POLSKA AKADEMIA NAUK, INSTYTUT MATEMATYCZNY

DISSERTATIONES MATHEMATICAE

(ROZPRAWY MATEMATYCZNE)

CCCXVII

NGUYEN BAC-VAN

Estimation and prediction in regression models with random explanatory variables

WARSZAWA 1992

Nguyen Bac-Van
Department of Mathematics
University of HoChiMinh City
227 Nguyen-van-Cu, Q.5, VietNam

Published by the Institute of Mathematics, Polish Academy of Sciences
Typeset in TEX at the Institute
Printed and bound by
drukarnia
herman & herman
02-240 Warszawa, ul. Jakobinów 23, tel: 846-79-66, tel/fax: 49-89-95

P R I N T E D I N P O L A N D

ISBN 83-85116-40-0 ISSN 0012-3862

CONTENTS

1991 *Mathematics Subject Classification*: Primary 62J02; Secondary 62F11.
Received June 11, 1990. Revised version April 19, 1991.

Abstract

The regression model $\{X(t), Y(t); t = 1, \ldots, n\}$ with random explanatory variable X is transformed by prescribing a partition $S_1, \ldots, S_k$ of the given domain S of X-values and specifying

$$\{X(1), \ldots, X(n)\} \cap S_i = \{X_{i1}, \ldots, X_{i\alpha(i)}\}, \qquad i = 1, \ldots, k\,.$$

Through the conditioning

$$\{\alpha(i) = a(i), i = 1, \ldots, k\}\,, \qquad \{X_{i1}, \ldots, X_{i\alpha(i)}; i = 1, \ldots, k\} = \{x_{11}, \ldots, x_{ka(k)}\}\,,$$

the initial model with i.i.d. pairs $(X(t), Y(t)), t = 1, \ldots, n$, becomes a conditional fixed-design $(x_{11}, \ldots, x_{ka(k)})$ model

$$\{Y_{ij}, i = 1, \ldots, k; j = 1, \ldots, a(i)\}$$

where the response variables Y_{ij} are independent and distributed according to the mixed conditional distribution $Q(\,\cdot\,, x_{ij})$ of Y given X at the observed value x_{ij}.

Afterwards, we investigate the case

$$(Q)E(Y' \mid x) = \sum_{i=1}^{k} b_i(x)\theta_i I_{S_i}(x), \qquad (Q)D(Y \mid x) = \sum_{i=1}^{k} d_i(x)\Sigma_i I_{S_i}(x)\,,$$

which arises when the conditional distribution law of Y given X changes as X passes from a domain S_i to another, whence Y follows a mixture of distributions. Then the general transformation gives the equivalent reduction to a conditional multivariate Behrens–Fisher model. We construct conditional generalized least squares estimators of $\theta' = (\theta_1' \vdots \cdots \vdots \theta_k')$ and predictors of $Y(n+1)$ given $X(n+1) = x \in S$. Through some condition imposed on the range of θ, the CGLS estimator and predictor are shown to enjoy local and global optimality.

Preface

There is a class of regression problems based on independent identically distributed (i.i.d.) bipartite observations $\big(X(t), Y(t)\big)$ on n items ($t = 1, \ldots, n$). One wishes to estimate the regression function $E(Y \mid X = x)$ or some other conditional location characteristic and predict the response value $Y(n+1)$ on a new item when the random explanatory value $X(n+1)$ is available. An example is the problem of determining the conditional means of body dimensions as a function of some main ones (e.g., height, etc.) in the standardization of clothes. Another one is the problem of predicting the wind direction at a height H, at a certain place and time, as a function of the wind direction at the level $H = 0$; in this example, the data is directional (see [7]), and the conditional median is an adequate location characteristic. That the random character of the explanatory variable may arise from the sampling scheme is exemplified in [6] by cross-sectional data.

In this work, we treat the problem from the parametric view point, which is reasonable when, by previous studies, the functional form of the regression is approximately known.

A useful data transformation is introduced by Theorems 1, 2. On the ground of the transformed data, by adequate conditioning, we manage to reduce the model with random explanatory variable to a conditional fixed-design model.

By this method, we work out the estimation and prediction problem in case the regression function is stepwise linear in a matrix parameter and the conditional dispersion matrix has a variance components structure. Theorem 3 gives the equivalent reduction of this case to a conditional fixed-design Behrens–Fisher model. Theorem 4 gives the asymptotic estimability of the regression parameter, and Theorem 5 the conditional unbiasedness and optimality of the conditional generalized least squares estimate (CGLSE).

The prediction problem is solved in Chapter 3 in connexion with the CGLS procedure. Theorem 6 gives the conditional unbiasedness and Theorems 7, 8 the local optimality of the CGLS predictor, whereas Theorem 9 states its global optimality.

All the results but the conditional estimability are finite sample ones; to attain the optimality of CGLS estimators and predictors some restriction has to be imposed on the parameter range (Assumption 2).

The systematic use of mixed conditional distributions allows us to avoid nuisance null sets in conditional statements (preceded by the prefix (Q)), whereas the use of linear mappings enables us to shorten the reasoning in proving the optimality of GLSE (Lemma 4).

We hope that this work gives some answer to the question often raised by practitioners in statistics: how to apply the least squares procedure to random explanatory variable regression problems and what quality of its own one may expect in finite samples.

I. A data transformation preserving the conditional distribution and localizing the explanatory variable

1. Introduction. The conditional probability distribution of a random variable given another one is an object of investigation in a class of regression problems. By a random variable, abbreviated as r.v., we always mean an $\mathcal{F}$-measurable function on the basic probability space $(\Omega, \mathcal{F}, P)$ to a general Borel space. Consider a pair of r.v.'s (X, Y). Their range spaces are respectively $(H, \mathcal{A})$, $(K, \mathcal{C})$, the σ-algebras $\mathcal{A}, \mathcal{C}$ being generated by the class of open sets in the corresponding space. We henceforth assume the existence of a mixed conditional distribution $Q(C, x)$ of Y given X. By definition, $Q(\,\cdot\,, x)$ is an $\mathcal{A}$-measurable probability on $\mathcal{C}$ such that for almost all $x \in H$

$$Q(C, x) = P(Y \in C \mid x)\,,$$

where $P(\,\cdot \mid x)$ is a conditional probability given X at x.

In regression analysis, on the ground of the data

$$\{\big(X(t), Y(t)\big)\,;\, t = 1, \ldots, n\} \tag{1}$$

consisting of n i.i.d. observations on the pair (X, Y), the unknown measure-valued function $Q(\,\cdot\,, x)$ is to be investigated. Let $S \in \mathcal{A}$ be a prescribed domain with $0 < P(X \in S) < 1$, and

$$\{X(1), \ldots, X(n)\} \cap S = \{X(T_1), X(T_2) \ldots\}, \qquad T_1 < T_2 < \ldots$$

The family

$$\{[X(T_1), Y(T_1)], \ldots\}$$

is obtained from the data (1) by a so-called S-transformation. The range of the new explanatory variable is now localized in S.

In this chapter we show that, through conditioning on their definition domains, S-transformations preserve the mixed conditional distribution of $Y(\,\cdot\,)$ given $X(\,\cdot\,)$ and the independence of the pairs $(X(\,\cdot\,), Y(\,\cdot\,))$.

By choosing suitably $k \geq 1$ disjoint domains $S_1, \ldots, S_k$ and performing all S_i-transformations, by adequate conditioning, we afterwards build a conditional fixed-design model in which the new response variables are independent and follow the distributions $Q(\,\cdot\,, x_{ij})$ corresponding to observed values x_{ij} falling into

S_i. This conditional model enables us to carry out the regression estimation easier, because in every restricted domain S_i, the main characteristics of $Q(\,\cdot\,, x_{ij})$, at least approximately, often have simple parametric representation.

For the role of conditioning in statistical inference, we refer to the comments in [2].

2. Theorems on data transformation. Let us introduce, once for all, some notations and conventions, beside the ones used in Section 1.

- $\mathbb{N}$ = set of all positive integers;
- $t = 1, \ldots$; $Z(t) = Z_\omega(t) = (X(t), Y(t)) = (X_\omega(t), Y_\omega(t))$ are pairs of r.v.'s; $Z = (X, Y) = (X(1), Y(1))$; $S_1, \ldots, S_m \in \mathcal{A}$, disjoint, $m \geq 1$; $0 < P(X{\in}S_i) < 1$ $(i = 1, \ldots, m)$.
- In case of $P(X{\in}S_1) + \ldots + P(X{\in}S_m) = 1$, we only consider $S_1 + \ldots + S_m = H$.

(2) $\{T_{i1}, \ldots\} = \{t \in \{1, \ldots, n\} : X(t) \in S_i\}, \quad T_{i1} < T_{i2} < \ldots, \quad i = 1, \ldots, m\,;$

T_{ij} is called a *falling time* (in S_i).

$$Z_{ij} = Z(T_{ij}) = [X(T_{ij}), Y(T_{ij})] = (X_{ij}, Y_{ij}).$$

$f_h(\,\cdot\,)$ $(h = 1, \ldots)$ is a zero-or-one-valued arbitrary function of a certain group of falling times T_{ij} $(i = 1, \ldots, m;\ j = 1, \ldots)$.

- For fixed x,
$$I(C, x) = \begin{cases} 1 & \text{when } x \in C, \\ 0 & \text{otherwise} \end{cases}$$
is the probability measure on C degenerate at x.
- Moreover, $I_B(\,\cdot\,)$ = the indicator of the set B.
- The prefix (Q) put just before some symbol will emphasize that the object so symbolized is determined by the mixed conditional distribution $Q(\,\cdot\,, x)$.
- P_B = the conditional probability measure on $(B, \mathcal{F}|B)$ given the event B.
- $P_B(\,\cdot \mid A)$ = the conditional probability measure given A with respect to the probability space $(B, \mathcal{F}|B, P_B)$.
- $P_B(C \mid X)$, $E_B(F \mid X)$ = the conditional probability of C (resp., expectation of F) given the function X with respect to the probability space $(B, \mathcal{F}|B, P_B)$.
It is easy to check that

$$P_B(\,\cdot \mid A) = P_{B\cap A}(\,\cdot\,) \quad \text{for every} \quad B, A \in \mathcal{F}, \qquad (3)$$
$$= P_A(\,\cdot\,) \quad \text{when } A \subset B\,,$$

provided $P(A \cap B) > 0$.

For fixed positive integers $a(1), \ldots, a(k)$, the family

$$\big(Z_{11}, \ldots, Z_{1a(1)}, \ldots, Z_{k1}, \ldots, Z_{ka(k)}\big)$$

has specific distribution features described in the following two theorems. Theorem 1 is crucial for our further investigation.

THEOREM 1. *Let* $Z(1), \ldots, Z(n)$ *be i.i.d.*,

$$I_{\Omega_h}(\omega) = f_h(\ldots, T_{ij}(\omega), \ldots), \quad h = 1, \ldots, u,$$
$$\Omega^* = \Omega_1 + \ldots + \Omega_u .$$

Suppose that $\Omega_1, \ldots, \Omega_u$ *are non-empty and, for some fixed* $a(1), \ldots, a(k) \in \mathbb{N}$, $1 \le k \le m$, *the family*

$$(T_{11}, \ldots, T_{1a(1)}, \ldots, T_{k1}, \ldots, T_{ka(k)})$$

is defined on the trajectory $\{X_\omega(1), \ldots, X_\omega(n)\}$ *for every* $\omega \in \Omega^*$. *Let*

$$s_i = a(1) + \ldots + a(i), \quad i = 1, \ldots, k .$$

Then

(i) $(Z_{11}, \ldots, Z_{1a(1)}, \ldots, Z_{k1}, \ldots, Z_{ka(k)})$ *is a* P_{Ω^*}*-independent family and is* P_{Ω^*}*-independent of the family*

$$(T_{11}, \ldots, T_{1a(1)}, \ldots, T_{k1}, \ldots, T_{ka(k)}),$$

(ii) *the* P_{Ω^*}*-distribution of* Z_{ij} *coincides with the* $P_{(X \in S_i)}$*-distribution of* Z $(i = 1, \ldots, k;\ j = 1, \ldots, a(i))$,

(iii) *for any given* $A \in \mathcal{A}^{s_k}$, *the product measure*

$$R(\,\cdot\,; x_{11}, \ldots, x_{ka(k)}) = \bigtimes_{i=1}^{k} \bigtimes_{j=1}^{a(i)} [I(\,\cdot\,, x_{ij}) \times Q(\,\cdot\,, x_{ij})] \tag{4}$$
$$(x_{ij} \in H,\ i = 1, \ldots, k;\ j = 1, \ldots, a(i))$$

is a mixed conditional distribution of the family $(Z_{ij},\ i = 1, \ldots, k;\ j = 1, \ldots, a(i))$ *given the family* $W = (X_{ij},\ i = 1, \ldots, k;\ j = 1, \ldots, a(i))$ *with respect to the probability space*

$$\{\Omega^* \cap (W \in A), \mathcal{F} \mid \Omega^* \cap (W \in A), P_{\Omega^* \cap (W \in A)}\},$$

provided

$$P\{[X(1), \ldots, X(s_k)] \in A \cap (S_1^{a(1)} \times \ldots \times S_k^{a(k)})\} > 0 .$$

Later, $R(\,\cdot\,,\,\cdot\,)$ is called the $(Q)P_{\Omega^* \cap (W \in A)}$*-mixed conditional distribution.*

Remark 1. Here are two examples of Ω^*: Let

$$\Omega' = \Omega'(a, n) = \{\omega : a(i) = \#S_i \cap \{X(1), \ldots, X(n)\},\ i \le k\}, \tag{5}$$

$m = k + 1$, $S_m = H - (S_1 + \ldots + S_k)$, $a(m) = n - a(1) - \ldots - a(k)$. Then

$$I_{\Omega'}(\omega) = f_1(T_{ij},\ i = 1, \ldots, m;\ j = 1, \ldots, a(i)),$$

and Theorem 1 applies to $\Omega^* = \Omega'$ and the family $(Z_{ij},\ i = 1, \ldots, k;\ j = 1, \ldots, a(i))$.

Further, consider the i.i.d. $Z(1), \ldots, Z(n+1)$. Let

$$\Omega_h = \{X_\omega(n+1) \in S_h\} \cap \Omega'(a, n), \quad a'(h) = a(h) + 1, \quad h = 1, \ldots, k .$$

Then

$$I_{\Omega_h}(\omega) = f_h(T_{ha'(h)}; T_{11}, \ldots, T_{1a(1)}, \ldots, T_{ma(m)}) \, .$$

Theorem 1 applies to the sequence $Z(1), \ldots, Z(n+1)$ and $\Omega^* = \Omega_1 + \ldots + \Omega_k$ and $(Z_{ij},\, i = 1, \ldots, k;\; j = 1, \ldots, a(i))$. This Ω^* will be used in a later chapter for prediction purposes, when knowing X_{n+1} one predicts Y_{n+1}.

THEOREM 2. *Suppose that the infinite sequence of i.i.d. pairs* $\big(X(1), Y(1)\big), \ldots$ *is defined on* $(\Omega, \mathcal{F}, P)$. *Let*

$$(6) \qquad (T_{i1}, \ldots) = \{t \ge 1 \,:\, X(t) \in S_i\} \, , \qquad T_{i1} < T_{i2} < \ldots;\; i = 1, \ldots, k;\; k \le m; Z_{ij} = Z(T_{ij}) \, .$$

Then $T_{11}, \ldots, \ldots, T_{k1}, \ldots$ *are* P*-almost surely finite*, *and*

(i) *the infinite family* $(Z_{11}, Z_{12}, \ldots, \ldots, Z_{k1}, \ldots)$ *is a* P*-independent family and is* P*-independent of the family* $(T_{11}, T_{12}, \ldots, \ldots, T_{k1}, \ldots)$,

(ii) *the* P*-distribution of* Z_{ij} $(i = 1, \ldots, k; j = 1, 2, \ldots)$ *coincides with the* $P_{(X \in S_i)}$*-distribution of* Z,

(iii) *for arbitrary fixed* $a(1), \ldots, a(k) \in \mathbb{N}$, *the product measure* (4) *is a mixed conditional distribution of the family* $(Z_{ij}\,, i = 1, \ldots, k;\; j = 1, \ldots a(i))$ *given* $W = (X_{ij},\, i = 1, \ldots, k;\; j = 1, \ldots, a(i))$ *with respect to* $(\Omega, \mathcal{F}, P)$.

Remark 2. If the infinite sequence $Z(1), Z(2), \ldots$ is defined on $(\Omega, \mathcal{F}, P)$, Theorem 1 is a consequence of Theorem 2(i),(ii). However, to get full generality and to explain comprehensively the structure of Ω^*, we shall later prove Theorem 1 assuming only that the finite sequence $Z(1), \ldots, Z(n)$ is defined on $(\Omega, \mathcal{F}, P)$, because there are examples of spaces $(\Omega, \mathcal{F}, P)$ on which the infinite sequence can never be defined.

3. Proofs of the theorems. To prove the theorems we need some lemmas.

LEMMA 1. *Suppose* $F(\omega), X(\omega)$ *are r.v.'s. Let* X *have the range space* $(H, \mathcal{A})$, *let* F *be extended real-valued and let* $E(F)$ *exist. Let* $S \in \mathcal{A}$, $P(X \in S) > 0$. *Then*

$$E_{(X \in S)}(F \mid x) = E(F \mid x)$$

for every $x \in S - J$, *where* J *is some set such that* $P(X \in J) = 0$. *In particular*, *the product measure* $I(\,\cdot\,, x) \times Q(\,\cdot\,, x) = (I \times Q)(\,\cdot\,, x)$ *is also a mixed conditional distribution of* $Z = (X, Y)$ *given* X *with respect to the probability measure* $P_{(X \in S)}$.

Proof. $E(F \mid X) = g(X)$ P-almost surely, defined by

$$E[g(X)\, I_B(X)] = E\{F(\omega)\, I_B[X(\omega)]\} \, , \qquad \forall B \in \mathcal{A} \, .$$

Upon replacing B by BS,

$$E[g(X)\, I_B(X)\, I_S(X)] = E[F(\omega)\, I_B(X)\, I_S(X)] \, ,$$

or

$$E_{(X \in S)}[g(X)\, I_B(X)] = E_{(X \in S)}[F(\omega)\, I_B(X)], \qquad \forall B \in \mathcal{A} \, .$$

Hence, $g(X) = E_{(X\in S)}(F \mid X)$, $P_{(X\in S)}$-a.s.

Now, the product measure $I(\,\cdot\,,x) \times Q(\,\cdot\,,x)$ is an $\mathcal{A}$-measurable probability on $\mathcal{A}\times\mathcal{C}$, and satisfies the condition $I(B,X)Q(C,X) = P\{(X,Y) \in B\times C \mid X\}$ a.s., for every $B \in \mathcal{A}$, $C \in \mathcal{C}$. By extension to the σ-field $\mathcal{A}\times\mathcal{C}$, we have $(I\times Q)(D,X) = P(Z \in D \mid X)$ a.s. $(\bmod P)$ for every $D \in \mathcal{A}\times\mathcal{C}$. Hence, $(I\times Q)(D,X) = P_{(X\in S)}(Z\in D \mid X)$ a.s. $(\bmod P_{(X\in S)})$. Lemma 1 is proved.

LEMMA 2. *Let* $(F_1,G_1),\dots,(F_p,G_p)$ *be independent pairs of r.v.'s and let* $G_1,\dots,G_p$ *be extended real-valued, integrable. Then*

$$E(G_1\dots G_p \mid F_1,\dots,F_p) = E(G_1\mid F_1)\dots E(G_p\mid F_p) \quad a.s.$$

P r o o f. $E(G_i \mid F_i)$, $i=1,\dots,p$, are defined by

$$E[E(G_i\mid F_i)I_{B_i}(F_i)] = E[G_iI_{B_i}(F_i)]$$

for every $B_i\in\mathcal{A}_i$, the Borel field in the range space of F_i. By the independence of (F_i,G_i)'s we get

$$E[E(G_1\mid F_1)\dots E(G_p\mid F_p)I_{B_1\times\ldots\times B_p}(F_1,\dots,F_p)] \\ = E[G_1\dots G_pI_{B_1\times\ldots\times B_p}(F_1,\dots,F_p)]\,,$$

and by extension of signed measures we have

$$E[E(G_1\mid F_1)\dots E(G_p\mid F_p)I_B(F_1,\dots,F_p)] = E[G_1\dots G_pI_B(F_1,\dots,F_p)]\,,$$

for every $B\in\mathcal{A}_1\times\ldots\times\mathcal{A}_p$. This completes the proof of Lemma 2.

In the sequel, the enumerations such as $i=1,\dots,k$; $j=1,\dots,a(i)$ will be simply written $i\le k$, $j\le a(i)$.

For arbitrary but fixed $a(1),\dots,a(k)\in\mathbb{N}$, let

$$M = \{\{n_{11},\dots,n_{1a(1)}\},\dots,\{n_{k1},\dots,n_{ka(k)}\}\}$$

where $n_{11}<\ldots<n_{1a(1)},\dots,n_{k1}<\ldots<n_{ka(k)}$, i.e. M is any collection of k disjoint sets in $\mathbb{N}$. For T_{ij} defined in (2) or (6), let

$$A_0 = \{T_{ij} = n_{ij},\, i=1,\dots,k;\, j=1,\dots,a(i)\}\subset\Omega\,. \tag{7}$$

LEMMA 3. $A_0\in\mathcal{F}$ *and, if* $P(A_0)>0$, *then for any* $B_{ij}\in\mathcal{A}\times\mathcal{C}$

$$P_{A_0}\{Z_{ij}\in B_{ij},\, i\le k,\, j\le a(i)\} = \prod_{i=1}^{k}\prod_{j=1}^{a(i)} P_{(X\in S_i)}(Z\in B_{ij}). \tag{8}$$

P r o o f. Let

$$A' = \{X(n_{ij})\in S_i,\, i\le k,\, j\le a(i)\},$$
$$A'' = \bigcap_{i=1}^{k}\{X(t)\notin S_i,\, \forall t\in\{1,\dots,n_{ia(i)}\}-M\}\,.$$

Then A' and A'' are independent events, $P(A')>0$, $A_0 = A'\cap A''$, $P(A_0) = P(A')P(A'')$. For the particular collection

$$\{n_{11}=1,\dots,n_{1a(1)}=a(1);\dots;n_{k1}=s_{k-1}+1,\dots,n_{ka(k)}=s_k\}\,,$$

$P(A_0) = P(A') > 0$. When $P(A_0) > 0$,

$$P_{A_0}\{Z_{ij} \in B_{ij},\, i \le k,\, j \le a(i)\}$$
$$= P^{-1}(A')P^{-1}(A'')P\{A' \cap [Z_{ij} \in B_{ij},\, i \le k,\, j \le a(i)]\}P(A'')\,.$$

By noting that $Z(n_{11}), \dots, Z(n_{ka(k)})$ are independent versions of Z, we get (8). Lemma 3 is proved.

Proof of Theorem 1. We proceed in several steps.

1° Consider

$$I_{\Omega_h}(\omega) = f_h(\dots, T_{1t(1)}; \dots; \dots, T_{mt(m)})$$

where $t(i) = \max j$ for all T_{ij} which are arguments of $f_h(\,\cdot\,)$, $t(i) = 0$ if no such T_{ij} exist. For every $\omega \in \Omega_h$, $T_{1t(1)}, \dots, T_{mt(m)}$ are defined on the trajectory $\{X_\omega(1), \dots, X_\omega(n)\}$, and so are $T_{1a(1)}, \dots, T_{ka(k)}$ by the assumption. If $a(i) > t(i)$ for some i, we define

$$f_h^*(\dots; T_{i1}, \dots, T_{it(i)}, \dots, T_{ia(i)}; \dots) \equiv f_h(\dots; \dots, T_{it(i)}; \dots)\,,$$

and thus we can also consider that all T_{ij}, $j = 1, \dots, t(i)$, are arguments of $f_h(\,\cdot\,)$; now also $I_{\Omega_h} = f_h^*(\,\cdot\,)$, and the situation is brought back to the case $a(i) \le t(i)$. Thus, we shall henceforth consider

$$I_{\Omega_h} = f_h(T_{11}, \dots, T_{1t(1)}; \dots; T_{m1}, \dots, T_{mt(m)})\,, \tag{9}$$

where $t(i) \ge a(i)$, $\forall i = 1, \dots, k$.

2° Let us show that $P(\Omega_h) > 0$, $h = 1, \dots, u$. By (9) we have the finite decomposition

$$\Omega_h = \sum (T_{11} = n_{11}, \dots; \dots; \dots, T_{mt(m)} = n_{mt(m)}) = \sum A_0 \tag{10}$$

where $\sum$ extends over all collections $(n_{11}, \dots, n_{1t(1)}; \dots; \dots, n_{mt(m)}) \in f_h^{-1}(\{1\})$. Because Ω_h is non-empty, there exists an ω belonging to some summand $(T_{11} = n_{11}, \dots; \dots; \dots, T_{mt(m)} = n_{mt(m)})$. Then, on the corresponding trajectory $\{X_\omega(1), \dots, X_\omega(n)\}$, $n_{11}, \dots, n_{mt(m)}$ are really happening falling times. Hence by naming

$$S = S_1 + \dots + S_m,$$
$$(n_1 < \dots < n_m) = \{n_{1t(1)}, \dots, n_{mt(m)}\},$$
$$n_1 = n_{i_1 t(i_1)}, \dots, n_{m-1} = n_{i_{m-1} t(i_{m-1})}, \dots,$$
$$M_1 = \{1, \dots, n_1\} - \{n_{11}, \dots; \dots; \dots, n_{mt(m)}\},$$
$$M_2 = \{n_1 + 1, \dots, n_2\} - \{n_{11}, \dots; \dots; \dots, n_{mt(m)}\}, \dots,$$
$$M_m = \{n_{m-1} + 1, \dots, n_m\} - \{n_{11}, \dots; \dots; \dots, n_{mt(m)}\}\,,$$

we have

$$P(\Omega_h) \ge P(T_{11} = n_{11}, \dots, T_{mt(m)} = n_{mt(m)})$$
$$= P\{X(n_{ij}) \in S_i,\, i \le m,\, j \le t(i)\} \cdot P\{X(t) \notin S,\, t \in M_1\} \cdot \delta\,.$$

The first and second factors are positive, unless $P\{X(t) \notin S\} = 0$, but then $H - S = \emptyset$ by our convention, hence $M_1 = \emptyset$ and the second factor would not exist. The third factor

$$\delta = P\{X(t) \in [S_{i_1} \cup (H - S)],\, t \in M_2\} \\ \times \ldots \times P\{X(t) \in S_{i_1} \cup \ldots \cup S_{i_{m-1}} \cup (H - S),\, t \in M_m\}$$

is positive. Hence $P(\Omega_h) > 0$.

3° For every $B_{ij} \in \mathcal{A} \times \mathcal{C}$, from (10) and (3) it follows that

$$P_{\Omega_h}\{Z_{ij} \in B_{ij},\, i \le m,\, j \le t(i)\} \\ = \sum P_{\Omega_h}(A_0) P_{A_0}\{Z_{ij} \in B_{ij},\, i \le m,\, j \le t(i)\}\,,$$

where $\sum$ extends over those terms of (10) with $P(A_0) > 0$. Because of $\sum P_{\Omega_h}(A_0) = 1$, (8) entails

$$P_{\Omega_h}\{Z_{ij} \in B_{ij},\, i \le m,\, j \le t(i)\} = \prod_{i=1}^{m} \prod_{j=1}^{t(i)} P_{(X \in S_i)}(Z \in B_{ij})\,. \tag{11}$$

Now, consider arbitrary $q_{ij} \in \{1, \ldots, n\}$ and the event

$$A_q = \{T_{ij} = q_{ij},\, i \le m,\, j \le t(i)\}\,.$$

Let us prove that

$$P_{\Omega_h}\{Z_{ij} \in B_{ij},\, T_{ij} = q_{ij}, i \le m, j \le t(i)\} \\ = P_{\Omega_h}\{Z_{ij} \in B_{ij}, i \le m, j \le t(i)\} \cdot P_{\Omega_h}\{T_{ij} = q_{ij}, i \le m, j \le t(i)\}\,. \tag{12}$$

This is trivial if $P(\Omega_h \cap A_q) = 0$. In the case $P(\Omega_h \cap A_q) > 0$, from (9) it follows that there exists an ω such that

$$1 = I_{\Omega_h}(\omega) = f_h[T_{11}(\omega), \ldots, T_{mt(m)}(\omega)] = f_h[q_{11}, \ldots, q_{mt(m)}]\,,$$

i.e. $A_q \subset \Omega_h$; from (3) we have $P_{\Omega_h}(\,\cdot \mid A_q) = P_{A_q}(\,\cdot\,)$, and by (8) and (11) we have

$$P_{\Omega_h}\{Z_{ij} \in B_{ij},\, i \le m,\, j \le t(i) \mid A_q\} = P_{\Omega_h}\{Z_{ij} \in B_{ij},\, i \le m,\, j \le t(i)\}\,.$$

Thus (12) holds. (12) means that the families $\{Z_{ij},\, i \le m,\, j \le t(i)\}$ and $\{T_{ij},\, i \le m,\, j \le t(i)\}$ are P_{Ω_h}-independent one of another, hence so are their subfamilies $\{Z_{ij},\, i \le k,\, j \le a(i)\}$ and $\{T_{ij},\, i \le k,\, j \le a(i)\}$. Thus we get

$$P_{\Omega_h}\{Z_{ij} \in B_{ij},\, T_{ij} = q_{ij}, i \le k, j \le a(i)\} \\ = P_{\Omega_h}\{Z_{ij} \in B_{ij}, i \le k, j \le a(i)\} \cdot P_{\Omega_h}\{T_{ij} = q_{ij}, i \le k, j \le a(i)\}\,. \tag{13}$$

Now, in (11) let $B_{ij} = H \times K$ unless $i \le k,\, j \le a(i)$; we get

$$P_{\Omega_h}\{Z_{ij} \in B_{ij}, i \le k, j \le a(i)\} = \prod_{i=1}^{k} \prod_{j=1}^{a(i)} P_{(X \in S_i)}(Z \in B_{ij})\,, \tag{14}$$

and after multiplying both sides by $P(\Omega_h)$ and summing up over $h = 1, \dots, u$, we have

$$(15) \qquad P_{\Omega^*}\{Z_{ij} \in B_{ij}, i \le k, j \le a(i)\} = \prod_{i=1}^{k} \prod_{j=1}^{a(i)} P_{(X \in S_i)}(Z \in B_{ij}).$$

Considering (14) and (15), the same multiplication and summation performed over (13) entails

$$(16) \qquad P_{\Omega^*}\{Z_{ij} \in B_{ij}, T_{ij} = q_{ij}, i \le k, j \le a(i)\} \\ = P_{\Omega^*}\{Z_{ij} \in B_{ij}, i \le k, j \le a(i)\} \cdot P_{\Omega^*}\{T_{ij} = q_{ij}, i \le k, j \le a(i)\}\,.$$

(15) and (16) prove parts (i) and (ii) of Theorem 1.

4° For any fixed point $w = (x_{11}, \dots, x_{1a(1)}; \dots; \dots, x_{ka(k)})$ in the range space H^{s_k} of the family $W = (X_{11}, \dots, X_{ka(k)})$ the function $R(\,\cdot\,, w)$ in (4) is a probability measure on the product σ-field $(\mathcal{A} \times \mathcal{C})^{s_k}$.

From (4) it follows that, for any fixed D belonging to the semi-algebra

$$(17) \qquad \Big\{D\,:\,D = \mathop{\times}_{i=1}^{k} \mathop{\times}_{j=1}^{a(i)} (A_{ij} \times C_{ij})\,;\,A_{ij} \in \mathcal{A},\,C_{ij} \in \mathcal{C}\Big\}\,,$$

$R(D, w)$ is an $\mathcal{A}^{s_k}$-measurable function in w satisfying

$$(18) \qquad R(D, W) = \prod_{i=1}^{k} \prod_{j=1}^{a(i)} I(A_{ij}, X_{ij}) Q(C_{ij}, X_{ij}) \\ = \prod_{i=1}^{k} \prod_{j=1}^{a(i)} P_{\Omega^*}\{Z_{ij} \in A_{ij} \times C_{ij} \mid X_{ij}\} \quad \text{a.s.}$$

Indeed, by (15), the P_{Ω^*}-distribution of Z_{ij} coincides with the $P_{(X \in S_i)}$-distribution of Z; hence a mixed conditional distribution of Z_{ij} given X_{ij} with respect to P_{Ω^*} is specified by the same function as that of Z given X with respect to $P_{(X \in S_i)}$, which is $I(\,\cdot, x) \times Q(\,\cdot\,, x)$ by Lemma 1. Thus (18) is justified; it is rewritten, by Lemma 2, as

$$(19) \qquad R(D, W) = P_{\Omega^*}\{\big(Z_{11}, \dots, Z_{ka(k)}\big) \in D \mid W\} \quad \text{a.s.}$$

Now, the class of sets D of $(\mathcal{A} \times \mathcal{C})^{s_k}$ such that $R(D, w)$ is an $\mathcal{A}^{s_k}$-measurable function in w satisfying (19) is monotone, closed under countable disjoint unions, and contains the semi-algebra (17), hence it coincides with $(\mathcal{A} \times \mathcal{C})^{s_k}$. Further, by applying Lemma 1, we can in (19) replace P_{Ω^*} by $P_{\Omega^* \cap (W \in A)}$ provided $P_{\Omega^* \cap (W \in A)} > 0$. Thus $R(D, w)$ is a mixed conditional distribution of $(Z_{11}, \dots, Z_{ka(k)})$ given W with respect to $P_{\Omega^* \cap (W \in A)}$, and part (iii) of Theorem 1 is proved. To finish the proof of Theorem 1 we should check that

$$P_{\Omega^*}(W \in A) = P_{\{[X(1), \dots, X(s_k)] \in S_1^{a(1)} \times \dots \times S_k^{a(k)}\}}\{[X(1), \dots, X(s_k)] \in A\}\,.$$

This follows from Theorem 1(i), (ii) by noting that the P_{Ω^*}-distribution of W coincides with the conditional distribution of $\{X(1), \ldots, X(s_k)\}$ given

$$\{X(1) \in S_1, \ldots, X(s_1) \in S_1, \ldots, X(s_{k-1}+1) \in S_k, \ldots, X(s_k) \in S_k\}.$$

Theorem 1 is proved.

Proof of Theorem 2.

1° It is simple to prove

$$P(T_{ij} < \infty, i = 1, \ldots, k; j = 1, 2, \ldots) = 1.$$

2° When considering the infinite sequence $Z(1), Z(2), \ldots$ we have, for arbitrary but fixed $a(1), \ldots, a(k) \in \mathbb{N}$,

$$1 = P(T_{11} < \infty, \ldots, T_{ka(k)} < \infty) = \sum P(A_0),$$

where the sum is extended over all sets A_0 as in (7) with $P(A_0) > 0$ that correspond to all possible collections M in Lemma 3. Then, with the same summation range we have

$$P\{Z_{ij} \in B_{ij},\, i \le k,\, j \le a(i)\} = \sum P(A_0) P_{A_0}\{Z_{ij} \in B_{ij},\, i \le k,\, j \le a(i)\}.$$

From (8) we get

$$\begin{aligned} &P\{Z_{ij} \in B_{ij},\, i \le k,\, j \le a(i) \mid T_{ij} = n_{ij},\, i \le k,\, j \le a(i)\} \\ &\qquad = P\{Z_{ij} \in B_{ij},\, i \le k,\, j \le a(i)\} = \prod_{i=1}^{k} \prod_{j=1}^{a(i)} P_{(X \in S_i)}(Z \in B_{ij}), \end{aligned}$$

which proves parts (i) and (ii) of Theorem 2.

3° Reconsider step 4° of the proof of Theorem 1; there we used the probability space $(\Omega^*, \mathcal{F}|\Omega^*, P_{\Omega^*})$. Here, in view of assertions (i) and (ii) of Theorem 2, we have to use the probability space $(\Omega, \mathcal{F}, P)$. This is the only change needed to produce the proof of part (iii) of Theorem 2; except this change, this proof copies word by word step 4° of the proof of Theorem 1.

4. Interpretation of the theorems. The subdomains $S_i, \ldots, S_k$ being suitably prescribed, now let

$$a(i) = \# S_i \cap \{X(1), \ldots, X(n)\} = \#\{X_{i1}, \ldots, X_{ia(i)}\},$$

and let $(x_{i1}, \ldots, x_{ia(i)})$ be the observed value of $(X_{i1}, \ldots, X_{ia(i)})$ corresponding to the data (1), $i = 1, \ldots, k$.

The point $a = (a(1), \ldots, a(k))$ and the observed value

$$w = (x_{11}, \ldots, x_{1a(1)}; \ldots; x_{k1}, \ldots, x_{ka(k)}),$$

corresponding to a definite data (1), are called respectively an *elementary data situation* and an *instantaneous data state*.

Theorem 1, applied to $\Omega^* = \Omega'$ in (5), now states that on the ground of any elementary situation a, at any instantaneous state w, the transformed model

$$(Y_{11}, \dots, Y_{1a(1)}; \dots; \dots, Y_{ka(k)})$$

is a fixed-design $(x_{11}, \dots, x_{ka(k)})$ model, with the response observations $Y_{11}, \dots$ $\dots, Y_{ka(k)}$ being independent and distributed according to $Q(\,\cdot\,, x_{11}), \dots$ $\dots, Q(\,\cdot\,, x_{ka(k)})$, respectively.

In particular, the random explanatory variables model $(X(1), Y(1)), \dots$ $\dots, (X(n), Y(n))$, with $f[X(t), \theta]$ as parametric conditional location function (median, expectation, ...) of $Y(t)$ given $X(t)$ $(t = 1, \dots, n)$, now becomes the fixed-design model $Y_{11}, \dots, Y_{ka(k)}$ with $f(x_{ij}, \theta)$ as location characteristic of Y_{ij} $(i = 1, \dots, k;\ j = 1, \dots, a(i))$.

Indeed, e.g., let $(K, \mathcal{C}) = (\mathbb{R}^r, \mathcal{B}^r)$, $Y' = (y_1, \dots, y_r)$. Let $f_h(x, \theta)$ $(h = 1, \dots, r)$ be the conditional median of y_h with respect to the mixed conditional distribution $Q(\,\cdot\,, x)$, which depends on a matrix parameter θ. The vector function

$$[f(x, \theta)]' = \big(f_1(x, \theta), \dots, f_r(x, \theta)\big)$$

is called the *marginal median row-vector* of Y' according to $Q(\,\cdot\,, x)$. Then, by Theorem 1(iii), with respect to the $(Q)P_{\Omega'}$-mixed conditional distribution given w, the variables $Y_{11}, \dots, Y_{ka(k)}$ are independent and Y_{ij} has the marginal median vector $f(x_{ij}, \theta)$ $(i = 1, \dots, k;\ j = 1, \dots, a(i))$.

In particular, at least approximately, for $x \in S_1 + \dots + S_k$, $f(x, \theta)$ may have the polygonal structure

$$[f(x, \theta)]' = \sum_{i=1}^{k} b_i(x)\theta_i I_{S_i}(x)\,,$$

where the row-vector functions $b_i(\,\cdot\,)$ are known, the unknown matrix parameter $\theta' = (\theta_1' \vdots \dots \vdots \theta_k')$ may be subject to linear constraints, e.g., in order to ensure the junction of the zones $f(x, \theta) = [b_i(x)\theta_i]'$ corresponding to several S_i, if one would wish to have a continuous, or smooth to some order, regression surface (or curve).

Further, we suppose every component y_h has a probability density under $Q(\,\cdot\,, x)$, which takes a positive value $\varphi_h(x)$ at the median f_h. Then setting

$$\delta_{gh} = [(Q)P\{y_g \le f_g,\, y_h \le f_h \mid x\} - \tfrac{1}{4}]/(\varphi_g(x)\varphi_h(x))$$

as the $Q(\,\cdot\,, x)$-conditional association coefficient between y_g, y_h, we can consider the conditional association matrix

$$\Gamma(x) = (\delta_{gh}(x))_{g,h=1,\dots,r}$$

(see also [8]). We suppose $\Gamma(x)$ equals a constant unknown matrix Γ_i for every $x \in S_i$. We can then state that with respect to the $(Q)P_{\Omega'}$-mixed conditional distribution given w, the r.v.'s Y_{ij} are independent and have marginal median vectors $\theta_i' b_i'(x_{ij})$ and association matrices Γ_i $(i = 1, \dots, k;\ j = 1, \dots, a(i))$. Here

$b'_i(x_{ij}) = [b_i(x_{ij}]'$. We shall frequently use the notations:

$$
\begin{aligned}
&U'_i = (Y_{i1} \vdots \ldots \vdots Y_{ia(i)}), \qquad U' = (U'_1 \vdots \ldots \vdots U'_k),\\
&B'_i = (b'_i(x_{i1}) \vdots \ldots \vdots b'_i(x_{ia(i)})),\\
&B = B(w) = \operatorname{diag}(B_1, \ldots, B_k) \qquad \text{(block diagonal matrix)},\\
&I_{a(i)} = \text{ unit } a(i) \times a(i)\text{-matrix},\\
&U = (u_{pq})_{s_k, r}, \qquad \overrightarrow{U} = (u_{11} \ldots u_{s_k 1} \ldots u_{1r} \ldots u_{s_k r})'.
\end{aligned} \tag{20}
$$

Let $(Q)M_{\Omega'}(U|w)$ be the matrix whose (p, q)th element is the $(Q)P_{\Omega'}(\cdot \mid w)$-conditional median of u_{pq} ($p = 1, \ldots, s_k$; $q = 1, \ldots, r$). Let $(Q)A_{\Omega'}(U|w)$ be the $(Q)P_{\Omega'}(\cdot \mid w)$-conditional association matrix of $\overrightarrow{U}$. We proceed to its computation. Write

$$Y'_{ij} = (y_{ij1}, \ldots, y_{ijr}) \,.$$

Then $U_i = (U_{i1} \vdots \ldots \vdots U_{ir})$, where $U_{ig} = (y_{i1g}, \ldots, y_{ia(i)g})$ ($i = 1, \ldots, k; g = 1, \ldots \ldots, r$). For $i, t = 1, \ldots, k$, we denote by $(Q)A_{\Omega'}(U_{ig}, U_{th} \mid w)$ the $a(i) \times a(t)$-matrix whose (j, v)-element ($j = 1, \ldots, a(i)$; $v = 1, \ldots, a(t)$) is the $(Q)P_{\Omega'}(\cdot \mid w)$-conditional association coefficient between the components y_{ijg}, y_{tvh} of U_{ig}, U_{th} respectively. We also use this notation for other couples of vectors, e.g. we define

$$A_{gh} = (Q)A_{\Omega'}(U_{(g)}, U_{(h)} \mid w), \qquad (g, h = 1, \ldots, r)\,,$$

where $U_{(g)}$ is the gth column vector of U. From the conditional association matrix definition, considering $(\overrightarrow{U})' = (U'_{(1)}, \ldots, U'_{(r)})$, we get $(Q)A_{\Omega'}(U \mid w) = \big(A_{gh}\big)_{g,h=1,\ldots,r}$. Moreover, noting that $U'_{(g)} = (U'_{1g}, \ldots, U'_{kg})$, we have

$$A_{gh} = \{(Q)A_{\Omega'}\big(U_{ig}, U_{th} \mid w\big)\}_{i,t=1,\ldots,k}.$$

From the definition of $(Q)A_{\Omega'}\big(U_{ig}, U_{th} \mid w\big)$ and by the $(Q)P_{\Omega'}(\cdot \mid w)$-conditional independence of $U_1, \ldots, U_k$ according to Theorem 1(iii), we see that

$$A_{gh} = \operatorname{diag}(F_{gh}(i),\, i = 1, \ldots, k)$$

where $F_{gh}(i) = (Q)A_{\Omega'}\big(U_{ig}, U_{ih} \mid w\big)$; also by Theorem 1(iii), $F_{gh}(i)$ is diagonal. On the other hand, by Theorem 1(iii)

$$(Q)A_{\Omega'}\big(Y_{ij} \mid w\big) = (Q)A\big(Y \mid x_{ij}\big) = \varGamma_i;$$

let us write $\varGamma_i = \big(\gamma_{gh}(i)\big)_{g,h=1,\ldots,r}$. Because $U_i = \big(U_{i1} \vdots \ldots \vdots U_{ir}\big)$, we have

$$(Q)A_{\Omega'}\big(U_i \mid w\big) = \big(F_{gh}(i)\big)_{g,h=1,\ldots,r}\,.$$

Considering $U'_i = (Y_{i1} \vdots \ldots \vdots Y_{ia(i)})$, we see that

$$F_{gh}(i) = \operatorname{diag}(\gamma_{gh}(i), \ldots, \gamma_{gh}(i)) = \gamma_{gh}(i) I_{a(i)}.$$

Finally, we get

$$
\begin{aligned}
&(Q)M_{\Omega'}(U \mid w) = B\theta\,,\\
&(Q)A_{\Omega'}(U \mid w) = (\operatorname{diag}[\gamma_{gh}(1) I_{a(1)}, \ldots, \gamma_{gh}(k) I_{a(k)}])_{g,h=1,\ldots,r}\,.
\end{aligned}
$$

This means that, so to speak, U follows a generalized conditional linear model. This enables us to use the LAD method (see [8]) for estimating θ.

II. Conditional linear models and estimation of regression parameters

5. Introduction. In Section 4, by performing all S_i-transformations ($i = 1,\dots,k$), the original data (1) are transformed into the set

$$\{(X_{i1},Y_{i1}),\dots,(X_{i\alpha(i)},Y_{i\alpha(i)});\ i=1,\dots,k\}$$

where $\alpha(i)$ is the random number of occurrences of X-value in S_i in the course of n observations. The new model is one with random numbers of observations $\alpha(i)$. In [1] a conception of treatment of such a set of observations was presented.

In this chapter, on the basis of this transformed model, we study the estimation of the regression parameter in the following case:

$$(K,\mathcal{C}) = (\mathbb{R}^r,\mathcal{B}^r), \qquad E(Y'Y) = E\|Y\|^2 < \infty\,,$$

$Q(\,\cdot\,,x)$ is previously chosen so that $(Q)E(YY' \mid x) = \int_{\mathbb{R}^r} yy'\,Q(dy,x)$ exists and is finite for every $x \in H$; $S_1,\dots,S_k$ being prescribed, for every $x \in S_1 + \dots + S_k$

$$\begin{cases} (Q)E(Y' \mid x) = \sum_{i=1}^k b_i(x)\theta_i I_{S_i}(x)\,,\\ (Q)D(Y \mid x) = \sum_{i=1}^k d_i(x)\Sigma_i I_{S_i}(x)\,, \end{cases} \tag{21}$$

where $d_i(\,\cdot\,) \ge 0$, $b_i(\,\cdot\,)$ are known functions, the $l_i \times r$-matrices θ_i and positive semidefinite $\Sigma_i = \big(\sigma_{gh}(i)\big)_{r,r}$ are unknown, there may be linear constraints on $\theta' = (\theta_1' \vdots \dots \vdots \theta_k')$ for the reason explained in Section 4.

The first mean structure (21) arises when we approximate the unknown regression function $(Q)E(Y \mid x)$ by linear parametric functions in every domain S_i.

Example 1. X, Y are real-valued, $k = 2$, $S_1 = [u_0,u_1]$, $S_2 = (u_1,u_2]$, and $(Q)E(Y \mid X)$ is approximated by two segments of parabola with common tangent at their common point:

$$\begin{gathered} (Q)E(Y \mid X) = b_i(X)\theta_i \quad \text{for } X \in S_i,\ i = 1,2\,;\\ b_1(X) = (X^2,\,X,\,1)\,, \qquad b_2(X) = (X^2,\,1)\,,\\ \theta_1' = (a_1,\,b_1,\,c_1)\,, \qquad \theta_2' = (a_2,\,c_2)\,. \end{gathered}$$

The constraints are

$$\begin{gathered} a_1u_1^2 + b_1u_1 + c_1 = a_2u_1^2 + c_2\,,\\ 2a_1u_1 + b_1 = 2a_2u_1\,, \end{gathered}$$

which can be written in the form $C\theta = 0$ or $\theta \in \Theta = \operatorname{Ker} C$, where

$$C = \begin{pmatrix} u_1^2 & u_1 & 1 & -u_1^2 & -1\\ 2u_1 & 1 & 0 & -2u_1 & 0 \end{pmatrix}.$$

The structure (21) may also arise in case the conditional distribution form of Y changes when X passes from a domain S_i to another. Indeed, for simplicity consider $S_1 + \ldots + S_k = H$; then Y is represented by

$$Y = \sum_{i=1}^{k} Y_i I_{(X\in S_i)}, \qquad Y_i = Y I_{(X\in S_i)}\,,$$

where Y_i is the representative of Y when X stays in S_i. For example (see [6]), Y is average foot length, X is age of a boy; (X, Y) is a bivariate measurement made one time per child within a group of children of different ages; $S_1, \ldots, S_k$ are different growth periods, Y_i is the average foot length of a boy in the ith growth period. Set

$$P(Y_i < y \mid X \in S_i) = F_i(y)\,, \qquad y \in \mathbb{R}^r,\ i = 1, \ldots, k\,.$$

Then Y follows the mixture of distributions

$$P(Y < y) = \sum_{i=1}^{k} q_i F_i(y), \quad \text{where} \quad q_i = P(X \in S_i)\,.$$

If every representative Y_i has the distribution structure

$$(Q)E(Y' \mid X) = b_i(X)\theta_i\,, \qquad (Q)D(Y_i \mid X) = d_i(X)\Sigma_i\,,$$

when X stays in S_i $(i = 1, \ldots, k)$, Y will follow (21).

The proof of existence of the model (21) can be outlined as follows: let ξ be a r.v. Then there exists an $\mathbb{R}^r$-valued random function $\zeta_i(\xi, x)$, $x \in H$, satisfying

$$\begin{aligned} &E\zeta_i(\xi, x) = 0,\\ &D\zeta_i(\xi, x) = E\zeta_i(\xi, x)\zeta_i'(\xi, x) = d_i(x)\Sigma_i \quad (i = 1, \ldots, k). \end{aligned}$$

Let ξ be independent of X. Then there exist determinations of $E(\zeta_i(\xi, X) \mid X)$ and $E\{\zeta_i(\xi, X)\zeta_i'(\xi, X)|X\}$ respectively, such that

$$\begin{aligned} E\{\zeta_i(\xi, X) \mid X = x\} &= E\zeta_i(\xi, x) = 0\,,\\ D\{\zeta_i(\xi, X) \mid X = x\} &= E\{\zeta_i(\xi, X)\zeta_i'(\xi, X) \mid X = x\}\\ &= E\zeta_i(\xi, x)\zeta_i'(\xi, x) = d_i(x)\Sigma_i. \end{aligned}$$

Let $Y_i' = \zeta_i'(\xi, X) + b_i(X)\theta_i$ $(i = 1, \ldots, k)$. We get

$$E(Y_i' \mid x) = b_i(x)\theta_i, \qquad D(Y_i \mid x) = d_i(x)\Sigma_i\,.$$

Then $Y = \sum_{i=1}^{k} Y_i I_{(X\in S_i)}$ satisfies (21).

Hereafter, all notations already introduced in Section 2 and (20), §4, are to be kept in mind; moreover, the following standard matrix notations will be used throughout:

- $M_{s\times r}$ = the linear space of all real $s \times r$-matrices,
- $f'(\,\cdot\,)$ = the transpose of a matrix-valued function $f(\,\cdot\,)$,
- B^- = an arbitrary g-inverse of a matrix B,
- $\|C\|^2 = \sum_{i,j} c_{ij}^2, \quad |C| = \sum_{i,j} |c_{ij}|$ for any real matrix $C = (c_{ij})$,

- $\|Y\|_F^2 = Y'FY$ for any $r \times 1$-vector Y and positive semidefinite (PSD) $r \times r$-matrix F,
- $\mathcal{M}(B) =$ the vector space generated by the column vectors of a matrix B,
- $\mathcal{M}_{\mathcal{L}}^r =$ the set of all $s \times r$-matrices with r columns belonging to $\mathcal{L} \subset \mathbb{R}^s$,
- $\overrightarrow{\eta} = (\eta_1' \dots \eta_r')'$ for $\eta = (\eta_1 \vdots \dots \vdots \eta_r) \in M_{s\times r}$,
- $D\eta = D\overrightarrow{\eta}$ for any random matrix η,
- $B \otimes C =$ the Kronecker product of matrices B, C,
- $B \geq C$ means $B - C$ is a PSD matrix,
- $p^{\beta_1,\dots,\beta_l} =$ the probability distribution of the family of r.v.'s $\beta_1, \dots, \beta_l$.

Finally, set

$$\begin{aligned} A^* &= \{w \,:\, \operatorname{Rank} B(w) = l_1 + \dots + l_k\}\,, \\ \Omega^* &= \text{ the general event in Theorem 1.} \end{aligned} \tag{22}$$

6. Conditional generalized least squares estimators (CGLSE)

THEOREM 3. *The domains $S_1, \dots, S_k$ being prescribed*,

(i) *if the underlying distribution of (X, Y) satisfies* (21), *then U follows the conditional linear model*

$$\begin{cases} (Q)E_{\Omega^*\cap(W\in A)}(U \mid w) = B\theta\,, \\ (Q)D_{\Omega^*\cap(W\in A)}(U \mid w) = \{\operatorname{diag}(\sigma_{gh}(1)V_1, \dots, \sigma_{gh}(k)V_k)\}_{r,r} \end{cases} \tag{23}$$

for every $w \in A \subset S_1^{a(1)} \times \dots \times S_k^{a(k)}$, where

$$V_i = \operatorname{diag}[d_i(x_{i1}), \dots, d_i(x_{ia(i)})] \quad (i = 1, \dots, k)\,,$$

(ii) *conversely*, *if the basic probability space $(\Omega, \mathcal{F}, P)$ is the sample probability space of a sequence $Z(1), \dots, Z(n')$ with $n' \geq n$, or of an infinite one, and if* (23) *is fulfilled for every A as in Theorem* 1(iii), *then* (21) *follows*, *i.e. we have*

$$(Q)E(Y' \mid x) = \sum_{i=1}^k b_i(x)\theta_i I_{S_i}(x)\,, \qquad (Q)D(Y \mid x) = \sum_{i=1}^k d_i(x)\Sigma_i I_{S_i}(x)\,.$$

P r o o f. (i) For any random matrix $F' = (F_1' \vdots \dots \vdots F_k')$ where $F_i = (F_{i1} \vdots \dots \vdots F_{ir})$ are $a(i) \times r$-matrices ($i = 1, \dots, k$), we can check

$$DF = \big(C_{gh}\big)_{g,h=1,\dots,r} \tag{24}$$

with $C_{gh} = \{\operatorname{Cov}\big(F_{ig}, F_{jh}\big)\}_{i,j=1,\dots,k}$. Then, recall $U_i' = (Y_{i1} \vdots \dots \vdots Y_{ia(i)})$ and write $Y_{ij}' = (y_{ij1}, \dots, y_{ijr})$. We have

$$(Q)D_{\Omega^*\cap(W\in A)}(U_i \mid w) = \big(G_{gh}\big)_{r,r}$$

where

$$G_{gh} = \{(Q)\operatorname{Cov}_{\Omega^*\cap(W\in A)}(y_{ijg}, y_{iph} \mid w)\}_{j,p=1,\dots,a(i)} \quad (i = 1, \dots, k)\,.$$

By (21) and Theorem 1(iii) we have

$$(25)\qquad (Q)D_{\Omega^*\cap(W\in A)}(Y_{ij}\mid w)=d_i(x_{ij})\Sigma_i,$$

$$G_{gh}=\mathrm{diag}\{d_i(x_{ij})\sigma_{gh}(i);\ j=1,\dots,a(i)\}=\sigma_{gh}(i)V_i\,.$$

On the other hand, write $U_i=(U_{i1}\vdots\dots\vdots U_{ir})$. Then we have

$$(Q)D_{\Omega^*\cap(W\in A)}(U_i\mid w)=\{(Q)\mathrm{Cov}_{\Omega^*\cap(W\in A)}(U_{ig},U_{ih}\mid w)\}_{r,r}$$

whence

$$(Q)\mathrm{Cov}_{\Omega^*\cap(W\in A)}(U_{ig},U_{ih}\mid w)=G_{gh}=\sigma_{gh}(i)V_i\,.$$

By applying (24) to $U'=(U'_1\vdots\dots\vdots U'_k)$, $U_i=(U_{i1}\vdots\dots\vdots U_{ir})$, we get

$$(Q)D_{\Omega^*\cap(W\in A)}(U\mid w)=\big(D_{gh}\big)_{r,r}$$

where

$$D_{gh}=\{(Q)\mathrm{Cov}_{\Omega^*\cap(W\in A)}(U_{ig},U_{jh}\mid w)\}_{i,j=1,\dots,k}$$

and by Theorem 1(iii), from the above, it follows that $D_{gh}=\mathrm{diag}\{\sigma_{gh}(i)V_i,\ i=1,\dots,k\}$, i.e. the second relation (23); the first one is immediate.

(ii) Define $(Q)D(Y\mid X)=v(X)$. Then by Theorem 1(iii) we have

$$(Q)D_{\Omega^*\cap(W\in A)}(Y_{ij}\mid W)=v(X_{ij})\,.$$

Let U satisfy (23). The matrix identification process in (i) entails (25), i.e.

$$(26)\qquad v(X_{ij})=d_i(X_{ij})\Sigma_i\,.$$

Now, the range of W when ω varies over Ω^* contains the one when ω varies over any set A_0 as in (10), §3, in particular, when ω varies over the set

$$A_1=(T_{11}=1,\dots,T_{1t(1)}=t(1);\dots;\dots,T_{mt(m)}=t(1)+\dots+t(m))\,.$$

For $\omega\in A_1$,

$$(X_{11},\dots,X_{1t(1)};\dots;\dots,X_{mt(m)})$$
$$\equiv\{X(1),\dots,X(t(1));\dots;\dots,X(t(1)+\dots+t(m))\}\,.$$

When ω varies over A_1, the range of this family is $S_1^{t(1)}\times\dots\times S_m^{t(m)}$ if $(\Omega,\mathcal{F},P)$ is the sample probability space of the sequence

$$(X(1),Y(1)),\dots,(X(n'),Y(n'))\qquad(n'\geq n)$$

or of an infinite one. Since (see Section 3) $a(i)\leq t(i)$, $i=1,\dots,k$, $k\leq m$,

$$W=(X_{11},\dots,X_{1a(1)};\dots;\dots,X_{ka(k)})$$

is a subfamily of the preceding family; hence when ω varies over Ω^* the range of X_{ij} is S_i $(i=1,\dots,k;\ j=1,\dots,a(i))$. By assumption, (23) is fulfilled for every A as in Theorem 1(iii); we take $A=S_1^{a(1)}\times\dots\times S_k^{a(k)}$. Then in (26) the range of X_{ij} is S_i, i.e. $v(x)=d_i(x)\Sigma_i$ for every $x\in S_i$, $i=1,\dots,k$. Thus the second equality (21) follows. We get the first one by the same reasoning. Theorem 3 is proved.

By this theorem, the estimation of the parameters θ, Σ_i $(i = 1, \ldots, k)$ in (21) is brought back to the one in the conditional linear model (23). We shall first seek the best estimator of θ in the following fixed-design model.

Definition 1. The random $s \times r$-matrix η is said to follow an *r-multivariate Behrens-Fisher model* if $E\|\eta\|^2 < \infty$ and

$$
\begin{cases} E\eta = B\theta\,, \\ D\eta = \{\operatorname{diag}(\sigma_{gh}(i)V_i,\ i = 1, \ldots, k)\}_{g,h=1,\ldots,r}, \end{cases} \tag{27}
$$

where

$$
\theta' = (\theta_1' \vdots \ldots \vdots \theta_k'), \qquad B = \operatorname{diag}(B_1, \ldots, B_k)\,,
$$
$$
s = a(1) + \ldots + a(k), \qquad l = l_1 + \ldots + l_k\,.
$$

Here the $a(i) \times l_i$-matrices B_i and positive semidefinite $a(i) \times a(i)$-matrices V_i are known, whereas the $l_i \times r$-matrices θ_i and PSD $\Sigma_i = \big(\sigma_{gh}(i)\big)_{r,r}$ are unknown $(i = 1, \ldots, k)$; the range of θ being any given set $\Theta \subset M_{l\times r}$.

When $r = 1$ and $\Theta = M_{l\times 1}$, we get the model already examined in [3].

The generalized least squares estimator for θ will be given in Lemma 4. For further considerations we introduce the following notations and assumptions. Let $\mathcal{D}$ be the range of $E\eta$ in (27) and

$$
\mathcal{D}_0 = \{\mu - \widehat{\mu} : \mu, \widehat{\mu} \in \mathcal{D}\}.
$$

Assumption 1. Either

(a) there exist linear subspaces $\mathcal{L}_i \subset \mathcal{M}(B_i)$ $(i = 1, \ldots, k)$ such that $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}$ is the linear hull of $\mathcal{D}_0$ in $M_{s\times r}$, or

(b) $\Sigma_1 = \ldots = \Sigma_k$ and there exists a linear subspace $\mathcal{L} \subset \mathcal{M}(B)$ such that $\mathcal{M}^r_{\mathcal{L}}$ is the linear hull of $\mathcal{D}_0$ in $M_{s\times r}$.

Let $V_1, \ldots, V_k$ be positive definite and G the orthoprojector on $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}$ of $M_{s\times r}$ endowed with the scalar product

$$
(\zeta_1, \zeta_2) = \overrightarrow{\zeta_1}[I_r \otimes \operatorname{diag}^{-1}(V_1, \ldots, V_k)]\overrightarrow{\zeta_2}, \qquad \zeta_1, \zeta_2 \in M_{s\times r}\,. \tag{28}
$$

Let $\Phi : M_{s\times r} \to M_{p\times q}$, $\Psi : M_{l\times r} \to M_{p\times q}$ be linear operators into some space $M_{p\times q}$. We say that $\Psi\theta$ is *estimable in the model* (27) if there exists an ILUE or LUE for $\Psi\theta$ (inhomogeneous linear unbiased estimator, resp. linear unbiased estimator).

Then, using G under Assumption 1(a) and again denoting by G the orthoprojector of $M_{s\times r}$ on $\mathcal{M}^r_{\mathcal{L}}$ under Assumption 1(b), we have the following lemma.

Lemma 4. (i) *In* (27), *a BILUE* (*best inhomogeneous linear unbiased estimator*) *for $\Phi E\eta$ exists and is given by*

$$
\Phi G\eta + \Phi(I_s - G)\mu_0
$$

where μ_0 is any fixed element of $\mathcal{D}$.

(ii) *If B is injective as a linear mapping $M_{l\times r} \to M_{s\times r}$, then $\Psi\theta$ is estimable and*

$$\Psi B^- G\eta + \Psi B^-(I_s - G)B\theta_0\,,$$

where θ_0 is arbitrarily fixed $\in \Theta$, is a BILUE for $\Psi\theta$.

(iii) *θ is estimable in* (27) *if*

$$\operatorname{Rank} B = l_1 + \ldots + l_k\,, \quad \textit{or} \quad \operatorname{Rank} B_i = l_i \ (i = 1, \ldots, k)\,.$$

When $k = 1$, Lemma 4(i) gives a result of Theorem 2.1.3 in [4] (Band I). By abuse of notations, we use e.g. $\Phi G\eta$ to designate the image by Φ of the element obtained either by the product matrix $G\eta$ or by G acting on η. Similarly $B^- G\eta$ is the image of η by the product mapping of G and the mapping $B^- : M_{s\times r} \to M_{l\times r}$.

Proof. (i) We proceed in several steps. First let us prove (i) under Assumption 1(a).

1) Let G_i be any $a(i)\times a(i)$ projection matrix of $\mathbb{R}^{a(i)}$ endowed with the scalar product

$$(y_1, y_2) = y_1' V_i^{-1} y_2\,, \qquad y_1, y_2 \in \mathbb{R}^{a(i)}\,.$$

Then $G_iG_i = G_i$, $G_i'V_i^{-1} = V_i^{-1}G_i$ $(i = 1, \ldots, k)$. Let $G = \operatorname{diag}(G_1, \ldots, G_k)$. Then

$$GG = G, \quad G' \operatorname{diag}^{-1}(V_1, \ldots, V_k) = \operatorname{diag}^{-1}(V_1, \ldots, V_k)G\,. \tag{29}$$

We have $G_iV_iG_i' = G_iV_i$, hence

$$G_i\sigma_{gh}(i)V_iG_i' = G_i\sigma_{gh}(i)V_i \quad (i = 1, \ldots, k)\,,$$

which entails

$$(I_r \otimes G)D\eta(I_{sr} - (I_r \otimes G))' = 0\,,$$

for $D\eta$ has an expression as in (27).

2) Let $T\eta + c$ be any inhomogeneous linear function $M_{s\times r} \to M_{p\times q}$, where T is any linear operator; then

$$\overrightarrow{T\eta} = L\overrightarrow{\eta}, \quad L \in M_{pq\times sr}\,.$$

Write $\eta = G\eta + (I_s - G)\eta$. We have $T\eta + c = TG\eta + T(I_s - G)\eta + c$, or

$$L\overrightarrow{\eta} + \overrightarrow{c} = L(I_s \otimes G)\overrightarrow{\eta} + L(I_{sr} - (I_r \otimes G))\overrightarrow{\eta} + \overrightarrow{c}\,.$$

Further,

$$\begin{aligned}&\operatorname{Cov}\{L(I_r \otimes G)\overrightarrow{\eta}, L(I_{sr} - (I_r \otimes G))\overrightarrow{\eta}\}\\ &\qquad\qquad = L(I_r \otimes G)D\overrightarrow{\eta}(I_{sr} - (I_r \otimes G))'L' = 0\,.\end{aligned}$$

Now, if γ_1, γ_2 are two random vectors in the same space, with $E||\gamma_1||^2 < \infty$, $E||\gamma_2||^2 < \infty$, we have the elementary formula

$$D(\gamma_1 + \gamma_2) = D(\gamma_1) + D(\gamma_2) + \operatorname{Cov}(\gamma_1, \gamma_2) + \operatorname{Cov}(\gamma_2, \gamma_1)\,,$$

and, when $\mathrm{Cov}(\gamma_1,\gamma_2)=0$, we have

$$D(\gamma_1+\gamma_2)=D(\gamma_1)+D(\gamma_2)\geq D(\gamma_1)\,.$$

Applying this, we get

$$D(L\overrightarrow{\eta}+\overrightarrow{c})=D(L\overrightarrow{\eta})\geq D\big(L(I_r\otimes G)\overrightarrow{\eta}\big)\,.$$

We can also obtain this inequality by applying the well-known Lehmann–Scheffé lemma ([5]; see also [4], Band III, Satz A.3.2, p. 329). Equivalently, we have

$$D(T\eta+c)\geq D(TG\eta). \tag{30}$$

3) Moreover, let $T\eta+c$ be any ILUE for $\Phi E\eta$. Then $E(T\eta+c)=TE\eta+c=\Phi E\eta$ for every value of $E\eta$ in its range $\mathcal{D}$, which entails $T\lambda=\Phi\lambda,\ \forall\lambda\in\mathcal{D}_0$, and hence by Assumption 1(a),

$$T\lambda=\Phi\lambda,\qquad \forall\lambda\in\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}\,. \tag{31}$$

4) Now, choose for an orthoprojector G_i the $a(i)\times a(i)$ projection matrix of $\mathbb{R}^{a(i)}$ on $\mathcal{L}_i$ $(i=1,\ldots,k)$. G, satisfying (29), is an $s\times s$ projection matrix of $\mathbb{R}^s$ endowed with the scalar product

$$(y,z)=y'\,\mathrm{diag}^{-1}(V_1,\ldots,V_k)z,\qquad y,z\in\mathbb{R}^s\,.$$

Hence G is also an orthoprojector of $M_{s\times r}$ endowed with the scalar product(28).

Let $z'=(z_1'\vdots\ldots\vdots z_k')$, $z_i\in\mathbb{R}^{a(i)}$. Considering the range $\mathcal{L}_i$ of G_iz_i when z_i varies over $\mathbb{R}^{a(i)}$ $(i=1,\ldots,k)$, the range of Gz when z varies over $\mathbb{R}^s=\mathbb{R}^{a(1)}\times\ldots\times\mathbb{R}^{a(k)}$ is $\mathcal{L}_1\times\ldots\times\mathcal{L}_k$, hence the range of $G\zeta$ when ζ varies over $M_{s\times r}$ is $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}$. Thus (31) is equivalent to $TG=\Phi G$. Therefore, by (30), for any ILUE $T\eta+c$ for $\Phi E\eta$ we get

$$D(T\eta+c)\geq D(\Phi G\eta)\,.$$

5) Because G projects $M_{s\times r}$ on $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}\supset\mathcal{D}_0$, $G\lambda=\lambda$, $\forall\lambda\in\mathcal{D}_0$. Hence for any fixed $\mu_0\in\mathcal{D}$, an ILUE for $\Phi E\eta$ is the function $\Phi G\eta+\Phi(I_s-G)\mu_0$ because

$$\begin{aligned}E[\Phi G\eta+\Phi(I_s-G)\mu_0]&=\Phi G(E\eta-\mu_0)+\Phi\mu_0\\&=\Phi(E\eta-\mu_0)+\Phi\mu_0=\Phi E\eta\,.\end{aligned}$$

Moreover, it is a BLUE because

$$D(T\eta+c)\geq D(\Phi G\eta+\Phi(I_s-G)\mu_0)$$

for every ILUE $T\eta+c$.

We have thus proved (i) only by means of Assumption 1(a) that the linear hull of the translate of the range of $E\eta$ has the form $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}$ where $\mathcal{L}_i$ is any subspace of $\mathbb{R}^{a(i)}$; neither the representation $E\eta=B\theta$ nor the particular assumption $B=\mathrm{diag}(B_1,\ldots,B_k)$ is needed.

Now consider $k=1$. Then $\mathcal{L}_1=\mathcal{L}$ is a linear subspace of $\mathbb{R}^s$ such that $\mathcal{M}^r_{\mathcal{L}}$ is the linear hull of $\mathcal{D}_0$; set $\Sigma_1=\Sigma=\big(\sigma_{gh}\big)_{r,r}$, $V_1=V$; we get the assertion (i) in the case

$$D\eta = \big(\sigma_{gh}V\big)_{g,h=1,\dots,r} = \Sigma \otimes V\,,$$

$G =$ the orthoprojector on $\mathcal{M}^r_{\mathcal{L}}$ of $M_{s\times r}$ endowed with the scalar product

$$(\zeta_1,\zeta_2) = \overrightarrow{\zeta_1}(I_r \otimes V^{-1})\overrightarrow{\zeta_2}, \qquad \zeta_1,\zeta_2 \in M_{s\times r}\,.$$

Therefore (i) follows under Assumption 1(b), because in that case

$$\begin{aligned} &\Sigma_1 = \,\dots = \Sigma_k = \Sigma \qquad \text{say,}\\ &D\eta = \{\mathrm{diag}(\sigma_{gh}V_i,\ i=1,\dots,k)\}_{r,r} = \Sigma \otimes V \end{aligned}$$

by renaming $\mathrm{diag}(V_1,\dots,V_k) = V$.

(ii) The equation $B\theta = E\eta$ is consistent in θ for $E\eta \in \{B\theta : \theta \in \Theta\}$. The solution $\theta = B^-E\eta$ is unique for B is injective. Thus any g-inverse B^- is a linear mapping $M_{s\times r} \to M_{l\times r}$ which is a linear extension on $M_{s\times r}$ of the inverse mapping of B, and $\Psi\theta = \Psi B^-E\eta$. Then, by (i), for any fixed $\theta_0 \in \Theta$

$$\Psi B^- G\eta + \Psi B^-(I_s - G)B\theta_0$$

is a BILUE for $\Psi\theta$, and (ii) is proved.

(iii) The following assertions are successively equivalent:

1) B is injective as a linear mapping $M_{l\times r} \to M_{s\times r}$,

2) B is injective as a linear mapping $\mathbb{R}^l \to \mathbb{R}^s$, or, equivalently, Ker $B = 0_{l\times 1}$, or $\dim \mathcal{M}(B) = l$ or $\mathrm{Rank}\, B = l$,

3) $\mathrm{Ker}\, B_i = 0_{l_i\times 1}$ $(i=1,\dots,k)$,

4) $\dim \mathcal{M}(B_i)$ or $\mathrm{Rank}\, B_i = l_i$ $(i=1,\dots,k)$.

This completes the proof of Lemma 4. From Lemma 5 it will follow that the BILUE in Lemma 4 are just the generalized least squares estimates.

We consider a random $s\times r$-matrix η and linear operators $\Phi : M_{s\times r} \to M_{p\times q}$, $\Psi : M_{l\times r} \to M_{p\times q}$; we have the following lemma.

Lemma 5. *Consider a general model η in which the range of $E\eta$ is some set $\mathcal{D} \subset M_{s\times r}$, $\mathcal{D}_0 = \{\mu - \widehat{\mu} : \mu, \widehat{\mu} \in \mathcal{D}\}$, and $\mathcal{E}$ is any linear subspace of $M_{s\times r}$ containing $\mathcal{D}_0$. Let $M_{s\times r}$ be endowed with a scalar product $q(y,z)$, and let $\|\cdot\|_q$ be the induced norm. Let μ_0 be any fixed element of $\mathcal{D}$. Then*

(i) *for given $Y \in M_{s\times r}$, there exists a unique element $\overline{p} = \overline{p}(Y)$ of the affine manifold $\mathcal{E} + \mu_0$ such that*

$$\|Y - \overline{p}\|_q \le \|Y - \mu\|_q$$

for every $\mu \in \mathcal{D}$, and more generally, for every $\mu \in \mathcal{E} + \mu_0$,

(ii) *if G_e is the orthoprojector of $M_{s\times r}$ on $\mathcal{E}$, we have $\overline{p}(Y) = G_eY + (I - G_e)\mu_0$ where I is the identity mapping $M_{s\times r} \to M_{s\times r}$,*

(iii) *$\Phi\overline{p}(\eta) = \Phi G_e\eta + \Phi(I - G_e)\mu_0$ is an ILUE for $\Phi E\eta$,*

(iv) *when* $E\eta$ *has a parametric structure* $E\eta = B\theta$, $\theta \in \Theta \subset M_{l\times r}$, *and the* $s \times l$*-matrix* B *defines an injective mapping* $B : M_{l\times r} \to M_{s\times r}$, *then*

$$\Psi B^{-}\overline{p}(\eta) = \Psi B^{-}G_e\eta + \Psi B^{-}(I - G_e)\mu_0$$

is an ILUE for $\Psi\theta$. *By abuse of notation, e.g.,* $\Psi B^{-}G_e$ *denotes the product of the mappings* G_e, B^{-}, Ψ.

Proof. (i) For $Y \in M_{s\times r}$,

$$Y = Y_0 + Z_0, \qquad Y_0 \in \mathcal{E}, \qquad Z_0 \perp_q \mathcal{E}$$

(i.e. Z_0 is orthogonal to $\mathcal{E}$ with respect to the scalar product q). For μ_0 arbitrarily fixed $\in \mathcal{D}$,

$$\mu_0 = e_0 + f_0, \qquad e_0 \in \mathcal{E}, \qquad f_0 \perp_q \mathcal{E}\,.$$

For μ arbitrarily chosen in $\mathcal{D}$, or more generally, in $\mathcal{E} + \mu_0$, we have

$$\mu = e_1 + f_0, \qquad e_1 = e_0 + (\mu - \mu_0) \in \mathcal{E}, \qquad f_0 \perp_q \mathcal{E}\,.$$

Now, $Y_0 + f_0 = (Y_0 - e_0) + \mu_0 \in \mathcal{E} + \mu_0$ and

$$Y - \mu = (Y_0 - e_1) + \big(Y - (Y_0 + f_0)\big)\,,$$
$$Y_0 - e_1 \in \mathcal{E}, \qquad Y - (Y_0 + f_0) = Z_0 - f_0 \perp_q \mathcal{E}\,.$$

Hence, by setting $\overline{p}(Y) = \overline{p} = Y_0 + f_0$, we have $\|Y - \mu\|_q \ge \|Y - \overline{p}\|_q$, with equality iff $e_1 = Y_0$, i.e. iff $\mu = Y_0 + f_0 = \overline{p}$. Thus (i) is proved.

(ii) $Y_0 = G_eY$, $e_0 = G_e\mu_0$, $f_0 = \mu_0 - e_0 = (I - G_e)\mu_0$, $\overline{p} = \overline{p}(Y) = G_eY + (I - G_e)\mu_0$.

(iii)

$$E\Phi\overline{p}(\eta) = \Phi G_e(E\eta - \mu_0) + \Phi\mu_0 = \Phi(E\eta - \mu_0) + \Phi\mu_0 = \Phi E\eta\,.$$

(iv) $\theta = B^{-}E\eta$, and $\Psi\theta = \Psi B^{-}E\eta$ if B is injective. This completes the proof of Lemma 5.

Lemmas 4 and 5, applied to the conditional linear model (23), will give conditional generalized least squares estimators for θ. Besides, we must ensure the conditional estimability of θ.

7. Conditional estimability. In connection with the conditional linear model (23), following the corresponding definitions in fixed-design models (see [4]), we introduce the following

DEFINITION 2. A function $L(w)\overrightarrow{U} + c(w)$, where $L(\,\cdot\,)$, $c(\,\cdot\,)$ are matrix-valued and $\mathcal{A}^{s_k}$-measurable in w, is called a *conditionally inhomogeneous linear unbiased estimator* (CILUE) for $\overrightarrow{\theta}$ if

$$(Q)E_{\Omega^*\cap(W\in A)}\{L(w)\overrightarrow{U} + c(w) \mid w\} = \overrightarrow{\theta}$$

for every $w \in A$ and every underlying distribution of (X, Y) fulfilling (21). When $c(w) \equiv 0$ this estimator is symbolized by CLUE. θ is called *conditionally estimable* in the model (23) if $\overrightarrow{\theta}$ has a CILUE or CLUE.

By Lemma 4(iii), whether there are constraints on θ or not, θ is conditionally estimable when $\operatorname{Rank} B(w) = l_1 + \ldots + l_k$. By (5), §2, with A^* as in (22), after letting $\Omega^* = \Omega'$ in (23), consider

$$\Omega_n = \sum_{l_i \le a(i), i=1,\ldots,k;\, a(1)+\ldots+a(k) \le n} \Omega'(a,n) \cap (W \in A^*)\,. \tag{32}$$

Ω_n is the conditional estimability domain for θ on the ground of n observations on (X, Y).

In this section we shall prove that when the number of observations is sufficiently large, it is practically sure that θ is conditionally estimable. We begin with the following

LEMMA 6. *Let $\beta_1, \ldots, \beta_l$ be i.i.d. random $l \times 1$-vectors, and $C_l' = (\beta_1 \vdots \ldots \vdots \beta_l)$. Then $\det C_l = 0$ a.s. iff the probability distribution of β_1 is concentrated in some proper subspace of $\mathbb{R}^l$.*

P r o o f. Write $f(\beta_1, \ldots, \beta_l) = |\det C_l| \ge 0$. Then $\det C_l = 0$ a.s. iff $Ef(\beta_1, \ldots, \beta_l) = 0$, or, iff $E\{f(\beta_1, \ldots, \beta_l) \mid \beta_2, \ldots, \beta_l\} = 0$ a.s. This is equivalent to

$$E\{f(\beta_1, \ldots, \beta_l) \mid z_2, \ldots, z_l\} = 0$$

or to

$$Ef(\beta_1, z_2, \ldots, z_l) = 0$$

for $P^{\beta_2,\ldots,\beta_l}$-almost all values $(z_2, \ldots, z_l)$. This, in turn, is equivalent to

$$f(\beta_1, z_2, \ldots, z_l) = 0 \quad \text{or} \quad \det(\beta_1 \vdots z_2 \vdots \ldots \vdots z_l) = 0 \quad P^{\beta_1}\text{-a.s.}$$

for $P^{\beta_2,\ldots,\beta_l}$-almost all values $(z_2, \ldots, z_l)$.

The sufficiency part of the lemma is evident. Let us prove the necessity part. We reason by induction. For $l = 1$, the assertion is true. Suppose it is true for $l - 1$; let $\det C_l = 0$ a.s. There are two possibilities:

1) $P\{\operatorname{Rank}(\beta_2 \vdots \ldots \vdots \beta_l) = l - 1\} > 0$. Then there exists $(z_2, \ldots, z_l)$ such that both

$$\operatorname{Rank}(z_2 \vdots \ldots \vdots z_l) = l - 1 \quad \text{and} \quad \det(\beta_1 \vdots z_2 \vdots \ldots \vdots z_l) = 0 \quad P^{\beta_1}\text{-a.s.},$$

hence β_1 lies a.s. in the subspace generated by $z_2, \ldots, z_l$.

2) $P\{\operatorname{Rank}(\beta_2 \vdots \ldots \vdots \beta_l) < l - 1\} = 1$. Then all $(l-1) \times (l-1)$-submatrices of $(\beta_2 \vdots \ldots \vdots \beta_l)$ are a.s. degenerate. By the induction hypothesis the projection of β_2 on any coordinate hyperplane is a.s. focussed in some proper subspace of this hyperplane, hence β_2 lies a.s. in some $(l-2)$-dimensional subspace of $\mathbb{R}^l$.

Lemma 6 is proved.

Corollary 1. *We have* $P(\Omega^* \cap (W \in A^*)) > 0$ *for every set* Ω^* *as in Theorem* 1 *provided* $a(i) \geq l_i$ $(i = 1, \ldots, k)$ *and the* $P_{(X\in S_i)}$*-distribution of* $b_i'(X)$ *is not concentrated in any proper subspace of* $\mathbb{R}^{l_i}$ $(i = 1, \ldots, k)$.

The last assumption is equivalent to the following: There exists no constant non-null $l_i \times 1$-vector γ_i such that $b_i(x)\gamma_i = 0$ for P^X-almost all values $x \in S_i$ $(i = 1, \ldots, k)$. Such a requirement is practically always fulfilled.

Proof. By Lemma 4(iii), (20), §4, and because $a(i) \geq l_i$, we have

$$\begin{aligned}(W \in A^*) &= \{\operatorname{Rank}\bigl(b_i'(X_{i1})\vdots \ldots \vdots b_i'(X_{ia(i)})\bigr) = l_i; i = 1, \ldots, k\} \\ &\supset \{\det\bigl(b_i'(X_{i1})\vdots \ldots \vdots b_i'(X_{il_i})\bigr) \neq 0; i = 1, \ldots, k\}.\end{aligned}$$

By Theorem 1(i) and (ii), the events on the last right-hand side are P_{Ω^*}-independent for $i = 1, \ldots, k$, the r.v.'s $b_i'(X_{i1}), \ldots, b_i'(X_{il_i})$ are P_{Ω^*}-i.i.d., and their common P_{Ω^*}-distribution coincides with the $P_{(X\in S_i)}$-distribution of $b_i(X)$. Hence by Lemma 6, it follows that $P_{\Omega^*}(W \in A^*) > 0$, i.e. $P(\Omega^* \cap (W \in A^*)) > 0$.

Corollary 2. *Let* $\beta_1, \beta_2, \ldots$ *be a sequence of i.i.d. random* $l \times 1$*-vectors and* $C_n' = (\beta_1 \vdots \ldots \vdots \beta_n)$. *Then*

(i) $P\{\operatorname{Rank} C_n = l\} \uparrow 1$ *as* $n \uparrow \infty$, *if the probability distribution of* β_1 *is not concentrated in any proper subspace of* $\mathbb{R}^l$,

(ii) $P\{\operatorname{Rank} C_n = l\} = 0$ *for every* $n \geq 1$ *otherwise.*

Proof. (i) By Lemma 6, under the assumption of (i), $P(\det C_l = 0) < 1$. For every $k = 1, 2, \ldots$

$$\{\operatorname{Rank} C_{kl} \leq l-1\} \subset \{\operatorname{Rank} C_l \leq l-1,\ \operatorname{Rank}(\beta_{l+1}\vdots \ldots \vdots \beta_{2l}) \leq l-1, \ldots\}.$$

Hence $P\{\operatorname{Rank} C_{kl} \leq l-1\} \leq [P(\operatorname{Rank} C_l \leq l-1)]^k$, and $P\{\operatorname{Rank} C_{kl} \leq l-1\} \to 0$ as $k \to \infty$. But $P(\operatorname{Rank} C_n \leq l-1)$ decreases as n increases, and (i) is proved.

(ii) Note that if $n \geq l$ then

$$\{\operatorname{Rank} C_n \leq l-1\} = \bigcap_{1\leq i_1 < \ldots < i_l \leq n} \{\operatorname{Rank}(\beta_{i_1}\vdots \ldots \vdots \beta_{i_l}) \leq l-1\}.$$

Hence $P(\operatorname{Rank} C_l \leq l-1) = 1$ entails $P(\operatorname{Rank} C_n \leq l-1) = 1$. Thus (ii) follows.

Theorem 4. *If* n *increases along the infinite sequence of i.i.d. r.v.'s* $X(1), \ldots, X(n), \ldots$, *then the sequence* $\{\Omega_n\}$ *given by* (32) *is non-decreasing. Moreover, if the* $P_{(X\in S_i)}$*-distribution of* $b_i'(X)$ *is not concentrated in any proper subspace of* $\mathbb{R}^{l_i}$ $(i = 1, \ldots, k)$, *then*

$$P(\lim_{n\uparrow\infty} \Omega_n) = 1.$$

Proof. In view of Lemma 4(iii), we see that in (32)

$$w \in A^* \quad \text{means} \quad \operatorname{Rank}(b_i'(x_{i1})\vdots \ldots \vdots b_i'(x_{ia(i)})) = l_i$$

for $i = 1, \ldots, k$. Hence, for known $a = (a(1), \ldots, a(k))$, $A^* = A^*(a)$ is a known set.

Consider any $\omega \in \Omega_n$ and the corresponding trajectory $X_\omega(1), \ldots$ Define

$$a(i) = \#S_i \cap \{X_\omega(1), \ldots, X_\omega(n)\} = \#\{X_{i1}, \ldots, X_{ia(i)}\}\,.$$

Then we must have $a(i) \geq l_i$ $(i = 1, \ldots, k)$, and $\omega \in \Omega'(a, n)$, $(X_{11}, \ldots \ldots, X_{ka(k)}) = W(\omega) \in A^*(a)$. Hence

$$\operatorname{Rank}\big(b_i'(X_{i1}) \vdots \ldots \vdots b_i'(X_{ia(i)})\big) = l_i \quad (i = 1, \ldots, k)\,.$$

Set

$$a'(i) = \#S_i \cap \{X_\omega(1), \ldots, X_\omega(n+1)\}\,.$$

Then $a'(i) \geq a(i)$, $a'(i) \geq l_i$ $(i = 1, \ldots, k)$, and a fortiori

$$\operatorname{Rank}\big(b_i'(X_{i1}) \vdots \ldots \vdots b_i'(X_{ia'(i)})\big) = l_i \quad (i = 1, \ldots, k)\,.$$

Hence $\omega \in \Omega'(a', n+1)$ and $(X_{11}, \ldots, X_{1a'(1)}; \ldots; \ldots, X_{ka'(k)}) \in A^*(a')$, i.e. $\omega \in \Omega_{n+1}$. Therefore $\Omega_n \subset \Omega_{n+1}$. To prove the second part of Theorem 4 note that for given $a(1), \ldots, a(k)$, the $P_{\Omega'(a,n)}$-distribution of $W = (X_{11}, \ldots, X_{ka(k)})$ by Theorem 1 coincides with the P-distribution of W by Theorem 2, hence $P_{\Omega'(a,n)}(W \in A^*) = P(W \in A^*)$. By Corollary 2 we have

$$P(W \in A^*) = \prod_{i=1}^{k} P\{\operatorname{Rank}\big(b_i'(X_{i1}) \vdots \ldots \vdots b_i'(X_{ia(i)})\big) = l_i\} \uparrow 1$$

as $a(i) \uparrow \infty$ $(i = 1, \ldots, k)$, provided the P-distribution of $b_i'(X_{ij})$ $(j = 1, \ldots, a(i))$ or the $P_{(X \in S_i)}$-distribution of $b_i'(X)$ is not concentrated in any proper subspace of $\mathbb{R}^{l_i}$ $(i = 1, \ldots, k)$. Hence

$$\forall \varepsilon > 0,\ \exists (\triangle_1, \ldots, \triangle_k),\ \forall a = (a(1), \ldots, a(k)) \geq \triangle = (\triangle_1, \ldots, \triangle_k)\,,$$
$$P_{\Omega'(a,n)}(W \in A^*) \geq 1 - \varepsilon\,.$$

But, with the same summation range as in (32), we have

$$P(\Omega_n) = \sum P(\Omega'(a,n)) P_{\Omega'(a,n)}(W \in A^*)\,,$$

hence as $a \geq \triangle$ we get the inequality

$$P(\Omega_n) \geq (1 - \varepsilon) P\Big(\sum \Omega'(a,n)\Big)\,.$$

By (5), §2, we have

$$\sum \Omega'(a,n) = \{\omega : \#S_i \cap \{X(1), \ldots, X(n)\} \geq l_i;\ i = 1, \ldots, k\}\,,$$

accordingly $P(\sum \Omega'(a,n)) \to 1$ as $n \to \infty$. Thus $\liminf_{n\to\infty} P(\Omega_n) \geq 1 - \varepsilon$, $\forall \varepsilon > 0$, hence $\lim_{n\to\infty} P(\Omega_n) = P(\lim_{n\to\infty} \Omega_n) = 1$. Theorem 4 is proved.

Remark 3. For the conditional estimability domain Ω_n of θ, always $\Omega_n \subset \sum \Omega'(a,n)$. If the $P_{(X \in S_i)}$-distribution of $b_i'(X)$ has a density with respect to the

Lebesgue measure in $\mathbb{R}^{l_i}$, then

$$P_{\Omega^*}\{\det\bigl(b_i'(X_{i1})\vdots\ldots\vdots b_i'(X_{il_i})\bigr)=0\}=0\,,\qquad (i=1,\ldots,k)\,.$$

Then, for the reason explained in the proof of Corollary 1, we have $P_{\Omega^*}(W\in A^*)=1$, in particular $P_{\Omega'(a,n)}(W\in A^*)=1$ if $a(i)\ge l_i$ $(i=1,\ldots,k)$, and thus $P(\Omega_n)=P(\sum\Omega'(a,n))$, i.e. in this case θ is conditionally estimable in almost the whole $\sum\Omega'(a,n)$.

8. Properties of the CGLSE. S_i-transformations acting on the data (1) split the initial model $\{(X(t),Y(t)),\ t=1,\ldots,n\}$ satisfying (21) into a system of conditional models (23), or, roughly speaking, into a system of "infinitesimal" models. In this section, we show that, by joining together instantaneous states w, several data situations a to recover the most part of Ω, from the properties of CGLSE in "infinitesimal" linear models we arrive at global properties.

Consider the model (23) in which the range of θ is some fixed set $\Theta\subset M_{l\times r}$, $\theta'=(\theta_1'\vdots\ldots\vdots\theta_k')$, $\Theta_0=\{\theta-\widetilde{\theta}\,:\,\theta,\widetilde{\theta}\in\Theta\}$,

$$\begin{aligned}
\Delta&=\text{the linear hull of }\Theta_0\text{ in }M_{l\times r}\,,\\
\Theta_1&=\{\theta_1\,:\,\exists(\theta_2,\ldots,\theta_k),\theta\in\Theta\},\ldots,\\
\Theta_k&=\{\theta_k\,:\,\exists(\theta_1,\ldots,\theta_{k-1}),\theta\in\Theta\}\,,\\
\Theta_{i0}&=\{\theta_i-\widetilde{\theta}_i\,:\,\theta_i,\widetilde{\theta}_i\in\Theta_i\}\,,\\
\Delta_{i0}&=\text{the linear hull of }\Theta_{i0}\text{ in }M_{l_i\times r}\qquad (i=1,\ldots,k)\,,\\
\mathcal{D}=\mathcal{D}(w)&=\{B\theta\,:\,\theta\in\Theta\}\,,\\
\mathcal{D}_0&=\{B(\theta-\widetilde{\theta})\,:\,\theta,\widetilde{\theta}\in\Theta\}=\{B\delta\,:\,\delta\in\Theta_0\}\,.
\end{aligned}$$

Like Assumption 1, we shall here make use of

Assumption 2. For every $w\in A^*$, either

(a) there exist linear subspaces

$$\mathcal{L}_i=\mathcal{L}_i(w)\subset\mathcal{M}(B_i)\subset\mathbb{R}^{a(i)}\qquad (i=1,\ldots,k)$$

such that $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}$ is the linear hull of $\mathcal{D}_0$ in $M_{s_k\times r}$, or

(b) $\Sigma_1=\ldots=\Sigma_k$ and there exists a linear subspace $\mathcal{L}=\mathcal{L}(w)\subset\mathcal{M}(B)\subset\mathbb{R}^{s_k}$ such that $\mathcal{M}^r_{\mathcal{L}}$ is the linear hull of $\mathcal{D}_0$ in $M_{s_k\times r}$.

When there are no constraints on θ, i.e. $\Theta=M_{l\times r}$, or, more generally, when there are only constraints separately on each θ_i of the form

$$\Delta=\{(\delta_1'\vdots\ldots\vdots\delta_k')':\delta_i\in\Delta_{i0},i=1,\ldots,k\}\,,$$

$$\overrightarrow{\Delta_{i0}}=\{\overrightarrow{\delta_i}:\delta_i\in\Delta_{i0}\}=\mathop{\times}_{h=1}^{r}\Lambda_i\,,$$

where Λ_i is a linear subspace of $\mathbb{R}^{l_i}$ $(i=1,\ldots,k)$, Assumption 2(a) is satisfied.

Indeed, write $\delta_i = \left(\delta_{i1} \vdots \ldots \vdots \delta_{ir}\right) \in \varDelta_{i0}$ or $\delta_{ih} \in \varLambda_i$ $(\forall h = 1, \ldots, r)$, and $\mathcal{L}_i = \{B_i \delta_{ih} : \delta_{ih} \in \varLambda_i\}$. The linear hull of $\mathcal{D}_0$ is then

$$\begin{aligned}\{B\delta : \delta \in \varDelta\} &= \{B\delta : \delta = (\delta_1' \vdots \ldots \vdots \delta_k')', \delta_i \in \varDelta_{i0},\ i = 1, \ldots, k\} \\ &= \{B\delta : B = \operatorname{diag}(B_1, \ldots, B_k),\ B_i\delta_{ih} \in \mathcal{L}_i,\ i = 1, \ldots, k; h = 1, \ldots, r\} \\ &= \mathcal{M}^r_{\mathcal{L}_1 \times \ldots \times \mathcal{L}_k}\end{aligned}$$

because the hth column of $B\delta$ is $B(\delta_{1h}' \ldots \delta_{kh}')' = ((B_1\delta_{1h})' \ldots (B_k\delta_{kh})')'$ and varies over $\mathcal{L}_1 \times \ldots \times \mathcal{L}_k$ $(h = 1, \ldots, r)$.

Further, consider constraints imposed on θ to ensure in (21), §5, the junction of the zones $(Q)E(Y' \mid x) = b_i(x)\theta_i$ for several S_i $(i = 1, \ldots, k)$. Write

$$\begin{aligned}\theta_i &= (\lambda_{i1} \vdots \ldots \vdots \lambda_{ir}) \quad (i = 1, \ldots, k)\,, \\ \theta &= (\lambda_1 \vdots \ldots \vdots \lambda_r)\,, \quad \lambda_h' = (\lambda_{1h}' \ldots \lambda_{kh}') \quad (h = 1, \ldots, r)\,.\end{aligned}$$

Let L_{ij} be the common boundary of the domains S_i and S_j; the junction requirements are

$$b_i(x)\theta_i = b_j(x)\theta_j\,, \quad \forall x \in L_{ij}\,,$$

or, equivalently,

$$b_i(x)\lambda_{ih} = b_j(x)\lambda_{jh}\,, \quad h = 1, \ldots, r,\ \forall x \in L_{ij}\,.$$

We can analogously impose conditions that several shreds $y = b_i(x)\theta_i$ $(i = 1, \ldots, k)$ of the regression surface have common tangent hyperplane at every point of their common boundaries. Thus these constraints are imposed in exactly the same manner on every λ_h $(h = 1, \ldots, r)$, i.e., to sum up, the junction constraints are $\lambda_h \in \varGamma$ $(h = 1, \ldots, r)$, where $\varGamma$, independent of h, is some linear subspace of $\mathbb{R}^l$. Therefore, when the junction constraints are imposed, $\varTheta$ has the property that the linear hull $\varDelta$ of $\varTheta_0$ is of the form

$$\overrightarrow{\varDelta} = \{\overrightarrow{\delta} : \delta \in \varDelta\} = \bigtimes_{h=1}^{r} \varGamma$$

where $\varGamma$ is some linear subspace of $\mathbb{R}^l$. Let $\varTheta$ have this property. Consider

$$\delta = (\delta_1 \vdots \ldots \vdots \delta_r) \in \varDelta\,, \qquad \mathcal{L} = \{B\delta_h : \delta_h \in \varGamma\}\,.$$

Then the linear hull of $\mathcal{D}_0$ in $M_{s_k \times r}$ is

$$\{B\delta : \delta \in \varDelta\} = \{(B\delta_1 \vdots \ldots \vdots B\delta_r) : \delta_h \in \varGamma,\ h = 1, \ldots, r\} = \mathcal{M}^r_{\mathcal{L}}\,.$$

Hence, in the case $\varSigma_1 = \ldots = \varSigma_k$, when θ is subject to the junction constraints, Assumption 2(b) is satisfied.

Let $d_1(\cdot), \ldots, d_k(\cdot)$ in (21),§5, be positive functions, let $V_1, \ldots, V_k$ be defined in Theorem 3, and let $M_{s_k \times r}$ be endowed with the scalar product

$$\overrightarrow{\zeta_1}[I_r \otimes \operatorname{diag}^{-1}(V_1, \ldots, V_k)]\overrightarrow{\zeta_2}, \quad \zeta_1, \zeta_2 \in M_{s_k \times r}\,. \tag{33}$$

Let $G_e = G_e(w)$ be the orthoprojector of $M_{s_k \times r}$ on an arbitrary linear subspace $\mathcal{E}$ of $M_{s_k \times r}$ containing $\mathcal{D}_0$. By the same symbol $G = G(w)$ we denote the orthoprojector of $M_{s_k\times r}$ on $\mathcal{M}^r_{\mathcal{L}_1\times\ldots\times\mathcal{L}_k}$ or $\mathcal{M}^r_{\mathcal{L}}$ according as Assumption 2(a) or 2(b) is used. Let θ_0 be some fixed element of Θ, and I the identity mapping of $M_{s_k\times r}$. Consider

$$\overline{\theta}(U,w)=B^-(w)G_e(w)U+B^-(w)(I-G_e(w))B(w)\theta_0\,,$$

where the products are those of the mappings $B^-, G_e, I-G_e, B$; moreover, B^-G_eU, e.g., is the image of $U\in M_{s_k\times r}$ by the mapping B^-G_e; and let

$$\widehat{\theta}(U,w)=B^-(w)G(w)U+B^-(w)\big(I_{s_k}-G(w)\big)B(w)\theta_0\,,$$

where the products are simply those of the matrices $B^-, G, U, \ldots$ By applying Lemmas 4, 5 to the model (23) for every fixed $w\in A^*$, we see that $\overline{\theta}(U,w)$, $\widehat{\theta}(U,w)$ are ILUE for θ. Now, in Theorem 1, $W=\{X_{11},\ldots,X_{1a(1)};\ldots;\ldots,X_{ka(k)}\}$, the pair (U,W) is defined on Ω^*, the functions $\overline{\theta}(U,W)$, $\widehat{\theta}(U,W)$ are defined on $\Omega^*\cap(W\in A^*)$. In particular, letting $\Omega^*=\Omega'$ (see (5), §2), we have (see §4)

$$W=\{S_i\cap\{X(1),\ldots,X(n)\},\ i=1,\ldots,k\}\,.$$

U being paired with W, the functions $\widehat{\theta}(U,W)$, $\overline{\theta}(U,W)$ are defined on Ω_n (see (32), §7), they are the CGLSE of θ; their nice properties in finite sample are stated in

Theorem 5. *Suppose the conditions of Corollary* 1 *are satisfied.*

(i) $\overline{\theta}(U,W)$ *is locally and globally unbiased*, *i.e.*

$$(Q)E_{\Omega^*\cap(W\in A^*)}\{\overline{\theta}(U,W)\mid W\}=\theta\,,\quad E_{\Omega_n}\overline{\theta}(U,W)=\theta\quad \textit{for every } \theta\in\Theta\,.$$

(ii) *Under Assumption* 2, $\widehat{\theta}(U,W)$ *is locally and globally optimal*, *i.e. for every* $M_{l\times r}$*-valued function* $\varphi(U,W)$ *such that* $\overrightarrow{\varphi}(U,w)=\lambda(w)\overrightarrow{U}+\delta(w)$ *is an ILUE for* $\overrightarrow{\theta}$ *in the model* (23) (see Definition 2)

$$(Q)D_{\Omega^*\cap(W\in A^*)}\{\varphi(U,W)\mid W\}\geq(Q)D_{\Omega^*\cap(W\in A^*)}\{\widehat{\theta}(U,W)\mid W\}\,,$$
$$D_{\Omega_n}\varphi(U,W)\geq D_{\Omega_n}\widehat{\theta}(U,W)\,.$$

These properties are valid provided for every $s_k\leq n$ *the following regularity conditions are fulfilled for some particular g-inverse* B^-:

Setting $\nu=\{1,\ldots,s_k\}$, $X(\nu)=\{X(1),\ldots,X(s_k)\}$, $\mathcal{S}=S_1^{a(1)}\times\ldots\times S_k^{a(k)}$, *let*

1) $E\|Y\|^2<\infty$,

2) $E_{\{X(\nu)\in\mathcal{S}\}}\|B^-(X(\nu))G(X(\nu))\|^2\cdot\sum_{i=1}^k\|Y(s_i)\|^2<\infty$,

3) $E_{\{X(\nu)\in\mathcal{S}\}}\|B^-(X(\nu))G(X(\nu))B(X(\nu))\|^2<\infty$,

4) $E_{\{X(\nu)\in\mathcal{S}\}}\|B^-(X(\nu))B(X(\nu))\|^2<\infty$,

5) $E_{\{X(\nu)\in\mathcal{S}\}}\|\lambda(X(\nu))\|^2\cdot\sum_{i=1}^k\|Y(s_i)\|^2<\infty$,

6) $E_{\{X(\nu)\in\mathcal{S}\}}\|\delta(X(\nu))\|^2<\infty$,

7) $E_{\{X(\nu)\in\mathcal{S}\}}|[I_r\otimes B^-(X(\nu))]G_e(X(\nu))|\cdot\sum_{i=1}^{k}|Y(s_i)|<\infty$,

8) $E_{\{X(\nu)\in\mathcal{S}\}}|[I_r\otimes B^-(X(\nu))]G_e(X(\nu))[I_r\otimes B(X(\nu))]|<\infty$,

where, by abuse of notation, the symbol G_e *in* 7), 8) *also denotes the* $s_k r\times s_k r$*-matrix associated with the linear operator* $G_e: M_{s_k\times r}\to M_{s_k\times r}$.

Proof. (i) By Lemma 5, at any instantaneous state $w\in A^*$, $\overline{\theta}(U,w)$ is unbiased for θ in (23), i.e.

$$(Q)E_{\Omega^*\cap(W\in A^*)}\{\overline{\theta}(U,w)\mid w\}=\theta$$

or, by Theorem 1(iii),

$$(Q)E_{\Omega^*\cap(W\in A^*)}\{\overline{\theta}(U,W)\mid w\}=\theta,\quad \text{or}$$
$$(Q)E_{\Omega^*\cap(W\in A^*)}\{\overline{\theta}(U,W)\mid W\}=\theta,\quad \forall\theta\in\Theta\,.$$

Hence for every set Ω^* in Theorem 1

$$E_{\Omega^*\cap(W\in A^*)}\overline{\theta}(U,W)=\theta\,.$$

Letting $\Omega^*=\Omega'=\Omega'(a,n)$, then multiplying both sides by $P(\Omega'\cap(W\in A^*))$ and summing up over the range as in (32), §7, we get $E_{\Omega_n}\overline{\theta}(U,W)=\theta$, $\forall\theta\in\Theta$.

(ii) By Lemma 5, at every $w\in A^*$, under Assumption 2, $\widehat{\theta}(U,w)$ is a BILUE for θ in (23), i.e.

$$(Q)D_{\Omega^*\cap(W\in A^*)}\{\varphi(U,w)\mid w\}\geq(Q)D_{\Omega^*\cap(W\in A^*)}\{\widehat{\theta}(U,w)\mid w\}\,.$$

Then, by the same reason as above,

$$(Q)D_{\Omega^*\cap(W\in A^*)}\{\varphi(U,W)\mid W\}\geq(Q)D_{\Omega^*\cap(W\in A^*)}\{\widehat{\theta}(U,W)\mid W\}\,.$$

Moreover, we get as in (i)

$$E_{\Omega^*\cap(W\in A^*)}\varphi(U,W)=\theta\,,\quad E_{\Omega^*\cap(W\in A^*)}\widehat{\theta}(U,W)=\theta\,,\quad\forall\theta\in\Theta\,.$$

The conditional dispersion matrices of $\varphi(U,W)$ and $\widehat{\theta}(U,W)$ are the conditional expectations of the expressions $(\overrightarrow{\varphi}-\overrightarrow{\theta})(\overrightarrow{\varphi}-\overrightarrow{\theta})'$ and $(\overrightarrow{\widehat{\theta}}-\overrightarrow{\theta})(\overrightarrow{\widehat{\theta}}-\overrightarrow{\theta})'$ respectively. Then, by proceeding as in (i) we get

$$D_{\Omega^*\cap(W\in A^*)}\varphi(U,W)\geq D_{\Omega^*\cap(W\in A^*)}\widehat{\theta}(U,W)\,,$$

and $D_{\Omega_n}\varphi(U,W)\geq D_{\Omega_n}\widehat{\theta}(U,W)$ because $E_{\Omega_n}\varphi(U,W)=\theta, E_{\Omega_n}\widehat{\theta}(U,W)=\theta$, $\forall\theta\in\Theta$.

To make the reasoning rigorous, we must check the existence and finiteness of the expectations of $\overrightarrow{\widehat{\theta}}\left(\overrightarrow{\widehat{\theta}}\right)'$, $\overrightarrow{\varphi}(\overrightarrow{\varphi})'$, $\overline{\theta}$. Because $Z(s_{i-1}+1),\dots,Z(s_i)$ are i.i.d.

with respect to $P_{\{X(\nu)\in\mathcal{S}\}}$, the regularity condition 2) entails

$$E_{\{X(\nu)\in\mathcal{S}\}}\|B^-(X(\nu))G(X(\nu))\|^2\cdot\sum_{t=1}^{s_k}\|Y(t)\|^2<\infty\,,$$

hence by Theorem 1(i),(ii)

$$E_{\Omega^*}\|B^-(W)G(W)\|^2\cdot\|U\|^2<\infty\,,$$

a fortiori

$$E_{\Omega^*\cap(W\in A^*)}\|B^-(W)G(W)U\|^2<\infty\,.$$

Similarly, 3) and 4) entail

$$E_{\Omega^*\cap(W\in A^*)}\|B^-(W)B(W)\theta_0\|^2<\infty\,,$$
$$E_{\Omega^*\cap(W\in A^*)}\|B^-(W)G(W)B(W)\theta_0\|^2<\infty\,.$$

Therefore

$$E_{\Omega^*\cap(W\in A^*)}\|\widehat{\theta}(U,W)\|^2<\infty,\qquad E_{\Omega^*\cap(W\in A^*)}\Big|\overrightarrow{\theta}\left(\overrightarrow{\theta}\right)'\Big|<\infty\,,$$

for every Ω^*, in particular for $\Omega^*=\Omega'(a,n)$, $a=(a(1),\dots,a(k))\ge(l_1,\dots,l_k)$. Hence

$$E_{\Omega_n}\Big|\overrightarrow{\theta}(U,W)\Big(\overrightarrow{\theta}\Big)'(U,W)\Big|<\infty\,.$$

Similarly, $E_{\Omega_n}|\overrightarrow{\varphi}(U,W)(\overrightarrow{\varphi})'(U,W)|<\infty$, $E_{\Omega_n}|\overline{\theta}(U,W)|<\infty$. Thus all expectations, variances and covariances involved exist and are finite. Theorem 5 is proved.

Theorems 4, 5 show that the CGLSE $\overline{\theta}$, $\widehat{\theta}$ are unbiased, $\widehat{\theta}$ has the smallest (in the PSD ordering sense) dispersion matrix on the domains Ω_n that enlarge to become almost the whole original space Ω when the sample size n increases.

Remark 4. Let G_0 and G^* be respectively the orthoprojectors of $M_{s_k\times r}$ and of the subspace $\mathcal{E}$, endowed with the scalar product (33), on the linear hull $\mathcal{T}$ of $\mathcal{D}_0$ in $M_{s_k\times r}$. It can be checked that $G_0=G^*G_e$. Consider two CGLSE of $\mu=E_{\Omega^*\cap(W\in A^*)}(U\mid w)$, $\mu\in\mathcal{D}$, based respectively on G_e, G_0 :

$$\tau=G_eU+(I-G_e)\mu_0,\qquad \tau_0=G_0U+(I-G_0)\mu_0\,,$$

where $\mu_0=\mu_0(w)$ is arbitrarily chosen in $\mathcal{D}=\mathcal{D}(w)$. Since $\mu_0-\mu\in\mathcal{D}_0\subset\mathcal{T}\subset\mathcal{E}\subset M_{s_k\times r}$, we have

$$\begin{aligned}\tau-\mu&=G_e(U-\mu_0)+\mu_0-\mu=G_e(U-\mu_0)+G_e(\mu_0-\mu)=G_e(U-\mu)\,,\\ \tau_0-\mu&=G_0(U-\mu)=G^*G_e(U-\mu)\,,\\ \tau_0-\mu&=G^*(\tau-\mu)\,.\end{aligned}$$

Hence, by the projection property, with the scalar product (33),

$$(\tau_0-\mu,\tau_0-\mu)\le(\tau-\mu,\tau-\mu)\,.$$

Because $w \in A^*$, $B = B(w)$ defines an injective mapping from $M_{l\times r}$ onto $\mathcal{M}(B) \subset M_{s_k\times r}$. Let us define on $M_{l\times r}$ a scalar product by $(\theta, \widetilde{\theta})_q = (\mu, \widetilde{\mu})$, where $(\mu, \widetilde{\mu})$ is given by (33), $\mu = B\theta$, $\widetilde{\mu} = B\widetilde{\theta}$. We have

$$\begin{aligned}
(\theta, \widetilde{\theta})_q &= (B\theta, B\widetilde{\theta}) = (\overrightarrow{B\theta})'[I_r \otimes \operatorname{diag}^{-1}(V_1, \ldots, V_k)](\overrightarrow{B\widetilde{\theta}}) \\
&= [(I_r \otimes B)\overrightarrow{\theta}]'[I_r \otimes \operatorname{diag}(V_1^{-1}, \ldots, V_k^{-1})](I_r \otimes B)\overrightarrow{\widetilde{\theta}} \\
&= (\overrightarrow{\theta})'[I_r \otimes B' \operatorname{diag}(V_1^{-1}, \ldots, V_k^{-1})B]\overrightarrow{\widetilde{\theta}} \\
&= (\overrightarrow{\theta})'[I_r \otimes \operatorname{diag}(B_i'V_i^{-1}B_i,\, i = 1, \ldots, k)]\overrightarrow{\widetilde{\theta}}\,.
\end{aligned}$$

Define $q = q(w) = [I_r \otimes \operatorname{diag}(B_i'V_iB_i,\, i = 1, \ldots, k)]$, a positive definite $lr \times lr$-matrix. Then

$$(\theta, \widetilde{\theta})_q = (\overrightarrow{\theta})'q\overrightarrow{\widetilde{\theta}}\,.$$

Consider two CGLSE of θ, which are CILUE by Theorem 5(i),

$$\begin{aligned}
\overline{\theta} &= B^-G_eU + B^-(I - G_e)\mu_0 = B^-\tau\,, \\
\theta^* &= B^-G_0U + B^-(I - G_0)\mu_0 = B^-\tau_0\,,
\end{aligned}$$

where $\mu_0 = B\theta_0$. Then, for every $\theta \in \Theta$, $\theta = B^-\mu$, $\overline{\theta} - \theta = B^-\tau - B^-\mu = B^-(\tau - \mu)$, $\theta^* - \theta = B^-(\tau_0 - \mu)$, we get

$$(\theta^* - \theta, \theta^* - \theta)_q = (\tau_0 - \mu, \tau_0 - \mu) \le (\tau - \mu, \tau - \mu) = (\overline{\theta} - \theta, \overline{\theta} - \theta)_q$$

or

$$(\overrightarrow{\theta^*} - \overrightarrow{\theta})'q(\overrightarrow{\theta^*} - \overrightarrow{\theta}) \le (\overrightarrow{\overline{\theta}} - \overrightarrow{\theta})'q(\overrightarrow{\overline{\theta}} - \overrightarrow{\theta})\,, \qquad \forall \theta \in \Theta\,.$$

Thus, the conditional quadratic loss function, given the instantaneous state w, based on $q = q(w)$, of the CGLSE θ^* corresponding to the linear hull $\mathcal{T}$ of $\mathcal{D}_0$ is uniformly not greater than that of the CGLSE $\overline{\theta}$ corresponding to any subspace $\mathcal{E}$ containing $\mathcal{D}_0$. This fact is to be taken into consideration when $\mathcal{T}$ is an arbitrary linear subspace of $M_{s_k\times r}$, especially when Assumption 2 is not satisfied, and then, by Theorem 5(i), we have to view different CGLSE of θ.

III. Prediction of the response variable

9. Introduction. We start from the data-bases $\{(X(t), Y(t)), t = 1, \ldots, n\}$ of n items. On a new item, the $(n+1)$th, say, the observed value $X(n+1)$ is available, and one must predict the response value $Y(n+1)$ on the ground of $\{(X(t), Y(t)), t = 1, \ldots, n; X(n+1)\}$.

Suppose on the range H of X the disjoint domains $S_1, \ldots, S_k$ are prescribed. The sum $S = S_1 + \ldots + S_k$ represents the relevant domain of explanatory values. We shall exhibit a predictor ψ only when the observed value x of $X(n+1)$ falls in S. Thus ψ is some function of

$$\{(X(1),Y(1)),\ldots,(X(n),Y(n)),x\} \quad \text{where } x\in S\,.$$

If the exact mixed conditional distribution $Q(\,\cdot\,,x)$, or at least a conditional location function $f(x,\theta)$ of $Y(n+1)$ given $X(n+1)$ were known, the predictor value would be $f(x,\theta)$. Thus the prediction is directed by $Q(\,\cdot\,,x)$, hence it would be reasonable to perform all S_i-transformations preserving $Q(\,\cdot\,,x)$ and to construct a good estimator $\widehat{\theta}$ of θ on the basis of the transformed data $\{(X_{ij},Y_{ij}), i = 1,\ldots,k; j = 1,\ldots,a(i)\}$, or (U,W), then to predict $Y(n+1)$ by $f[x,\widehat{\theta}(U,W)]$ when the functional form $f(\,\cdot\,,\,\cdot\,)$ is known.

The adequacy of the predictor ψ will be assessed by $\|\psi-Y(n+1)\|^2\cdot I_S(X(n+1))$, with the Euclidean norm $\|\cdot\|$; the factor I_S reminds us not to carry out the prediction when $X(n+1)\notin S$.

10. Predictors connected with the CGLSE. For studying the properties of predictors, we introduce

Definition 3. Consider a function

$$\psi(x,U,w) = L(x,w)\overrightarrow{U} + c(x,w)$$

where $L(\,\cdot\,,\,\cdot\,), c(\,\cdot\,,\,\cdot\,)$ are matrix-valued and $\mathcal{A}\times\mathcal{A}^{s_k}$-measurable in $(x,w)\in S\times A$; the regularity conditions for the existence of

$$E_{\Omega'(a,n)\cap(W\in A)}\psi(X(n+1),U,W)I_S(X(n+1))$$

are to be satisfied. Then $\psi(X(n+1),U,W)$ is called a *conditional inhomogeneous linear unbiased predictor* (CILUP) for $Y(n+1)$ if, for every envisaged underlying distribution of (X,Y),

$$\begin{aligned}(Q)E_{\Omega'\cap(W\in A)}&\{\psi(X(n+1),U,W)I_S(X(n+1)) \mid W, X(n+1)\}\\ &= (Q)E_{\Omega'\cap(W\in A)}\{Y(n+1)I_S(X(n+1)) \mid W, X(n+1)\}\,.\end{aligned}$$

The prefix (Q) will be justified below, by Lemma 7.

This definition is stated from the viewpoint of predicting the r.v. $Y(n+1)$ by a random predictor $\psi(X(n+1),U,W)$, after performing all S_i-transformations, at every random instantaneous state W, $X(n+1)$ being observed only in S; the basic space used is $\Omega'(a,n)\cap(W\in A)$, i.e. we always start from an elementary situation a, with A as described in Theorem 1(iii). The following lemmas give an equivalent form of Definition 3.

Lemma 7. *Choose* (see Remark 1, §2)

$$\Omega^* = \overline{\Omega} = \{X(n+1)\in S\}\cap\Omega'(a,n) = \Omega_1+\ldots+\Omega_k$$

where

$$\Omega_h = \{X(n+1)\in S_h\}\cap\Omega', \qquad h=1,\ldots,k\,.$$

Let $g(Z(n+1),U,W)$ be any matrix-valued function such that

$$E_{\Omega'\cap(W\in A)}g(Z(n+1),U,W)I_S(X(n+1)) \quad \textit{exists.}$$

Then a.s.

$$E_{\overline{\Omega}\cap(W\in A)}\{g(Z(n+1),U,W) \mid U,W,X(n+1)\} = \sum_{h=1}^{k} I_{S_h}(X(n+1))E_{\Omega_h\cap(W\in A)}\{g(Z(n+1),U,W) \mid U,W,X(n+1)\}\,.$$

In particular, for $\omega \in \{X(n+1)\in S\}$,

$$(Q)E_{\Omega'\cap(W\in A)}\{g(Z(n+1),U,W)I_S(X(n+1)) \mid W,X(n+1)\} = (Q)E_{\overline{\Omega}\cap(W\in A)}\{g(Z(n+1),U,W) \mid W,X(n+1)\}\,,$$

where both sides are determined by $Q(\,\cdot\,,\,\cdot\,)$.

Proof. For every set $\Gamma \in \mathcal{C}^{s_k}\times\mathcal{A}^{s_k}\times\mathcal{A}$, consider

$$E_{\Omega_h\cap(W\in A)}\{\Phi_h(U,W,X(n+1))I_\Gamma(U,W,X(n+1))\} = E_{\Omega_h\cap(W\in A)}\{g(Z(n+1),U,W)I_\Gamma(U,W,X(n+1))\}$$

where $\Phi_h = E_{\Omega_h\cap(W\in A)}\{g \mid U,W,X(n+1)\}$. Set $\Phi=\sum_{h=1}^{k}\Phi_h I_{S_h}(X(n+1))$. Then

$$E_{\overline{\Omega}\cap(W\in A)}\{\Phi I_\Gamma(U,W,X(n+1))\} = E_{\overline{\Omega}\cap(W\in A)}\{gI_\Gamma\}$$

and the first a.s. equality is proved.

By applying Lemma 1 to the space $\Omega'\cap(W\in A)$ instead of Ω, we get

$$E_{\overline{\Omega}\cap(W\in A)}\{g(Z(n+1),U,W) \mid W,X(n+1)\} = E_{\Omega'\cap(W\in A)}\{g(Z(n+1),U,W)I_S(X(n+1)) \mid W,X(n+1)\}$$

a.s. mod $P_{\overline{\Omega}\cap(W\in A)}$ because

$$\overline{\Omega}\cap(W\in A) = (\Omega'\cap(W\in A))\cap\{(X(n+1),W)\in S\times A\}\,.$$

On the other hand, set $a'(h)=a(h)+1$; we get

$$E_{\Omega_h\cap(W\in A)}\{g(Z(n+1),U,W) \mid W,X(n+1)\} = E_{\Omega_h\cap(W\in A)}\{g(Z_{ha'(h)},U,W) \mid W,X_{ha'(h)}\}$$

(for notations, see §2). By Theorem 1(iii) and Remark 1, the last expectation can be determined by $Q(\,\cdot\,,\,\cdot\,)$; hence, because the first equality just proved applies as well to

$$E_{\overline{\Omega}\cap(W\in A)}\{g(Z(n+1),U,W) \mid W,X(n+1)\}\,,$$

this conditional expectation can also be determined by $Q(\,\cdot\,,\,\cdot\,)$, and we get the second equality of Lemma 7 where the left-hand side is determined by $Q(\,\cdot\,,\,\cdot\,)$ on the set $\{X(n+1)\in S\}$.

LEMMA 8. *The unbiasedness in Definition* 3 *is equivalent to*

$$(Q)E_{\overline{\Omega}\cap(W\in A)}\{\psi(X(n+1),U,W)\mid W,X(n+1)\}$$
$$= (Q)E\{Y(n+1)|X(n+1)\} \quad \textit{on the set } \{X(n+1)\in S\}\,.$$

In other words, a CILUP for $Y(n+1)$ in the sense of Definition 3 is also a CILUE for the regression function value at $X(n+1)\in S$, and conversely.

Proof. By the decomposition in Lemma 7,

$$(Q)E_{\Omega'\cap(W\in A)}\{Y(n+1)I_S(X(n+1))\mid W,X(n+1)\}$$
$$= (Q)E_{\overline{\Omega}\cap(W\in A)}\{Y(n+1)\mid W,X(n+1)\}$$
$$= \sum_{h=1}^{k} I_{S_h}(X(n+1))(Q)E_{\Omega_h\cap(W\in A)}\{Y_{ha'(h)}\mid W,X_{ha'(h)}\}\,,$$

where, as in the proof of Lemma 7, $a'(h)=a(h)+1$. By Theorem 1(iii), writing $p(x)=(Q)E(Y\mid X=x)$, we get

$$(Q)E_{\Omega_h\cap(W\in A)}\{Y_{ha'(h)}\mid W,X_{ha'(h)}\}=p(X_{ha'(h)})=p(X(n+1))\,;$$

indeed, here $\omega\in\Omega_h=\{X(n+1)\in S_h\}\cap\Omega'$, and so $X(n+1)=X_{ha'(h)}$. Thus

$$(Q)E_{\Omega'\cap(W\in A)}\{Y(n+1)I_S(X(n+1))\mid W,X(n+1)\}$$
$$=\sum_{h=1}^{k} I_{S_h}(X(n+1))p(X(n+1))=p(X(n+1))$$

provided $X(n+1)\in S$. But $p(X(n+1))=(Q)E(Y(n+1)\mid X(n+1))$ for $p(x)=(Q)E(Y\mid X=x)=(Q)E(Y(n+1)\mid X(n+1)=x)$, hence the relation in Definition 3 coincides with the one of Lemma 8, in view of Lemma 7. Lemma 8 is proved.

If the structure assumptions (21) are satisfied, then

$$(Q)E(Y'\mid x)=f'(x,\theta)=\sum_{i=1}^{k} b_i(x)I_{S_i}(x)\theta_i=b(x)\theta$$

where $b(x)=\big(b_1(x)I_{S_1}(x)\dots b_k(x)I_{S_k}(x)\big)$, an $1\times l$-matrix. Then a CILUP for $Y'(n+1)$ is also a CILUE of $b(x)\theta$ when $X(n+1)=x\in S$, and conversely. From the model (23), by Lemma 5(iv), $b(x)\overline{\theta}(U,w)$ (see Theorem 5) is an ILUE for $b(x)\theta$ for any given x. Hence we are led to consider the CGLS (conditional generalized least squares) predictors

$$\overline{Y'}(n+1)=b(X(n+1))\overline{\theta}(U,W)$$

connected with the CGLSE $\overline{\theta}(U,W)$ of θ by Theorem 5.

11. Properties of CGLS predictors

THEOREM 6. $\overline{Y'}(n+1)$ *is a CILUP for* $Y'(n+1)$ *in the sense of Definition* 3.

P r o o f. By Lemma 7 and its proof, in view of Corollary 1 we have $P(\Omega_h \cap (W \in A^*)) > 0$, for $h = 1, \ldots, k$, and

$$(Q)E_{\overline{\Omega}\cap(W\in A^*)}\{b(X(n+1)\overline{\theta}(U,W) \mid W, X(n+1)\}$$
$$= b(X(n+1)) \sum_{h=1}^{k} I_{S_h}(X(n+1))(Q)E_{\Omega_h\cap(W\in A^*)}\{\overline{\theta}(U,W) \mid W, X_{ha'(h)}\}\,.$$

By Theorem 1(iii)

$$(Q)E_{\Omega_h\cap(W\in A^*)}\{\overline{\theta}(U,W) \mid W, X_{ha'(h)}\} = (Q)E_{\Omega_h\cap(W\in A^*)}\{\overline{\theta}(U,W) \mid W\}\,.$$

By Theorem 5(i), applied to Ω_h, the right-hand side equals θ. Hence we have

$$(Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\overline{Y'}(n+1) \mid W, X(n+1)\} = b(X(n+1))\theta$$
$$= (Q)E(Y'(n+1) \mid X(n+1)) \quad \text{for } X(n+1) \in S\,.$$

From Lemma 8, Theorem 6 follows.

THEOREM 7. *Denote by* F *an arbitrary non-random positive semidefinite* $r \times r$*-matrix. Under the conditions of Theorem* 5 *together with Assumption* 2,

$$\widehat{Y'}(n+1) = b(X(n+1))\widehat{\theta}(U,W)$$

is a locally optimal estimator for $(Q)E(Y'(n+1) \mid X(n+1))$ *in the following sense*: *For every CILUE* $\Psi(X(n+1),U,W)$ *of* $(Q)E(Y(n+1) \mid X(n+1))$ *in the sense of Lemma* 8,

(i) $$(Q)D_{\overline{\Omega}\cap(W\in A^*)}\{\widehat{Y}(n+1) \mid W, X(n+1)\}$$
$$\le (Q)D_{\overline{\Omega}\cap(W\in A^*)}\{\Psi(X(n+1),U,W) \mid W, X(n+1)\},$$

(ii) $$(Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\|\widehat{Y}(n+1) - (Q)E(Y(n+1) \mid X(n+1))\|_F^2 \mid W, X(n+1)\}$$
$$\le (Q)W_{\overline{\Omega}\cap(W\in A^*)}\{\|\Psi(X(n+1),U,W)$$
$$-(Q)E(Y(n+1) \mid X(n+1))\|_F^2 \mid W, X(n+1)\}$$

provided the regularity conditions, *like those of Theorem 5*, *ensuring the existence of conditional dispersion matrices*, *are satisfied.*

P r o o f. (i) Define $\Psi = \Psi(X(n+1),U,W)$, $b = b(X(n+1))$, $\widehat{\theta} = \widehat{\theta}(U,W)$,

$$\mathcal{T} = \mathcal{T}(X(n+1),U,W) = (\Psi - \overrightarrow{b\theta})(\Psi' - b\theta) - (\overrightarrow{b\widehat{\theta}} - \overrightarrow{b\theta})(b\widehat{\theta} - b\theta)\,.$$

By Lemma 7, at any given value w, x_{n+1} of $(W, X(n+1))$ with $x_{n+1} \in S$,

(34) $$(Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\mathcal{T}(X(n+1),U,W) \mid w, x_{n+1}\}$$
$$= \sum_{h=1}^{k} I_{S_h}(x_{n+1})(Q)E_{\Omega_h\cap(W\in A^*)}\{\mathcal{T}(X(n+1),U,W) \mid w, x_{n+1}\}\,.$$

Applying Theorem 1(iii), integrating with respect to the mixed conditional distribution given $(W, X_{ha'(h)}) = (w, x_{n+1})$, we get

$$\begin{aligned}(Q)E_{\Omega_h\cap(W\in A^*)}&\{\mathcal{T}\,(X(n+1),U,W)\mid w,x_{n+1}\}\\&= (Q)E_{\Omega_h\cap(W\in A^*)}\{\mathcal{T}(x_{n+1},U,w)\mid w,x_{n+1}\}\,.\end{aligned}$$

Apply Theorem 1(iii) again to get

$$\begin{aligned}(Q)E_{\Omega_h\cap(W\in A^*)}&\{\mathcal{T}\,(x_{n+1},U,w)\mid w,x_{n+1}\}\\&= (Q)E_{\Omega_h\cap(W\in A^*)}\{\mathcal{T}\,(x_{n+1},U,w)\mid w\}\,.\end{aligned}$$

Apply Theorem 1(iii) to both Ω_h and $\overline{\Omega}$,

$$(Q)E_{\Omega_h\cap(W\in A^*)}\{\mathcal{T}\,(x_{n+1},U,w)\mid w\} = (Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\mathcal{T}\,(x_{n+1},U,w)\mid w\}\,.$$

Hence, replacing this expectation on the right-hand side of (34) gives

$$\begin{aligned}(Q)E_{\overline{\Omega}\cap(W\in A^*)}&\{\mathcal{T}\,(X(n+1),U,W)\mid w,x_{n+1}\}\\&= (Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\mathcal{T}\,(x_{n+1},U,w)\mid w\}\end{aligned}$$

because $x_{n+1}\in S$. Similarly,

$$\begin{aligned}(Q)E_{\overline{\Omega}\cap(W\in A^*)}&\{\Psi(x_{n+1},U,w)\mid w\}\\&= (Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\Psi(X(n+1),U,W)\mid w,x_{n+1}\}\\&= (Q)E\{Y(n+1)\mid X(n+1)=x_{n+1}\} = \big(b(x_{n+1})\theta\big)'\,,\end{aligned}$$

by the relation of Lemma 8, i.e. $\Psi(x_{n+1},U,w)$ is an ILUE for $\overrightarrow{b(x_{n+1})\theta}$ in the model (23); but $b(x_{n+1})\widehat{\theta}(U,w)$ is a BILUE for $b(x_{n+1})\theta$ in (23) by Lemma 4(ii), hence, for every $w\in A^*$, $x_{n+1}\in S$,

$$(Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\mathcal{T}\,(x_{n+1},U,w)\mid w\}\geq 0\,.$$

Thus, from the above we have

$$(Q)E_{\overline{\Omega}\cap(W\in A^*)}\{\mathcal{T}\,(X(n+1),U,W)\mid W,X(n+1)\}\geq 0\,,$$

which proves (i).

(ii) Observe that, for any two $r\times 1$ random vectors ξ,ζ with zero mean and $E\|\xi\|^2<\infty$, $E\|\zeta\|^2<\infty$,

$$\begin{aligned}E\|\xi\|_F^2 &= E(\xi' F\xi) = E\,\mathrm{Trace}\,\xi' F\xi = \mathrm{Trace}\,FE(\xi\xi') = \mathrm{Trace}\,FD\xi\,,\\ E\|\xi\|_F^2 - E\|\zeta\|_F^2 &= \mathrm{Trace}\,F(D\xi - D\zeta)\geq 0 \quad \text{if } D\xi\geq D\zeta\,.\end{aligned}$$

This argument, when applied to conditional expectations and dispersion matrices, proves (ii). Theorem 7 is proved.

Before passing to the optimality of $\widehat{Y}(n+1)$ as a predictor, we need two lemmas.

LEMMA 9. *Let ξ_1, ξ_2, ξ_3 be r.v.'s, with ξ_1 matrix-valued. Suppose (ξ_1, ξ_2) is independent of ξ_3, and $E\xi_1$ exists. Then*

$$E(\xi_1 \mid \xi_2, \xi_3) = E(\xi_1 \mid \xi_2) \qquad P^{\xi_2,\xi_3}\text{-}a.s.$$

The standard proof is omitted.

LEMMA 10. *For every $\mathbb{R}^r$-valued, $\big(\mathcal{B}^r, \mathcal{A} \times \mathcal{B}^{rs_k} \times \mathcal{A}^{s_k}\big)$-measurable function $\Psi(x, u, w)$ and non-random $F \geq 0$, almost surely*

$$\begin{aligned} &E_{\overline{\Omega}\cap(W\in A)}\{\|\Psi(X(n+1), U, W) - Y(n+1)\|_F^2 \mid W, X(n+1)\} \\ &\qquad = E_{\overline{\Omega}\cap(W\in A)}\{\|\Psi(X(n+1), U, W) \\ &\qquad\qquad -p(X(n+1))\|_F^2 \mid W, X(n+1)\} + u(X(n+1)) \end{aligned}$$

where

$$p(X) = (Q)E(Y \mid X), \qquad u(X) = (Q)E\{\|Y - p(X)\|_F^2 \mid X\}\,.$$

Proof. Define $(y_1, y_2) = y_1' F y_2$ for $y_1, y_2 \in \mathbb{R}^r$. Then

$$\|y_1 + y_2\|_F^2 = \|y_1\|_F^2 + \|y_2\|_F^2 + 2(y_1, y_2)\,.$$

We have

$$\begin{aligned} &E_{\overline{\Omega}\cap(W\in A)}\{\Psi(X(n+1), U, W) - p(X(n+1))\,, \\ &\qquad\qquad p(X(n+1)) - Y(n+1) \mid U, W, X(n+1)\} \\ &= \{\Psi(X(n+1), U, W) - p(X(n+1)), \\ &\qquad\qquad E_{\overline{\Omega}\cap(W\in A)}[p(X(n+1) - Y(n+1)) \mid U, W, X(n+1)]\} \end{aligned}$$

almost surely, because $\Psi(X(n+1), U, W), p(X(n+1))$ are Borel functions of $(X(n+1), U, W)$. By Lemma 7,

$$\begin{aligned} &E_{\overline{\Omega}\cap(W\in A)}\{p(X(n+1)) - Y(n+1) \mid U, W, X(n+1)\} \\ &\qquad = \sum_{h=1}^{k} I_{S_h}(X(n+1)) E_{\Omega_h\cap(W\in A)}\{p(X_{ha'(h)}) - Y_{ha'(h)} \mid U, W, X_{ha'(h)}\} \end{aligned}$$

almost surely. By Lemma 1, applied to Ω_h,

$$\begin{aligned} &E_{\Omega_h\cap(W\in A)}\{p(X_{ha'(h)}) - Y_{ha'(h)} \mid U, W, X_{ha'(h)}\} \\ &\qquad = E_{\Omega_h}\{p(X_{ha'(h)}) - Y_{ha'(h)} \mid U, W, X_{ha'(h)}\} \end{aligned}$$

$P_{\Omega_h\cap(W\in A)}$-almost surely, because

$$(W \in A) = \{(U, W, X_{ha'(h)}) \in K^{s_k} \times A \times S_h\}\,.$$

By Theorem 1(i), $(X_{ha'(h)}, Y_{ha'(h)})$ is P_{Ω_h}-independent of (U, W), hence, by Lemma 9,

$$\begin{aligned} &E_{\Omega_h}\{p(X_{ha'(h)}) - Y_{ha'(h)} \mid U, W, X_{ha'(h)}\} \\ &\qquad = E_{\Omega_h}\{p(X_{ha'(h)}) - Y_{ha'(h)} \mid X_{ha'(h)}\} \quad \text{a.s.} \end{aligned}$$

By Theorem 1(ii), the P_{Ω_h}-distribution of $(X_{ha'(h)}, Y_{ha'(h)})$ coincides with the $P_{\{X\in S_h\}}$-distribution of (X,Y), which implies, according to Lemma 1,

$$E_{\Omega_h}\{Y_{ha'(h)} \mid X_{ha'(h)}\} = p(X_{ha'(h)}) \quad \text{a.s.}$$

To sum up, we get

$$E_{\overline{\Omega}\cap(W\in A)}\{\Psi - p, p - Y(n+1) \mid U, W, X(n+1)\} = 0 \quad \text{a.s.}$$

Therefore

$$\begin{aligned} &E_{\overline{\Omega}\cap(W\in A)}\{\|\Psi(X(n+1),U,W) - Y(n+1)\|_F^2 \mid U,W,X(n+1)\} \\ &\quad = E_{\overline{\Omega}\cap(W\in A)}\{\|\Psi(X(n+1),U,W) - p(X(n+1))\|_F^2 \mid U,W,X(n+1)\} \\ &\qquad + E_{\overline{\Omega}\cap(W\in A)}\{\|p(X(n+1)) - Y(n+1)\|_F^2 \mid U,W,X(n+1)\} \quad \text{a.s.} \end{aligned}$$

By taking $E_{\overline{\Omega}\cap(W\in A)}\{\,\cdot \mid W, X(n+1)\}$, we obtain

$$\begin{aligned} &E_{\overline{\Omega}\cap(W\in A)}\{\|\Psi(X(n+1),U,W) - Y(n+1)\|_F^2 \mid W,X(n+1)\} \\ &\qquad = E_{\overline{\Omega}\cap(W\in A)}\{\|\Psi - p\|_F^2 \mid W,X(n+1)\} \\ &\qquad\quad + E_{\overline{\Omega}\cap(W\in A)}\{\|p - Y(n+1)\|_F^2 \mid W,X(n+1)\} \end{aligned}$$

almost surely. By applying the decomposition in Lemma 7 to the second summand of the right-hand side, we get

$$\sum_{h=1}^{k} I_{S_h}(X(n+1))(Q)E_{\Omega_h\cap(W\in A)}\{\|p(X_{ha'(h)}) - Y_{ha'(h)}\|_F^2 \mid W, X_{ha'(h)}\},$$

which equals $u(X(n+1))$ on $\overline{\Omega}\cap(W\in A)$ since, by Theorem 1(iii),

$$\begin{aligned} &(Q)E_{\Omega_h\cap(W\in A)}\{\|p(X_{ha'(h)}) - Y_{ha'(h)}\|_F^2 \mid W, X_{ha'(h)}\} \\ &\qquad = u(X_{ha'(h)}) = u(X(n+1)) \quad \text{on } \Omega_h\cap(W\in A). \end{aligned}$$

Thus Lemma 10 is proved.

THEOREM 8. *Under the conditions of Theorem* 7, $\widehat{Y}(n+1)$ *is a locally optimal predictor for* $Y(n+1)$ *in the following sense*: *for every CILUP* $\Psi(X(n+1),U,W)$ *in the sense of Definition* 3, *almost surely*

$$\begin{aligned} &E_{\overline{\Omega}\cap(W\in A^*)}\{\|\widehat{Y}(n+1) - Y(n+1)\|_F^2 \mid W,X(n+1)\} \\ &\qquad \le E_{\overline{\Omega}\cap(W\in A^*)}\{\|\Psi(X(n+1),U,W) - Y(n+1)\|_F^2 \mid W,X(n+1)\}. \end{aligned}$$

Proof. Theorem 8 follows from Theorem 7(ii) and Lemma 10.

THEOREM 9. *Under the conditions of Theorem* 7, $\widehat{Y}(n+1)$ *is a globally optimal predictor for* $Y(n+1)$ *in the following sense*: *for every CILUP* Ψ

$$\begin{aligned} &E_{\Omega_n}\{\|\widehat{Y}(n+1) - Y(n+1)\|_F^2 I_S(X(n+1))\} \\ &\qquad \le E_{\Omega_n}\{\|\Psi(X(n+1),U,W) - Y(n+1)\|_F^2 I_S(X(n+1))\}. \end{aligned}$$

P r o o f. From Theorem 8,

$$E_{\overline{\Omega}\cap(W\in A^*)}\{\|\widehat{Y}(n+1)-Y(n+1)\|_F^2\} \le E_{\overline{\Omega}\cap(W\in A^*)}\{\|\Psi(X(n+1),U,W)-Y(n+1)\|_F^2\}\,.$$

By multiplying by $P(\overline{\Omega}\cap(W\in A^*))$ and summing over the range as in (32), §7, we obtain

$$E_{\{X(n+1)\in S\}\cap\Omega_n}\{\|\widehat{Y}(n+1)-Y(n+1)\|_F^2\} \le E_{\{X(n+1)\in S\}\cap\Omega_n}\{\|\Psi(X(n+1),U,W)-Y(n+1)\|_F^2\}$$

because $\overline{\Omega}=\{X(n+1)\in S\}\cap\Omega'$. This is equivalent to what is to be proved.

Acknowledgment

This paper was presented at the mathematical statistics seminar of the Institute of Mathematics, Polish Academy of Sciences, in Warsaw, Poland, on January 11th and February 15th, 1990. The author would like to thank Professor R. Zieliński and members of the seminar for helpful comments and discussions. The author is also indebted to the referee for important hints on the presentation.

References

[1] N. Bac-Van, *On the statistical analysis of a random number of observations*, Acta Math. Vietnam. 13 (1) (1988), 55–61.

[2] D. R. Cox, *Some aspects of conditional and asymptotic inference: a review*, Sankhyā Ser. A 50 (1988), 314–337.

[3] H. Drygas, *Best quadratic unbiased estimation in variance-covariance component models*, Math. Operationsforsch. Statist. Ser. Statist. 8 (1977), 211–231.

[4] K. M. S. Humak, *Statistische Methoden der Modellbildung*, Band I, Akademie-Verlag, Berlin 1977, Band III, Akademie-Verlag, Berlin 1984.

[5] E. L. Lehmann and H. Scheffé, *Completeness, similar regions, and unbiased estimation. I*, Sankhyā 10 (1950), 305–340.

[6] Y. P. Mack and H.-G. Müller, *Derivative estimation in nonparametric regression with random predictor variable*, Sankhyā Ser. A 51 (1989), 59–72.

[7] K. V. Mardia, *Statistics of directional data*, Academic Press, London 1972.

[8] C. R. Rao, *Methodology based on the L_1-norm, in statistical inference*, Sankhyā Ser. A 50 (1988), 289–313.