The Annals of Statistics
1976, Vol. 4, No. 4, 779–787

GAUSS–MARKOV ESTIMATION FOR MULTIVARIATE LINEAR MODELS WITH MISSING OBSERVATIONS[1]

BY HILMAR DRYGAS

Johann Wolfgang Goethe-Universität

In this note we discuss multivariate linear models from the coordinate-free point of view, as earlier done by Eaton (1970). We generalize the result of Eaton by allowing for missing observations. This leads to models of the kind $EY \in L$, Cov $Y \in \{P(I \otimes \Sigma)P'\}$ where P is a diagonal mapping. The paper starts by deriving the conditions for existence of Gauss–Markov estimators (GME) of EY in models where the covariance-mappings are not necessarily nonsingular. These conditions are then applied to the above models if Σ runs either over all PSD-mappings or over all diagonal PSD-mappings. In the latter case L must be of the form $L = L_1 \times \cdots \times L_p$ while in the general case some further conditions on the L_i must be met. (If $P = I$, then $L_i = L_j$ must hold for all i, j; this is equivalent to the result obtained by Eaton). Examples show that these conditions are satisfied only under rather exceptional conditions.

1. Introduction and notation. Consider a multivariate linear model with n uncorrelated observations on a p-dimensional random vector y with common covariance-matrix Σ. An interesting question concerns the conditions under which a Gauss–Markov estimator (GME) of the mathematical expectation of the y's is independent of Σ. This problem has already been investigated in a recent paper by Eaton [3]. Eaton's result can equivalently be formulated as follows: If Y is the $n \times p$-matrix of the observations of the y's and $EY \in L$, then a GME exists iff the following holds: There exists a linear subspace $L_0 \subseteqq R^n$ such that $L = L_0 \times \cdots \times L_0$, where a $n \times p$-matrix A is an element of $L_0 \times \cdots \times L_0$ if and only if the ith column of A belongs to L_0, $i = 1, 2, \cdots, p$. Since L_0 can be represented in the form $L_0 = X_0 R^k = \{X_0 \vartheta : \vartheta \in R^k\}$ for some integer $k \leqq n$ and some $n \times k$-matrix X_0, it is clear that L can be represented in the form $L = \{X_0 \Pi : \Pi$ a matrix order $k \times p\}$, which is exactly Eaton's result. A generalization of Eaton's result is required for the following problem: Consider the situation that not at all points of observations (belonging to different rows) have all variables (belonging to different columns) been observed. In the univariate case, e.g., this means that instead of $\eta_1, \cdots, \eta_m, \cdots, \eta_n$ only $\eta_1, \cdots, \eta_m$ is observed, so instead of $y = (\eta_1, \cdots \eta_n)'$ only $Py = \hat{y} = (\eta_1, \cdots, \eta_m, 0, 0, \cdots, 0)$ is known, where $P = (p_{ij}; i, j = 1, 2, \cdots, n)$ and $p_{ij} = \delta_{ij}$ if

Received September 1972; revised June 1975.

[1] This work was done when the author was a research worker at "Sonderforschungsbereich 21 (Ökonometrie und Unternehmensforschung)" at the University of Bonn.

AMS 1970 *subject classifications.* 62F10, 62J05.

Key words and phrases. Multivariate statistics, linear models, regression analysis, Gauss–Markov estimation, missing observations.

$i, j \leqq m$ and $p_{ij} = 0$ otherwise. P is a symmetric, idempotent matrix (mapping) and, moreover, if $Ey \in L$, then $E\hat{y} \in PL$.

In extending this example to the multivariate case, let $Y = (y_1 \cdots y_p)$ be a random $n \times p$-matrix such that $EY \in L$, where L is a linear subspace of the vector-space V of all $n \times p$-matrices. If we assume that not all vectors y_i are observed at any point of observation we are led to the consideration of projections P_i $(i = 1, 2, \cdots, p)$ such that with

$$PY = \operatorname{diag}(P_1, \cdots, P_p)Y = (P_1 y_1, \cdots, P_p y_p) \tag{1.1}$$

only $\hat{Y} = PY$ is observed. Moreover, assuming that observations at different points are uncorrelated we have that

$$E\hat{Y} \in PL \equiv \hat{L} \subseteqq PV, \qquad \operatorname{Cov} \hat{Y} \in \{P(I \otimes \Sigma)P'\},$$

i.e., in view of that fact that $(I \otimes \Sigma)A$ is defined to be equal to $A\Sigma$, we have $(\operatorname{Cov} \hat{Y})A = P((P'A)\Sigma)$ for some positive semidefinite (PSD) matrix Σ of order $p \times p$.

Again, the question arises under what conditions on L and hence $\hat{L}$ a GME which is independent of Σ exists. To answer this question, in Section 3 it is first assumed that Σ runs over all PSD-matrices which are diagonal. In this case it turns out that necessary and sufficient for the existence of a GME is that $\hat{L}$ is of the form (see Theorem 3.2)

$$PV \cap \hat{L} = L_1 \times \cdots \times L_p, \tag{1.2}$$

where the L_i are well-defined linear subspaces of R^n and again $A \in L_1 \times \cdots \times L_p$ holds if and only if the ith column of A belongs to L_i and this should be true for all indices $i = 1, 2, \cdots, p$. If Σ is allowed to run over all PSD-matrices, then some additional conditions must be added to (1.2), namely (see Theorem 3.3 and Remark 3.4):

$$P_i P_j(L_j) \subseteqq L_i \tag{1.3}$$

for $i, j = 1, 2, \cdots, p$.

In the sequel we will make use of the following notation: If M_α, $\alpha \in A$ are subsets of the Euclidian vector space V then $\sum_{\alpha \in A} M_\alpha$ denotes the smallest linear subspace $V_0 \subseteqq V$ such that $M_\alpha \subseteqq V_0$ for all $\alpha \in A$. We denote by $N^\perp$ the orthogonal complement of a subset N of a given vector space. The relations $(\sum_{\alpha \in A} M_\alpha)^\perp = \bigcap_{\alpha \in A} M_\alpha{}^\perp$ and $(\bigcap_{\alpha \in A} M_\alpha)^\perp = \sum_{\alpha \in A} M_\alpha{}^\perp$ are then easy to prove if all M_α are linear subspaces of the Euclidian vector space V. If $X\colon \Theta \to V$ is a linear mapping, Θ, V being Euclidian vector spaces, then we will several times in this paper use the Farkas' theorem (see, e.g., Drygas [1, page 288]) $(X\Theta_0)^\perp = X'^{-1}(\Theta_0{}^\perp)$ and $(X^{-1}(V_0))^\perp = X'V_0{}^\perp$ valid for linear subspaces $V_0 \subseteqq V$, $\Theta_0 \subseteqq \Theta$.

2. The existence of Gauss–Markov estimators. Let V be an Euclidian vector-space, endowed with the inner product $\langle\ ,\ \rangle$ and let (Ω, F, P) be a probability-space. Consider moreover a linear subspace $L \subseteqq V$ and a set Θ of symmetric

PSD mappings from V to V. We assume that (Ω, F, P) is so large that to any $Q \in \Theta$ there exists a V-valued random vector y such that $\operatorname{Cov} y = Q$. By the model $M(L, \Theta)$ we will then mean the set of all V-valued random vectors y such that $Ey \in L$, $\operatorname{Cov} y \in \Theta$. Let $G: V \to V$ be a linear mapping, then Gy is called Gauss–Markov estimator (GME) of Ey in the model $M(L, \Theta)$ if Gy is a best linear unbiased estimator (BLUE) of Ey in the model $M(L, \{Q\})$ for all $Q \in \Theta$. For a more detailed discussion of these concepts see Eaton [3, 4], Drygas [1, 2].

Eaton [3] has solved the problem of the existence of a GME in the model $M(L, \Theta)$ provided all $Q \in \Theta$ are positive definite. In a recent but still unpublished paper [4], Eaton has extended his results to the case where singular covariance matrices are allowed, too. He so follows Kruskal [5] who already had admitted the case of a singular covariance matrix.

In Drygas [2, pages 309–310] it has been shown that Gy is GME in the model $M(L, \{Q\})$ if and only if (i) $Ga = a \ \forall\, a \in L$ and (ii) $GQa = 0 \ \forall\, a \in L^\perp$. Since $0 = QL^\perp \cap L$ for any PSD Q, the existence of a GME is granted if Θ is a one element-set. Consequently if Θ is an arbitrary set of symmetric PSD-mappings, Gy is GME in the model $M(L, \Theta)$ if and only if (i) $Ga = a \ \forall\, a \in L$ and (ii) $Ga = 0 \ \forall\, a \in \sum_{Q \in \Theta} QL^\perp$. This proves part (a) of the following theorem:

2.1 Theorem. *Let the model $M(L, \Theta)$ be given. Then the following is true*:

(a) *A* GME *for this model exists if and only if*

$$L \cap \left(\textstyle\sum_{Q \in \Theta} QL^\perp\right) = 0 . \tag{2.1}$$

(b) *If a* GME *for the model $M(L, \Theta_0)$, $\Theta_0 \subseteqq \Theta$ exists, and moreover,*

$$QL^\perp \subseteqq \textstyle\sum_{Q \in \Theta_0} QV + L \quad \forall\, Q \in \Theta , \tag{2.2}$$

then there exists a GME *for the model $M(L, \Theta)$ if and only if*

$$\textstyle\bigcap_{Q \in \Theta_0} Q^{-1}(L) \subseteqq Q^{-1}(L) \quad \forall\, Q \in \Theta . \tag{2.3}$$

Proof. By the Farkas theorem, (2.3) is equivalent to $QL^\perp \subseteqq \sum_{Q \in \Theta_0} QL^\perp$ for all $Q \in \Theta$. Let a GME for the model $M(L, \Theta)$ exist and $z \in QL^\perp$, $Q \in \Theta$. Then by assumption (2.2) $Z = Z_1 + Z_2$; $Z_1 \in L$, $Z_2 \in \sum_{Q \in \Theta_0} QV$. But $\sum_{Q \in \Theta_0} QV \subseteqq L + \sum_{Q \in \Theta_0} QL^\perp$ since this by again using Farkas' theorem turns out to be equivalent to $\bigcap_{Q \in \Theta_0} Q^{-1}(L) \cap L^\perp \subseteqq \bigcap_{Q \in \Theta_0} Q^{-1}(0)$ which is trivially true. So $Z_2 \in \sum_{Q \in \Theta_0} QL^\perp$ can be assumed without any qualification. But $Z_1 = Z - Z_2 \in L \cap (\sum_{Q \in \Theta} QL^\perp) = 0$ since a GME for the model $M(L, \Theta)$ is assumed to exist. This implies $Z = Z_2 \in \sum_{Q \in \Theta_0} QL^\perp$. If on the other hand (2.3) is met and a GME for the model $M(L, \Theta_0)$ exists and, moreover, $Z \in (\sum_{Q \in \Theta} QL^\perp) \cap L$, then by (2.3) $Z \in (\sum_{Q \in \Theta_0} QL^\perp) \cap L = 0$ by part (a) of the theorem. So, again, by part (a) of the theorem, a GME for the model $M(L, \Theta)$ exists, too. $\square$

2.2 Example. Let y be a random vector of order $n \times 1$ such that $Ey = 0$, $\operatorname{Cov} y = \sigma_1^2 W_1 + \cdots + \sigma_m^2 W_m$, the W_i being PSD-matrices of order $n \times n$. This is a simple example of a variance-component model as, e.g., considered by Seely [7], [8]. In the context of estimating the σ_i^2 it has turned out to be convenient to consider the random $n \times n$ matrix $Y = yy'$ whose expectation is equal to $\sum_{i=1}^m \sigma_i^2 W_i$ and whose covariance-mapping under assumptions of normality (at least as far as the moments up to fourth order are concerned) is equal to

(2.4) $$\operatorname{Cov} Y = 2 \sum_{i,j=1}^m \sigma_i^2\sigma_j^2(W_i \otimes W_j)\,, \qquad \text{i.e.,}$$

(2.4a) $$\begin{aligned}\operatorname{Var}(\operatorname{tr}(A\cdot Y)) &= \operatorname{tr}((\operatorname{Cov} Y\cdot A)\cdot A)\\ &= 2\sum_{i,j=1}^m \sigma_i^2\sigma_j^2 \operatorname{tr}(W_i A W_j A)\,, \qquad A \text{ of order } n\times n\,.\end{aligned}$$

So here L is the linear space generated by $W_1, \cdots, W_m$ while Θ is the set of all mappings on the set V of $n \times n$-matrices which are of the form (2.4). Again the question may arise when a GME of EY (which of course is a quadratic estimator of y) exists. We show that $W_j W_k W_i + W_i W_k W_j \in L$ for $i, j, k = 1, 2, \cdots, m$, is a sufficient condition for the existence of a GME. Again this condition is met if $W \in L$ implies $W^2 \in L$ or if the W_i are idempotent and pairwise orthogonal, i.e., $W_i W_j = \delta_{ij} W_i$, $i = 1, 2, \cdots, m$. Indeed $W, V \in L$ then implies $(W + V)^2 - V^2 - W^2 = VW + WV \in L$ and thus

$$\begin{aligned}W_j W_k W_i + W_i W_k W_j &\\ = \tfrac{1}{2}[&W_j(W_i W_k + W_k W_i) + (W_i W_k + W_k W_i)W_j\\ &+ W_i(W_j W_k + W_k W_j) + (W_j W_k + W_k W_j)W_i\\ &- W_k(W_i W_j + W_j W_i) - (W_i W_j + W_j W_i)W_k] \in L\,.\end{aligned}$$

The proof of our assertion runs as follows: Let

$$Z \in L \cap \sum_{Q\in\Theta} QL^\perp\,, \qquad \text{i.e.,}$$

(2.5) $$Z = \sum_{i,j=1}^m b_{ij}(W_i \otimes W_j)A = \sum_{i,j=1}^m b_{ij} W_i A W_j = \sum_{k=1}^m \gamma_k W_k\,,$$

where $A \in L^\perp$, i.e., $\operatorname{tr}(AW_i) = 0$, $i = 1, 2, \cdots, m$. Then, as $b_{ij} = b_{ji}$ without restricting generality,

(2.6) $$\begin{aligned}\operatorname{tr}(Z^2) &= \sum_{k,i,j} \operatorname{tr}(b_{ij} W_i A W_j \gamma_k W_k)\\ &= \sum_{k,i\leq j} \gamma_k b_{ij} \operatorname{tr}((W_j W_k W_i + W_i W_k W_j)A) = 0\,,\end{aligned}$$

since $W_j W_k W_i + W_i W_k W_j \in L$ and $A \in L^\perp$, so Z must vanish. □

2.3 Example. Let us consider the model $M(L, P\Theta_1 P')$ where $L \subseteqq V$, $P: V \to V$ is a linear mapping and Θ_1 is a set of symmetric PSD mappings on V such that $I \in \Theta_1$. Since always $(PQ_1P')(V) \subseteqq PV = PP'V$, Theorem 2.1 can be applied with $\Theta_0 = \{PIP'\} = \{PP'\}$. By Theorem 2.1, (2.3) a GME for the model $M(L, \Theta)$ exists if and only if

(2.7) $$\bigcap_{Q\in\Theta_0} Q^{-1}(L) = (PP')^{-1}(L) \subseteqq Q^{-1}(L) = (PQ_1P')^{-1}(L) \quad \forall\, Q_1 \in \Theta_1\,.$$

This again is the case if and only if

(2.8) $$(PP')^{-1}(L \cap PV) \subseteqq (PQ_1P')^{-1}(L \cap PV) \quad \forall\, Q_1 \in \Theta_1$$

for $(PQ_1P')^{-1}(L) = (PQ_1P')^{-1}(L \cap PV)$. Since the relation (2.8) is linear Q_1 it is enough to have the relation for a set of mappings Q_1 which generate Θ_1.

3. Multivariate linear models with missing observations. In order to treat the problem of the existence of a GME in a multivariate linear model with missing observations we will introduce some technical concepts. Let V_1 and V_2 be finite-dimensional Euclidian vector spaces with inner products $(\ ,\ )_1$ and $(\ ,\ )_2$, respectively and let $V = L(V_1, V_2)$ be the set of all linear mappings from V_2 to V_1. V is a vector space which becomes Euclidian by introducing the inner product $\langle A, B\rangle = \operatorname{tr}(AB')$.

Let $B = \{x_1, \cdots, x_p\}$ be a basis of the vector space V_2 and let M_i, $i = 1, 2, \cdots, p$ be subsets of V_1. We define the *Cartesian product* M of the sets M_i with respect to the basis B, in symbols $M = (M_1 \times \cdots \times M_p)_B$ as the set of all $A \in V$ such that $Ax_i \in M_i$ for $i = 1, 2, \cdots, p$. If $0 \in M_i$, $i = 1, 2, \cdots, p$ and B is an orthonormal basis (ON-basis), then it is readily verified that $((M_1 \times \cdots \times M_p)_B)^\perp = (M_1^\perp \times \cdots \times M_p^\perp)_B$. The property that an element $A \in V$ belongs to the Cartesian product $(M_1 \times \cdots \times M_p)_B$ depends in general on the choice of the basis B of V_2. If, however, $M_i = M_0$ and M_0 is a linear subspace of V_1, then the statement $A \in (M_1 \times \cdots \times M_p)_B$ is independent of the choice of the basis B of V_2.

Let P_i $(i = 1, 2, \cdots, p)$ be linear mappings from V_1 to V_1 and let $B = \{x_i; 1 \leqq i \leqq p\}$ be a basis of V_2. If $A \in V$, we define the mapping $P = \operatorname{diag}(P_i; 1 \leqq i \leqq p)$ or more briefly $P = \operatorname{diag}(P_i)$, from V to V by reference to $PAx_i = P_iAx_i$; $i = 1, 2, \cdots, p$. P is said to be a diagonal mapping. If $P_i = P_0$ for $i = 1, 2, \cdots, p$, then evidently $\operatorname{diag}(P_i) = P_0 \otimes I$, where I is the identity mapping of V_2. If B is an ON-basic of V_2, then $(\operatorname{diag}(P_i))' = \operatorname{diag}(P_i')$. Moreover, the relation $(\operatorname{diag}(P_i))(M_1 \times \cdots \times M_p)_B = (P_1M_1 \times \cdots \times P_pM_p)_B$ holds.

Let us now consider the models $M(L, P\tilde{\Theta}_1P')$, and $M(L, P\Theta_1P')$, where L is a linear subspace of V, $P = \operatorname{diag}(P_i)$ is a diaganal mapping with respect to some fixed ON-basis $B = \{x_i; i \leqq i \leqq p\}$ of V_2. Before defining the sets Θ_1 and Θ_1 let us call a mapping $\Sigma: V_2 \to V_2$ diagonal with respect to (w.r.t.) B if $\Sigma x_i = \lambda_i x_i$ for some $\lambda_i \in R$, $i = 1, 2, \cdots, p$. Then define:

$$\tilde{\Theta}_1 = \{Q_1 = I_{V_1} \otimes \Sigma : \Sigma \text{ PSD}, \ \Sigma \text{ diagonal w.r.t. } B\}\,; \tag{3.1}$$

$$\Theta_1 = \{Q_1 = I_{V_1} \otimes \Sigma : \Sigma : V_2 \to V_2, \ \text{PSD}\}\,. \tag{3.2}$$

The consideration of the models $M(L, P\tilde{\Theta}_1P')$ and $M(L, P\Theta_1P')$ is motivated by the problem of missing observations in multivariate linear models. In this case $B = \{l_i, 1 \leqq i \leqq p\}$ where l_i is the ith unit-vector of R^p and P has same additional properties. These properties, however, are not relevent at this stage of investigation, and they are not necessary to establish the general result. They will be introduced and used at a later stage of the paper.

The following remarks will be useful in the sequel: Let Δ_{ij} be linear mappings from V to V, which are defined by reference to:

$$\Delta_{ij}Ax_k = \delta_{ik}Ax_j + \delta_{jk}Ax_i\,. \tag{3.3}$$

In other words, if we represent V as matrices of order $n \times p$, then $\Delta_{ii} A$ is the matrix which doubles the ith column of A, while all other columns are annihilated. Moreover, for $i \neq j$ $\Delta_{ij} A$, is the matrix obtained from A by interchanging the ith and the jth column and by annihilating all other columns.

It is obvious that $\Delta_{ij}: V \to V$ is a symmetric mapping. Moreover, Δ_{ii} is PSD. If $i \neq j$, Δ_{ij} is in general not, but $\Delta_{ij} = (\lambda(I_{V_1} \otimes I_{V_2}) + \Delta_{ij}) - \lambda(I_{V_1} \otimes I_{V_2})$, $\lambda > 1$, which means that Δ_{ij} can be represented as the difference of two PSD-mappings belonging to Θ_1. Moreover, the Δ_{ii} $(i = 1, 2, \cdots, p)$ do not only belong to $\tilde{\Theta}_1$, they also form a generating system of $\tilde{\Theta}_1$, since $\frac{1}{2}(\sum_{i=1}^{p} \lambda_i \Delta_{ii}) = I_{V_1} \otimes A$, if A is diagonal and $Ax_i = \lambda_i x_i$. A similar result holds for Θ_1, Θ_1 is generated by Δ_{ij}, $i, j = 1, 2, \cdots$, since $\frac{1}{2}(\sum_{i,j=1}^{p} \sigma_{ij} \Delta_{ij}) = I \otimes \Sigma$, if $\Sigma x_i \doteq \sigma_{i1} x_1 + \cdots + \sigma_{ip} x_p$.

3.1 LEMMA. *$P\Delta_{ij} P' = \Delta_{ij} PP'$ if $i = j$.*

PROOF.

$$P\Delta_{ij} P' A x_k = \delta_{ik} P_k P_j' A x_j + \delta_{jk} P_k P_i' A x_i ,$$

while

$$\Delta_{ij} PP' A x_k = \delta_{ik} P_j P_j' A x_j + \delta_{jk} P_i P_i' A x_i .$$

If $k \neq i, j$ both expressions vanish and if $i = k$ and $j = k$ the two expressions are obviously identical. Note however that this in general is not true if $k = i$ or $k = j$ and $i \neq j$. □

Let us now define

$$L_i = \{ l \in V_1 : \exists\, A \in L \cap PV \wedge A x_i = 1 \} . \tag{3.4}$$

Evidently the L_i are linear subspaces (of V_1) because L is such and, moreover, $L \cap PV \subseteqq (L_1 \times \cdots \times L_p)_B$. After having defined L_i we can formulate our first theorem:

3.2 THEOREM. *A GME for the model $M(L, \tilde{\Theta})$ exists if and only if*

$$L \cap PV = (L_1 \times \cdots \times L_p)_B . \tag{3.5}$$

PROOF. Since $I \in \tilde{\Theta}_1$ we can apply the results of Example 2.3. By (2.8) a GME for the model in consideration exists if and only if

$$(PP')^{-1}(L \cap PV) \subseteqq (PQ_1P')^{-1}(L \cap PV) \quad \forall\, Q_1 \in \tilde{\Theta}_1 . \tag{3.6}$$

Since $\{\Delta_{ii}; i = 1, 2, \cdots, p\}$ is a generating system of Θ_1, (3.6) is equivalent to

$$(PP')^{-1}(L \cap PV) \subseteqq (P\Delta_{ii}P')^{-1}(L \cap PV) = (\Delta_{ii}PP')^{-1}(L \cap PV) \tag{3.7}$$

for $i = 1, 2, \cdots, p$. From this we want to infer that $(L_1 \times \cdots \times L_p)_B \subseteqq L \cap PV$. If $A \in (L_1 \times \cdots \times L_p)_B$, then $Ax_i = A_i x_i$; for some $A_i \in L \cap PV$. Let $A_i = PP'B_i$, $B_i \in V$. Then $B_i \in (PP)^{-1}(L \cap PV)$. Consequently by (3.7) $B_i \in (\Delta_{ii}PP')^{-1}(L \cap PV)$, i.e., $\Delta_{ii}PP'B_i = \Delta_{ii}A_i \in L \cap PV$. But evidently $\Delta_{ii}A = \Delta_{ii}A_i$ and $A = \Delta_{11}A + \cdots + \Delta_{pp}A \in L \cap PV$.

If on the other hand $L \cap PV = (L_1 \times \cdots \times L_p)_B$, then let $A_0 \in (PP')^{-1}(L \cap PV)$, i.e., $A = PP'A_0 \in L \cap PV$ and $\Delta_{ii}A \in (L_1 \times \cdots \times L_p)_B = L \cap PV$ for $\Delta_{ii}Ax_i = 2Ax_i \in L_i$. This shows $A_0 \in (\Delta_{ii}PP')^{-1}(L \cap PV)$. □

3.3 THEOREM. *A* GME *for the model* $M(L, \Theta)$ *exists if and only if*

(a) $L \cap PV = (L_1 \times \cdots \times L_p)_B$

and

(b) $P_i P_j'((P_j P_j')^{-1}(L_j)) \subseteqq L_i$; $i = 1, 2, \cdots, p$.

PROOF. Since $I \in \Theta_1$ again example 2.3 applies. Since $\{\Delta_{ij}(i, j = 1, \cdots, p)\}$ is a generating system of Θ_1, we get by (2.8)

$$\forall_{i,j}: (PP')^{-1}(L \cap PV) \subseteqq (P\Delta_{ij}P)^{-1}(L \cap PV) \tag{3.8}$$

as a necessary and sufficient condition for the existence of a GME. By a preceding theorem this is equivalent to

$$\forall_{i,j}: (PP')^{-1}((L_1 \times \cdots \times L_p)_B) \subseteqq (P\Delta_{ij}P')^{-1}((L_1 \times \cdots \times L_p)_B \tag{3.9}$$

and

$$L \cap PV = (L_1 \times \cdots \times L_p)_B .$$

This is still to be shown to be equivalent to condition (b) of the theorem. Therefore, first let (b) hold and $A \in (PP')^{-1}(L \cap PV)$. Since (a) already holds then $PP'Ax_k \in L_k$, i.e., $Ax_k \in (P_k P_k')^{-1}(L_k)$ holds for all k. Then $P\Delta_{ij}P'Ax_k = \delta_{ik}P_iP_k'Ax_j + \delta_{jk}P_jP_k'Ax_i \in L_k$ has to be shown. If $k \neq i, j$ this is trivial, since $0 \in K_k$. If $k = i$ or $k = j$ this means $P_jP_i'Ax_i \in L_j$ and $P_iP_j'Ax_j \in L_i$, respectively. But in view of (b) both conditions are true.

On the other hand if $l_j \in (P_jP_j')^{-1}(L_j)$ let $A \in V$ be arbitrary but such that $PP'A \in L$ and $Ax_j = l_j$ and so $P_jP_j'Ax_j = P_jP_j'l_j \in L_j$ e.g., A can be defined by reference to $Ax_k = \delta_{jk}l_j$. Then (3.9) implies $P\Delta_{ij}P'A \in L \cap PV$, in particular $P\Delta_{ij}P'Ax_i = P_iP_j'Ax_j \in L_i$, i.e., (b). □

3.4 REMARK. The condition (b) of the preceding theorem is considerably simplified if $P_i' = P_i^+$, i.e., $P_iP_i'P_i = P_i$ and $L \subseteqq PV$, i.e., $L_j \subseteqq P_j(V_1)$, $j = 1, 2, \cdots, p$. This condition is met for the model of missing observations, since then $P_i^2 = P_i' = P_i$. In this case $P_iP_j' = P_iP_j'P_jP_j'$ and it is easy to prove that $P_iP_j'((P_jP_j')^{-1}(L_j)) = P_iP_j'L_j$. Thus condition (b) of the preceding theorem then simplifies to $P_iP_j'L_j \subseteqq L_i$ for $i, j, = 1, 2, \cdots, p$.

3.5 REMARK. Theorem 3.2 and 3.3 give necessary and sufficient conditions for the existence of a GME. Moreover, however, the question arises how a GME can be actually determined if the conditions of the theorems are satisfied. It is the purpose of this remark to show how this actually can be done. Let us make the additional assumptions $P' = P^+$, $L = (L_1 \times \cdots \times L_p)_B \subseteqq PV \leftrightarrow L_i \subseteqq P_iV_1 \forall_i$. These assumptions are satisfied in the case of missing observations because then L appearing in our theorem has to be replaced by $PL \subseteqq PV$. If $P_iP_j'L_j \subseteqq L_i$, $i, j = 1, 2, \cdots, p$, determine linear mappings $G_i: V_1 \to V_1$ such that G_iy_i is GME of Ey_i in the model $M(L_i, \{I_{V_1}\})$ or in the model $M(L_i, \{P_iP_i'\})$. A simple choice of G_i may be P_{L_i}, the orthogonal projection on L_i. Then $(\mathrm{diag}\,(G_i))Y \equiv GY$, $Y \in V$ is GME of EY in the model $M((L_1 \times \cdots \times L_p)_B, \Theta)$

and by Drygas [2, page 309].

(3.10) $$\operatorname{Cov}(GY) = G\operatorname{Cov}(Y) = (\operatorname{diag}(G_iP_i))((I \otimes \Sigma)P'),$$ i.e.,

(3.10a) $$(\operatorname{Cov} GY)Ax_i = \sigma_{i1}G_iP_iP_j'Ax_1 + \cdots \sigma_{ip}G_iP_iP_p'Ax_p,$$

$i = 1, 2, \cdots, p$. The GME-property is easy to prove since $((L_1 \times \cdots \times L_p)_B)^\perp = ((L_1^\perp \times \cdots \times L_p^\perp))_B$ and $P_iP_j'L_j \subseteqq L_j$ implies that $P_jP_i'L_i^\perp \subseteqq L_j^\perp$. This result means that the multivariate model is decomposed into p univariate models. A similar result can also be obtained without the assumptions $L \subseteqq PV$, $P' = P^+$.

3.6 Example. The conditions (a) and (b) practically reduce the problem to the case $p = 2$ since condition (b) of Theorem 3.4 is a condition on pairs of "coordinates." So let us consider this case. If we have two variables we have four kinds of points of observations: points at which both variables are observed, points at which exactly one variable is observed and finally points where both variables are not observed. The latter is uninteresting and can be neglected. By suitably arranging the observations the matrices P_1 and P_2 will be of the following form:

(3.11) $$P_1 = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad P_2 = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & I_3 \end{pmatrix},$$

where the I_i are unit matrices of appropriate orders. Then

(3.12) $$P_1P_2 = P_2P_1 = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Let us represent L_1 and L_2 in the form

(3.13) $$L_1 = \begin{pmatrix} Z_{11} \\ Z_{12} \\ 0 \end{pmatrix} \theta_1, \qquad L_2 = \begin{pmatrix} Z_{21} \\ 0 \\ Z_{23} \end{pmatrix} \theta_2,$$

where Z_{ij} is a matrix of order $n_j \times k_i$ and $\theta_i = R^{k_i}$; $j = 1, 2, 3$ and $i = 1, 2$. Then there exists a GME for the model $M(L_1 \times L_2, \Theta)$ iff

(3.14) $$\begin{pmatrix} Z_{11} \\ 0 \\ 0 \end{pmatrix} \theta_1 \subseteqq \begin{pmatrix} Z_{21} \\ 0 \\ Z_{23} \end{pmatrix} \theta_2 \quad \text{and} \quad \begin{pmatrix} Z_{21} \\ 0 \\ 0 \end{pmatrix} \theta_2 \subseteqq \begin{pmatrix} Z_{11} \\ Z_{12} \\ 0 \end{pmatrix} \theta_1,$$

i.e., to any $v_1 \in \theta_1$ there must be a $v_2 \in \theta_2$ such that $Z_{11}v_1 = Z_{21}v_2$ and $Z_{23}v_2 = 0$, and similarly for the second inclusion. If, moreover, $k_1 = k_2$ and $Z_{11} = Z_{21}$ (i.e., if a GME for the common observations of the two "endogeneous" variables exists), this is equivalent to $Z_{11}'R^{n_1} \cap Z_{23}'R^{n_3} = Z_{11}'R^{n_1} \cap Z_{12}'R^{n_2} = 0$ by the complementarity conditions, see [6, page 101]. If, e.g., Z_{23} has full column rank then (3.14) implies that $Z_{11} = 0$ (and similarly if Z_{12} has full column rank then $Z_{21} = 0$). In this case Gauss–Markov estimation is possible but the common observations are treated as if they were not made. In particular if there is no common observation, GM-estimation is always possible.

3.7 EXAMPLE. Let $k_1 = k_2 = 1$ and $Z_{ij} = (z_{ij}^{(k)})$; $k = 1, 2, \cdots, n_j$. By condition (3.14) we have to investigate the relation $z_{23}^{(k)} v_2 = 0$ for all k which implies that either $Z_{23} = 0$ or $Z_{11} = 0$. Similarly from $z_{12}^{(k)} V_2 = 0$ it follows that either $Z_{12} = 0$ or $Z_{21} = 0$. So a GME exists iff either the common observations or the observations belonging to one of the "endogeneous" variables are irrelevant.

4. Conclusion. If not all components of a multivariate random vector in a linear model are observed simultaneously then there is little hope that a GME of the expectation exists. Only in very special situations can an optimal estimator which is independent of the intercorrelations of the "endogeneous" variables be computed.

Acknowledgments. I am indebted to J. Kmenta, P. Schönfeld and M. Deistler for their suggestions and comments. Moreover, the helpful comments of the two referees have considerably improved the presentation of the paper.

REFERENCES

[1] DRYGAS, H. (1969). On a generalization of the Farkas theorem. *Unternehmensforschung* **13** 283–290.

[2] DRYGAS, H. (1971). Consistency of the least squares and Gauss-Markov estimators in regression models. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **17** 309–326.

[3] EATON, M. L. (1970). Gauss-Markov estimation for multivariate linear models: A coordinate-free approach. *Ann. Math. Statist.* **41** 528–538.

[4] EATON, M. L. (1972). A note on the Gauss-Markov theorem. Preprint of the *Institute of Mathematical Statistics, University of Copenhagen*, March.

[5] KRUSKAL, W. (1968). When are Gauss-Markov and least squares estimators identical? A coordinate-free approach. *Ann. Math. Statist.* **39** 70–75.

[6] SEBER, G. A. F. (1966). The linear hypothesis, a general theory. *Charles Griffins Statist. Monogr.* No. 19, London.

[7] SEELY, J. (1970). Linear spaces and unbiased estimation. *Ann. Math. Statist.* **41** 1725–1734.

[8] SEELY, J. (1970). Linear spaces and unbiased estimation-application to the mixed linear model. *Ann. Math. Statist.* **41** 1735–1744.

MATHEMATISCHES SEMINAR
INSTITUT FÜR ANGEWANDTE MATHEMATIK
6 FRANKFURT AM MAIN
ROBERT-MAYER-STRASSE 10
FEDERAL REPUBLIC OF GERMANY