Multivariate Analysis and Its Applications
IMS Lecture Notes – Monograph Series (1994) Volume 24

NORMAL LINEAR MODELS WITH LATTICE CONDITIONAL INDEPENDENCE RESTRICTIONS*

By Steen A. Andersson and Michael D. Perlman
Indiana University and University of Washington

It is shown that each multivariate normal model determined by lattice conditional independence (LCI) restrictions on the covariances may be extended in a natural way to a normal linear model with corresponding lattice restrictions on the means. For these extended models it remains true that the likelihood function (LF) and parameter space (PS) can be factored into the products of conditional LF's and PS's, respectively, each factor being the LF or PS of an ordinary multivariate normal linear regression model, from which maximum likelihood estimators and likelihood ratio test statistics are readily obtained. This extends the classical MANOVA and GMANOVA models, where the linear restrictions on the means are less general but where no restrictions are imposed on the covariances. It is shown how a collection of nonnested dependent normal linear regression models may be combined into a single linear model by imposing a parsimonious set of LCI restrictions.

1. Introduction. This paper is part of an ongoing study of the structure and analysis of multivariate normal statistical models defined by algebraic conditions on the means and/or covariances.

Because conditional independence (CI) plays an increasingly important role in statistical model building, it is of interest to study CI models with tractable statistical properties and to develop methods for testing one such model against another. Andersson and Perlman (1993a, b) have introduced a class of multivariate normal models defined by pairwise *lattice conditional*

AMS 1980 Subject Classifications: Primary 62H12, Secondary 62J05.

Key words and phrases: Distributive lattice, generalized block-triangular matrices, join-irreducible elements, likelihood function, likelihood ratio test, linear regression model, maximum likelihood estimator, multivariate normal distribution, nonnested regression models, pairwise conditional independence.

* Research supported in part by the Danish Research Council, by U.S. National Science Foundation Grant No. 89-02211, and by U.S. National Security Agency Grant No. MDA 904-92-H-3083.

independence (LCI) restrictions on the covariance structure.[1] The LCI model $\mathrm{N}(\mathcal{K})$ is defined to be the set of all nonsingular normal distributions $N(0,\Sigma)$ on $\mathbb{R}^I$ with mean 0 such that

$$x_L \perp\!\!\!\perp x_M \mid x_{L\cap M}, \qquad \forall\, L, M \in \mathcal{K}, \tag{1.1}$$

i.e., x_L and x_M are CI given $x_{L\cap M}$[2]. Here $\mathcal{K}$ is a ring of subsets (hence a finite distributive lattice) of the finite index set I such that $\phi, I \in \mathcal{K}$ and, for $K \in \mathcal{K}, x_K$ is the coordinate projection of $x \in \mathbb{R}^I$ onto $\mathbb{R}^K$. Let $\mathrm{P}(\mathcal{K})$ denote the set of all $I \times I$ positive definite covariance matrices Σ that satisfy the CI restrictions (1.1) determined by $\mathcal{K}$[3].

For the model $\mathrm{N}(\mathcal{K})$, the likelihood function (LF) and parameter space (PS) can be factored into the products of conditional LF's and PS's, respectively, each factor being the LF or PS of an ordinary multivariate normal linear regression model for which MLE's and LRT's may be obtained by standard linear methods. The products are indexed by $J(\mathcal{K})$, the poset of *join-irreducible* elements of $\mathcal{K}$. The collection of regression parameters is called the family of *$\mathcal{K}$-parameters* of the covariance matrix Σ, and these uniquely determine Σ when $\Sigma \in \mathrm{P}(\mathcal{K})$. These definitions and results will be reviewed in Section 2.

Example 1.1. Let $I = 123$ and $\mathcal{K} = \{\phi, 1, 12, 13, 123\}$ — cf. Figure 1.1. Under the LCI model $\mathrm{N}(\mathcal{K}), x_2 \perp\!\!\!\perp x_3 \mid x_1$. The join-irreducible elements of $\mathcal{K}$ are 1, 12, and 13, and the $\mathcal{K}$-parameters of the 3×3 covariance matrix Σ are given by

$$\left(\Sigma_{11},\ \Sigma_{21}\Sigma_{11}^{-1},\ \Sigma_{22\cdot 1},\ \Sigma_{31}\Sigma_{11}^{-1},\ \Sigma_{33\cdot 1}\right), \tag{1.2}$$

which uniquely determine Σ under the LCI model. If $x \sim N(0,\Sigma)$ then the density of x can be factored as follows:

$$f(x) = f(x_1)f(x_2 \mid x_1)f(x_3 \mid x_1), \tag{1.3}$$

where $f(x_2 \mid x_1)$ and $f(x_3 \mid x_1)$ are ordinary normal linear regression models, from which the MLE's of the $\mathcal{K}$-parameters in (1.1) easily can be derived. ∎

In the general case, many of the statistical properties of the LCI model $\mathrm{N}(\mathcal{K})$ are derived via the properties of an algebra $\mathrm{M}(\mathcal{K})$ of generalized block-triangular $I \times I$ matrices with additional structure determined by the lattice $\mathcal{K}$ (see Section 3). The group of nonsingular matrices in this algebra preserves

---

[1] The LCI restrictions arise naturally in the analysis of non-monotone multivariate missing data patterns — see Andersson and Perlman (1991).

[2] If $L \cap M = \phi$ then this condition reduces to $x_L \perp\!\!\!\perp x_M$, i.e., x_L and x_M are independent.

[3] When $\mathcal{K}$ is a chain, (1.1) is trivially satisfied and $\mathrm{P}(\mathcal{K}) = \mathrm{P}(I)$, the set of all $I \times I$ positive definite covariance matrices. In Andersson and Perlman (1991), $\mathrm{P}(\mathcal{K})$ and $\mathrm{M}(\mathcal{K})$ were denoted by $\mathrm{P}_{\mathcal{K}}(I)$ and $\mathrm{M}_{\mathcal{K}}(I)$, respectively.

and acts transitively on the class of all covariance matrices in the model $\mathrm{N}(\mathcal{K})$. For example, if $\mathcal{K}$ is the lattice in Figure 1.1 then $\mathrm{M}(\mathcal{K})$ consists of all 3×3 matrices of the form

$$A = \begin{pmatrix} a_{11} & 0 & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{pmatrix}. \tag{1.4}$$

Each LCI model $\mathrm{N}(\mathcal{K})$ is determined by CI restrictions on the covariances. In the present paper (see Section 4) we show that the LCI model consisting of n i.i.d. observations from $N(0, \Sigma) \in \mathrm{N}(\mathcal{K})$ may be extended in the following natural way to a normal linear model $\mathrm{N}(U, \mathcal{K})$, called a *$\mathcal{K}$-linear model*, determined by corresponding restrictions on the means as well.

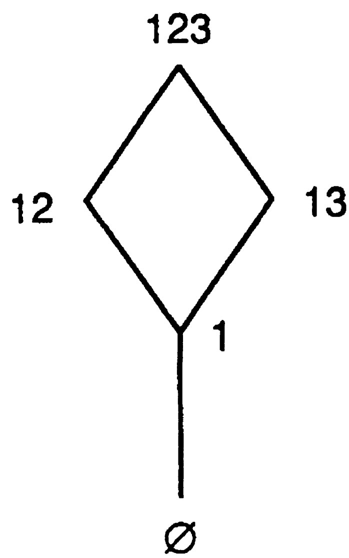

Figure 1.1
The lattice $\mathcal{K}$

The model $\mathrm{N}(U, \mathcal{K})$ consists of one observation of the $I \times N$ random matrix $y \sim N(\mu, \Sigma \otimes I_N)$, where $N = \{1, \cdots, n\}$ and 1_N is the $N \times N$ identity matrix. It is assumed that $\Sigma \in \mathrm{P}(\mathcal{K})$ and $\mu \in U$, where U is assumed to satisfy the condition $\mathrm{M}(\mathcal{K})U \subseteq U$, i.e., U is an *$\mathrm{M}(\mathcal{K})$-invariant linear subspace* (often abbreviated to *$\mathcal{K}$-subspace*) of the observation space $\mathrm{M}(I \times N)$, the set of all $I \times N$ matrices. This generalizes the classical MANOVA and GMANOVA models where $\mathcal{K}$ is a chain, $\mathrm{M}(\mathcal{K})$ is an algebra of block-triangular matrices in the usual sense, and Σ is unrestricted (see Andersson, Marden and Perlman (1994).

With $I = 123$ and $\mathcal{K} =$ the lattice in Figure 1.1, Figure 1.2 gives an example of a $\mathcal{K}$-subspace $U \subseteq \mathrm{M}(I \times N)$. In this example, unshaded regions indicate blocks of 0's while shaded regions indicate unrestricted blocks. With $\mathrm{M}(\mathcal{K})$ consisting of all matrices of the form (1.4), it is easily verified that the condition $\mathrm{M}(\mathcal{K})U \subseteq U$ is satisfied. Other examples of $\mathrm{M}(\mathcal{K})$-subspaces appear in Example 6.2.

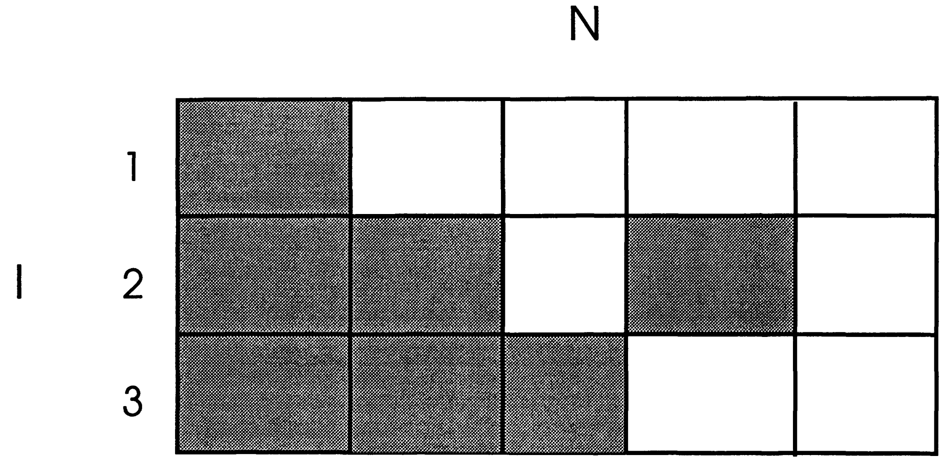

Figure 1.2
An $\mathrm{M}(\mathcal{K})$-invariant subspace U of $\mathrm{M}(I \times N)$

Of course, while the linear conditions imposed on the matrix of means μ by the $\mathcal{K}$-subspace U are more general that those that occur in MANOVA or GMANOVA (as seen in Figure 1.2), the $\mathcal{K}$-linear model $\mathrm{N}(U, \mathcal{K})$ imposes non-trivial LCI restrictions on the covariance Σ.

For the model $\mathrm{N}(U, \mathcal{K})$ it is again true (Section 5) that the LF and PS can be factored into the products (again indexed by $J(\mathcal{K})$) of conditional LF's and PS's, respectively, each factor being the LF or PS of an ordinary multivariate normal linear regression model, from which the MLE's are readily obtained. Furthermore, as will be shown elsewhere, it is then straightforward to obtain the likelihood ratio (LR) statistic for testing between two $\mathcal{K}$-subspaces $U_0 \subseteq U$ for the mean μ under the assumption $\Sigma \in \mathrm{P}(\mathcal{K})$. This generalizes the classical MANOVA and GMANOVA testing problems in several directions, yet should have a similarly tractable solution.

Lastly, in applications one may be presented with a collection of nonnested, possibly dependent normal linear regression models that determine a subspace U of $\mathrm{M}(I \times N)$. In Section 6 we show how to determine the minimal lattice $\mathcal{K}(U)$ such that these regression models can be combined into a single $\mathcal{K}(U)$-linear model $\mathrm{N}(U, \mathcal{K}(U))$. The procedure is similar to that in Andersson and Perlman (1991), where it is shown how a nonnested missing data pattern determines a minimal LCI model such that the likelihood function factors into a product of linear regression models.

2. Factorization of the Likelihood Function under the LCI Model $\mathrm{N}(\mathcal{K})$. For any finite index sets S and T let $\mathrm{M}(S \times T)$ denote the vector space of all real $S \times T$ matrices and $\mathrm{P}(S)$ the cone of all real positive definite $S \times S$ matrices. The factorizations of the LF and PS of the LCI model $\mathrm{N}(\mathcal{K})$ require that each $y \in \mathrm{M}(I \times N)$ and each $\Sigma \in \mathrm{P}(\mathcal{K}) \subseteq \mathrm{P}(I)$ be partitioned according to the poset $J(\mathcal{K})$ of join-irreducible elements of the lattice $\mathcal{K}$, whose definition we now review (see also Andersson and Perlman (1993a)).

For $K \in \mathcal{K},\ K \neq \phi$, define

$$\begin{aligned}\langle K\rangle &:= \cup(K' \in \mathcal{K} \mid K' \subset K),\\ [K] &:= K\backslash\langle K\rangle,\end{aligned}$$

where $\subset$ indicates strict inclusion, so that K is the disjoint union

$$K = \langle K\rangle\dot{\cup}[K], \tag{2.1}$$

The poset $J(\mathcal{K})$ of *join-irreducible* elements of $\mathcal{K}$ is then defined as follows:

$$J(\mathcal{K}) := \{K \in \mathcal{K} \mid K \neq \phi, \langle K\rangle \subset K\}.$$

Then every set $K \in \mathcal{K}$ may be decomposed as the disjoint union

$$K = \dot{\cup}([K'] \mid K' \in J(\mathcal{K}), K' \subseteq K) \tag{2.2}$$

(cf. Proposition 2.1 of [AP] (1993a)); in particular,

$$I = \dot{\cup}([K] \mid K \in J(\mathcal{K})). \tag{2.3}$$

Thus, each matrix $y \in M(I \times N)$ may be uniquely partitioned according to the decomposition (2.3) as

$$y = (y_{[K]} \mid K \in J(\mathcal{K})), \tag{2.4}$$

where, for any subset $S \subseteq I,\ y_S \in M(S \times N)$ is the $S \times N$ submatrix of y.

For any $\Sigma \in \mathbf{P}(I)$ and any $S \subseteq I$, let $\Sigma_S \in \mathbf{P}(S)$ denote the $S \times S$ submatrix of Σ and let Σ_S^{-1} denote $(\Sigma_S)^{-1}$. For $K \in \mathcal{K}$ partition Σ_K according to (2.1) as follows:

$$\Sigma_K = \begin{pmatrix} \Sigma_{\langle K\rangle} & \Sigma_{\langle K]} \\ \Sigma_{[K\rangle} & \Sigma_{[K]} \end{pmatrix}, \tag{2.5}$$

so $\Sigma_{\langle K\rangle} \in \mathrm{P}(\langle K\rangle), \Sigma_{[K]} \in \mathrm{P}([K]),\ \Sigma_{[K\rangle} \in \mathrm{M}([K] \times \langle K\rangle)$, and $\Sigma_{\langle K]} = (\Sigma_{[K\rangle})^t$.

Furthermore, define

$$\Sigma_{[K]\cdot} \equiv \Sigma_{[K]\cdot\langle K\rangle} := \Sigma_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}\Sigma_{\langle K]} \in \mathbf{P}([K]) \tag{2.6}$$

and let $\Sigma_{[K]\cdot}^{-1}$ denote $(\Sigma_{[K]\cdot})^{-1}$. It follows from Theorem 2.1 (ii) of Andersson and Perlman (1993a) that for every $y \in \mathrm{M}(I \times N)$ and $\Sigma \in \mathrm{P}(\mathcal{K})$,

$$\operatorname{tr}(\Sigma^{-1}yy^t) = \sum(\operatorname{tr}(\Sigma_{[K]\cdot}^{-1}(y_{[K]} - \Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1}y_{\langle K\rangle})(\cdots)^t) \mid K \in J(\mathcal{K})). \tag{2.7}$$

For $\Sigma \in \mathbf{P}(I)$, the family of matrices

$$((\Sigma_{[K\rangle}\Sigma_{\langle K\rangle}^{-1},\ \Sigma_{[K]\cdot}) \mid K \in J(\mathcal{K})) =: ((R_{[K\rangle},\ \Lambda_{[K]}) \mid K \in J(\mathcal{K})) \tag{2.8}$$

is the family of *$\mathcal{K}$-parameters* of Σ. By Theorem 2.2 of Andersson and Perlman (1993a) the mapping

$$\begin{aligned} \mathrm{P}(\mathcal{K}) &\to \times\, (\mathrm{M}([K] \times \langle K\rangle) \times \mathrm{P}([K]) \mid K \in J(\mathcal{K})) \\ \Sigma &\to \quad ((R_{[K\rangle}, \Lambda_{[K]}) \mid K \in J(\mathcal{K})) \end{aligned} \tag{2.9}$$

is a bijection, hence represents the parameter space $\mathrm{P}(\mathcal{K})$ as a product of the spaces of the $\mathcal{K}$-parameters. Thus, every $\Sigma \in \mathrm{P}(\mathcal{K})$ is uniquely determined by its $\mathcal{K}$-parameters.

Now consider n independent, identically distributed (i.i.d.) observations $x_1, \cdots, x_n$ from the LCI model $\mathrm{N}(\mathcal{K})$ and denote the matrix of observations by y, i.e.,

$$y := (x_1, \cdots, x_n) \in \mathrm{M}(I \times N), \tag{2.10}$$

where $N = \{1, \cdots, n\}$. For $K \in J(\mathcal{K})$ partition y_K according to (2.1) as follows:

$$y_K = \begin{pmatrix} y_{\langle K\rangle} \\ y_{[K]} \end{pmatrix}. \tag{2.11}$$

By (2.7) and Lemma 2.5 of Andersson and Perlman (1993a), the likelihood function for the model $\mathrm{N}(\mathcal{K})$ has the following factorization:

$$\begin{aligned} \mathrm{P}(\mathcal{K}) \times \mathrm{M}(I \times N) &\to (0, \infty) \\ (\Sigma,\ y) &\to (\det(\Sigma))^{-n/2} \cdot \exp(-\mathrm{tr}(\Sigma^{-1}yy^t)/2) \\ = \Pi((\det(\Lambda_{[K]}))^{-n/2} &\cdot \exp(-\mathrm{tr}(\Lambda_{[K]}^{-1}(y_{[K]} - R_{[K\rangle}y_{\langle K\rangle})(\cdots)^t)/2) \mid K \in J(\mathcal{K})). \end{aligned} \tag{2.12}$$

Note that the K-th factor in this product is the density for the conditional distribution of $y_{[K]}$ given $y_{\langle K\rangle}$, which is that of a standard normal linear regression model. Furthermore, the parameter space $\mathrm{P}(\mathcal{K})$ has the factorization (2.9). It follows that the MLE $\widehat{\Sigma}(y)$ exists and is unique for a.e. y iff

$$n \geq \max\{|\langle K\rangle| + |[K]| \mid K \in J(\mathcal{K})\} \equiv \max\{|K| \mid K \in J(\mathcal{K})\}, \tag{2.13}$$

where $|K|$ denotes the number of elements in K. In this case the $\mathcal{K}$-parameters of $\widehat{\Sigma}(y)$ are determined explicitly from the usual formulas for regression estimators (Andersson and Perlman (1993a), §3.1), then $\widehat{\Sigma}(y)$ itself may be reconstructed from these $\mathcal{K}$-parameters by the algorithm in Andersson and Perlman (1993a), §2.7.

3. Generalized Block-triangular Matrices with Lattice Structure. For any matrix $A \in \mathrm{M}(I) := \mathrm{M}(I \times I)$ and any two subsets $L, M \in$

$J(\mathcal{K})$, let $A_{[LM]}$ denote the $[L] \times [M]$ submatrix of A. Thus each $A \in \mathrm{M}(I)$ may be partitioned into blocks according to the decomposition (2.3) as

$$A = (A_{[LM]} \mid L, M \in J(\mathcal{K})), \tag{3.1}$$

By Proposition 2.3 of Andersson and Perlman (1993a), for any $A \in \mathrm{M}(I)$ the following three conditions on A are equivalent:

(i) $\forall\, x \in \mathbb{R}^I, \forall\, L \in \mathcal{K} : x_L = 0 \Rightarrow\ (Ax)_L = 0;$

(ii) $\forall\, x \in \mathbb{R}^I, \forall\, L \in \mathcal{K} : (Ax)_L = A_L x_L;$

(iii) $\forall\, L, M \in J(\mathcal{K}) :\ M \not\subseteq L \Rightarrow A_{[LM]} = 0.$

Let $\mathrm{M}(\mathcal{K})$ denote the set of all $A \in \mathrm{M}(I)$ that satisfy the equivalent conditions (i), (ii), (iii) (also see Footnote 3). By (i), $\mathrm{M}(\mathcal{K})$ is a matrix algebra, i.e., is closed under addition and multiplication, and contains the identity 1_I. If $\mathcal{K}$ is a chain of subsets of I (i.e., $\mathcal{K}$ is totally ordered under inclusion), then by (3.1) and (iii), $\mathrm{M}(\mathcal{K})$ is just the algebra of all lower block-triangular matrices. For general $\mathcal{K}$, however, $\mathrm{M}(\mathcal{K})$ must satisfy additional restrictions determined by (i)–(iii) (e.g., see (1.4)).

For $K \in \mathcal{K}$ and $A \in \mathrm{M}(I)$ let A_K denote the $K \times K$ submatrix of A and partition A according to (2.1) and (2.5) as follows:

$$A_K = \begin{pmatrix} A_{\langle K \rangle} & A_{\langle K]} \\ A_{[K \rangle} & A_{[K]} \end{pmatrix}; \tag{3.2}$$

note that $A_{[K]} = A_{[KK]}$ when $K \in J(\mathcal{K})$. By (ii), if $A \in \mathrm{M}(\mathcal{K})$ then $\forall\, K \in J(\mathcal{K}), \forall\, x \in \mathbb{R}^I$,

$$A_{\langle K]} = 0, \tag{3.3}$$

$$(Ax)_{[K]} = A_{[K]} x_{[K]} + A_{[K \rangle} x_{\langle K \rangle}. \tag{3.4}$$

Furthermore, it follows from (iii) that the linear mapping

$$\begin{gathered} \mathrm{M}(\mathcal{K}) \to \times\, (\mathrm{M}([K] \times \langle K \rangle) \times \mathrm{M}([K]) \mid K \in J(\mathcal{K})) \\ A \to ((A_{[K\rangle},\ A_{[K]}) \mid K \in J(\mathcal{K})) \end{gathered} \tag{3.5}$$

is a bijection.

4. Mean Value Hypotheses under LCI Restrictions. We now show how the LCI model consisting of n i.i.d. observations from $N(0, \Sigma) \in \mathrm{N}(\mathcal{K})$ may be extended to a $\mathcal{K}$-linear model $\mathrm{N}(U, \mathcal{K})$ with corresponding linear restrictions on the means.

A subspace $V \subseteq \mathrm{M}(S \times N)$ is called a *MANOVA subspace* if $\mathrm{M}(S)V \subseteq V$. Since $1_S \in \mathrm{M}(S)$, this is equivalent to the condition $\mathrm{M}(S)V = V$. Note that any subspace of the form

$$V = \{\beta Z \mid \beta \in \mathrm{M}(S \times T)\} \tag{4.1}$$

is a MANOVA subspace, where T is a finite index set and $Z \in \mathrm{M}(T \times N)$ is a fixed design matrix. It can be shown (cf. Andersson, Marden, and Perlman (1994)) that *every* MANOVA subspace V of $\mathrm{M}(S \times N)$ has the form (4.1) for some T and Z. It then follows that for any MANOVA subspace V there exists a unique projection matrix $P \in \mathrm{M}(N)$ (i.e., $P = P^2 = P^t$) such that $V = \mathrm{M}(S \times N)P$; note that $\dim(V) = |S| \operatorname{rank}(P) = |S| \operatorname{tr}(P)$. [$P$ is simply the projection onto the row space $\mathrm{Row}(Z)$ of Z.]

DEFINITION 4.1. *A subspace $U \subseteq \mathrm{M}(I \times N)$ is an $\mathrm{M}(\mathcal{K})$-invariant subspace ($\equiv \mathcal{K}$-subspace) if*

$$\mathrm{M}(\mathcal{K})U \subseteq U, \tag{4.2}$$

or, equivalently, if $\mathrm{M}(\mathcal{K})U = U$. The corresponding $\mathcal{K}$-linear model $\mathrm{N}(U, \mathcal{K})$ is defined as

$$\mathrm{N}(U,\mathcal{K}) = \{N(\mu, \Sigma \otimes 1_N) \mid \mu \in U,\, \Sigma \in \mathrm{P}(\mathcal{K})\}. \tag{4.3}$$

Thus, under the model $\mathrm{N}(U,\mathcal{K})$ we observe y as in (2.10), where the columns $x_1, \cdots, x_n$ are independent (but no longer identically distributed) normal random vectors with common covariance matrix $\Sigma \in \mathrm{M}(\mathcal{K})$, while $E(y) \equiv \mu \in U$.

For any subspace $U \subseteq \mathrm{M}(I \times N)$ and $K \in J(\mathcal{K})$ define the two subspaces (cf. (2.11))

$$\begin{aligned} U_{[K]} &= \{y_{[K]} \mid y \in U\} \subseteq \mathrm{M}([K] \times N), \\ U_{\langle K \rangle} &= \{y_{\langle K \rangle} \mid y \in U\} \subseteq \mathrm{M}(\langle K \rangle \times N). \end{aligned}$$

Thus we have the natural linear embedding

$$U \to \times (U_{[K]} \mid K \in J(\mathcal{K})), \tag{4.4}$$

$$y \to (y_{[K]} \mid K \in J(\mathcal{K})). \tag{4.5}$$

Note that $A \in \mathrm{M}(\mathcal{K})$ acts on $\times(U_{[K]} \mid K \in J(\mathcal{K}))$ in accordance with the partitionings (3.1) and (4.5). The following basic characterization of a $\mathcal{K}$-subspace can be derived from (3.3) and (3.4) (see Andersson, Marden, and Perlman (1994) for the case where $\mathcal{K}$ is a chain):

THEOREM 4.2. *A subspace $U \subseteq \mathrm{M}(I \times N)$ is a $\mathcal{K}$-subspace if and only if it satisfies the following three conditions:*

(i) $U = \times(U_{[K]} \mid K \in J(\mathcal{K}))$, *i.e.*, (4.4) *is a bijection*;

(ii) $\forall K \in J(\mathcal{K}), U_{[K]}$ *is a MANOVA subspace of* $\mathrm{M}([K] \times N)$;

(iii) $\forall K \in J(\mathcal{K})$, $\mathrm{M}([K] \times \langle K \rangle)U_{\langle K \rangle} \subseteq U_{[K]}$.

REMARK 4.3. By (2.2) with K replaced by $\langle K\rangle$, conditions (i), (ii), and (iii) in Theorem 4.2 are equivalent to conditions (i) and (iv);

(iv) $\forall K, K' \in J(\mathcal{K})$ with $K' \subseteq K$, $\mathrm{M}([K] \times [K'])U_{[K']} \subseteq U_{[K]}$.

REMARK 4.4. By (4.1) and by (2.2) with K replaced by $\langle K\rangle$, conditions (ii) and (iii) of Theorem 4.2 may stated in the following equivalent forms:

(ii)′ $\forall K \in J(\mathcal{K})$ there exists a finite index set T_K and a fixed design matrix $Z_K \in \mathrm{M}(T_K \times N)$ such that $U_{[K]} = \{\beta Z_K \mid \beta \in \mathrm{M}([K] \times T_K)\}$.

(iii)′ $\forall K, K' \in J(\mathcal{K})$ with $K' \subset K$, $\mathrm{Row}(Z_{K'}) \subseteq \mathrm{Row}(Z_K)$; i.e., $\exists W \equiv W(Z_K, Z_{K'}) \in \mathrm{M}(T_{K'} \times T_K)$ such that $Z_{K'} = WZ_K$.

5. Factorization of the likelihood function under the $\mathcal{K}$-linear model $\mathrm{N}(U,\mathcal{K})$.

For any $y, \mu \in \mathrm{M}(I \times N)$ and $\Sigma \in \mathrm{P}(\mathcal{K})$, extend the identity (2.7) as follows:

$$\operatorname{tr}(\Sigma^{-1}(y-\mu)(y-\mu)^t) =$$
$$\sum(\operatorname{tr}(\Sigma^{-1}_{[K]\cdot}(y_{[K]} - \mu_{[K]} - \Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}(y_{\langle K\rangle} - \mu_{\langle K\rangle}))(\cdots)^t) \mid K \in J(\mathcal{K})). \quad (5.1)$$

For any pair $(\mu, \Sigma) \in \mathrm{M}(I \times N) \times \mathrm{P}(I)$, the family of matrices

$$\begin{aligned}&((\mu_{[K]} - \Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle}\mu_{\langle K\rangle},\ \Sigma_{[K\rangle}\Sigma^{-1}_{\langle K\rangle},\ \Sigma_{[K]\cdot}) \mid K \in J(\mathcal{K}))\\ &\quad =: ((\xi_{[K]},\ R_{[K\rangle},\ \Lambda_{[K]}) \mid K \in J(\mathcal{K}))\end{aligned} \quad (5.2)$$

is called the family of $\mathcal{K}$-*parameters* of (μ, Σ). From (2.9) and Theorem 4.2 (i) and (iii) we obtain the following factorization of the parameter space of the $\mathcal{K}$-linear model $\mathrm{N}(U,\mathcal{K})$.

THEOREM 5.1. *Let U be a $\mathcal{K}$-subspace of $\mathrm{M}(I \times N)$. Then the mapping*

$$\begin{array}{rcl} U \times \mathrm{P}(\mathcal{K}) & \to & \times(U_{[K]} \times \mathrm{M}([K] \times \langle K\rangle) \times \mathrm{P}([K]) \mid K \in J(\mathcal{K}))\\ (\mu, \Sigma) & \to & ((\xi_{[K]}, R_{[K\rangle}, \Lambda_{[K]}) \mid K \in J(\mathcal{K}))\end{array} \quad (5.3)$$

is a bijection, hence represents the parameter space $U \times \mathrm{P}(\mathcal{K})$ of the $\mathcal{K}$-linear model $\mathrm{N}(U,\mathcal{K})$ as a product of the spaces of the $\mathcal{K}$-parameters. Thus, every pair $(\mu,\Sigma) \in U \times \mathrm{P}(\mathcal{K})$ is uniquely determined by its $\mathcal{K}$-parameters. ∎

If U is a $\mathcal{K}$-subspace of $\mathrm{M}(I \times N)$, for each $K \in J(\mathcal{K})$ let $P_K \in \mathrm{M}(N)$ denote the projection matrix corresponding to the MANOVA subspace $U_{[K]}$ and set $Q_K := 1_N - P_K$. By (iii) of Theorem 4.2, $\mu \in U \Rightarrow \xi_{[K]} \in U_{[K]}$. Thus, by the orthogonality of P_K and Q_K,

$$\begin{aligned}&\operatorname{tr}(\Lambda^{-1}_{[K]}(y_{[K]} - \xi_{[K]} - R_{[K\rangle}y_{\langle K\rangle})(\cdots)^t)\\ &= \operatorname{tr}(\Lambda^{-1}_{[K]}(y_{[K]}P_K - \xi_{[K]} - R_{[K\rangle}y_{\langle K\rangle}P_K)(\cdots)^t)\\ &\quad + \operatorname{tr}(\Lambda^{-1}_{[K]}(y_{[K]}Q_K - R_{[K\rangle}y_{\langle K\rangle}Q_K)(\cdots)^t).\end{aligned} \quad (5.4)$$

Then (5.1), (5.4), and Lemma 2.5 of Andersson and Perlman (1993a) yield the following basic result:

THEOREM 5.2. *The likelihood function for the $\mathcal{K}$-linear model* $\mathrm{N}(U,\mathcal{K})$ *has the following factorization:*

$$\begin{aligned}(U\times \mathrm{P}(\mathcal{K}))\times \mathrm{M}(I\times N) &\rightarrow (0,\infty)\\ ((\mu,\Sigma),y) &\rightarrow (\det(\Sigma))^{-n/2}\cdot\exp(-\mathrm{tr}(\Sigma^{-1}(y-\mu)(\cdots)^t)/2)\\ =\Pi((\det(\Lambda_{[K]}))^{-n/2}&\cdot\exp(-\mathrm{tr}(\Lambda_{[K]}^{-1}(y_{[K]}P_K-\xi_{[K]}-R_{[K\rangle}y_{\langle K\rangle}P_K)(\cdots)^t)/2)\\ &\cdot\exp(-\mathrm{tr}(\Lambda_{[K]}^{-1}(y_{[K]}Q_K-R_{[K\rangle}y_{\langle K\rangle}Q_K)(\cdots)^t)/2)\mid K\in J(\mathcal{K}))).\end{aligned} \tag{5.5}$$

∎

Once again the K-th factor in this product is the density for the conditional distribution of $y_{[K]}$ given $y_{\langle K\rangle}$ and is seen to be the LF of a standard multivariate normal linear regression model. Since the parameter space $U\times\mathrm{P}(\mathcal{K})$ has the factorization (5.3), it follows that the MLE $(\hat{\mu},\widehat{\Sigma})\equiv(\hat{\mu}(y),\widehat{\Sigma}(y))$ exists and is unique for a.e. y iff

$$n\geq\max\{|K|+p_K\mid K\in J(\mathcal{K})\}, \tag{5.6}$$

where $p_K:=\mathrm{tr}(P_K)=\dim(U_{[K]})/|[K]|$. In this case the $\mathcal{K}$-parameters $((\hat{\xi}_{[K]},\widehat{R}_{[K\rangle},\hat{\Lambda}_{[K]})\mid K\in J(\mathcal{K}))$ of $(\hat{\mu},\widehat{\Sigma})$ are determined from the usual formulas for regression estimators (compare to Andersson and Perlman (1993a), §3.1 and Andersson and Perlman (1991), §3.4):

$$\widehat{R}_{[K\rangle}=y_{[K]}Q_Ky^t_{\langle K\rangle}(y_{\langle K\rangle}Q_Ky^t_{\langle K\rangle})^{-1} \tag{5.7}$$

$$n\hat{\Lambda}_{[K]}=y_{[K]}Q_Ky^t_{[K]}-y_{[K]}Q_Ky^t_{\langle K\rangle}(y_{\langle K\rangle}Q_Ky^t_{\langle K\rangle})^{-1}y_{\langle K\rangle}Q_Ky^t_{[K]} \tag{5.8}$$

$$\hat{\xi}_{[K]}=y_{[K]}P_K-y_{[K]}Q_Ky^t_{\langle K\rangle}(y_{\langle K\rangle}Q_Ky^t_{\langle K\rangle})^{-1}y_{\langle K\rangle}P_K \tag{5.9}$$

Finally, $(\hat{\mu},\widehat{\Sigma})$ itself may be reconstructed from these $\mathcal{K}$-parameters by the algorithm given in Andersson and Perlman (1991), §3.3.

In general, the MLE $(\hat{\mu},\widehat{\Sigma})$, is not a sufficient statistic for the model $\mathrm{N}(U,\mathcal{K})$, which is a curved exponential family. Nevertheless, it will be shown elsewhere that for two $\mathcal{K}$-subspaces $U_0\subseteq U$, the LR statistic Q for testing H_0: $\mu\in U_0$ vs. H: $\mu\in U$ has a simple expression, its exact central distribution can be obtained in terms of its moments, and the asymptotic central distribution of $-2\log Q$ as $n\to\infty$ is χ^2 with $\dim(U)-\dim(U_0)$ degrees of freedom. The classical MANOVA testing problem is obtained when $\mathcal{K}=\{\phi,I\}$, while the GMANOVA testing problem is a testing problem of this form with $\mathcal{K}=\{\phi,K,I\}$, where $\phi\subset K\subset I$.

6. Construction of a $\mathcal{K}$-linear Regression Model. In applications, one may be presented with a collection of nonnested, possibly dependent normal linear regression models which together determine a linear subspace U of the observation space (= mean value space) $\mathrm{M}(I \times N)$ (see Example 6.2). We now show how to determine the minimal lattice $\mathcal{K}(U)$ such that these regression models can be combined into a single $\mathcal{K}(U)$-linear model $\mathrm{N}(U, \mathcal{K}(U))$. This model is parsimonious in the sense that if the given regression models can also be combined into a single $\mathcal{M}$-linear model $\mathrm{N}(U, \mathcal{M})$, then necessarily $\mathcal{K}(U) \subseteq \mathcal{M}$, hence $\mathrm{N}(U, \mathcal{K}(U))$ imposes a minimal set of conditional independence restrictions on the covariance structure.

As above, let $y \in \mathrm{M}(I \times N)$ be a matrix of observations, where I and N are finite index sets. Suppose that we are given a decomposition

$$I = \dot{\cup}(S \mid S \in \mathcal{S}), \tag{6.1}$$

where each $S \neq \phi$, and a corresponding family $(U_S \mid S \in \mathcal{S})$ of MANOVA subspaces, where $U_S \subseteq \mathrm{M}(S \times N)$. If we now set

$$U = \times(U_S \mid S \in \mathcal{S}) \subseteq \mathrm{M}(I \times N), \tag{6.2}$$

then we wish to determine $\mathcal{K}(U)$, the minimal ring of subsets of the finite index set I such that $\phi, I \in \mathcal{K}(U)$ and such that U is a $\mathcal{K}(U)$-subspace of $\mathrm{M}(I \times N)$.

Define the partial ordering[4] $\leq$ on $\mathcal{S}$ as follows:

$$S' \leq S \qquad \text{if} \quad \mathrm{M}(S \times S')U_{S'} \subseteq U_S. \tag{6.3}$$

A nonempty subset $\mathcal{H}$ of the poset $\mathcal{S}$ is called *hereditary* if $S \in \mathcal{H}$ and $S' \leq S$ imply that $S' \in \mathcal{H}$. Now define $\mathcal{K}(U)$ to be the set of all hereditary subsets of S together with the empty subset ϕ. It follows from the fundamental correspondence between the class of all finite posets and the class of all finite distributive lattices[5] that $\mathcal{K}(U)$ is a finite distributive lattice such that

$$J(\mathcal{K}(U)) = \{K_S \mid S \in \mathcal{S}\}, \tag{6.4}$$

where

$$K_S = \dot{\cup}(S' \in \mathcal{S} \mid S' \leq S). \tag{6.5}$$

---

[4] **Clearly, $\leq$ is reflexive and transitive on $\mathcal{S}$. we may also assume that $\leq$ is antisymmetric —-otherwise replace $\mathcal{S}$ by the set of equivalence classes induced by the equivalence relation $S \sim S'$ if $S \leq S'$ and $S' \leq S$. Note that when each U_S is defined by a design matrix Z_S as in (6.10), it follows by (6.11) that $S \sim S'$ iff $\mathrm{Row}(Z_S) = \mathrm{Row}(Z_{S'})$.**

[5] **See Grätzer (1978, pp.61–62), Theorem 9 and Corollary 10, or Andersson (1990), Theorem 3.2 (ii).**

Furthermore, it is readily verified from (6.5) that

$$[K_S] = S. \tag{6.6}$$

Also, by (6.1) each member of $\mathcal{K}(U)$ is identified with a subset of I, so that $\mathcal{K}(U)$ may be considered as a ring of subsets of I.

By applying Theorem 4.2, Remark 4.3, (6.3), and the fact that $K_{S'} \subseteq K_S$ iff $S' \leq S$, it can be shown that U is a $\mathcal{K}(U)$-subspace of $\mathrm{M}(I \times N)$. Define

$$\mathbb{K} = \{\mathcal{K} \subseteq 2^I \mid \mathcal{K} \text{ is a ring, } U \text{ is a } \mathcal{K}\text{-subspace of } \mathrm{M}(I \times N)\}.$$

We now show that $\mathcal{K}(U)$ *is the unique minimal member of* $\mathbb{K}$.

First, suppose there exists $\mathcal{L} \in \mathbb{K}$ such that $\mathcal{L} \subset \mathcal{K}(U)$. Then there exists $L \in J(\mathcal{L})$ such that

$$[L] = \dot{\cup}(S \mid S \in \mathcal{D}), \tag{6.7}$$

where $\mathcal{D} \subseteq S$ and $|\mathcal{D}| \geq 2$. By (6.2) and (6.7),

$$U_{[L]} = \times(U_S \mid S \in \mathcal{D}). \tag{6.8}$$

Since $U_{[L]}$ is a MANOVA subspace, i.e., $\mathrm{M}([L])U_{[L]} \subseteq U_{[L]}$, it now may be shown that $S' \leq S$ for every pair $S, S' \in \mathcal{D}$. But this contradicts the assumption that $\leq$ is antisymmetric on S (see Footnote 4), hence $\mathcal{L} = \mathcal{K}(U)$, so $\mathcal{K}(U)$ is a minimal member of $\mathbb{K}$. Finally, the uniqueness of $\mathcal{K}(U)$ follows from the (non-trivial) relation

$$\mathrm{M}(\mathcal{K}_1 \cap \mathcal{K}_2) = \mathrm{Alg}\left(\mathrm{M}(\mathcal{K}_1), \mathrm{M}(\mathcal{K}_2)\right), \tag{6.9}$$

(the smallest algebra containing $\mathrm{M}(\mathcal{K}_1)$ and $\mathrm{M}(\mathcal{K}_2)$), valid for any rings $\mathcal{K}_1$ and $\mathcal{K}_2$ of subsets of I.

REMARK 6.1. In applications, $\mathcal{K}(U)$ is most easily determined as the ring generated by $\{K_S \mid S \in \mathcal{S}\}$ and ϕ, where K_S is given by (6.5) and (6.3). Also, each MANOVA subspace U_S usually is specified in the form

$$U_S = \{\beta Z_S \mid \beta \in \mathrm{M}(S \times T_S)\}, \tag{6.10}$$

where T_S is a finite index set and $Z_S \in \mathrm{M}(T_S \times N)$ is a fixed design matrix. In this case (6.3) can be expressed equivalently as

$$S' \leq S \quad \text{if} \quad \mathrm{Row}\,(Z_{S'}) \subseteq \mathrm{Row}\,(Z_S). \tag{6.11}$$

EXAMPLE 6.2. Take $I = 123$ (see Example 1.1) and denote the matrix of observations by

$$y \equiv \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} \in \mathrm{M}(I \times N).$$

Assume that the columns of y are independent and normally distributed with common unknown covariance matrix $\Sigma \in \mathrm{P}(I)$. Suppose that the row vectors y_1, y_2, y_3 are known to satisfy the following *nested* set of ordinary linear regression equations: for $j = 1, \cdots, n$,

$$E(y_{1j}) = \beta_{11} t_{1j}, \tag{6.12}$$

$$E(y_{2j}) = \beta_{21} t_{1j} + \beta_{22} t_{2j}, \tag{6.13}$$

$$E(y_{3j}) = \beta_{31} t_{1j} + \beta_{32} t_{2j} + \beta_{33} t_{3j}. \tag{6.14}$$

Here $\mathcal{S} = (S_1, S_2, S_3)$, where $S_i = i$, and the corresponding design matrices (cf. (6.10)) are

$$Z_{S_1} = (t_{11} \quad \cdots \quad t_{1n}), \tag{6.15}$$

$$Z_{S_2} = \begin{pmatrix} t_{11} & \cdots & t_{1n} \\ t_{21} & \cdots & t_{2n} \end{pmatrix}, \tag{6.16}$$

$$Z_{S_3} = \begin{pmatrix} t_{11} & \cdots & t_{1n} \\ t_{21} & \cdots & t_{2n} \\ t_{31} & \cdots & t_{3n} \end{pmatrix}. \tag{6.17}$$

We assume that the row vectors $(t_{\alpha 1} \cdots t_{\alpha n})$, $\alpha = 1, 2, 3$ are linearly independent. The relations (6.12), (6.13), and (6.14) determine a linear subspace $U \subseteq \mathrm{M}(I \times N)$ in which the matrix of means $E(y)$ is assumed to lie.

In order to combine the three (dependent) regression models (6.12)–(6.14) into a single $\mathcal{K}$-linear model by imposing a minimal set of LCI restrictions on Σ, apply (6.5) and (6.11) to find that

$$K_{S_1} = 1, \qquad K_{S_2} = 12, \qquad K_{S_3} = 123. \tag{6.18}$$

Thus $\{K_S \mid S \in \mathcal{S}\}$ is a chain, so $\mathcal{K}(U) = \{K_S \mid S \in \mathcal{S}\} \cup \{\phi\}$ (see Remark 6.1), which is also chain. It is easy to apply Theorem 4.2 and Remark 4.4 to verify that U is a $\mathcal{K}(U)$-subspace. The MLE's of the $\beta_{i\alpha}$ and Σ under the $\mathcal{K}(U)$-linear model $\mathrm{N}(U, \mathcal{K}(U))$ are obtained using (5.7)–(5.9). Note that in this example, as well as for any *nested* set of linear regression models, $\mathrm{P}(\mathcal{K}(U)) = \mathrm{P}(I)$, so no nontrivial CI restrictions are imposed on Σ under the model $\mathrm{N}(U, \mathcal{K}(U))$ (see Footnote 3).

Now suppose that (6.14) is replaced by

$$E(y_{3j}) = \beta_{31} t_{1j} + \beta_{33} t_{3j}; \tag{6.19}$$

equivalently, we add the assumption that $\beta_{32} = 0$. Here the three regression models determined by (6.12), (6.13) and (6.19) are *nonnested* and determine a proper subspace $U_0 \subset U$. We construct $\mathcal{K}(U_0)$ as above, but now (6.18) becomes

$$K_{S_1} = 1, \quad K_{S_2} = 12, \qquad K_{S_3} = 13, \tag{6.20}$$

so $\{K_S \mid S \in \mathcal{S}\}$ no longer is a chain. It is easily seen that $\mathcal{K}(U_0)$ is the lattice in Figure 1.1 and that U_0 is a $\mathcal{K}(U_0)$-subspace. The MLE's of the $\beta_{i\alpha}$ and Σ under the $\mathcal{K}(U_0)$-linear model $\mathrm{N}(U_0, \mathcal{K}(U_0))$ are again obtained using (5.7)–(5.9), where the projection matrices P_{123} and Q_{123} corresponding to Row (Z_{S_3}) in (6.17) must be replaced by P_{13} and Q_{13} corresponding to Row (Z_{S_3}) in (6.12):

$$Z_{S_3} = \begin{pmatrix} t_{11} & \cdots & t_{1n} \\ t_{31} & \cdots & t_{3n} \end{pmatrix}. \tag{6.21}$$

Since $\mathcal{K}(U_0)$ is not a chain, $\mathrm{P}(\mathcal{K}(U_0)) \subset \mathrm{P}(I)$, so a nontrivial CI restriction now must be imposed on Σ to combine the regression models given by (6.12), (6.13) and (6.19) into the $\mathcal{K}(U_0)$-linear model $\mathrm{N}(U_0, \mathcal{K}(U_0))$.

REFERENCES

ANDERSSON, S. A. (1990). The lattice structure of orthogonal linear models and orthogonal variance component models, *Scand. J. Statist.* **17**, 287–319.

ANDERSSON, S. A., MARDEN, J. I. and PERLMAN, M. D. (1994). Totally ordered normal linear models and testing problems. Monograph to appear in DMV-Seminar, Birkhäuser, Basel.

ANDERSSON, S. A. and PERLMAN, M. D. (1991). Lattice-ordered conditional independence models for missing data, *Statist. Prob. Letters* **12**, 465–486.

ANDERSSON, S.A. and PERLMAN, M.D. (1993a). Lattice models for conditional independence in a multivariate normal distribution. *Ann. Statist.* **21**, 1318–1358.

ANDERSSON, S.A. and PERLMAN, M.D. (1993b). Testing lattice conditional independence models, (Submitted for publication).

GRÄTZER, G. (1978). *General Lattice Theory.* Birkhäuser, Basel.

DEPARTMENT OF MATHEMATICS
INDIANA UNIVERSITY
BLOOMINGTON, IN 47405
USA

DEPARTMENT OF STATISTICS
UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195
USA