beta} = a_1\gamma_1 + a_2\gamma_2 + \ldots + a_{j+1}\gamma_{j+1},$$

that is, that

$$\mathbf{c}'\mathbf{V} = (a_1, a_2, \ldots, a_{j+1}, 0, 0, \ldots, 0).$$

To make $\mathbf{c}'\boldsymbol{\beta}$ estimable we must choose $\mathbf{x}_{n+1}$ so that $\mathbf{z}' = \mathbf{x}'_{n+1}\mathbf{V}$, has the form

$$\mathbf{z}' = l(z_1, z_2, \ldots, z_j, 1, 0, 0, \ldots, 0),$$

because we must make γ_{j+1} estimable, and the only way of doing so is to have the $(j+1)$th component of $\mathbf{z}'$ non-zero and the subsequent components zero.

For the canonical version of the problem, therefore, we may ignore $\gamma_{j+2}, \gamma_{j+3}, \ldots, \gamma_k$ completely and suppose that we are given initially a matrix $(\mathbf{Z}^{(1)}\,\mathbf{0})$ of explanatory variables, where $\mathbf{Z}^{(1)}$ is an $n \times j$ matrix of rank j and $\mathbf{0}$ is *the zero column vector*. We have to choose a new set of values

$$l(z_1, z_2, \ldots, z_j, 1) = l(\mathbf{z}', 1)$$

of the explanatory variables in order to minimize

$$\mathbf{a}'\left[\begin{pmatrix} \mathbf{Z}^{(1)} & \mathbf{0} \\ l\mathbf{z}' & l \end{pmatrix}'\begin{pmatrix} \mathbf{Z}^{(1)} & \mathbf{0} \\ l\mathbf{z}' & l \end{pmatrix}\right]^{-1}\mathbf{a},$$

which is the variance of the least-squares estimator of $\mathbf{a}'\boldsymbol{\gamma} = \mathbf{c}'\boldsymbol{\beta}$.

Write $\mathbf{U} = \mathbf{Z}^{(1)'}\mathbf{Z}^{(1)}$. Then

$$\begin{pmatrix} \mathbf{Z}^{(1)} & l\mathbf{z} \\ \mathbf{0}' & l \end{pmatrix}\begin{pmatrix} \mathbf{Z}^{(1)} & \mathbf{0} \\ l\mathbf{z}' & l \end{pmatrix} = \begin{pmatrix} \mathbf{U} + l^2\mathbf{z}\mathbf{z}' & l^2\mathbf{z} \\ l^2\mathbf{z}' & l^2 \end{pmatrix}$$

and the inverse of this matrix is

$$\begin{pmatrix} \mathbf{U}^{-1} & -\mathbf{U}^{-1}\mathbf{z} \\ -\mathbf{z}'\mathbf{U}^{-1} & (1/l^2) + \mathbf{z}'\mathbf{U}^{-1}\mathbf{z} \end{pmatrix}.$$

Hence

$$\operatorname{var}(\mathbf{c}'\boldsymbol{\beta}_{n+1}) = \mathbf{a}'\begin{pmatrix} \mathbf{U}^{-1} & -\mathbf{U}^{-1}\mathbf{z} \\ -\mathbf{z}'\mathbf{U}^{-1} & (1/l^2) + \mathbf{z}'\mathbf{U}^{-1}\mathbf{z} \end{pmatrix}\mathbf{a}$$

$$= (\boldsymbol{\alpha}'\, a_{j+1})\begin{pmatrix} \mathbf{U}^{-1} & -\mathbf{U}^{-1}\mathbf{z} \\ -\mathbf{z}'\mathbf{U}^{-1} & (1/l^2) + \mathbf{z}'\mathbf{U}^{-1}\mathbf{z} \end{pmatrix}\begin{pmatrix} \boldsymbol{\alpha} \\ a_{j+1} \end{pmatrix}$$

where $\mathbf{a}' = (\boldsymbol{\alpha}'\, a_{j+1})$, i.e., $\boldsymbol{\alpha}' = (a_1, a_2, \ldots, a_j)$.

This simplifies to

$$\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_{n+1}) = (a_{j+1}\mathbf{z} - \boldsymbol{\alpha})'\mathbf{U}^{-1}(a_{j+1}\mathbf{z} - \boldsymbol{\alpha}) + \frac{a_{j+1}^2}{l^2}$$

Now the condition $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} = b^2$ is equivalent to $l^2(\mathbf{z}'\mathbf{z}+1) = b^2$ and so

$$\operatorname{var}(\mathbf{c}'\hat{\boldsymbol{\beta}}_{n+1}) = (a_{j+1}\mathbf{z}-\boldsymbol{\alpha})'\mathbf{U}^{-1}(a_{j+1}\mathbf{z}-\boldsymbol{\alpha})+a_{j+1}^2\frac{(1+\mathbf{z}'\mathbf{z})}{b^2}.$$

The problem is to choose $\mathbf{z}$ to minimize this expression.

Differentiation yields

$$a_{j+1}\mathbf{U}^{-1}(a_{j+1}\mathbf{z}-\boldsymbol{\alpha})+\frac{a_{j+1}^2}{b^2}\mathbf{z} = \mathbf{0}$$

or

$$a_{j+1}^2\left[\frac{1}{b^2}\mathbf{I}+\mathbf{U}^{-1}\right]\mathbf{z} = a_{j+1}\mathbf{U}^{-1}\boldsymbol{\alpha}$$

or

$$\mathbf{z} = \left(\mathbf{I}+\frac{1}{b^2}\mathbf{U}\right)^{-1}\left[\frac{1}{a_{j+1}}\boldsymbol{\alpha}\right]. \tag{4.4}$$

(Note that a_{j+1} is non-zero since we assumed $\mathbf{c}'\boldsymbol{\beta}$ to be inestimable initially.)

Equation (4.4) is equivalent to

$$\begin{pmatrix}\mathbf{z}\\1\end{pmatrix} = \frac{1}{a_{j+1}}\begin{pmatrix}\mathbf{I}+b^{-2}\mathbf{U} & \mathbf{0}\\ \mathbf{0}' & 1\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{\alpha}\\ a_{j+1}\end{pmatrix}$$

or to

$$\begin{pmatrix}\mathbf{z}\\1\\0\\\vdots\\0\end{pmatrix} = \frac{1}{a_{j+1}}\begin{pmatrix}\mathbf{I}+b^{-2}\mathbf{U} & \mathbf{0}\\ \mathbf{0}' & \mathbf{I}\end{pmatrix}^{-1}\begin{pmatrix}\boldsymbol{\alpha}\\ a_{j+1}\\0\\\vdots\\0\end{pmatrix}, \tag{4.5}$$

where each vector is of order $k\times 1$.

Reverting now to the original model, (4.5) says that

$$\mathbf{V}'\mathbf{x}_{n+1} \propto \begin{pmatrix}\mathbf{I}+b^{-2}\mathbf{U} & \mathbf{0}\\ \mathbf{0}' & \mathbf{I}\end{pmatrix}^{-1}\mathbf{V}'\mathbf{c}$$

that is,

$$\mathbf{x}_{n+1} \propto \left[\mathbf{V}\begin{pmatrix}\mathbf{I}+b^{-2}\mathbf{U} & \mathbf{0}\\ \mathbf{0}' & \mathbf{I}\end{pmatrix}^{-1}\mathbf{V}'\right]\mathbf{c}.$$

As may be easily verified, the matrix in square brackets in this final equation is $[\mathbf{I}+b^{-2}\mathbf{X}'\mathbf{X}]^{-1}$.

We have now completed a proof of the following.

Theorem 2. Given the model (1.2) and the condition $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} = b^2$ satisfied by a new set, $\mathbf{x}_{n+1}$, of values of the explanatory variables, the optimum direction of $\mathbf{x}_{n+1}$ for improving the precision of estimation of $\mathbf{c}'\boldsymbol{\beta}$, is that of the vector $(\mathbf{I}+b^{-2}\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$. Here it is assumed that the error associated with the new observation is uncorrelated with errors in the original model and has the same variance as each of them. No condition is imposed on rank $(\mathbf{X})$.

5. REMARKS

5.1. One of the implications of Theorem 2 seems worthy of comment. The larger is b^2, that is, the larger we can take the length of the new vector $\mathbf{x}_{n+1}$, the nearer to the direction of $\mathbf{c}$ should we take its direction. This is intuitively acceptable. If b^2 is very large relative to the components of $\mathbf{X}'\mathbf{X}$, the potential information about $\mathbf{c}'\boldsymbol{\beta}$ in the new observations dominates that contained in the original observations, so that we should act almost as if we were starting with no information at all about $\mathbf{c}'\boldsymbol{\beta}$. If, on the other hand, b^2 is small, it becomes much more important to make optimum use of the information about $\mathbf{c}'\boldsymbol{\beta}$ in the original observations when choosing the direction of $\mathbf{x}_{n+1}$.

5.2. It is natural to enquire how the result stated in Theorem 2 is affected if the error in the new observation taken at $\mathbf{x}_{n+1}$ has a variance different from that in each of the original data.

Suppose, then, that the variance of the new observation is $\lambda\sigma^2$, that is, λ times the variance of each of the original errors. To highlight the effect of this we shall take $b^2 = 1$ and suppose that $\mathbf{x}'_{n+1}\mathbf{x}_{n+1} = 1$. The model for the new observation,

$$y_{n+1} = \mathbf{x}'_{n+1}\boldsymbol{\beta} + \epsilon_{n+1},$$

is equivalent to

$$\frac{y_{n+1}}{\sqrt{\lambda}} = \left(\frac{1}{\sqrt{\lambda}}\mathbf{x}_{n+1}\right)'\boldsymbol{\beta} + \frac{\epsilon_{n+1}}{\sqrt{\lambda}},$$

which we may rewrite as

$$y^*_{n+1} = \mathbf{x}^{*\prime}_{n+1}\boldsymbol{\beta} + \epsilon^*_{n+1},$$

where ϵ^*_{n+1} has variance σ^2 and $\mathbf{x}^*_{n+1}$ is subject to the condition $\mathbf{x}^{*\prime}_{n+1}\mathbf{x}_{n+1} = \lambda^{-1}$. By Theorem 2 the optimum direction for $\mathbf{x}^*_{n+1}$, (and so for $\mathbf{x}_{n+1}$ which is proportional to $\mathbf{x}^*_{n+1}$) is that of the vector $(\mathbf{I}+\lambda\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$. Thus the smaller the variance of the new observation, the more nearly should we take $\mathbf{x}_{n+1}$ in the direction of $\mathbf{c}$. This certainly conforms with the intuitively obvious result that if the new observation has no error at all, we must take $\mathbf{x}_{n+1}$ in the direction of $\mathbf{c}$!

5.3. Theorem 2 tells us that if we have no information about $\boldsymbol{\beta}$ at all, then the best direction for a set of explanatory variables in order to estimate $\mathbf{c}'\boldsymbol{\beta}$ is that of $\mathbf{c}$, for we can apply the theorem simply by taking $\mathbf{X} = \mathbf{0}$. This is a result which could scarcely be described as surprising. However we can apply the theorem in a slightly less obvious way, when we have prior information about $\boldsymbol{\beta}$ in the form of an unbiased estimator $\tilde{\boldsymbol{\beta}}$ whose variance matrix is known; in other words the summary of our prior information is

$$\tilde{\boldsymbol{\beta}} = \boldsymbol{\beta} + \mathbf{u}, \tag{5.3.1}$$

where $E(\mathbf{u}) = \mathbf{0}$ and $\operatorname{var}(\mathbf{u}) = \sigma^2\boldsymbol{\Sigma}$, $\boldsymbol{\Sigma}$ being a known positive definite matrix, and σ^2 a known constant [in fact the variance of the new observation (see below)]. We propose to take an observation on a dependent variable y at a set of values $\mathbf{x}' = (x_1, x_2, \ldots, x_k)$ of explanatory variables in order to increase our information about a particular linear function $\mathbf{c}'\boldsymbol{\beta}$ as, for example, if we wish to know more about a particular coefficient β_i. In what direction should we choose $\mathbf{x}$, when $\mathbf{x}'\mathbf{x} = 1$ and the error in the new observation is uncorrelated with $\mathbf{u}$ and has variance σ^2? Theorem 2 can be applied as follows.

Since $\boldsymbol{\Sigma}$ is positive definite, it can be expressed in the form

$$\boldsymbol{\Sigma} = \mathbf{P}\mathbf{P}',$$

where $\mathbf{P}$ is non-singular. We use the standard trick of writing (5.3.1) in the form

$$\mathbf{P}^{-1}\tilde{\boldsymbol{\beta}} = \mathbf{P}^{-1}\boldsymbol{\beta}+\mathbf{P}^{-1}\mathbf{u}$$

or

$$\mathbf{Y} = \mathbf{P}^{-1}\boldsymbol{\beta}+\mathbf{v}$$

where $\mathbf{Y} = \mathbf{P}^{-1}\tilde{\boldsymbol{\beta}}$ and $\mathbf{v} = \mathbf{P}^{-1}\mathbf{u}$ so that $\operatorname{var}\mathbf{v} = \sigma^2\mathbf{I}$.

Thus our prior knowledge in (5.3.1) is equivalent to that contained in observations $Y_1, Y_2, \ldots, Y_k$ at values of the explanatory variables given by the rows of $\mathbf{P}^{-1}$, those observations having uncorrelated errors with a common variance. By Theorem 2 the optimum direction for $\mathbf{x}$ is that of

$$[\mathbf{I}+(\mathbf{P}')^{-1}\mathbf{P}^{-1}]^{-1}\mathbf{c} = (\mathbf{I}+\boldsymbol{\Sigma}^{-1})^{-1}\mathbf{c}.$$
$$= (\mathbf{I}+\boldsymbol{\Sigma})^{-1}\boldsymbol{\Sigma}\mathbf{c}.$$

5.4. There are various obvious questions which we have left unanswered. How do we choose $\mathbf{x}_{n+1}$ when one of the explanatory variables is the constant variable 1? The answer will depend on what restrictions there are on the choice of values of the other explanatory variables. If we have to choose several sets of explanatory variables for further observations, how do we choose them? Clearly one possibility is to choose them sequentially, optimizing at each stage; but this sequential choice may not be optimum. What happens if the error in the new observation is correlated with the previous errors? And so on.

5.5. It is scarcely necessary to remark that, from the point of view of the econometrician such questions are often of purely academic interest, since values of explanatory variables are determined for him by the system and he has no freedom of choice whatsoever. Our main object has been to illustrate some additional possible advantages of looking at a particular orthogonal version of the linear model, as suggested by Kendall (1957), namely that derived from consideration of the latent roots and vectors of $\mathbf{X}'\mathbf{X}$. We do not suggest that this version is invariably the most advantageous either from a theoretical or from a practical viewpoint. However, it may be found useful for other problems, as it clearly is for those considered here.

REFERENCES

FARRAR, D. E. and GLAUBER, R. R. (1967). Multicollinearity in regression analysis: the problem revisited. *Rev. Econ. Statist.*, **XLIX**, 92–107.

JOHNSTON, J. (1963). *Econometric Methods*, pp. 201–207. New York: McGraw-Hill.

KARLIN, S. and STUDDEN, W. J. (1966). Optimal experimental design. *Ann. Math. Statist.*, **37**, 783–816.

KENDALL, M. G. (1957). *A Course in Multivariate Analysis*, pp. 70–75. London: Griffin.

KIEFER, J. (1958). On non-randomised optimality and randomised non-optimality of experimental designs. *Ann. Math. Statist.*, **29**, 675–699.

KIEFER, J. and WOLFOWITZ, J. (1960). The equivalence of two extremum problems. *Can. J. Math.*, **12**, 363–366.

MALINVAUD, E. (1966). *Statistical Methods of Econometrics*, pp. 187–192. Amsterdam: North Holland.

RAO, C. R. (1965). *Linear Statistical Inference and its Applications*, p. 181. New York: Wiley.

SCHEFFÉ, H. (1959). *The Analysis of Variance*, pp. 13–15. New York: Wiley.