MIVQUE and Maximum Likelihood Estimation for Multivariate Linear Models with Incomplete Observations

David Causeur
Agrocampus Rennes, France

Abstract

The problem of estimating the parameters of multivariate linear models in the context of an arbitrary pattern of missing data is addressed in the present paper. While this problem is frequently handled by EM strategies, we propose a Gauss-Markov approach based on an initial linearization of the covariance of the model. A complete class of quadratic estimators is first exhibited in order to derive locally Minimum Variance Quadratic Unbiased Estimators (MIVQUE) of the variance parameters. Apart from the interest in locally MIVQUE itself, this approach gives more insight into maximum likelihood estimation. Indeed, an iterated version of MIVQUE is proposed as an alternative to EM to calculate the maximum likelihood estimators. Finally, MIVQUE and maximum likelihood estimation are compared by simulations.

Keywords and phrases. Incomplete observations, MIVQUE, multivariate linear models.

1 Introduction

Irrespective of the underlying model, estimation in the presence of incomplete observations has been addressed in the last four decades mainly through a large class of algorithmic tools, which are more or less derived from the EM algorithm primarily defined by Dempster et al. (1977). History of this popular class of algorithms, whose use goes beyond the scope of missing-data problems, is thoroughly reviewed by Meng and van Dyk (1997). A very extensive review of the most recent developments concerning the EM algorithms, especially dedicated to incomplete multivariate data, can be found in Schafer (1997). In some models with many parameters, the use of iterative optimization techniques to find the maximum likelihood estimators can however be cumbersome due to the complexity or eventually
the ridge form of the likelihood. In those situations, it can be crucial to use alternative estimation procedures at least to find out relevant starting points for the iterative algorithms.

In some traditional fields of applications for EM, such as the estimation of the variance components in mixed models, there are attractive alternative strategies that can essentially be seen as reformulations in the framework of quadratic estimation of the Gauss-Markov approach, usually presented in the context of linear estimation. Minimum Variance Quadratic Unbiased Estimation (MIVQUE), which is mainly developed by Rao (1970, 1971a, 1971b) and Rao (1977) for the estimation of variance components in mixed models, consists of finding the minimum variance estimator among all quadratic unbiased estimators of the variance components. Rao and Kleffe (1988) provide a detailed overview of the results obtained by MIVQUE approaches for the analysis of variance components.

While the EM approaches rely on the use of functional analysis theory to derive the maximum likelihood estimator (MLE), the developments induced by MIVQUE are based on a geometrical investigation of the covariance structure of the model, which is sometimes called the coordinate-free point of view. Though MIVQUE and MLE are technically different estimators, Harville (1977) has bridged the gap between the two by showing that the MLE could be seen as infinitely iterated versions of MIVQUE. This furthermore leads to novel algorithms to compute the MLE. As a difference between the EM and iterated MIVQUE approaches, it may be noted that the latter coordinate-free approach does not belong to the class of data augmentation algorithms, in which the imputation step can turn out to be sensitive. In our context, at each step, the iterated MIVQUE of the covariance is updated by solving a linear system which is provided explicitly later.

Some attempts to extend the debate between EM and MIVQUE to the context of incomplete multivariate data have emerged on the basis of Seely’s (1970, 1971) algebraic characterization of optimality in the theory of linear estimation. For instance, Drygas (1976) gives conditions for the existence of Gauss-Markov estimators for the expectation parameters of multivariate linear models. However, the complexity of the calculations has probably discouraged any further extension of Drygas’ (1976) approach towards more constructive theorems. The major aim of this paper is to show that explicit construction of MIVQUEs is possible in the special case of a multivariate linear model with an arbitrary pattern of missing data, which is assumed to be missing at random in the sense defined by Rubin (1976). Among the
main consequences of this result, a sequence of estimators converging to the MLE is also provided on the basis of Harville’s (1977) proposal for iterated versions of MIVQUE.

In Section 2, the paper introduces the theoretical background required for the main results obtained thereafter. The review in this section focuses on general theorems concerning MIVQUE which have mainly been used in the context of optimal estimation of variance components. In particular, the Gauss-Markov strategy for variance parameters is properly defined, and the concept of completeness of a class of quadratic estimators is introduced.

Section 3 is devoted to the presentation of the model in the presence of missing data. In particular, specific notations are introduced to deal with all the possible patterns of missing data, and a linearized version of the covariance is proposed as a basis for the upcoming developments.

In Section 4, Seely’s (1970, 1971) general results on the theory of linear estimation are applied to our problem to show that, as soon as one observation is incomplete, no uniformly MIVQUE can be found for any linear combination of the variance parameters. This result provides motivation for the search of locally MIVQUE.

The Gauss-Markov approach for the estimation of the variance parameters is presented in Section 5. The key result, which enables an explicit derivation of locally MIVQUE, is the characterization of a complete class of quadratic estimators. A general expression for the unbiased estimators of an arbitrary linear combination of the variance parameters in the complete class is deduced and the locally MIVQUE is provided.

Section 6 investigates the contributions of the Gauss-Markov approach for the derivation of the MLE. In particular, a new estimation algorithm is proposed. It extends the iterated version of MIVQUE, which was presented by Harville (1977) in the context of optimal estimation of variance components, to the missing data framework. Note that, in some cases of highly unbalanced pattern of missing data, the estimators obtained at each step of the iterative procedure may not be nonnegative definite, and hence are projected to the cone of nonnegative definite matrices, which may destroy their optimality property.

Finally, the MIVQUE and maximum likelihood strategies are compared in Section 7 by simulation studies.
2 Theoretical Background

Let $Y$ denote a random $n$-vector assumed to be normally distributed with null expectation and positive definite covariance matrix $V$. The parametrization of $V$ considered in the following is used by Rao and Kleffe (1988) as a linear structure of the covariance especially suited for mixed linear models. Without further reference to mixed linear models, let us assume that there exists known symmetric $n \times n$ matrices $V_k$, $k = 1, \ldots, r$ and an $r$-vector $\theta = (\theta_1, \theta_2, \ldots, \theta_r)'$ of unknown parameters called variance components such that $V = \sum_{k=1}^{r} \theta_k V_k$. Note that, due to the positive definiteness of $V$, the vector of variance components is restricted to the convex parametric set $\Theta = \{\theta \in \mathbb{R}^r, \sum_{k=1}^{r} \theta_k V_k \text{ is positive definite}\}$. Let $V$ stands for the linear subspace of the $n \times n$ symmetric matrices spanned by $V_k$, $k = 1, \ldots, r$. It will also be assumed without loss of generality for our purpose that $V$ contains the identity matrix $I_n$. Dealing with the estimation of linear combinations $l'\theta$, $l \in \mathbb{R}^r$, of the variance components, the Gauss-Markov strategy will then focus on estimators defined as quadratic forms $Y'AY$, where $A$ is a symmetric $n \times n$ matrix.

**Definition 2.1.** A quadratic estimator $Y'AY$ of $l'\theta$ ($l \in \mathbb{R}^r$), where $A$ is a symmetric $n \times n$ matrix, is said to be a locally [resp. uniformly] Minimum Variance Quadratic Unbiased Estimator (MIVQUE) if the conditions $(C_1)$ and $(C_2)$ [resp. $(C_2')$] are satisfied:

- $(C_1)$: $E_{\theta}(Y'AY) = l'\theta$, for all $\theta \in \Theta$.
- $(C_2)$: There exists $\theta_0 \in \Theta$ such that, for all $n \times n$ symmetric matrix $B$ with $E_{\theta}(Y'BY) = l'\theta$, for all $\theta \in \Theta$, $\text{Var}_{\theta_0}(Y'AY) \leq \text{Var}_{\theta_0}(Y'BY)$.
- $(C_2')$: For all $\theta_0 \in \Theta$, for all $n \times n$ symmetric matrix $B$ with $E_{\theta}(Y'BY) = l'\theta$, for all $\theta \in \Theta$, $\text{Var}_{\theta_0}(Y'AY) \leq \text{Var}_{\theta_0}(Y'BY)$.

In this context, the first problem that had been addressed, initially by Seely (1970, 1971), was the conditions on $V$ for the existence of uniformly MIVQUEs. In particular, in the Gauss-Markov approach, the introduction of the notions of quadratic subspaces of the linear space of $n \times n$ symmetric matrices has resulted in a geometrical characterization of the situations in which uniform optimality can be reached. As noted by Malley (1994),
these quadratic subspaces are in fact revisited versions of the special Jordan algebras in the particular framework of optimal estimation of variance components. The following definition of a special Jordan algebra serves our purpose.

**Definition 2.2.** A special Jordan algebra \( \mathcal{J} \) of the linear space of \( n \times n \) symmetric matrices is a linear subspace such that for all \( (A, B) \in \mathcal{J}^2 \), \( A \ast B = \frac{1}{2}(AB + BA) \in \mathcal{J} \). In addition, note that \( \ast \) is sometimes called the Jordan product.

A trivial example of a special Jordan algebra is the linear subspace spanned by \((I_n, J_n)\), where \( J_n \) is the \( n \times n \) matrix with all elements equal to 1. Less trivial examples can be found in Marin and Dhorne (2003), who study the Gauss-Markov estimation of variance components for autoregressive linear models.

The following theorem, due to Seely (1971), provides the conditions for the existence of uniformly optimal estimators of all the variance parameters.

**Theorem 2.1.** There exists a uniformly MIVQUE of \( l' \theta \), for all \( \theta \in \mathbb{R}^r \), if and only if \( \mathcal{V} \) is a special Jordan algebra of the linear space of \( n \times n \) symmetric matrices. Under this assumption, the random vector \((Y'V_1Y, Y'V_2Y, \ldots, Y'V_rY)\) forms a complete sufficient statistic for \( \theta \).

As a consequence, the uniformly MIVQUE coincides with the MLE of \( \theta \) when \( \mathcal{V} \) is closed under the Jordan product. The former key result obtained by Seely (1971) was exploited by Drygas (1976) to explore the conditions of existence of uniformly optimal estimators in multivariate linear models with incomplete observations.

When \( \mathcal{V} \) is not closed under the Jordan product, Kleffe (1977) gives more insight into the characterization of locally MIVQUE by introducing the concept of complete classes of quadratic estimators. This is a revisited version of the complete class of estimators presented by Ferguson (1967) in the general context of the decision theory.

**Definition 2.3.** Let \( \mathcal{C} \) denote a set of quadratic forms \( Y'AY \), where \( A \) is a symmetric \( n \times n \) matrix. If, for all linear combinations \( l' \theta \) of the variance components such that there exists an unbiased estimator of \( l' \theta \) in \( \mathcal{C} \), there exists an unbiased estimator in \( \mathcal{C} \) with a lower variance, then \( \mathcal{C} \) is called a complete class of estimators for the variance components.
The definition of completeness for classes of estimators aims therefore at restricting the search for optimal estimators to relevant subsets of the overall class of quadratic estimators. As an extension of Seely’s (1971) first results, Kleffe (1977) provides a characterization of the complete class of estimators for the variance components based on the algebraic properties of $V$.

**Theorem 2.2.** If $A$ is a special Jordan algebra of the linear space of symmetric $n \times n$ matrices containing $V$, then $\mathcal{C} = \{Y'AY, A \in A\}$ is a complete class of estimators for the variance components.

While Seely’s (1971) results were setting the algebraic conditions for the existence of uniformly MIVQUE, Kleffe’s (1977) preceding theorem gives more constructive elements for the derivation of locally MIVQUE.

### 3 Incomplete Observations in Multivariate Regression Models

Consider a random $q$-vector $(Y_1, Y_2, \ldots, Y_q)'$ assumed to be normally distributed with expectation $xB$, where $x$ is a $1 \times p$ matrix of observed predictors and $B = (\beta_1, \beta_2, \ldots, \beta_q)$ is a $p \times q$ matrix of unknown parameters, and with positive definite covariance matrix $\Sigma = (\sigma_{ij})_{1 \leq i,j \leq q}$. The following sections are devoted to the problem of inference on $\Sigma$ for an arbitrarily complex pattern of missing data, where every combination of observed and unobserved variables is potentially represented in the sample. If $\Lambda = (\lambda_{ij})_{1 \leq i,j \leq q}$ stands for an arbitrary $q \times q$ symmetric matrix, by analogy with the Gauss-Markov approach for variance parameters (see e.g., Rao and Kleffe, 1988), we will focus on linear combinations $tr(\Lambda \Sigma)$ of the variance parameters, where $tr(\cdot)$ stands for the usual trace operator.

**Remark 3.1.** As in the complete-case theory of linear models, estimation of variance parameters under the multivariate linear regression model is only modified, relative to the case $B = 0$, by the introduction of orthogonal projection kernels in the quadratic forms instead of identity kernels. Therefore, although the results are presented in the following in the simpler context of a null expectation, the general situation of a linear regression model is always mentioned as a direct corollary.

#### 3.1. Patterns of missing data

In order to define properly a pattern of missing data, it is convenient to index the set of possible combinations of observed variables.
Table 1. Two examples of pattern of missing data for $q = 3$. Missing values are indicated by the symbol ‘?‘, observed values by ‘□‘.

<table>
<thead>
<tr>
<th>Sampling item</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
<th>$Y_1$</th>
<th>$Y_2$</th>
<th>$Y_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>□</td>
<td>?</td>
<td>?</td>
<td>□</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>□</td>
<td>?</td>
<td>?</td>
<td>□</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>□</td>
<td>□</td>
<td>?</td>
<td>□</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>4</td>
<td>□</td>
<td>?</td>
<td>□</td>
<td>□</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>5</td>
<td>□</td>
<td>?</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>?</td>
</tr>
<tr>
<td>6</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>7</td>
<td>□</td>
<td>□</td>
<td>?</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>8</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>9</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
<tr>
<td>10</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
<td>□</td>
</tr>
</tbody>
</table>

Let $S_k = \{(i_1, \ldots, i_k), 1 \leq i_1 < i_2 < \ldots < i_k \leq q\}, k = 1, \ldots, q$, denote the set of $k$-combinations of the first $q$ non-null integers in ascending order, and $S = \bigcup_{k=1}^{q} S_k$. To be completely unambiguous in the definition of $S$, the vectors in $S$ are, up to now, ordered increasingly first relative to their length, and, for vectors with equal lengths, relative to the first element in $s$, then relative to the second element, and so on. For instance, if $q = 3$, $S$ will be described in its increasingly ordered form as follows:

$$S = \{(1) < (2) < (3) < (1, 2) < (1, 3) < (2, 3) < (1, 2, 3)\}.$$ 

In the following, for all $s$ in $S$, $c_s$ stands for the rank of $s$ relative to the former definition of an ordering relation on $S$.

A pattern of missing data is then completely described by the vector $n = (n_s)_{s\in S}$ of the smallest sample sizes $n_s$ for which the combination of variables with indices in $s$ is observed.

Examples. In the case $q = 3$, the left pattern of missing data displayed in Table 1 corresponds to $n = (2, 0, 2, 1, 3, 1, 1)'$. Special patterns of missing data, called monotone by Rubin (1974) because of their nested structure, were widely studied as these patterns enable a closed form expression of the MLE and also because they are useful to implement monotone versions of the data augmentation steps used in some EM algorithms. If, for all $s$ in $S$, $k_s$ denotes the number of elements in $s$, monotone patterns correspond to special configurations $n$ with

(i) $n_s = 0$ for all $s$ in $\hat{S}$ with $k_s = 1$, except one, say $n_{(1)} > 0$,

(ii) $n_s = 0$ for all $s$ in $S$ with $k_s = 2$, except one including $(1)$, say $n_{(1,2)} > 0$, and so on until the whole set of variables is observed. For instance, the right pattern of missing data in Table 1 is monotone and corresponds to $n = (3, 0, 0, 5, 0, 0, 2)$. 


Moreover, in order to ensure the estimability of the covariance parameters, it is assumed that, for all \( 1 \leq i \leq j \leq q \), the variables \( Y_i \) and \( Y_j \) are observed at least once together, i.e., \( \sum_{s \in S_{ij}} n_s > 0 \), where \( S_{ij} = \{ s \in S, \{ i, j \} \subseteq s \} \). In the context of the multivariate linear regression model, it is furthermore assumed that, for all \( 1 \leq i \leq j \leq q \), and \( s \in S_{ij} \) with \( n_s \neq 0 \), \( n_s \geq p \).

3.2. Settings for the Gauss-Markov estimation. For \( 1 \leq j \leq q \) and \( s \) in \( S \) with \( j \in s \), call \( Y_j^{(s)} \) the \( n_s \)-vector of observed value for \( Y_j \) on the smallest sample for which the variables with indices in \( s \) are observed. The usual sampling independence is assumed among elements in \( Y_j^{(s)} \).

Estimation strategies are therefore based on the \( n \times n \) matrix \( V \) of incomplete data, in ascending order of the ranks of \( s \) of incomplete data, in ascending order of the \( k_s \) indices of the variables. For instance, in the case \( q = 3 \), \( V \) is partitioned into 12 sub-vectors as follows:

\[
Y = \left( Y_1^{(n_1)^\prime}, Y_2^{(n_2)^\prime}, Y_3^{(n_3)^\prime}, Y_1^{(n_12)^\prime}, Y_1^{(n_13)^\prime}, Y_2^{(n_13)^\prime}, \ldots, Y_2^{(n_123)^\prime}, Y_3^{(n_123)^\prime} \right)^\prime.
\]

If, for some \( s \) in \( S \), \( n_s = 0 \), the \( k_s \) related sub-vectors \( Y_{j,n_s} \), \( j \in s \), are removed from the vector \( Y \) without consequences on the next developments.

The induced structure of the \( n \times n \) variance matrix \( V \) of \( Y \) is deduced in a straightforward manner from the former arbitrary partitioning of \( Y \) into \( \sum_{i=1}^q i^{(q)} = q2^{q-1} \) sub-vectors. Indeed, \( V \) is a block-diagonal matrix whose \( i \)th block, \( i = 1, \ldots, 2^q - 1 \) is \( \Sigma_{s_i} \otimes I_{n_s} \), where \( \Sigma_{s_i} \) is the \( k_{s_i} \times k_{s_i} \) sub-matrix of \( \Sigma \) containing covariances of the variables with indices in \( s_i \).

Now, for \( 1 \leq k \leq q \), let \( \mathcal{E}_k \) denote the set of \( k \times k \) symmetric matrices \( E_{lm}^{(k)} \), with \( 1 \leq l \leq m \leq k \), whose components \((l, m)\) and \((m, l)\) are equal to 1 and the other components are all null. In other words, the \( k(k+1)/2 \) matrices in \( \mathcal{E}_k \) form the canonical basis for the linear space of symmetric \( k \times k \) matrices. For \( s \) in \( S \), let \( \mathcal{V}_s \) denote the set of \( n \times n \) block-diagonal matrices \( V_{lm}^{(s)} \), with \( 1 \leq l \leq m \leq k_s \), whose \( i \)th diagonal block \( V_{lm}^{(s)}[i, i], i = 1, \ldots, 2^q - 1 \), is defined as follows: if \( i \neq c_s \) and \( c_s = i \) then \( V_{lm}^{(s)}[i, i] \) is the null \( k_{s_i,n_{s_i}} \times k_{s_i,n_{s_i}} \) matrix, and if \( i = c_s \), \( V_{lm}^{(s)}[i, i] = E_{lm}^{(k_s)} \otimes I_{n_s} \).
Using the preceding notations, the random $n$-vector $Y$ is normally distributed with zero expectation and positive definite covariance $V$ such that:

$$V = \sum_{i=1}^{q} \sum_{j=i}^{q} \sigma_{ij} V_{ij},$$

where, for $1 \leq i \leq j \leq q$, $V_{ij} = \sum_{s \in S_{ij}} V^{(s)}_{ij}$.

If $\sigma$ stands for the $q(q+1)/2$-vector whose generic element is $\sigma_{ij}$ for $1 \leq i \leq j \leq q$, note that, as a result of the positive definiteness of $\Sigma$, $\sigma$ is restricted to the convex parametric space

$$\Omega = \left\{ \sigma \in \mathbb{R}^{q(q+1)/2}, \sum_{i=1}^{q} \sum_{j=i}^{q} \sigma_{ij} V_{ij} \text{ is positive definite} \right\}.$$

**Remark 3.2.** Similarly, in the context of a multivariate linear regression model, the $n_s \times p$ matrix containing the $n_s$ observations of $x$ on the smallest sample for which the variables with indices in $s$ are observed is denoted $x^{(s)}$, and it is assumed that $\text{rank}(x^{(s)}) = p$. Moreover, for $s \in S$, call $X^{(s)}$ the $k_s n_s \times pq$ matrix partitioned into $n_s \times p$ blocks $X^{(s)}[i,j], i = 1, \ldots, k_s, j = 1, \ldots, q$ such that $X^{(s)}[i,j] = x^{(s)}$ if $j$ is the index of the $i$th element of $s$ and 0 elsewhere. Now call $X$ the $n \times pq$ matrix whose $i$th $k_s n_s \times pq$ submatrix is $X^{(s_i)}$, with $c_{s_i} = i$. With the former notations, if furthermore $\beta = (\beta_1', \beta_2', \ldots, \beta_q')'$, $\mathbb{E}(Y) = X \beta$.

For instance, in the case $q = 2$, the matrix $X$ is defined as follows:

$$X = \begin{bmatrix}
\begin{array}{ccc}
x^{(1)} & 0 & 0 \\
0 & x^{(2)} & 0 \\
x^{(12)} & 0 & x^{(12)}
\end{array}
\end{bmatrix}.$$

**4 Necessary and Sufficient Conditions for the Existence of Uniformly MIVQUE**

Let $\mathcal{V}$ denote the linear subspace spanned by $V_{ij}, 1 \leq i \leq j \leq q$. As a result of the theory recalled in Section 2, optimality of the estimation of the variance parameters can be investigated through the characterization of the algebraic properties of $\mathcal{V}$. These properties are directly inherited from
general rules for the derivations of the Jordan products between matrices \(V^{(s)}_{ij}\). The following lemma is therefore very useful for the proofs of the upcoming theorems.

**Lemma 4.1.** For all \((s, s')\) in \(S^2\) with \(s \neq s'\), for all \(1 \leq i \leq j \leq q\) and \(1 \leq k \leq l \leq q\), \(V^{(s)}_{ij} \ast V^{(s')}_{kl} = 0\). Further, for all \(s\) in \(S\), for all \(1 \leq i \leq j \leq q\) and \(1 \leq k \leq l \leq q\), with \(i \neq k\) and \(j \neq l\),

\[
V^{(s)}_{ij} \ast V^{(s)}_{kl} = \begin{cases} 
0 & \text{if } j \neq k \\
\frac{1}{2}V^{(s)}_{\min(i,l),\max(i,l)} & \text{if } j = k
\end{cases},
\]

\[
V^{(s)}_{ij} \ast V^{(s)}_{kj} = \frac{1}{2}V^{(s)}_{\min(i,k),\max(i,k)},
\]

\[
V^{(s)}_{ij} \ast V^{(s)}_{il} = \frac{1}{2}V^{(s)}_{\min(j,l),\max(j,l)},
\]

\[
V^{(s)}_{ij} \ast V^{(s)}_{ij} = \begin{cases} 
V^{(s)}_{ii} + V^{(s)}_{jj} & \text{if } i \neq j \\
V^{(s)}_{ii} & \text{if } i = j
\end{cases}.
\]

As a consequence of the preceding calculations, it can easily be checked that \(V\) is not closed under Jordan product. For instance, for all \(1 \leq i < j \leq q\), \(V_{ij} \ast V_{ij} = \sum_{s \in S_{ij}} V^{(s)}_{ij} + \sum_{s \in S_{ij}} V^{(s)}_{jj}\), which is generally not an element of \(V\). Consequently, according to Seely’s (1971) result recalled in Theorem 2.1, there generally does not exist uniformly MIVQUE for every linear combination of the variance parameters, which was also pointed out by Drygas (1976) and confirms our intuition.

However, the preceding calculations also enable a clear statement of the necessary and sufficient conditions on the pattern of missing data for the existence of uniformly MIVQUE for every linear combination of the variance parameter.

**Theorem 4.1.** For all symmetric \(q \times q\) matrix \(\Lambda\), there exists a uniformly MIVQUE of \(\theta = tr(\Lambda \Sigma)\) if and only if, for all \(s\) in \(S\), \(n_s = 0\), except for \(s = (1, 2, \ldots, q)\).

**Proof of Theorem 4.1.** Due to Seely (1971), a necessary and sufficient condition for the existence of uniformly MIVQUE for \(\theta = tr(\Lambda \Sigma)\), for all symmetric \(q \times q\) matrix \(\Lambda\), can be stated as the stability of \(V\) through the Jordan product, or equivalently:

\[
||P_{\mathcal{V}} V_{ij} \ast V_{kl}||^2 = 0, \text{ for all } 1 \leq i, j, k, l \leq q,
\]
where $|| \cdot ||^2$ is the canonical $L_2$-norm on the linear subspace of symmetric $n \times n$ matrices and $P_{V\perp}$ is the orthogonal projection operator on the linear subspace orthogonal to $V$. Moreover, Lemma 4.1 enables a closed form calculation of the former $L_2$-norms. For instance, for all $1 \leq i < j \leq q$,

$$
||P_{V\perp} V_{ij} \ast V_{ij}||^2 = 2 \sum_{s \in S_{ij}} n_s - \frac{[\sum_{s \in S_{ij}} n_s]^2}{\sum_{s \in S_{ij}} n_s} - \frac{[\sum_{s \in S_{ij}} n_s]^2}{\sum_{s \in S_{ij}} n_s}.
$$

Equating the preceding $L_2$-norms to 0 leads to the unique solution which is given in Theorem 4.1.

Therefore, as soon as one data is missing, no uniformly MIVQUE can be found for any linear combination of the variance parameters. The following sections are devoted to the characterization of locally MIVQUE through the derivation of a complete class of estimators, as introduced in Definition 2.3.

5 Gauss-Markov Estimation for the Variance Parameters

The first step in the Gauss-Markov approach consists of exhibiting a subclass of the set of quadratic estimators $Q = \{Y'AY, \text{ where } A \text{ is an } n \times n \text{ positive definite matrix}\}$, which is complete for the estimation of $\sigma$ in the sense of Definition 2.3. Once this complete class is properly defined, the next step focuses on the unbiased estimators of a particular linear combination $\theta$ of the variance parameters in the complete class and finally, among the unbiased estimators, the locally MIVQUE of $\theta$ is found.

5.1. Complete class of estimators. It has already been shown in Theorem 4.1 that $V$ is not closed under Jordan product. Therefore, according to Theorem 2.2, kernels of the quadratic forms in the complete class of estimators must be found in a linear subspace which contains $V$ and is closed under Jordan product.

**Theorem 5.1.** Let $Q_{ij}$, $1 \leq i \leq j \leq q$, denote the random vector with generic element $Q_{ij}^{(s)} = Y'V_{ij}^{(s)}Y$, $s \in S_{ij}$. Then

$$
C = \left\{ \sum_{1 \leq i \leq j \leq q} \gamma_{ij} Q_{ij}, \text{ where } \gamma_{ij} \in \mathbb{R}^{\text{card}S_{ij}} \right\}
$$

is a complete class of quadratic estimators for $\sigma$. 
Proof of Theorem 5.1. Closedness of the linear subspace spanned by \( \cup_{s \in S} V_s \) under Jordan product is immediate from Lemma 4.1. Therefore, this linear subspace, spanned by \( \left\{ V_{ij}^{(s)}, 1 \leq i \leq j \leq q, s \in S_{ij} \right\} \), contains obviously \( V \) and furthermore, is stable through the Jordan product. Finally, Theorem 5.1 follows directly from Theorem 2.2.

Note that an alternative expression for \( Q_{ij}^{(s)} \) is \( 2Y_i^{(s)j}Y_j^{(s)} \) for \( i < j \) and \( Y_i^{(s)j}Y_j^{(s)} \) for \( i = j \).

Remark 5.1. The extension of the former result to the multivariate linear model is easily obtained by considering that \( Q_{ij}^{(s)} = 2Y_i^{(s)j}(I_n - P^{(s)})Y_j^{(s)} \) if \( i < j \) and \( Q_{ij}^{(s)} = Q_{ii}^{(s)} = Y_i^{(s)j}(I_n - P^{(s)})Y_j^{(s)} \) if \( i = j \), where \( P^{(s)} = \hat{x}^{(s)}(\hat{x}^{(s)j}x^{(s)})^{-1}x^{(s)j} \) is the orthogonal projection matrix onto the linear subspace of \( \mathbb{R}^n_s \) spanned by the columns of \( x^{(s)} \).

Two illustrative examples are given below for more insight into the main results.

Example 5.1.1. The bivariate case. Suppose \( q = 2 \) and \( n_{(1)} > 0, n_{(2)} > 0, n_{(12)} > 0 \). Though apparently simple, Kotz et al. (2000) recall that many authors, since Anderson (1957), have proposed a sophisticated technique to solve the likelihood equations. We propose to examine this estimation problem with the Gauss-Markov approach. At this stage, Theorem 5.1 gives the complete class of estimators for \( \sigma = (\sigma_{11}, \sigma_{22}, \sigma_{12})' \):

\[
C = \left\{ \gamma_{11}Q_{11} + \gamma_{22}Q_{22} + \gamma_{12}Q_{12}, \gamma_{11} \in \mathbb{R}^2, \gamma_{22} \in \mathbb{R}^2, \gamma_{12} \in \mathbb{R} \right\},
\]

where \( Q_{11} = (Q_{11}^{(1)}, Q_{11}^{(12)})' \), \( Q_{22} = (Q_{22}^{(2)}, Q_{22}^{(12)})' \), \( Q_{12} = Q_{12}^{(12)} \), \( Q_{ii}^{(s)} = Y_i^{(s)}Y_i^{(s)} \) for all \( i = 1, 2 \) and \( s \in S_{ii} \), and \( Q_{12}^{(12)} = 2Y_1^{(n_{12})}Y_2^{(n_{12})} \).

Example 5.1.2. The monotone trivariate case. Suppose \( q = 3 \), \( n_s = 0 \) for all \( s \in \{(2), (3), (13), (23)\} \) and \( n_s > 0 \) for all \( s \in \{(1), (12), (13)\} \). This pattern corresponds to the given in the right part of Table 1. As recalled for instance by Schafer (1997), in the case of a monotone pattern of missing data, the MLEs for \( \sigma = (\sigma_{11}, \sigma_{22}, \sigma_{33}, \sigma_{12}, \sigma_{13}, \sigma_{23})' \) can be given explicitly. In this context, the complete class of estimators which is given by Theorem 5.1 is as follows:

\[
C = \left\{ \gamma_{11}Q_{11} + \gamma_{22}Q_{22} + \gamma_{33}Q_{33} + \gamma_{12}Q_{12} + \gamma_{13}Q_{13} + \gamma_{23}Q_{23}, \gamma_{11} \in \mathbb{R}^3, \gamma_{22}, \gamma_{12}, \gamma_{13}, \gamma_{23} \in \mathbb{R} \right\},
\]
where $Q_{11} = (Q_{11}^{(1)}, Q_{11}^{(12)}, Q_{11}^{(123)})'$, $Q_{22} = (Q_{22}^{(12)}, Q_{22}^{(123)})'$, $Q_{33} = Q_{33}^{(123)}$, $Q_{12} = (Q_{12}^{(12)}, Q_{12}^{(123)})'$, $Q_{13} = Q_{13}^{(123)}$ and $Q_{23} = Q_{23}^{(123)}$.

5.2. Unbiasedness. Let $\theta = \text{tr}(\Lambda \Sigma)$, where $\Lambda$ stands for a $q \times q$ symmetric matrix, denote an arbitrary linear combination of the variance parameters and $\hat{\theta} = \sum_{1 \leq i \leq j \leq q} \gamma_{ij} Q_{ij}$, an estimator of $\theta$ in the complete class $\mathcal{C}$. The estimator $\hat{\theta}$ is unbiased if, for all $\sigma$, $E(\hat{\theta}) = \theta$, which results in the following condition, according to the previous calculations:

$$
\sum_{1 \leq i \leq j \leq q} \sum_{s \in S_{ij}} n_s \gamma_{ij}^{(s)} \sigma_{ij} = \sum_{1 \leq i \leq j \leq q} \lambda_{ij} \sigma_{ij}, \text{ for all } \sigma,
$$
or equivalently,

$$
\sum_{s \in S_{ij}} n_s \gamma_{ij}^{(s)} = \lambda_{ij}, \text{ for all } 1 \leq i \leq j \leq q.
$$

Call

$$
S_{ij}^{*} = \{s \in S_{ij} \setminus \{(i, j)\}, n_s > 0\}, \text{ for all } 1 \leq i \leq j \leq q,
$$
and

$$
D = \{1 \leq i, j \leq q, S_{ij}^{*} \neq \emptyset\}.
$$

Up to the new parametrization $\alpha_{ij}^{(s)} = \alpha_{ji}^{(s)} = n_s (\gamma_{ij}^{(s)} - \gamma_{ij}^{(ij)})$, for all $(i, j) \in D$ and $s \in S_{ij}^{*}$, the general expression for quadratic unbiased estimators of $\theta$ in $\mathcal{C}$ can be deduced from the preceding conditions:

$$
\text{tr}(\Lambda \tilde{\Sigma}) + \sum_{(i,j) \in D} \sum_{s \in S_{ij}^{*}} \alpha_{ij}^{(s)} \left[\tilde{\sigma}_{ij}^{(s)} - \tilde{\sigma}_{ij}\right], \quad (5.1)
$$

where $\tilde{\Sigma} = (\tilde{\sigma}_{ij})_{1 \leq i,j \leq q}$ is the empirical estimator of $\Sigma$, namely $\tilde{\sigma}_{ii} = Q_{ii} / \sum_{s \in S_{ii}} n_s$ and for $i \neq j$, $\tilde{\sigma}_{ij} = 2^{-1} Q_{ij} / \sum_{s \in S_{ij}} n_s$. Equivalently, $\tilde{\sigma}_{ii}^{(s)} = Q_{ii}^{(s)} / n_s$ and for $i \neq j$, $\tilde{\sigma}_{ij}^{(s)} = 2^{-1} Q_{ij}^{(s)} / n_s$.

As commonly encountered in Gauss-Markov approaches, the unbiased estimators in $\mathcal{C}$ are additively corrected versions of the unbiased empirical estimator $\sum_{1 \leq i,j \leq q} \lambda_{ij} \tilde{\sigma}_{ij}$ of $\theta$, which relies on the estimation of each variance parameter $\sigma_{ij}$ by the mean squared error over all the samples for which $(i,j)$ is observed. The additive correction is then a linear combination of differences between such estimators of $\sigma_{ij}$ and other estimators derived on smaller samples.
Remark 5.2. In the context of the multivariate regression model, the general expression (5.1) is still valid but the empirical estimators of the variance parameters are corrected for bias by accounting for the appropriate degrees of freedom of the sum of squares, namely \( n_s \) replaced by \( n_s - p \).

Example 5.2.1. The bivariate case. In this special case, the complete class of estimators have been exhibited above. According to expression (5.1), the general form for unbiased estimators of one particular variance parameter \( \sigma_{ij} \) in this complete class is given by the following expression:

\[
\hat{\sigma}_{ij} = \tilde{\sigma}_{ij} + \alpha^{(12)}_{ij} \left[ \tilde{\sigma}^{(12)}_{ij} - \tilde{\sigma}_{11} \right] + \alpha^{(12)}_{22} \left[ \tilde{\sigma}^{(12)}_{22} - \tilde{\sigma}_{22} \right].
\]

Example 5.2.2. The monotone trivariate case. Expression (5.1) yields the following general form for unbiased estimators of one particular variance parameter \( \sigma_{ij} \) in the complete class:

\[
\hat{\sigma}_{ij} = \tilde{\sigma}_{ij} + \alpha^{(12)}_{ij} \left[ \tilde{\sigma}^{(12)}_{ij} - \tilde{\sigma}_{11} \right] + \alpha^{(123)}_{ij} \left[ \tilde{\sigma}^{(123)}_{ij} - \tilde{\sigma}_{11} \right] + \alpha^{(12)}_{22} \left[ \tilde{\sigma}^{(12)}_{22} - \tilde{\sigma}_{22} \right] \\
+ \alpha^{(123)}_{22} \left[ \tilde{\sigma}^{(123)}_{22} - \tilde{\sigma}_{22} \right] + 2\alpha^{(123)}_{ij} \left[ \tilde{\sigma}^{(123)}_{ij} - \tilde{\sigma}_{12} \right].
\]

5.3. Locally MIVQUE of the variance parameters. Suppose now \( \hat{\theta} \) is an unbiased estimator of \( \theta \) in \( C \). According to expression (5.1), there exists vectors \( \alpha_{ij} = (\alpha_{ij}^{(s)})_{s \in S_{ij}^*}, 1 \leq i \leq j \leq q, \) such that

\[
\hat{\theta} = \text{tr}(\Lambda \tilde{\Sigma}) + \sum_{(i,j) \in D} \sum_{s \in S_{ij}^*} \alpha_{ij}^{(s)} \left[ \tilde{\sigma}_{ij}^{(s)} - \tilde{\sigma}_{ij} \right].
\]

The following lemma provides the general expression for \( V(\theta; \alpha; \Sigma; n) = \text{Var}(\hat{\theta}) \).

Lemma 5.1. Let \( S_{ijkl} = \{s \in S, \{i, j, k, l\} \subseteq s\} \). Then

\[
\text{Var}(\hat{\theta}) = 2 \sum_{1 \leq i, j \leq q} \sum_{1 \leq k, l \leq q} \sigma_{ij} \sigma_{kl} \lambda_{ij} \lambda_{kl} \frac{\sum_{s \in S_{ijkl}} n_s}{\sum_{s \in S_{ij}} n_s \sum_{s \in S_{kl}} n_s} \\
+ 2 \sum_{(i, j) \in D} \sum_{(k, l) \in D} \alpha_{ij}^{(s)} P_{ijkl} \alpha_{ij}^{(s)} - 4 \sum_{(i, j) \in D} \sum_{1 \leq k, l \leq q} \lambda_{kl} \alpha_{ij}^{(s)} r_{ijkl},
\]

where
for $1 \leq i, j, k, l \leq q$, $P_{ijkl}$ stands for the matrix whose rows and columns are indexed by $s \in S^i_j$ and $s' \in S^k_l$ respectively, with generic element defined as follows:

$$P_{ijkl} = \left[ \frac{\delta_{s=s'}}{n_s} - \frac{\delta_{(k,l) \in s' \cap s'}}{\sum_{s \in S_k} n_s} - \frac{\delta_{(i,j) \in s' \cap s}}{\sum_{s \in S_j} n_s} + \frac{\sum_{s \in S_{ijkl}} n_s}{\sum_{s \in S_{ij}} n_s \sum_{s \in S_{kl}} n_s} \right] \sigma_{il} \sigma_{jk},$$

and with the usual notation $\delta_A = 1$ if $A$ is true and 0 elsewhere;

- for $1 \leq i, j, k, l \leq q$, $r_{ijkl}$ stands for the vector whose rows are indexed by $s \in S^i_j$, with generic element defined as follows:

$$r_{ijkl} = - \left[ \frac{\delta_{(k,l) \in S_k}}{\sum_{s \in S_k} n_s} - \frac{\sum_{s \in S_{ijkl}} n_s}{\sum_{s \in S_{ij}} n_s \sum_{s \in S_{kl}} n_s} \right] \sigma_{il} \sigma_{jk}.$$

The derivation of MIVQUE is therefore equivalent to searching for the solution of the following minimization problem:

$$a = \arg \min_{\alpha_{ij}, \ (i,j) \in D} v(\theta; \alpha; \Sigma; n). \quad (5.2)$$

Equating the partial derivatives of $v(\theta; \alpha; \Sigma; n)$ with respect to the vectors $\alpha_{ij}, \ (i,j) \in D$, gives an explicit expression for the locally MIVQUE of $\theta$.

**Theorem 5.2.** Let $a = (a_{ij})_{(i,j) \in D}$, with $a_{ij} = (a_{ij}^{(s)})_{s \in S^i_j}$, stand for the solution of the following system of linear equations

$$\sum_{(k,l) \in D} P_{ijkl} a_{kl} = \sum_{1 \leq k,l \leq q} r_{ijkl} \lambda_{kl}, \text{ for all } (i,j) \in D. \quad (5.3)$$

Then $\hat{\theta} = tr(\hat{\Sigma}) + \sum_{(i,j) \in D} \sum_{s \in S^i_j} a_{ij}^{(s)} \left[ \tilde{\sigma}_{ij}^{(s)} - \tilde{\sigma}_{ij} \right]$ is the locally MIVQUE for $\theta$.

In the following, the estimator $\hat{\theta}$ is denoted by MIVQUE($\theta, \Sigma$) and its variance by $v^*(\theta; \Sigma; n) = v(\theta; a; \Sigma; n)$.

In the special case of independence between the variables $Y_j, j = 1, \ldots, q$, namely when $\Sigma$ is a diagonal matrix, the locally MIVQUE of the variance parameters $\sigma_{ij}$ clearly coincide with the empirical estimators $\tilde{\sigma}_{ij}$.

As a consequence of the preceding theorem, conditions for existence and uniqueness of the locally MIVQUE depend on the regularity of the system...
(5.3). However, a complete study, in a general framework, of the conditions on the pattern $n$ of missing data under which the system (5.3) is invertible, appears to be tedious and goes beyond the scope of this paper.

**Example 5.3.1. The bivariate case.** Here $\mathcal{D} = \{(1,1),(2,2)\}$ and the system (5.3) becomes

$$
\begin{align*}
\begin{cases}
P_{11,11}\alpha_{11} + P_{11,22}\alpha_{22} &= r_{11,11}\lambda_{11} + r_{11,22}\lambda_{22} + 2r_{11,12}\lambda_{12}, \\
P_{22,11}\alpha_{11} + P_{22,22}\alpha_{22} &= r_{22,11}\lambda_{11} + r_{22,22}\lambda_{22} + 2r_{22,12}\lambda_{12}.
\end{cases}
\end{align*}
$$

In the present context, the vector $(\alpha_{11}, \alpha_{22})'$ is solution of the following full rank linear system

$$
\begin{pmatrix}
1 & \frac{\sigma^2_{12}}{\sigma^2_{11}} & -\frac{n(2)}{n(12)} & -\frac{n(12)}{n(1)+n(12)} \\
\frac{\sigma^2_{12}}{\sigma^2_{22}} & -\frac{n(1)}{n(1)+n(12)} & 1 & -\frac{n(12)}{n(1)+n(12)}
\end{pmatrix}
\begin{pmatrix}
\alpha_{11} \\
\alpha_{22}
\end{pmatrix}
= -
\begin{pmatrix}
\frac{\sigma^2_{12}}{\sigma^2_{11}} & -\frac{n(12)}{n(12) + n(112)} & \lambda_{12} + 2\frac{\sigma^2_{12}}{\sigma^2_{11}} \\
\frac{\sigma^2_{12}}{\sigma^2_{22}} & -\frac{n(12)}{n(12) + n(112)} & \lambda_{12} + 2\frac{\sigma^2_{12}}{\sigma^2_{22}}
\end{pmatrix}.
$$

Note that, in the present context, the locally MIVQUE only depends on the fractions $\sigma_{12}/\sigma_{11}$ and $\sigma_{12}/\sigma_{22}$, which, in the next section, will help exhibiting a sequence of estimators obtained iteratively and converging to the MLEs.

**Example 5.3.2. The monotone trivariate case.** As an illustration, we give here the closed form for the MIVQUE of the variance parameters $\sigma_{11}, \sigma_{22}$ and $\sigma_{12}$ obtained by solving the system (5.3):

$$
\begin{align*}
\hat{\sigma}_{11} &= \tilde{\sigma}_{11}, \\
\hat{\sigma}_{22} &= \tilde{\sigma}_{22} - \frac{n(12)}{n(12) + n(123)} \frac{\sigma^2_{12}}{\sigma^2_{11}} \left[ \tilde{\sigma}_{11}^{(12)} - \tilde{\sigma}_{11} \right] - \\
&\quad\frac{n(123)}{n(12) + n(123)} \sigma^2_{12} \left[ \tilde{\sigma}_{11}^{(123)} - \tilde{\sigma}_{11} \right], \\
\hat{\sigma}_{12} &= \tilde{\sigma}_{12} - \frac{n(12)}{n(12) + n(123)} \frac{\sigma^2_{12}}{\sigma^2_{11}} \left[ \tilde{\sigma}_{11}^{(12)} - \tilde{\sigma}_{11} \right] - \\
&\quad\frac{n(123)}{n(12) + n(123)} \sigma^2_{12} \left[ \tilde{\sigma}_{11}^{(123)} - \tilde{\sigma}_{11} \right], \\
\hat{\sigma}_{33} &= \tilde{\sigma}_{33} - \gamma' \left[ \begin{array}{c}
\tilde{\sigma}_{11}^{(123)} - \tilde{\sigma}_{11} \\
\tilde{\sigma}_{12}^{(123)} - \tilde{\sigma}_{12} \\
\tilde{\sigma}_{22}^{(123)} - \tilde{\sigma}_{22}
\end{array} \right] \gamma, \\
\hat{\sigma}_{13} &= \hat{\sigma}_{13} - \left[ \begin{array}{c}
\tilde{\sigma}_{11}^{(123)} - \tilde{\sigma}_{11} \\
\tilde{\sigma}_{12}^{(123)} - \tilde{\sigma}_{12} \\
\tilde{\sigma}_{22}^{(123)} - \tilde{\sigma}_{22}
\end{array} \right] \gamma,
\end{align*}
$$

or equivalently

$$
\begin{align*}
\begin{pmatrix}
\hat{\sigma}_{13} \\
\hat{\sigma}_{23}
\end{pmatrix}
= \begin{pmatrix}
\sigma_{13} \\
\sigma_{23}
\end{pmatrix}
- \begin{pmatrix}
\tilde{\sigma}_{11}^{(123)} - \tilde{\sigma}_{11} \\
\tilde{\sigma}_{12}^{(123)} - \tilde{\sigma}_{12} \\
\tilde{\sigma}_{22}^{(123)} - \tilde{\sigma}_{22}
\end{pmatrix} \gamma, \quad (5.4)
\end{align*}
$$
where $\gamma = (\sigma_{13.2}/\sigma_{1.2}, \sigma_{23.1}/\sigma_{2.1})'$.

In this special context, it can be noted that $MIVQUE(\sigma_{11}, \Sigma)$ is uniformly optimal since the expression of its variance does not depend on $\Sigma$.

6 Iterated MIVQUE and Maximum-likelihood Estimation

As it was mentioned above, MIVQUEs have also been used, beyond their intrinsic interest, as helpful tools to derive numerical algorithms for computing MLEs. Indeed, it is well known that iterative MIVQUE is similar to the restricted maximum likelihood estimators in the sense that, if the procedure converges to a point in the parameter space, then this point is necessarily a solution of the restricted likelihood equations. A comprehensive review of such a use for MIVQUE is given by Harville (1977) in the context of estimation of the variance components of mixed models. In some situations, computational concerns may occur since nonnegative definiteness of the MIVQUE is not guaranteed. A way to handle this problem, suggested by Rao and Kleffe (1988), is to project at each step the iterated MIVQUE to the cone of nonnegative definite matrices, which has no effect if the MIVQUE is nonnegative definite, but does not guarantee optimality of estimation otherwise.

6.1. Iterative MIVQUE.. Let $\Sigma_0$ denote a known $q \times q$ positive definite symmetric matrix and $MIVQUE(\Sigma, \Sigma_0)$ be the $q \times q$ symmetric matrix whose $(i, j)$ element is $MIVQUE(\sigma_{ij}, \Sigma_0)$ as defined in Theorem 5.2. For instance, when $\Sigma_0 = I_q$, $\hat{\Sigma} = \tilde{\Sigma} = MIVQUE(\Sigma, \Sigma_0)$ is the empirical variance-covariance estimator of $\Sigma$. Therefore, this choice for $\Sigma_0$ appears as a natural starting point to estimate $\Sigma$. In order to reduce arbitrariness of this starting point, the preceding procedure can be iterated, leading to a sequence $(\hat{\Sigma}_i)_{i \geq 0}$ of estimators of $\Sigma$ defined by the recurrence relation

$$\hat{\Sigma}_i = MIVQUE(\Sigma, \hat{\Sigma}_{i-1}), \quad i \geq 1.$$  

It can easily be seen for the empirical estimator $\hat{\Sigma}_1$ that, as recalled by Rao and Kleffe (1988), this procedure does not ensure that the estimators obtained at each stage are positive definite matrices. This can be a cause for non-convergence. Therefore, to avoid violations of the constraints on the parameter space, we propose to modify the iterative MIVQUE by projecting, at each stage, the iterated MIVQUE into the convex space $S_q^+$ of $q \times q$ symmetric positive definite matrices. The following lemma gives the explicit
expression for $P_{S_q^+}M$, where $M$ is an arbitrary symmetric matrix and $P_{S_q^+}$ is the projection operator onto $S_q^+$.

**Lemma 6.1.** For $q \geq 1$, let $M$ denote a $q \times q$ real symmetric matrix. Let $P_{S_q^+}M$ denote the matrix

$$P_{S_q^+}M = \arg \min_{S \in S_q^+} \text{tr} \left[ (M - S)(M - S)' \right].$$

Let $p$ denote the number of positive eigenvalues for $M$. If $p = 0$, $P_{S_q^+}M = 0$. Else,

$$P_{S_q^+}M = Q_p \Lambda_p Q_p',$

where $\Lambda_p$ stands for the $p \times p$ diagonal matrix of positive eigenvalues of $M$ in decreasing order and $Q_p$ is the corresponding $q \times p$ orthogonal matrix of normalized eigenvectors of $M$.

The following theorem can be seen as the missing-data version of Harville’s (1977) parallel between the MIVQUE and maximum likelihood approaches.

**Theorem 6.1.** Let $(\hat{\Sigma}_i)_{i \geq 0}$ denote the sequence of $q \times q$ symmetric positive definite matrices defined as follows

$$\hat{\Sigma}_0 = I_q,$$

$$\hat{\Sigma}_i = P_{S_q^+} \text{MIVQUE}(\Sigma, \hat{\Sigma}_{i-1}), \text{ for all } i \geq 1.$$

Then, if $\hat{\Sigma} = \lim_{i \to \infty} \hat{\Sigma}_i$ exists, $\hat{\Sigma}$ is a stationary point for the likelihood equations.

**Example 6.1.1. The bivariate case.** Simulated results are used here to illustrate Theorem 6.1. The simulation is based on 30000 samples of centered bivariate normal distribution with variances equal to one and correlation coefficient varying from 0.05 to 0.95. The missing data pattern is assumed to be $n = (n_1 = 6, n_2 = 8, n_{12} = 4)$.

For every correlation structure, the mean squared error, calculated over the 30000 simulated values, of each iterate of the MIVQUE of $\sigma_{12}$ is derived. The relative efficiency of the iterated MIVQUE is then obtained by dividing its mean squared error by that of the empirical estimator of $\sigma_{12}$ calculated from the $n_{12}$ items with joint observations of the two variables. These efficiencies are reported in Table 2.
Table 2. Relative efficiencies of the iterated MIVQUE of $\sigma_{12}$.

<table>
<thead>
<tr>
<th>Iteration number</th>
<th>Correlation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.05</td>
</tr>
<tr>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>1</td>
<td>0.96</td>
</tr>
<tr>
<td>2</td>
<td>0.72</td>
</tr>
<tr>
<td>3</td>
<td>1.13</td>
</tr>
<tr>
<td>4</td>
<td>1.02</td>
</tr>
<tr>
<td>5</td>
<td>1.25</td>
</tr>
<tr>
<td>6</td>
<td>1.13</td>
</tr>
<tr>
<td>7</td>
<td>1.30</td>
</tr>
<tr>
<td>8</td>
<td>1.18</td>
</tr>
<tr>
<td>9</td>
<td>1.32</td>
</tr>
<tr>
<td>10</td>
<td>1.20</td>
</tr>
<tr>
<td>11</td>
<td>1.33</td>
</tr>
<tr>
<td>12</td>
<td>1.21</td>
</tr>
<tr>
<td>13</td>
<td>1.34</td>
</tr>
<tr>
<td>14</td>
<td>1.22</td>
</tr>
<tr>
<td>15</td>
<td>1.34</td>
</tr>
</tbody>
</table>

Asymptotic 0.99 0.93 0.68 0.5 0.38

To give more insight into the behaviour of the relative efficiencies, the convergence of the sequence of iterated MIVQUE to the MLE is illustrated in Figure 1. First, it is important to note that for a weak correlation between the variables, a MIVQUE, iterated once or twice, shall be preferred to the MLE. However, in that context, the difference between the efficiency of the MLE and that of the MIVQUE, iterated once or twice, is mostly due to the small sample framework and is clearly reduced in asymptotic conditions. Nevertheless, as it was sometimes suggested by MIVQUE users for the estimation of variance components, a strategy of iterating once or twice turns out to be either equivalent to the maximum likelihood for strong correlation between the variables or even better than it for correlation close to zero. The simulation study also points out the gains, in terms of relative efficiency, that can be obtained by iterated MIVQUE relative to the commonly used strategy consisting in omitting the items with at least one missing observation.

Example 6.1.2. The monotone trivariate case. When the missing data pattern is monotone, Little and Rubin (1987, pp. 172-181) show that explicit MLEs of $\Sigma$ can be obtained by factoring the likelihood function. This closed-form estimator can also be found by solving the equation $\hat{\Sigma} =$
Figure 1: Relative efficiencies of the iterated MIVQUE of $\sigma_{12}$.

The MIVQUE($\Sigma, \hat{\Sigma}$). According to the expressions in (5.4), this consists of estimating the three parametric functions needed to derive the MIVQUEs of the variance parameters as follows

$$\hat{\sigma}_{12} = \frac{n_{(12)}\sigma^{(12)}_{12} + n_{(123)}\sigma^{(123)}_{12}}{n_{(12)}\sigma^{(12)}_{11} + n_{(123)}\sigma^{(123)}_{11}},$$

$$\hat{\gamma} = \left( \begin{array}{ccc} \hat{\sigma}^{(123)}_{11} & \hat{\sigma}^{(123)}_{12} & \hat{\sigma}^{(123)}_{13} \\ \hat{\sigma}^{(123)}_{12} & \hat{\sigma}^{(123)}_{12} & \hat{\sigma}^{(123)}_{22} \end{array} \right)^{-1} \left( \begin{array}{ccc} \hat{\sigma}^{(123)}_{13} \\ \hat{\sigma}^{(123)}_{13} \\ \hat{\sigma}^{(123)}_{23} \end{array} \right).$$

6.2. Large sample properties of the MLE. Detailed arguments for the large sample properties of the MIVQUE in the framework of the estimation
of variance components in mixed linear models (for instance its weak and strong consistency) are provided by Rao and Kleffe (1988). However, the application of their results to the missing-data context seems to be quite intractable due to the complexity, in general, of the system (5.3).

Nevertheless, concerning the MLEs, general results are well known, and can be found for instance in Anderson (1973), which states the asymptotic normality of these estimators. The remaining problem of deriving the asymptotic variance of the estimators is solved in the missing-data context by the calculation of the minimum variance reached by the MIVQUE.

The asymptotic conditions can be defined as follows. There exists \( f = (f_s)_{s \in S} \) with \( 0 \leq f_s \leq 1 \) for all \( s \in S \), and such that

\[
\sum_{s \in S} n_s \to \infty, \quad \frac{n_s}{\sum_{s \in S} n_s} \to f_s.
\]

(6.1)

The following theorem gives the asymptotic properties of the MLE of an arbitrary linear combination \( \theta \) of the variance parameters. It can be seen as an application of Anderson’s (1973) general result to the incomplete multivariate data problem.

**Theorem 6.2.** Let \( \hat{\theta} \) stand for the MLE of \( \theta \). Under the asymptotic conditions (6.1),

\[
\left( \sum_{s \in S} n_s \right)^{1/2} (\hat{\theta} - \theta) \to \mathcal{N}[0; v^*(\theta; \Sigma; f)].
\]

Practical consequences of the preceding theorem are numerous since it enables the simultaneous calculation of the MLE and that of its standard error.

**Example 6.2.1. The bivariate case.** In order to illustrate the large sample properties of the MLE, the asymptotic relative efficiency in the bivariate example presented above is given in the last line of Table 3. The difference between these asymptotic efficiencies and the finite sample efficiencies indicated in the simulation results is of course due to the small sample sizes involved in our example.
In what follows, the properties of MIVQUE and maximum likelihood estimation are compared by simulations. The simulation consists of datasets with $p = 3, 5$ and $10$ variables with rows independently distributed as normal variables having mean $0$ and variance-covariance matrix $\Sigma$, where $\sigma_{ij} = \rho$, if $i \neq j$ and $1$ if $i = j$. For each value of $p$, $\rho$ takes values in $\{0.05, 0.1, 0.2, 0.5, 0.8, 0.9\}$. The patterns of missing data are defined as follows: $n_s = 10$ for $s$ containing all variables, $n_s = 50$ for $s$ containing pairs of variables, and $n_s = 0$ for $s$ corresponding to other combinations. For every pair $(p, \rho)$, $1000$ datasets are simulated.

MIVQUE and maximum likelihood are now compared through their performance in estimating the covariance matrix $\Sigma$. First, for each data set, the parameters are estimated using the EM algorithm implemented for incomplete normal data in the procedure \texttt{norm} available in the software \texttt{R} (2004). The default options are left unchanged and the number of iterations is limited to 1000. On the same data tables, MIVQUE is also used to estimate the parameters. The initial value for $\Sigma$ is given by its empirical estimate using all complete pairs of observations on the variables. Then MIVQUE is iterated three times. Finally, the efficiencies of the different estimation strategies are compared by the mean squared error calculated as follows

$$
MSE = \frac{1}{1000} \sum_{k=1}^{1000} \text{tr} \left[ \left( \hat{\Sigma}_{(k)} - \Sigma \right)^2 \right],
$$

where $\hat{\Sigma}_{(k)}$ is the estimated value of $\Sigma$ for the $k$th dataset.

Table 3 reports the relative efficiency of each of the MLE and the iterated MIVQUE, calculated as the ratio between its MSE and that of the empirical estimator which handles missing data by pairwise deletion. Furthermore, Figure 2 gives a graphical display of the content of Table 3.

The first remark about the results of the simulations concerns the general aspect of the curves in Figure 2. Both maximum likelihood and MIVQUE are increasingly better beneficial relative to the empirical estimator as the correlations between the variables increase. For example, when $\rho = 0.9$, using EM or MIVQUE can result in reduction of the MSE by up to 95% of the MSE of the empirical estimator.
MIVQUE and MLE in multivariate linear models

Table 3. Relative efficiencies of the maximum likelihood estimation and the iterated MIVQUE.

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.50</th>
<th>0.80</th>
<th>0.90</th>
<th>Time (s) for 1000 data tables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML</td>
<td>1.01</td>
<td>1.00</td>
<td>0.96</td>
<td>0.90</td>
<td>0.76</td>
<td>0.54</td>
<td>0.48</td>
<td>4 $\leq T \leq 7$</td>
</tr>
<tr>
<td></td>
<td>MIVQUE (1 iter.)</td>
<td>0.83</td>
<td>0.83</td>
<td>0.86</td>
<td>0.80</td>
<td>0.72</td>
<td>0.54</td>
<td>0.48</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>MIVQUE (2 iter.)</td>
<td>0.85</td>
<td>0.83</td>
<td>0.83</td>
<td>0.77</td>
<td>0.69</td>
<td>0.48</td>
<td>0.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIVQUE (3 iter.)</td>
<td>0.85</td>
<td>0.84</td>
<td>0.83</td>
<td>0.78</td>
<td>0.67</td>
<td>0.48</td>
<td>0.35</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.50</th>
<th>0.80</th>
<th>0.90</th>
<th>Time (s) for 1000 data tables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML</td>
<td>1.04</td>
<td>1.02</td>
<td>0.95</td>
<td>0.82</td>
<td>0.56</td>
<td>0.24</td>
<td>0.19</td>
<td>26 $\leq T \leq 39$</td>
</tr>
<tr>
<td></td>
<td>MIVQUE (1 iter.)</td>
<td>0.91</td>
<td>0.90</td>
<td>0.87</td>
<td>0.80</td>
<td>0.60</td>
<td>0.26</td>
<td>0.16</td>
<td>79</td>
</tr>
<tr>
<td></td>
<td>MIVQUE (2 iter.)</td>
<td>0.97</td>
<td>0.93</td>
<td>0.89</td>
<td>0.82</td>
<td>0.57</td>
<td>0.19</td>
<td>0.14</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIVQUE (3 iter.)</td>
<td>0.97</td>
<td>0.95</td>
<td>0.91</td>
<td>0.84</td>
<td>0.55</td>
<td>0.20</td>
<td>0.13</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\rho$</th>
<th>0.05</th>
<th>0.10</th>
<th>0.20</th>
<th>0.30</th>
<th>0.50</th>
<th>0.80</th>
<th>0.90</th>
<th>Time (s) for 1000 data tables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ML</td>
<td>1.06</td>
<td>1.02</td>
<td>0.91</td>
<td>0.75</td>
<td>0.42</td>
<td>0.11</td>
<td>0.06</td>
<td>510 $\leq T \leq 780$</td>
</tr>
<tr>
<td></td>
<td>MIVQUE (1 iter.)</td>
<td>0.95</td>
<td>0.92</td>
<td>0.85</td>
<td>0.78</td>
<td>0.54</td>
<td>0.15</td>
<td>0.07</td>
<td>240 $\leq T \leq 252$</td>
</tr>
<tr>
<td></td>
<td>MIVQUE (2 iter.)</td>
<td>1.09</td>
<td>1.04</td>
<td>0.99</td>
<td>0.86</td>
<td>0.49</td>
<td>0.12</td>
<td>0.05</td>
<td></td>
</tr>
<tr>
<td></td>
<td>MIVQUE (3 iter.)</td>
<td>1.08</td>
<td>1.04</td>
<td>0.96</td>
<td>0.82</td>
<td>0.51</td>
<td>0.13</td>
<td>0.05</td>
<td></td>
</tr>
</tbody>
</table>

Moreover, the difference in terms of relative efficiencies between the MIVQUE and the maximum likelihood estimator decreased when the number of variables increased. This can be due to the fact that, with the pattern of missing data we have chosen, the amount of information on each variable is much larger when $p = 10$ than when $p = 3$. This means that increasing $p$ makes the situation closer to asymptotic conditions. Concerning the comparison between MIVQUE and maximum likelihood itself, whatever may the number of variables be, it seems that in situations of low correlation between variables, MIVQUE, iterated once, performs better than MLE whereas in situations of high correlation, MLE can be slightly better. Note also that, when correlations are low, iteration of the MIVQUE makes its MSE close to that of MLE rapidly. However, convergence seems to be much slower when the correlations are large.

For the simulations, the time needed for calculating MIVQUE and MLEs have also been measured and the results are reported in Table 3. However, the comparison between the two approaches, in terms of time of calculation, is still very unfair since MIVQUE is currently implemented entirely in R.
Figure 2: Relative efficiencies of the maximum likelihood estimation and the iterated MIVQUE.

(2004) whereas the package norm used for the EM algorithm is optimized by calling auxiliary programs written in Fortran. Moreover, computation times are probably very sensitive to the pattern of missing data, and this requires a wider study including other patterns of missing data. Studies about the need for this important aspect of the comparison are currently in progress.

8 Concluding Remarks

This paper introduces a Gauss-Markov approach for the estimation of the parameters of multivariate linear regression models in the presence of
incomplete observations. Our framework is quite general since no particular assumption is made either on the missing pattern or on the correlation structure between the variables. The key result, which makes possible the derivation of the MIVQUEs, is the characterization of a complete class of quadratic estimators for the estimation of the variance parameters.

The MIVQUEs, which result from the Gauss-Markov approach are then shown to enjoy desirable properties, especially in non-asymptotic conditions relative to the MLEs. Furthermore, iterated versions of MIVQUE bring some novel solutions for the computation of the maximum likelihood estimators. Using these relationships between the Gauss-Markov approach and the maximum likelihood, large sample properties of the MLEs are studied and the asymptotic variances of the estimators are provided.

The estimation of the regression coefficients is not explicitly dealt with in this paper. The main results focus on the estimation of the variance parameters, which usually precedes the estimation of regression coefficients. However, the asymptotic properties of the estimators of the expectation parameters have to be addressed. Note that, in that context, the case of missing data both for the predictors and the $Y$ variables can easily be handled by estimating the covariance of the whole set of variables.

Computational issues of the MIVQUE will also be addressed elsewhere. However, it can be said that the stability of the calculations rely entirely on the conditioning of the linear system (5.3). In some situations, solving the system would require use of some numerical tools usually employed to tackle ill-conditioned problems. Furthermore, more sophisticated algebra has to be used to deal with the problem of possibly non-positive definite estimation of the variance matrix.

Acknowledgement. The author thanks a referee and a co-editor for their comments that have helped in improving the original manuscript.

References


D. Causeur


David Causeur
Laboratoire de Mathématiques Appliquées
Agrocampus Rennes
65 rue de Saint-Brieuc – CS 84215
35042 Rennes Cedex, France
E-mail: david.causeur@agrocampus-rennes.fr

Paper received July 2006; revised July 2006.