View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by Research Papers in Economics

Center for Economic Research

CANONICAL PARTITIONS IN THE RESTRICTED LINEAR MODEL

by

Ben Van Der Genugten

June 1997

Abstract

In the linear model $y = X\beta + \varepsilon$ under the restriction $C\beta = 0$ a canonical partition $C = [C_0; C_1]$ of the rows of C admits a simple representation of linear subspaces of values $\mu = X\beta$. Its use is shown for the identification, estimating and testing of linear combinations $D\beta$. Results are derived without imposing any rank conditions on X, C and D. Applications are in the field of experimental design with unbalanced data.

Keywords. Restricted linear model, canonical partition, identification, testing, estimation, experimental design.

JEL codes: C12, C13, C90

1 Introduction

Consider the linear model

$$y = \mu + \varepsilon, \quad \mu = X\beta, \quad E(\underline{\varepsilon}) = 0 \tag{1}$$

with $y \in \mathbb{R}^n$, deterministic $X \in \mathbb{R}^{n \times k}$, $\beta \in \mathbb{R}^k, \varepsilon \in \mathbb{R}^n$, where $\mathcal{L}(\varepsilon)$ does not depend on β. In the restricted model the range of β-values has to satisfy $C\beta = 0$ for some given $C \in \mathbb{R}^{p \times k}$; we write $\beta \in K_0 = \mathcal{N}(C)$. Such models arise in a natural way in experimental design. Restrictions can become quite complicated in designs with unbalanced data.

We are interested in statistical conclusions about identifiable $D\beta$ based on statistics obtained from LS (Least Squares). There is a huge literature available on this old problem. For an overview we refer to the textbooks Rao (1973), 4a and 4i and, more recently, Searle (1987), 5.6. Formulae for LS-statistics in terms of X, C, D are mostly generated by formal matrix manupulations involving generalized inverses. For more numerical aspects we refer to Björck (1996), 5.1.

In this paper we derive alternatives based on clear geometric interpretations. This becomes possible by using a canonical partition of the rows of C. This concept was introduced in Van Der Genugten (1997) in the restrictive context of testing $H_0 : D\beta = 0$. However, its application is far more general as will be shown below.

We give an outline of the paper. Section 2 is preliminary and treats the equivalence between identifiability, estimability and LS-uniqueness. Section 3 introduces the concept of a canonical partition and shows its relation with identifiability. Section 4 treats the LS-estimation of $D\beta$ and section 5 the usual F-test based on LS for $H_0 : D\beta = 0$. Finally, in section 6 we make a comparison with some results of Rao and Searle.

For conclusions about distributional properties of estimators and tests we need further assumptions about $\mathcal{L}(\varepsilon)$ in (1) like $\mathrm{Cov}(\varepsilon) = \sigma^2 I_n$ or even $\varepsilon \sim N_n(0, \sigma^2 I_n)$.

2 Identifiability

From an inferential point of view we should refuse estimates of non-identifiable functions of the unknown parameter. Within the context of the linear model $D\beta, \beta \in K_0$ is identifiable if (by definition) different values of $D\beta$ correspond to different sample distributions $\mathcal{L}(y)$. Due to the fact that β does not depend on $\mathcal{L}(\varepsilon)$ we have (see e.g. Van Der Genugten (1977) or Prakaso Rao (1992), 7.2):

Identifiability: $D\beta, \beta \in K_0$ is identifiable iff there exists a (necessarily linear) function $\nu : X(K_0) \to D(K_0)$ with $\nu(X\beta) = D\beta, \beta \in K_0$.

Roughly spoken, $D\beta$ is identifiable iff it is a function of $\mu = X\beta$.

Let $z_0 = P_0 y$ with P_0 the orthogonal projection matrix with respect to $X(K_0)$. Then z_0 is an unbiased estimator for $\mu \in X(K_0)$. From this it easily follows that $D\beta, \beta \in K_0$ is identifiable iff it can be estimated unbiasedly.

Let $b_0 = b_0(y)$ be LS-estimator of $\beta \in K_0$, i.e. minimizes $|y - X\beta|^2$ under $\beta \in K_0$. We call Db_0 LS-estimator of $D\beta, \beta \in K_0$. We say that Db_0 is unique if Db_0 does not depend on the choice of b_0 (for all $y \in \mathbb{R}^n$). In particular, $z_0 = Xb_0 = P_0 y$ is the unique LS-estimator of $\mu = X\beta, \beta \in K_0$. From this it easily follows that $D\beta, \beta \in K_0$ is identifiable iff its LS-estimator Db_0 is unique.

Note that the equivalence between identifiability, estimability and LS-uniqueness can be derived just by considering linear spaces and functions. No matrix calculations are needed at all.

A simple necessary and sufficient condition for identifiability of $D\beta, \beta \in K_0$ directly in terms of X and C is given by (see e.g. Rao (1973), 4i.2 (iii), p. 297):

$$\mathcal{R}(D') \subseteq \mathcal{R}(X' \;\; C'). \tag{2}$$

The relation (2) simply means that the rows of D are linear combinations of the rows of X and C.

Often rows of C do not contribute to identification. We formulate this trivial extension in the following way. Let C_0 be a submatrix of rows of C with

$$\mathcal{R}(X' \;\; C_0) = \mathcal{R}(X' \;\; C'). \tag{3}$$

Then $D\beta, \beta \in K_0$ is identifiable iff

$$\mathcal{R}(D') \subseteq \mathcal{R}(X' \;\; C_0'). \tag{4}$$

We need (3) and (4) for further reference.

The condition (2) or its generalization (4) cannot be verified straightforward in a numeric way. For this we use a g-inverse.

In general, for any $A \in \mathbb{R}^{n\times m}$ we call $A^- \in \mathbb{R}^{m\times n}$ g-inverse of A if $AA^-A = A$. Such a matrix acts as an inverse for appropriate matrices: $BA^-A = B$ iff $\mathcal{R}(B') \subseteq \mathcal{R}(A')$. Note that $\mathcal{R}(A') = \mathcal{R}((A^-A)')$.

In particular, set

$$[X; C_0]^- = [H_0 \;\; G_0]. \tag{5}$$

Here and in the following we write $[X; C_0] = [X' \;\; C_0']'$ for column partition. Let

$$J_0 = [X; C_0]^-[X; C_0] = H_0X + G_0C_0. \tag{6}$$

Clearly, (4) is fulfilled iff

$$D(I_k - J_0) = 0. \tag{7}$$

This condition can be verified easily in a numeric way.

Note that (7) holds for $D = J_0$. So $J_0\beta, \beta \in K_0$ generates the class of all identifiable linear combinations of $\beta \in K_0$ by premultiplication of matrices.

Furthermore, note that for identifiable $D\beta, \beta \in K_0$ a matrix representation of the linear function ν is given by DH_0 since

$$D\beta = DJ_0\beta = DH_0X\beta + G_0C_0\beta = DH_0\mu. \tag{8}$$

Finally, note that other candidates for generating all identifiable linear functions can be found as well. For example, since $\mathcal{R}(X'X) = \mathcal{R}(X')$ we may replace X by $X'X$ in the definition of J_0. The same can be done with C_0 and $C_0'C_0$.

3 Canonical partitions

The concept of canonical partitions was introduced in Van Der Genugten (1997) in the restricted context of testing.

We write $L_0 = X(K_0) = X(\mathcal{N}(C)) \subseteq L = X(\mathbb{R}^k) = \mathcal{R}(X)$. So L is the range of μ-values in the unrestricted model.

Let C^* be any submatrix of rows from C. Then (see e.g. Rao (1973), 1b.6 (iii), p. 28):

$$\begin{aligned}\dim L = r(X) \quad &\geq \dim X(\mathcal{N}(C^*)) = r(X; C^*) - r(C^*) \\ &\geq \dim L_0 = r(X; C) - r(C).\end{aligned}$$

Hence, there exists an (in general not uniquely determined) submatrix C_0 with a *maximum* number of rows from C such that

$$\dim L = r(X) = \dim X(\mathcal{N}(C_0)) = r(X; C_0) - r(C_0). \tag{9}$$

or, equivalently, $L = X(\mathcal{N}(C_0))$.

Given such C_0 we denote the submatrix of remaining rows by C_1. By reordering we may write $C = [C_0; C_1]$ without loss of generality. We call this a *canonical* partition of C. (It is possible that C_0 or C_1 is empty; we proceed with the general case that C_0 and C_1 are not empty.) The construction of C_0 is quite easy by inspecting the rows of C subsequently and adding a row to the already obtained rows if the rank condition (9) still holds for the augmented set.

Since (9) is equivalent to $\mathcal{R}(X') \cap \mathcal{R}(C_0') = \{0\}$ and since the number of rows in C_0 is maximal, it follows that the rows of C_1 belong to $\mathcal{R}(X' \ \ C_0')$ or

$$\mathcal{R}(C_1') \subseteq \mathcal{R}(X' \ \ C_0'). \tag{10}$$

This implies that (3) holds for the part C_0 of the canonical partition. We use (4)$-$(7) for this choice of C_0. In particular, it follows from (4), (10) and (7) that

$$C_1 J_0 = C_1. \tag{11}$$

Roughly spoken, the rows of C_0 help with identification and given C_0 the rows of C_1 generate the real restrictions.

The key of the canonical partition $[C_0; C_1]$ of C is that it admits a simple form of linear subspaces related with L_0 and L.

Theorem. Let L_1 be the orthogonal complement of L_0 with respect to L. Then

$$L_1 = \mathcal{R}(XH_0'C_1'). \tag{12}$$

Proof. With (11) and (6) we get

$$\begin{aligned} L_0 \ &= X(\mathcal{N}(C)) = X(\mathcal{N}(C_0; C_1)) = X(\mathcal{N}(C_0; C_1 J_0)) = \\ &= X(\mathcal{N}(C_0; C_1 H_0 X'X + C_1 G_0 C_0)) = X(\mathcal{N}(C_0); C_1 H_0 X'X)) = \\ &= \{\mu \in X(\mathcal{N}(C_0) : C_1 H_0 X'\mu = 0\} = \\ &= \{\mu \in L : C_1 H_0 X'\mu = 0\} = X(\mathcal{N}(C_1 H_0 X'X)) \end{aligned}$$

or

$$L_1 = \mathcal{R}((C_1 H_0 X')') = \mathcal{R}(XH_0'C_1'). \square$$

The dimensions of the linear spaces involved follow immediately from (3):

$$\left.\begin{array}{lcl} \dim L_0 & = & r(X;C) - r(C) = r(X;C_0) - r(C) \\ \dim L & = & r(X) = r(X;C_0) - r(C_0) \\ \dim L_1 & = & r(C) - r(C_0). \end{array}\right\} \quad (13)$$

We use the results in the following sections for obtaining expressions for LS-statistics related to estimation and testing.

4 Estimation

Let P, P_0, P_1 denote the orthogonal projection matrices belonging to $L = \mathcal{R}(X), L_0 = X(\mathcal{N}(C))$ and the orthogonal complement L_1 of L_0 with respect to L, respectively. Clearly,

$$P = X(X'X)^- X'. \quad (14)$$

With (12) this implies

$$P_1 = XH_0'C_1'(C_1H_0X'XH_0'C_1')^- C_1H_0X'. \quad (15)$$

From the definition of P_0 we have

$$P_0 = P - P_1. \quad (16)$$

From (14)$-$(16) explicit expressions for the LS-estimator Db_0 for identifiable $D\beta, \beta \in K_0$ are easily derived. With (8) we get $Db_0 = DH_0z_0 = DH_0P_0y$ and so

$$Db_0 = DM_0y \quad (17)$$

with

$$M_0 = H_0 P_0. \tag{18}$$

We can apply (17) for $D = X$ and $D = C$. This gives $Xb_0 = P_0 y$ and $Cb_0 = 0$. So, if we define

$$b_0 = M_0 y, \tag{19}$$

then b_0 is a LS-solution of $\beta \in K_0$.

Under the assumption $V(\varepsilon) = \sigma^2 I_n$ we get from (19) and (18)

$$\mathrm{Cov}(b_0) = \sigma^2 M_0 M_0' = \sigma^2 H_0 P_0 H_0'. \tag{20}$$

The error variance σ^2 is estimated unbiasedly by

$$\hat{\sigma}^2 = |e_0|^2 / \dim L_0^{\perp}, \tag{21}$$

where $e_0 = y - z_0 = (I_n - P_0)y$ stands for the LS-residual. Note that $\dim L_0^{\perp} = n - \dim L_0$ and

$$|e_0|^2 = |y|^2 - |z_0|^2 = y'y - y'P_0 y \tag{22}$$

5 Testing

Assume that $\varepsilon \sim N_n(0, \sigma^2 I_n)$.

The usual F-statistic for $H_0 : D\beta = 0$ against $H_1 : D\beta \neq 0$ for identifiable $D\beta, \beta \in K_0$, is given by

$$F = \frac{|z_{01}|^2 / \dim L_{01}}{|e_0|^2 / \dim L_0^{\perp}} \sim F^{\dim L_{01}}_{\dim L_0^{\perp}}(|\mu_{01}|^2 / \sigma^2). \tag{23}$$

Here $z_{01} = P_{01} y, \mu_{01} = P_{01} \mu$ with P_{01} the orthogonal projection matrix of L_{01}, by definition the orthogonal complement of $L_{00} = X(\mathcal{N}(C; D))$ with respect to $L_0 = X(\mathcal{N}(C))$. We derive expressions for $P_{01}, |z_{01}|^2$ and $\dim L_{01}$ appearing in (23).

From (4) it follows that $[C_0; D_1]$ with $D_1 = [C_1; D]$ is a canonical partition of $[C; D]$. So we can apply (12) to the orthogonal complement $L_{11} = L_{01} + L_1$ of L_{00} with respect to L:

$$L_{11} = \mathcal{R}(XH_0'D_1'). \tag{24}$$

Let P_{11} be the orthogonal projection matrix with respect to L_{11}. Then from (14) we get

$$P_{11} = (XH_0'D_1')(D_1H_0X'XH_0'D_1)^- D_1H_0X'. \tag{25}$$

Since $L_{01} \perp L_1$ this implies

$$P_{01} = P_{11} - P_1 \tag{26}$$

$$|z_{01}|^2 = y'P_{01}y = y'P_{11}y - y'P_1y. \tag{27}$$

Furthermore, with (13) we get $\dim L_{11} = r(C; D) - r(C_0)$ and so

$$\dim L_{01} = r(C; D) - r(C) \tag{28}$$

In Van Der Genugten (1977) a slight simplification for (26) is given.

6 A comparison

In Searle (1987), 5.6 two seperate cases are distinguished:

a) Restrictions involving estimable functions. This refers to the condition $\mathcal{R}(C') \subseteq \mathcal{R}(X')$. So in this case the part C_0 of the canonical partition of C is empty.

b) Restrictions involving non-estimable functions. This refers to the condition that $L_0 = L$. So in this case the part C_1 of the canonical partition of C is empty.

In fact, section 4 treats the general case without imposing any rank conditions on C.

The classic approach of Rao (1973), 4i.1 gives an alternative for (19). It can be given a nice geometric interpretation as well in the following way.

The vector $b_0 \in K_0$ is LS-estimator iff the corresponding LS-residual $e_0 = y - Xb_0$ is orthogonal to $L_0 = X(\mathcal{N}(C))$ or, equivalently,

$$X'e_0 \in \mathcal{N}(C)^{\perp} = \mathcal{R}(C'). \tag{29}$$

Note that (29) is equivalent to

$$Cb_0 = 0 \text{ and } X'y - X'Xb_0 = -C'd_0 \text{ for some } d_0$$

or

$$[X'X \quad C'; C \quad 0](b_0; d_0) = (X'y; 0) \text{ for some } d_0.$$

Hence,

$$b_0 = M_0^* y$$

with M_0^* the (1,1)-part of $[X'X \quad C'; C \quad 0]^-$. This gives the alternative expression $P_0 = XM_0^*X'$ for (16).

This approach lacks the nice interpretation (12) and gives no easy condition for identifiability of $D\beta, \beta \in K_0$ like (7). An alternative would be to use in (29) the fact that $\mathcal{R}(C') = \mathcal{R}(C'C)$ and that $Cb_0 = 0$ iff $C'Cb_0 = 0$, leading to

$$\begin{pmatrix} X'X + C'C & C'C \\ C'C & 0 \end{pmatrix} \begin{pmatrix} b_0 \\ d_0 \end{pmatrix} = \begin{pmatrix} X'y \\ 0 \end{pmatrix} \text{ for some } d_0$$

or

$$b_0 = M_0^{**} y$$

with M_0^{**} the (1,1)-part of $[X'X + C'C \;\; C'C; C'C \;\; 0]^-$. Then the expression for $X'X + C'C$ can also be used for the identifiability condition (7) (with $C_0 = C$). We simply take

$$[X; C]^- = ([X; C]'[X; C])^-[X; C]' = (X'X + C'C)^-[X' \;\; C']$$

leading to

$$J_0^* = [X; C]^-[X; C] = (X'X + C'C)^-(X'X + C'C).$$

So, in (7) J_0 may be replaced by J_0^*.

For other alternatives for J_0 using the fact that $\mathcal{R}(X') = \mathcal{R}(X'X)$ we refer to Van Der Genugten (1997). See also Van Der Genugten (1993) for a more elaborate discussion.

References

Björck, Å. (1996) - Numerical methods for least squares problems - Siam, Philadelphia.

Prakaso Rao, B.L.S. (1992) - Identifiability in Stochastic Distributions - Academic, New York.

Rao, C.R. (1973) - Linear Statistical Inference and Its Applications, 2^{nd} edit. - Wiley, New York.

Searle, S.R. (1987) - Linear Models for Unbalanced Data - Wiley, New York.

Van Der Genugten, B.B. (1977) - Identification in statistical inference - Statist. Neerlandica 31, p. 69-89.

Van Der Genugten, B.B. (1993) - Identification, estimating and testing in the restricted linear model - Research Memorandum FEW 632, Tilburg University.

Van Der Genugten, B.B. (1997) - Testing in the restricted linear model using canonical partitions - Linear Algebra and Its Applications (to appear).