Maximum Likelihood Estimation in Linear Models with Equi-Correlated Random Errors

By: Haimeng Zhang, M. Bhaskara Rao

This is the accepted version of the following article:

Zhang, H. and Rao, M. B. (2006). Maximum likelihood estimation in linear models with equi-correlated random errors. *Australian and New Zealand Journal of Statistics, 48(1)*, 79 – 93,

which has been published in final form at http://dx.doi.org/10.1111/j.1467-842X.2006.00427.x.

Abstract:

Necessary and sufficient conditions for the existence of maximum likelihood estimators of unknown parameters in linear models with equi-correlated random errors are presented. The basic technique we use is that these models are, first, orthogonally transformed into linear models with two variances, and then the maximum likelihood estimation problem is solved in the environment of transformed models. Our results generalize a result of Arnold, S. F. (1981)[The theory of linear models and multivariate analysis. Wiley, New York]. In addition, we give necessary and sufficient conditions for the existence of restricted maximum likelihood estimators of the parameters. The results of Birkes, D. & Wulff, S. (2003)[Existence of maximum likelihood estimates in normal variance-components models. J Statist Plann. Inference.113, 35–47] are compared with our results and differences are pointed out.

Keywords: equi-correlated variables | linear models | maximum likelihood estimator | multivariate normal distribution | restricted maximum likelihood estimator

Article:

1. Introduction

The linear model $\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$ is ubiquitous in many areas of statistics. The entity $\boldsymbol{Y}$ is a $n\times 1$ random vector, $\boldsymbol{X}$ is a $n\times m$ known design matrix, $\boldsymbol{\beta}$ is an unknown parameter vector in $\mathbb{R}^m$, and $\boldsymbol{\varepsilon}$ is the error vector normally distributed with mean vector 0 and dispersion matrix $\sigma^2\boldsymbol{I}n$, where $\sigma^2 > 0$ is unknown. If $m < n$ and rank $(\boldsymbol{X}) = m$, then maximum likelihood (ML) estimators of $\boldsymbol{\beta}$ and σ^2 do exist and have an explicit form. If $m < n$ and rank$(\boldsymbol{X}) < m$, ML estimators of $\boldsymbol{\beta}$ and σ^2 exist but the ML estimator of $\boldsymbol{\beta}$ is not unique. If $m\geq n$ and rank $(\boldsymbol{X}) = n$, ML estimators of $\boldsymbol{\beta}$ and σ^2 do not exist, and finally, if $m\geq n$ but rank $(\boldsymbol{X}) < n$, ML estimators of $\boldsymbol{\beta}$ and σ^2 exist but the ML estimator of $\boldsymbol{\beta}$ is not unique.

The moment we relax the assumption of independence of the components of the error vector **ε**, the ML estimation problem radically changes. In this note, we consider a linear model more general than the one postulated above:

$\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon},$ (1)

where $\boldsymbol{Y}$ is a $n\times 1$ random vector, $\boldsymbol{\mu}$ is an unknown vector belonging to a subspace V of $\mathbb{R}^n$, and $\boldsymbol{\varepsilon}$ is normally distributed with mean vector 0 and equi-correlated errors, i.e., its dispersion matrix is of the form $\sigma^2\boldsymbol{A}(\rho) = \sigma^2((1-\rho)\boldsymbol{I}_n+\rho\boldsymbol{J}_n)$ for some unknown $\sigma^2 > 0$, and $-1/(n-1) < \rho < 1$, where $\boldsymbol{I}_n$ is the $n\times n$ identity matrix and $\boldsymbol{J}_n=\boldsymbol{1}_n\boldsymbol{1}^\top_n$ for $\boldsymbol{1}_n= (1, 1, \ldots, 1)^\top$, the $n\times 1$ column vector with each entry equal to unity.

A motivation for the emergence of this problem arose when, some years ago, a statistician associated with the Milwaukee project presented the following query. A mother with low intelligence (IQ $\leq$ 60) was being monitored over a long period of time. As an example, assume she had 5 children, two males and three females. The mother was being given intensive schooling in problem-solving skills over time. The IQs of her children were measured from time to time. At a given instant of time when the IQ of the children were measured, the measurements Y_1, Y_2, Y_3, Y_4, and Y_5 were modelled as having a 5-variate normal distribution with mean vector $\boldsymbol{\mu}^\top= (\mu_1, \mu_2, \mu_3, \mu_4, \mu_5)$ and dispersion matrix $\sigma^2\boldsymbol{A}(\rho)$, where Y_1, Y_2 are the measurements on the male siblings and Y_3, Y_4, Y_5 are on female siblings. The model stipulated that $\mu_1=\mu_2$ and $\mu_3=\mu_4=\mu_5$. The question raised was whether or not ML estimates of $\boldsymbol{\mu}$, σ^2 and ρ exist. We will take up this problem in Section 2.

Arnold (1981, Section 14.9) (see also Arnold 1979) showed that if $\boldsymbol{1}_n\in V$, ML estimators of $\boldsymbol{\mu}$, σ^2, and ρ do not exist. In this note, we generalize his result by characterizing precisely when ML estimators of $\boldsymbol{\mu}$, σ^2, and ρ exist (see Theorem 1 in Section 2). The main technique we employ is as follows. We transform the model $\boldsymbol{Y}=\boldsymbol{\mu}+\boldsymbol{\varepsilon}$ into a linear model $\boldsymbol{Z}=\boldsymbol{\nu}+\boldsymbol{\varepsilon}^*$ with the dispersion matrix of $\boldsymbol{\varepsilon}^*$ given by $\mathrm{Disp}(\boldsymbol{\varepsilon}^*) = \mathrm{diag}(\sigma^2_1, \sigma^2_2\boldsymbol{I}_{n-1})$. Necessary and sufficient conditions are then provided for the existence of ML estimators of $\boldsymbol{\nu}$, σ^2_1, and σ^2_2, and hence those of $\boldsymbol{\mu}$, σ^2, and ρ (see Theorem 2 in Section 2).

This model comes under the realm of the heteroscedastic variances model, which is a special case of variance components model as presented in Rao & Kleffe (1988). The literature up to 1997 has been covered to a large extent in Rao (1997). Although ML estimation of variance components is extensively discussed in the literature, the existence of the estimates in a general framework is difficult to establish. However, in our case, we are able to identify precisely the circumstances under which the estimates exist.

It should be pointed out that there is considerable literature on estimating heteroscedastic variances by different techniques. Following the work of Hartley, Rao & Kiefer (1969), Hartley & Jayatillake (1973) considered ML estimation in the linear model $\boldsymbol{Y}=\boldsymbol{X\beta}^*+\boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon}$ has a multivariate normal distribution with mean vector 0 and dispersion matrix $\mathrm{Disp}(\boldsymbol{\varepsilon}) = \mathrm{diag}(\sigma^2_1,$

$\sigma^2_2, \ldots, \sigma^2 n$) with the proviso that $0 < \delta^2 i \leq \sigma^2 i$, $i = 1, 2, \ldots, n$, where the $\delta^2 i$s are known in advance. This model involves n observations and n variances. (If no restriction is imposed on the variances in this model, there would be no ML estimates.) They described a procedure for obtaining ML estimates of the model. They also looked at another case in which for each variance in the model more than one observation is available. In this case, there is no need to impose any restriction on variances in order to obtain ML estimates. In our case, we have only two variances σ^2_1 and σ^2_2, with one observation available for the variance σ^2_1 and ($n-1$) observations for σ^2_2. Unlike Hartley & Jayatillake (1973), we do not impose any conditions on the variances.

There is also some related literature. Neyman & Scott (1948) discuss ML estimation in a number of problems involving heteroscedastic variances. Other researchers, for example, Putter (1967), Rao (1970, 1972), Horn, Horn & Duncan (2005), Horn & Horn (1975), and Hartley & Rao (1967), have looked at various facets of heteroscedastic variance estimation. For a detailed discussion of the issues involved, see Wiorkowski (1975), Chaubey (1980), Rao & Kleffe (1988), and Rao & Rao (1998).

One could suggest that the original model (1) can be written as a variance components model as presented by Demidenko & Massam (1999). This will not work out. We can write (1) as $\boldsymbol{Y} = \boldsymbol{\mu} + 1_n\tilde{U} + \tilde{\boldsymbol{\varepsilon}}$ with $\tilde{\boldsymbol{\varepsilon}}$ normally distributed as $N_n(0, \tilde{\sigma}_0^2)$ and $\tilde{U}$ normally distributed as $N(0, \tilde{\sigma}_1^2)$ where $\tilde{\sigma}_0^2 = \sigma^2(1-\rho)\boldsymbol{I}_n$ and $\tilde{\sigma}_1^2 = \sigma^2\rho$. However, in our model, $\tilde{\sigma}_1^2$ may be negative. In the model considered by Demidenko & Massam (1999), $\tilde{\sigma}_1^2$ has to be non-negative. This crucial difference is reflected in the conclusions of Theorem 1 of this paper and Theorem 3.1 of Demidenko & Massam (1999). In Section 4, we present an example illustrating the differences in the results.

Following a recommendation of the referee, we also explore the connection between our model (1) and the variance components model considered by Birkes & Wulff (2003). For the ML estimation problem over the entire parameter space, the results we present for our model (1) cover all possible scenarios whereas the results of Birkes & Wulff (2003) when applied to model (1) cover only a subset. These differences are spelled out in Section 4.

This paper is organized as follows. In Section 2, we present the main results of the paper, which provide necessary and sufficient conditions for the existence of ML estimators. In Section 3, we take up the problem of existence of restricted ML estimators. A comparison of our results with the results of two related papers (Demidenko & Massam (1999) and Birkes & Wulff (2003)) is made in Section 4.

2. Main Results

Now we state our main result. Let V be a subspace of $\mathbb{R}^n$, with dimension dim(V), and let $\boldsymbol{\mu}$ be a vector in V.

Theorem 1 *Consider the problem of ML estimation in model (1).*

1 . *No ML estimators of* $\boldsymbol{\mu}$, σ^2, *and* ρ *exist in each of the following cases.*

   (a) $\mathbf{1}_n \in V$.

   (b) $\mathbf{1}_n \in V^{\perp}$ *and* $\dim(V) = n - 1$.

   (c) *Neither* $\mathbf{1}_n \in V$ *nor* $\mathbf{1}_n \in V^{\perp}$ *and* $\dim(V) < n - 1$.

2 . *ML estimators exist in all other cases, i.e.,*

   (d) $\mathbf{1}_n \in V^{\perp}$ and $\dim(V) < n - 1$;

   (e) *Neither* $\mathbf{1}_n \in V$ *nor* $\mathbf{1}_n \in V^{\perp}$ *and* $\dim(V) = n - 1$.

In the specific example from the Milwaukee project, in view of Theorem 1, no ML estimates exist.

Some comments are in order on Theorem 1 and the rest of this section. Non-existence of ML estimates means that there is a set of data scenarios $\boldsymbol{Y}$ with positive probability for which no ML estimates exist, while existence of ML estimates means that ML estimates exist almost surely. The case 1(a) needs a special mention, as ML estimates do not exist for any data scenario $\boldsymbol{Y}$.

To prove Theorem 1, we require the following lemmas. Proofs of these lemmas are trivial and therefore are omitted.

Lemma 1 *Let* U_1 *be a random variable with a continuous distribution function and support equal to* $(-\infty, +\infty)$, *and* $\boldsymbol{U}_2$ *be a random vector with a continuous distribution function and support equal to* $\mathbb{R}^m$. *In addition, we assume* U_1 *and* $\boldsymbol{U}_2$ *are independent. Let U be a random variable given by*

$$U = \left(U_1 - \boldsymbol{d}_1^{\top}\boldsymbol{U}_2\right)^2 - \left(\boldsymbol{d}_2^{\top}\boldsymbol{U}_2\right)^2$$

for non-zero real vectors $\boldsymbol{d}_1$ *and* $\boldsymbol{d}_2$ *each of order* $m \times 1$. *Then* $Pr(U < 0) > 0$.

Lemma 2 *If* $Y \sim N(\tau, \sigma^2)$, *where* $-\infty < \tau < \infty$ *and* $\sigma^2 > 0$ *are both unknown, ML estimators of* τ *and* σ^2 *based on a single observation of Y do not exist.*

Lemma 3 *Let* $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_n)^{\top}$ *be a multivariate normal random vector with unknown mean vector* $\boldsymbol{\mu}$ *and dispersion matrix* $\sigma^2 \boldsymbol{I}_n$, *where* $\boldsymbol{\mu} \in V$, *a subspace of* $\mathbb{R}^n$, $\sigma^2 > 0$ *unknown. Then the ML estimators of* $\boldsymbol{\mu}$ *and* σ^2 *exist if and only if* $\dim(V) \le n - 1$.

Now we transform the given linear model (1): $\boldsymbol{Y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$, with $\boldsymbol{\mu} \in V$ and $\dim(V) = m$, into a linear model with two variances. First, for $n \ge 2$, let $\boldsymbol{P}_n$ be a $(n - 1) \times n$ matrix such that the $n \times n$ matrix

$$C_n = \begin{pmatrix} (1/\sqrt{n})1_n^\top \\ P_n \end{pmatrix} \quad (2)$$

is orthogonal, that is $C^\top_n C_n = C_n C^\top_n = I_n$. Obviously, P_n has the following properties.

1 $P_n P^\top_n = I_{n-1}$;

2 $P_n A(\rho) P^\top_n = (1-\rho) I_{n-1}$;

3 $P^\top_n P_n = I_n - (1/n) J_n$.

One example of P_n is the well-known Helmert matrix (e.g. see Press (1982, pp. 13–14)). Now we define a column vector of random variables $Z \equiv (Z_1, Z_2, \ldots, Z_n)^\top$ by $Z = C_n Y$. The mean vector of Z, denoted by $\mathbf{v} \equiv (\nu_1, \mathbf{v}_{(2)})^\top \equiv (\nu_1, \nu_2, \ldots, \nu_n)^\top$, is given by $\nu = EZ = C_n\mu = ((1/\sqrt{n})1_n, P_n^\top)^\top \mu$, where, obviously, $\nu_1 = (1/\sqrt{n})1_n^\top \mu$ and $\mathbf{v}_{(2)} = P_n \boldsymbol{\mu}$. In addition, we denote the parameter space for $\mathbf{v}$ by $W \equiv \{\mathbf{v} = C_n \boldsymbol{\mu} : \boldsymbol{\mu} \in V\}$. Then $\dim(W) = \dim(V) = m$. The dispersion matrix of Z is given by

$$\sigma^2 C_n A(\rho) C_n^\top = \sigma^2 \begin{pmatrix} 1 + (n-1)\rho & 0 \\ 0 & (1-\rho) I_{n-1} \end{pmatrix}. \quad (3)$$

Therefore, $Z_1, Z_2, \ldots, Z_n$ are independent, $Z_1 \sim N(\nu_1, \sigma^2(1 + (n-1)\rho))$, and $Z_i \sim N(\nu_i, \sigma^2(1-\rho))$, $i = 2, 3, \ldots, n$. Note that finding ML estimators of $\mathbf{v}$, $\sigma^2_1 \equiv \sigma^2(1 + (n-1)\rho) > 0$, and $\sigma^2_2 \equiv \sigma^2(1-\rho) > 0$ is equivalent to finding ML estimators of $\boldsymbol{\mu}$, σ^2 and ρ. The transformed model

$$Z = \nu + \varepsilon^*, \quad (4)$$

with $\mathbf{v} \in W$ and $\boldsymbol{\varepsilon}^*$ having a multivariate normal distribution with mean vector zero and dispersion matrix $\mathrm{Disp}(\boldsymbol{\varepsilon}^*) = \mathrm{diag}(\sigma^2_1, \sigma^2_2 I_{n-1})$, has two variances σ^2_1 and σ^2_2.

If $m < n$, the subspace W of dimension m can be viewed as the intersection of some $(n-m)$ linearly independent hyperplanes of the form

$$H_1 = \left\{(\nu_1, \nu_2, \ldots, \nu_n)^\top \in \mathbb{R}^n; a_{11}\nu_1 + \cdots + a_{1n}\nu_n = 0\right\},$$
$$H_2 = \left\{(\nu_1, \nu_2, \ldots, \nu_n)^\top \in \mathbb{R}^n; a_{21}\nu_1 + \cdots + a_{2n}\nu_n = 0\right\},$$
$$\cdots$$
$$H_{n-m} = \left\{(\nu_1, \nu_2, \ldots, \nu_n)^\top \in \mathbb{R}^n; a_{n-m,1}\nu_1 + \cdots + a_{n-m,n}\nu_n = 0\right\}.$$

Linear independence means that the $(n-m) \times n$ matrix $A \equiv (a_{ij})$ is of rank $n-m$.

The subspace W of dimension m can also be the intersection of another set of $(n-m)$ linearly independent hyperplanes of the form

$$J_1 = \left\{(\nu_1, \nu_2, \ldots, \nu_n)^\top \in \mathbb{R}^n; b_{11}\nu_1 + \cdots + b_{1n}\nu_n = 0\right\},$$

$$J_2 = \left\{(\nu_1, \nu_2, \ldots, \nu_n)^\top \in \mathbb{R}^n; b_{21}\nu_1 + \cdots + b_{2n}\nu_n = 0\right\},$$

$$\ldots$$

$$J_{n-m} = \left\{(\nu_1, \nu_2, \ldots, \nu_n)^\top \in \mathbb{R}^n; b_{n-m,1}\nu_1 + \cdots + b_{n-m,n}\nu_n = 0\right\},$$

where the $(n-m) \times n$ matrix $\boldsymbol{B} \equiv (b_{ij})$ is of rank $n-m$. In fact, $\boldsymbol{B}$ can be obtained from $\boldsymbol{A}$ by a series of elementary row transformations. Whatever may be the intersection of these hyperplanes, it can be noted, for example, that $a_{i1} = 0$ for all $1 \le i \le n-m$ if and only if $b_{i1} = 0$ for all $1 \le i \le n-m$.

We now clearly spell out when ML estimators of $\mathbf{v}$, σ^2_1, and σ^2_2 in the transformed model (4) exist almost surely and when they do not.

Theorem 2 *Let* $\boldsymbol{Z} = (Z_1, Z_2, \ldots, Z_n)^\top$ *be a multivariate normal random vector with mean vector* $\mathbf{v} = (\nu_1, \nu_2, \ldots, \nu_n)^\top$ *and dispersion matrix* $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma^2_1, \sigma^2_2 \boldsymbol{I}_{n-1})$, *where* $\mathbf{v} \in W$, *a subspace of* , *and* $\sigma^2_1, \sigma^2_2 > 0$ *are unknown. Then a characterization of when ML estimators exist is given in*

Table 1 . Cases showing whether ML estimators exist.

Case	MLEs exist?
1. $\dim(W) = m = n$	No
2. $\dim(W) = m < n$	
A. $a_{i1} = 0$ for all $1 \le i \le n-m$	No
B. $a_{i1} \neq 0$ for some $1 \le i \le n-m$	
B_1. $\dim(W) = n-1$	
B_{11}. $W = \{(\nu_1, \ldots, \nu_n)^\top \in \mathbb{R}^n; \nu_1 = 0\}$	No
B_{12}. $W = \{(\nu_1, \ldots, \nu_n)^\top \in \mathbb{R}^n; \nu_1 = a_2\nu_2 + a_3\nu_3 + \cdots + a_r\nu_r$ for some $2 \le r \le n$ and constants $a_i \neq 0$, $i = 2, 3, \ldots, r$	Yes
B_2. $\dim(W) < n-1$	
B_{21}. One of the hyperplanes defining W is $\{(\nu_1, \ldots, \nu_n)^\top \in \mathbb{R}^n; \nu_1 = 0\}$	Yes
B_{22}. All other cases	No

Proof Cases 1 and 2A follow from Lemma 2. Cases B_{11} and B_{21} follow from Lemma 3. Therefore, it is sufficient to prove Cases B_{12} and B_{22}. First, we look at the case B_{12}. Let $\dim(W) = m = n-1$ and ν_1 be a non-trivial linear combination of some ν_i's. Without loss of generality, we assume $\nu_1 = a_2\nu_2 + a_3\nu_3 + \cdots + a_r\nu_r$ for some $2 \le r \le n$ and constants $a_i \neq 0$, $i = 2, 3, \ldots, r$. The log-likelihood of the data, up to a constant not depending on the unknown parameters, is then given by

$$\ln L = -\frac{1}{2}\ln\sigma_1^2 - \frac{n-1}{2}\ln\sigma_2^2 - \frac{(Z_1 - a_2\nu_2 - a_3\nu_3 - \cdots - a_r\nu_r)^2}{2\sigma_1^2} - \frac{\sum_{i=2}^{n}(Z_i - \nu_i)^2}{2\sigma_2^2}.$$

The likelihood equations simplify to

$$\sigma_1^2 = (Z_1 - a_2\nu_2 - a_3\nu_3 - \cdots - a_r\nu_r)^2; \quad (5)$$

$$\sigma_2^2 = \sum_{i=2}^{r}(Z_i - \nu_i)^2/(n-1); \quad (6)$$

$$\nu_i = Z_i; \quad \text{for } i = r+1, r+2, \ldots, n, \text{ and} \quad (7)$$

$$\frac{a_i}{Z_1 - a_2\nu_2 - a_3\nu_3 - \cdots - a_r\nu_r} + \frac{(Z_i - \nu_i)}{\sigma_2^2} = 0; \quad \text{for } i = 2, 3, \ldots, r, \quad (8)$$

which lead to $(Z_2 - \nu_2)/a_2 = (Z_3 - \nu_3)/a_3 = \cdots = (Z_r - \nu r)/a_r \equiv C$, say, a constant depending only on the data. Substituting the expressions for νi above in terms of C back to (8) with $i = 2$, one can get

$$C = \left(\frac{n-1}{n}\right)\frac{\sum_{i=2}^{r} a_i Z_i - Z_1}{\sum_{i=2}^{r} a_i^2},$$

$$\nu_i = Z_i - Ca_i, i = 2, 3, \ldots, r. \quad (9)$$

Therefore, ML estimators of $(\nu_2, \nu_3, \ldots, \nu n)^\top$, σ^2_1 and σ^2_2 exist and are given by (9) and (7), (5) and (6), respectively.

Next we consider Case B_{22}. We prove non-existence of ML estimators for the case $m = n - 2$. For arbitrary $m < n - 2$, the arguments used for $m = n - 2$ can be adapted to establish the non-existence. For the case $m = n - 2$, note that W is the intersection of two hyperplanes. In one hyperplane, we

will have for some $2 \le l \le n$ and $a_2, a_3, \ldots, a_l \ne 0$. For the other hyperplane, there exists some νj, $r + 1 \le j \le n$, which is a linear combination of some of $\nu_2, \nu_3, \ldots, \nu n$. Without loss of generality, we let $\nu j = \nu r_{+1}$, $r > l$. Relabel the components $\nu_2, \nu_3, \ldots, \nu n$, if necessary. We identify four cases.

Case I: (The case of no common component in the linear combinations)

$$\nu_1 = \sum_{i=2}^{l} a_i\nu_i, \quad \text{for some } 2 \le l \le n \text{ and } a_2, a_3, \ldots, a_l \ne 0,$$

$$\nu_{r+1} = \sum_{i=l+1}^{r} b_i\nu_i, \quad \text{with } r > l \text{ and } b_i \ne 0 \text{ for all } i\text{'s};$$

Case II: (The case of one common component in the linear combinations)

$$\nu_1 = \sum_{i=2}^{l} a_i \nu_i, \quad \text{for some } 2 \le l \le n \text{ and } a_2, a_3, \ldots, a_l \ne 0,$$

$$\nu_{r+1} = \sum_{i=l}^{r} b_i \nu_i, \quad \text{with } r \ge l \text{ and } b_i \ne 0 \text{ for all } i\text{'s};$$

Case III:

$$\nu_1 = \sum_{i=2}^{l} a_i \nu_i, \quad \text{for some } 2 \le l \le n \text{ and } a_2, a_3, \ldots, a_l \ne 0, \qquad \nu_{r+1} = 0;$$

Case IV: (The case of more than one common component in the linear combinations).

It is sufficient to consider Cases I and II since for Case III, we have $\nu_1 = \sum_{i=2}^{l} a_i \nu_i$ and $\nu_{r+1}= 0$ with $r \ge l$ and $a_i \ne 0$ for all i's. This case can be viewed as a special case of Case I but with $b_i= 0$ for all i's. Case IV can be easily transformed to Case II above. As an example, if there are two common variables ν_{l-1} and ν_l in the linear combinations for ν_1 and ν_{r+1}, that is $\nu_1 = \sum_{i=2}^{l} a_i \nu_i$, and $\nu_{r+1} = \sum_{i=l-1}^{r} b_i \nu_i$ with $a_i \ne 0$, $b_i \ne 0$ for all i's, then we may introduce a new parameter $\theta = a_{l-1}\nu_{l-1} + a_l \nu_l$, implying that $\nu_1 = \sum_{i=2}^{l-2} a_i \nu_i + \theta$ and $\nu_{r+1} = \frac{b_{l-1}}{a_{l-1}}\theta + (b_l - \frac{a_l}{a_{l-1}})\nu_l + \sum_{i=l+1}^{r} b_i \nu_i$, which is Case II.

Now consider Case I. We have $\nu_1 = \sum_{i=2}^{l} a_i \nu_i$, $\nu_{r+1} = \sum_{i=l+1}^{r} b_i \nu_i$, with $a_i \ne 0$, $b_i \ne 0$ for all i's. Note that the log-likelihood of the data, up to a constant not depending on unknown parameters, can be written as

$$\ln L = -\frac{1}{2}\ln \sigma_1^2 - \frac{n-1}{2}\ln \sigma_2^2 - \frac{\left(Z_1 - \sum_{i=2}^{l} a_i \nu_i\right)^2}{2\sigma_1^2} - \frac{\left(Z_{r+1} - \sum_{i=l+1}^{r} b_i \nu_i\right)^2 + \sum_{i=2}^{r}(Z_i - \nu_i)^2 + \sum_{i=r+2}^{n}(Z_i - \nu_i)^2}{2\sigma_2^2}.$$

Again, ML estimation of $(\nu_2, \nu_3, \ldots, \nu_r, \nu_{r+2}, \ldots, \nu_n)^\top$, σ^2_1, and σ^2_2 involve the following estimating equations

$$\sigma_1^2 = \left(Z_1 - \sum_{i=2}^{l} a_i \nu_i\right)^2, \quad \sigma_2^2 = \left(\left(Z_{r+1} - \sum_{i=l+1}^{r} b_i \nu_i\right)^2 + \sum_{i=2}^{r}(Z_i - \nu_i)^2\right) \Big/ (n-1);$$

and

$$\nu_i = Z_i, \quad \text{for } i = r+2, r+3, \ldots, n$$

$$\frac{a_i}{\sigma_1} + \frac{Z_i - \nu_i}{\sigma_2^2} = 0, i = 2, 3, \ldots, l, \tag{10}$$

$$b_i \left(Z_{r+1} - \sum_{k=l+1}^{r} b_k \nu_k \right) + (Z_i - \nu_i) = 0, \quad \text{for } i = l+1, \ldots, r. \tag{11}$$

We show that these equations have no solution. Note that (10) is equivalent to

$$\frac{Z_i - \nu_i}{a_i} = -\frac{\sigma_2^2}{\sigma_1} = C, i = 2, 3, \ldots, l, \tag{12}$$

for some common ratio C depending on the data. Therefore, we have

$$\sigma_1 = Z_1 - \sum_{i=2}^{l} a_i \nu_i = C \sum_{i=2}^{l} a_i^2 + \left(Z_1 - \sum_{i=2}^{l} a_i Z_i \right). \tag{13}$$

In addition, (11) can be written as $\boldsymbol{N} \cdot (\nu_{l+1}, \nu_{l+2}, \ldots, \nu_r)^\top = (Z_{l+1} + b_{l+1} Z_{r+1}, \ldots, Z_r + b_r Z_{r+1})^\top$, where

$$\boldsymbol{N} = \begin{pmatrix} 1 + b_{l+1}^2 & b_{l+1} b_{l+2} & \cdots & b_{l+1} b_r \\ b_{l+1} b_{l+2} & 1 + b_{l+2}^2 & \cdots & b_{l+2} b_r \\ \vdots & \vdots & & \vdots \\ b_{l+1} b_r & b_{l+2} b_r & \cdots & 1 + b_r^2 \end{pmatrix} = \boldsymbol{I}_{r-l} + (b_{l+1}, b_{l+2}, \ldots, b_r)^\top \cdot (b_{l+1}, b_{l+2}, \ldots, b_r),$$

with $\boldsymbol{I}_{r-l}$ being the $(r-l) \times (r-l)$ identity matrix. Obviously, $\boldsymbol{N}$ is a positive definite matrix with $\boldsymbol{N}^{-1} = \boldsymbol{I}_{r-l} - (1 + \sum_{i=l+1}^{r} b_i^2)^{-1} (b_{l+1}, b_{l+2}, \ldots, b_r)^\top \cdot (b_{l+1}, b_{l+2}, \ldots, b_r)$. Hence, we have $\nu_i = Z_i + b_i (Z_{r+1} - \sum_{i=l+1}^{r} b_i Z_i)/(1 + \sum_{i=l+1}^{r} b_i^2), i = l+1, \ldots, r$. Therefore,

$$\sigma_2^2 = \left(\frac{1}{n-1} \right) \left(\frac{1}{1 + \sum_{i=l+1}^{r} b_i^2} \right) \left(Z_{r+1} - \sum_{i=l+1}^{r} b_i Z_i \right)^2 + \frac{C^2}{n-1} \sum_{i=2}^{l} a_i^2. \tag{14}$$

Substituting (14) and (13) into the second equality of (12), we have

$$C^2 \left(\frac{n}{n-1} \sum_{i=2}^{l} a_i^2 \right) + C \left(Z_1 - \sum_{i=2}^{l} a_i Z_i \right) + \left(\frac{1}{n-1} \right) \left(\frac{1}{1 + \sum_{i=l+1}^{r} b_i^2} \right) \left(Z_{r+1} - \sum_{i=l+1}^{r} b_i Z_i \right)^2 = 0. \tag{15}$$

The discriminant of the quadratic equation in C in (15) is given by

$$\Delta = \left(Z_1 - \sum_{i=2}^{l} a_i Z_i\right)^2 - \frac{4n}{(n-1)^2} \frac{\sum_{i=2}^{l} a_i^2}{1 + \sum_{i=l+1}^{r} b_i^2} \left(Z_{r+1} - \sum_{i=l+1}^{r} b_i Z_i\right)^2.$$

Let U_1=Z_1, $\boldsymbol{U}_2$= $(Z_2, Z_3, \ldots, Z_{r+1})^{\top}$ in Lemma 1. This means that on a set of positive probability, (15) has no real solution. Consequently, no ML estimators of σ^2_1, σ^2_2, and the ν_is exist.

For Case II, we have $\nu_1 = \sum_{i=2}^{l} a_i \nu_i$, and $\nu_{r+1} = \sum_{i=l}^{r} b_i \nu_i$ with $a_i \neq 0$, $b_i \neq 0$ for all is. Then the log-likelihood of the data, up to a constant not depending on unknown parameters, can be written as

$$\ln L = -\frac{1}{2} \ln \sigma_1^2 - \frac{n-1}{2} \ln \sigma_2^2 - \frac{\left(Z_1 - \sum_{i=2}^{l} a_i \nu_i\right)^2}{2\sigma_1^2}$$
$$- \frac{\left(Z_{r+1} - \sum_{i=l}^{r} b_i \nu_i\right)^2 + \sum_{i=2}^{r} (Z_i - \nu_i)^2 + \sum_{i=r+2}^{n} (Z_i - \nu_i)^2}{2\sigma_2^2}.$$

Again, ML estimation of $(\nu_2, \nu_3, \ldots, \nu_r, \nu_{r+2}, \ldots, \nu_n)^{\top}$, σ^2_1, and σ^2_2 involve the following estimating equations

$$\sigma_1^2 = \left(Z_1 - \sum_{i=2}^{l} a_i \nu_i\right)^2, \quad \sigma_2^2 = \left(\left(Z_{r+1} - \sum_{i=l+1}^{r} b_i \nu_i\right)^2 + \sum_{i=2}^{r} (Z_i - \nu_i)^2\right) \Big/ (n-1),$$
$$\nu_i = Z_i, \quad \text{for } i = r+2, r+3, \ldots, n,$$

and

$$\frac{a_i}{\sigma_1} + \frac{Z_i - \nu_i}{\sigma_2^2} = 0, i = 2, 3, \ldots, l-1,$$
$$\frac{a_l}{\sigma_1} + \frac{b_l\left(Z_{r+1} - \sum_{i=l}^{r} b_i \nu_i\right) + (Z_l - \nu_l)}{\sigma_2^2} = 0,$$
$$b_i \left(Z_{r+1} - \sum_{k=l+1}^{r} b_k \nu_k\right) + (Z_i - \nu_i) = 0, \quad \text{for } i = l+1, \ldots, r.$$

Following the same idea used in Case I and after some tedious but straightforward calculations, one can easily get, for some common ratio Cdepending on the data,

$$\sigma_1 = C\left(\sum_{i=2}^{l} a_i^2 - \frac{a_l^2 b_l^2}{1+\sum_{i=l}^{r} b_i^2}\right) + \left(Z_1 - \sum_{i=2}^{l} a_i Z_i - \frac{a_l b_l}{1+\sum_{i=l}^{r} b_i^2}\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)\right),$$

$$\sigma_2^2 = \frac{1}{n-1}\left(\frac{C^2 a_l^2\left(1+\sum_{i=l+1}^{r} b_i^2\right) + \left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)^2}{\left(1+\sum_{i=l}^{r} b_i^2\right)} + C^2 \sum_{i=2}^{l-1} a_i^2\right),$$

and a quadratic equation in C

$$nC^2\left(\sum_{i=2}^{l} a_i^2 - \frac{a_l^2 b_l^2}{1+\sum_{i=l}^{r} b_i^2}\right) + (n-1)C\left(\left(Z_1 - \sum_{i=2}^{l} a_i Z_i\right) - \frac{a_l b_l}{1+\sum_{i=l}^{r} b_i^2}\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)\right) + \frac{\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)^2}{1+\sum_{i=l}^{r} b_i^2} = 0.$$

The discriminant is given by

$$\Delta = (n-1)^2\left(\left(Z_1 - \sum_{i=2}^{l} a_i Z_i\right) - \frac{a_l b_l}{1+\sum_{i=l}^{r} b_i^2}\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)\right)^2 - 4n\left(\sum_{i=2}^{l} a_i^2 - \frac{a_l^2 b_l^2}{1+\sum_{i=l}^{r} b_i^2}\right)\frac{\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)^2}{1+\sum_{i=l}^{r} b_i^2}.$$

Again, using Lemma 1 with $U_1 = Z_1$ and $\boldsymbol{U}_2 = (Z_2, Z_3, \ldots, Zr_{+1})^\top$, we conclude that ML estimators do not exist.

Example We present an example to illustrate the idea used in the proof. Let $\mathbf{Z} = (Z_1, Z_2, Z_3)^\top$ be a multivariate normal random vector with mean vector $\mathbf{\nu} = (\nu_1, \nu_1, \nu_1)^\top$ and dispersion matrix $\mathbf{\Sigma} = \mathrm{diag}(\sigma^2_1, \sigma^2_2, \sigma^2_2)$, where $\nu_1 \in \mathbb{R}$, and $\sigma^2_1, \sigma^2_2 > 0$ are all unknown. This is Case II discussed in the proof of Theorem 2. The log-likelihood of the data, up to a constant not depending on unknown parameters, is then given by

$$\Delta = (n-1)^2\left(\left(Z_1 - \sum_{i=2}^{l} a_i Z_i\right) - \frac{a_l b_l}{1+\sum_{i=l}^{r} b_i^2}\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)\right)^2 - 4n\left(\sum_{i=2}^{l} a_i^2 - \frac{a_l^2 b_l^2}{1+\sum_{i=l}^{r} b_i^2}\right)\frac{\left(Z_{r+1} - \sum_{i=l}^{r} b_i Z_i\right)^2}{1+\sum_{i=l}^{r} b_i^2}.$$

Obviously, ML estimation of ν_1, σ^2_1, and σ^2_2 involves the following equations; $\sigma^2_1 = (Z_1 - \nu_1)^2$, $\sigma^2_2 = ((Z_2 - \nu_1)^2 + (Z_3 - \nu_1)^2)/2$, $(Z_1 - \nu_1)/\sigma^2_1 + ((Z_2 - \nu_1) + (Z_3 - \nu_1))/\sigma^2_2 = 0$. Note that the last equation is equivalent to $(Z_2 - \nu_1) + (Z_3 - \nu_1) = -\sigma^2_2/\sigma_1 = C$, for some constant C depending only on the data. A simple calculation gives the following quadratic equation in C, $3C^2 + 2C(2Z_1 - Z_2 - Z_3) + (Z_2 - Z_3)^2 = 0$. The discriminant is then given by $\Delta = 4(2Z_1 - Z_2 - Z_3)^2 - 12(Z_2 - Z_3)^2$. Therefore, no ML estimate of ν_1 exists since there is a positive probability that $\Delta < 0$.

Now we are in a position to prove Theorem 1.

Proof of Theorem 1 If $\dim(V) = 0$, then it is obvious that ML estimates exist. We assume $\dim(V) = m > 0$. If $m < n$, then there exists a $n \times (n-m)$ matrix $\boldsymbol{M} \equiv (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \ldots, \boldsymbol{\alpha}_{n-m})$, such that

$$\boldsymbol{M}^\top \boldsymbol{\mu} = 0, \qquad \text{for all } \boldsymbol{\mu} \in V, \quad (16)$$

where $\boldsymbol{\alpha}_i = (\alpha_{i1}, \alpha_{i2}, \ldots, \alpha_{in})^\top \in V^\perp$, $i = 1, 2, \ldots, n-m$ and the $\boldsymbol{\alpha}_i$s are linearly independent. Note that $\boldsymbol{\nu} = \boldsymbol{C}_n \boldsymbol{\mu}$, then $\boldsymbol{\mu} = \boldsymbol{C}_n^{-1} \boldsymbol{\nu} = \boldsymbol{C}_n^T \boldsymbol{\nu} = \left((1/\sqrt{n})1_n, \boldsymbol{P}_n^\top\right)\boldsymbol{\nu}$. Substituting this into (16), one has

$$0 = \boldsymbol{M}^\top \left((1/\sqrt{n})1_n, \boldsymbol{P}_n^\top\right) \boldsymbol{\nu} = (1/\sqrt{n})\boldsymbol{M}^\top 1_n \nu_1 + \boldsymbol{M}^\top \boldsymbol{P}_n^\top \boldsymbol{\nu}_{(2)} \quad (17)$$

with $\boldsymbol{\nu}_{(2)} = (\nu_2, \nu_3, \ldots, \nu_n)^\top$. Finally, note that $\boldsymbol{1}_n$ can be uniquely decomposed as

$$1_n = \boldsymbol{a} + \boldsymbol{b}, \quad (18)$$

where $\boldsymbol{a} \in V$ and $\boldsymbol{b} \in V^\perp$.

1 If $\boldsymbol{1}_n \in V$ then $\boldsymbol{b} = \boldsymbol{0}$. We only consider the case when $m < n$ since it is obvious that ML estimators of $\boldsymbol{\nu}$, σ^2 and ρ do not exist when $m = n$. Note that $\boldsymbol{M}^\top \boldsymbol{1}_n = \boldsymbol{0}$ in (17). This is obviously Case 2A in Theorem 2, and hence the result follows. The proof here is similar to the one given by Arnold (1981, Section 14.9).

2 If $\boldsymbol{1}_n \in V^\perp$, then $\nu_1 = (1/\sqrt{n})1_n^\top \boldsymbol{\mu} = 0$. Obviously, the ML estimator of σ^2_1 is given by $\hat{\sigma}_1^2 = Z_1^2$, and from Cases B_{11} and B_{21} in Theorem 2, ML estimators of $\boldsymbol{\nu}_{(2)}$, σ^2_2 based on Z_2, Z_3, $\ldots$, Z_n exist if and only if $m < n - 1$.

3 Suppose neither $\boldsymbol{1}_n \in V$ nor $\boldsymbol{1}_n \in V^\perp$. Then $\dim(V) = m < n$, and $\boldsymbol{a} \neq \boldsymbol{0}$, $\boldsymbol{b} \neq \boldsymbol{0}$ from (17). First, we assume $m = n - 1$. Then $\boldsymbol{M} \in V^\perp$ is a $n \times 1$ column vector. Therefore, $\boldsymbol{M}^\top \cdot \boldsymbol{1}_n = \boldsymbol{M}^\top \cdot \boldsymbol{b} \neq 0$ and $\boldsymbol{M}^\top \cdot \boldsymbol{P}^\top_n \neq \boldsymbol{0}$, that is, from (17), ν_1 is a non-trivial linear combination of some $\nu_2, \nu_3, \ldots, \nu_n$. Obviously, this is Case B_{12} in Theorem 2, and therefore, ML estimates exist from Theorem 2. Now if $m < n - 1$, one can easily deduce that, for each $1 \leq i \leq n-m$, $\boldsymbol{\alpha}^\top_i \cdot \boldsymbol{P}^\top_n \neq \boldsymbol{0}$ and for at least one $1 \leq i \leq n-m$, $\boldsymbol{\alpha}^\top_i \cdot \boldsymbol{1}_{(n)} = \boldsymbol{\alpha}^\top_i \cdot \boldsymbol{b} \neq 0$ from assumptions. Obviously, this case comes under Case B_{22} in Theorem 2. Hence from Theorem 2 ML estimates do not exist.

Now we apply Theorem 1 to the linear model $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\boldsymbol{\varepsilon} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{A}(\rho))$, $\boldsymbol{X}$ is a $n \times m$ design matrix with full rank $m \leq n$, and $\boldsymbol{\beta} \in \mathbb{R}^m$ an unknown parameter. Denote $V = \{\boldsymbol{X}\boldsymbol{\beta} : \boldsymbol{\beta} \in \mathbb{R}^m\} \subset \mathbb{R}^n$. We have the following result.

Corollary 1

1 . *No ML estimators of* $\boldsymbol{\beta}$, σ^2, *and* ρ *exist in each of the following cases.*

(a) $\boldsymbol{1}_n \in V$.

(b) *Every column vector of* $\boldsymbol{X}$ *is orthogonal to* $\boldsymbol{1}_n$ *and* $m = n - 1$.

(c) *Neither* $\boldsymbol{1}_n \in V$ *nor* $\boldsymbol{1}_n \in V^{\perp}$ *and* $m < n - 1$.

2 . *ML estimators exist in all other cases.*

Proof 1(b) is trivial since $\boldsymbol{1}_n \in V^{\perp}$ if and only if $\boldsymbol{1}^{\top}_n \boldsymbol{X\beta} = 0$ for all $\boldsymbol{\beta} \in \mathbb{R}^m$ if and only if $\boldsymbol{1}^{\top}_n \boldsymbol{X} = \boldsymbol{0}$. Therefore, the result follows from Theorem 1.

Theorem 1 can be generalized. Let the dispersion matrix of $\boldsymbol{Y}$ in model (1) be given by $\sigma^2((1 - \rho)\, \boldsymbol{I}_n + \rho \boldsymbol{\xi\xi}^{\top})$, where $\boldsymbol{\xi}$ is a given $n \times 1$ column vector with $\boldsymbol{\xi}^{\top}\boldsymbol{\xi} > 1$ and $\sigma^2 > 0$. If $\boldsymbol{\xi} = \boldsymbol{1}_n$, the dispersion matrix will be $\sigma^2 \boldsymbol{A}(\rho)$. It can be shown that the dispersion matrix $\sigma^2((1 - \rho)\, \boldsymbol{I}_n + \rho \boldsymbol{\xi\xi}^{\top})$ is positive definite if and only if $-1/(\boldsymbol{\xi}^{\top}\boldsymbol{\xi} - 1) < \rho < 1$ and $\sigma^2 > 0$. In what follows, we assume that $\boldsymbol{\xi}^{\top}\boldsymbol{\xi} > 1$ and $-1/(\boldsymbol{\xi}^{\top}\boldsymbol{\xi} - 1) < \rho < 1$.

Theorem 1 can be rephrased to encompass this general framework of the dispersion matrix.

Theorem 3 *Consider the problem of ML estimation in Model (1) with dispersion matrix of* $\boldsymbol{Y}$ *given by* $\sigma^2 \tilde{\boldsymbol{A}}(\rho) = \sigma^2((1 - \rho)\boldsymbol{I}_n + \rho \boldsymbol{\xi\xi}^{\top})$.

1 *No ML estimators of* $\boldsymbol{\mu}$, σ^2, *and* ρ *exist in each of the following cases.*

(a) $\boldsymbol{\xi} \in V$.

(b) $\boldsymbol{\xi} \in V^{\perp}$ *and* $\dim(V) = n - 1$.

(c) *Neither* $\boldsymbol{\xi} \in V$ *nor* $\boldsymbol{\xi} \in V^{\perp}$ *and* $\dim(V) < n - 1$.

2 . *ML estimators exist in all other cases, i.e.,*

(d) $\boldsymbol{\xi} \in V^{\perp}$ *and* $\dim(V) < n - 1$;

(e) *Neither* $\boldsymbol{\xi} \in V$ *nor* $\boldsymbol{\xi} \in V^{\perp}$ *and* $\dim(V) = n - 1$.

A proof of Theorem 3 can be fashioned along the lines of the proof given for Theorem 1. The transformation (2) needs to be modified as

$$\tilde{\boldsymbol{C}}_n = \begin{pmatrix} \boldsymbol{\xi}^{\top}/\sqrt{\boldsymbol{\xi}^{\top}\boldsymbol{\xi}} \\ \tilde{\boldsymbol{P}}_n \end{pmatrix}. \quad (19)$$

Here, for $n \geq 2$, $\tilde{\boldsymbol{P}}_n$ is a $(n-1) \times n$ matrix such that the $n \times n$ matrix $\tilde{\boldsymbol{C}}_n$ is orthogonal. Obviously $\tilde{\boldsymbol{P}}_n$, has the same properties as those of $\boldsymbol{P}n$, *i.e.*,

1 $\tilde{\boldsymbol{P}}_n \tilde{\boldsymbol{P}}_n^\top = \boldsymbol{I}_{n-1}$

Therefore, Model (1) with dispersion matrix given by $\sigma^2((1-\rho)\,\boldsymbol{I}n + \rho\xi\xi^\top)$ can now be transformed into Model (4) with $\sigma^2{}_1 = \sigma^2\,(1 + (\xi^\top\xi - 1)\rho) > 0$ and $\sigma^2{}_2 = \sigma^2\,(1-\rho) > 0$.

3. Restricted Maximum Likelihood Estimates

It might be preferable to use the method of restricted maximum likelihood (REML) estimation, which is based on a linear transformation of the data. Assume that $\dim(V) = m \geq 1$, then there exists a $n \times m$ matrix $\boldsymbol{X}$ such that rank $(\boldsymbol{X}) = m$ and V is spanned by the column vectors of $\boldsymbol{X}$. We denote this by $V = \langle \boldsymbol{X} \rangle$. Consider a $n \times (n-m)$ matrix $\boldsymbol{Q}$ whose columns form an orthonormal basis for $V^\perp$. Note that $\boldsymbol{Q}$ satisfies $\boldsymbol{Q}^\top\boldsymbol{Q} = \boldsymbol{I}n_{-}m$ and $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{I}n - \boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top$. Also $\boldsymbol{Q}^\top\boldsymbol{Y}$ follows a $N_{(n-m)}(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^\top((1-\rho)\,\boldsymbol{I}n + \rho\boldsymbol{J}n)\,\boldsymbol{Q})$ distribution. First notice that if $\boldsymbol{1}n \in V$, then $\boldsymbol{Q}^\top\boldsymbol{J}n\boldsymbol{Q} = \boldsymbol{0}$, and so $\boldsymbol{Q}^\top\boldsymbol{Y}$ follows a $N_{(n-m)}(\boldsymbol{0}, \sigma^2\,(1-\rho)\,\boldsymbol{I}n_{-}m)$ distribution. Obviously, σ^2 and ρ are non-identifiable, that is, no REML estimates for σ^2 and ρ exist.

Now if $\mathbb{1}_n \notin V$, we can write $\boldsymbol{1}n = \boldsymbol{a} + \boldsymbol{b}$ where $\boldsymbol{a} \in V$ and $\boldsymbol{b} \in V^\perp$. Therefore, $\boldsymbol{Q}^\top((1-\rho)\,\boldsymbol{I}n + \rho\boldsymbol{J}n)\,\boldsymbol{Q} = (1-\rho)\,\boldsymbol{I}n_{-}m + \rho\boldsymbol{Q}^\top\boldsymbol{b}\boldsymbol{b}^\top\boldsymbol{Q}$. Hence, $\boldsymbol{Q}^\top\boldsymbol{Y}$ follows a $N_{(n-m)}(\boldsymbol{0}, \sigma^2\,((1-\rho)\,\boldsymbol{I}n_{-}m + \rho\boldsymbol{Q}^\top\boldsymbol{b}\boldsymbol{b}^\top\boldsymbol{Q}))$ distribution. Let $\boldsymbol{\xi} = \boldsymbol{Q}^\top\boldsymbol{b} \in \mathbb{R}^{n-m}$, and then apply the same type of transformation $\tilde{\boldsymbol{C}}_{n-m}$ as (19) above but with order of $(n-m) \times (n-m)$. After the transformation, the transformed data $\tilde{\boldsymbol{C}}_{n-m}\,\boldsymbol{Q}^\top\boldsymbol{Y}$ would follow $N_{(n-m)}(\boldsymbol{0}, \boldsymbol{\Sigma})$ with the dispersion matrix $\boldsymbol{\Sigma}$ given by $\boldsymbol{\Sigma} = \sigma^2$ diag $(1 + (\xi^\top\xi - 1)\rho, (1-\rho)\,\boldsymbol{I}n_{-}m_{-1})$. First note that $\xi^\top\xi = \boldsymbol{b}^\top\boldsymbol{Q}\boldsymbol{Q}^\top\boldsymbol{b} = \boldsymbol{b}^\top\boldsymbol{b} - \boldsymbol{b}^\top\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{b} = \boldsymbol{b}^\top\boldsymbol{b} \leq \boldsymbol{1}^\top n\boldsymbol{1}n = n$. Under the condition of $-1/(n-1) < \rho < 1$, it follows that $1 + (\xi^\top\xi - 1)\rho > 0$. Hence if $n \leq m + 1$ (or more accurately, $n = m + 1$ since $n > m$), no REML estimates exist for $\tilde{\sigma}_1^2 = \sigma^2(1 + (\boldsymbol{\xi}^\top\boldsymbol{\xi} - 1)\rho) > 0$ and $\tilde{\sigma}_2^2 = \sigma^2(1-\rho) > 0$. On the other hand, if $n > m + 1$, REML estimates exist for $\tilde{\sigma}_1^2$ and $\tilde{\sigma}_2^2$, or equivalently, for σ^2 and ρ, exist. In summary, we have the following theorem for REML estimates.

Theorem 4 *Consider the problem of REML estimation in model (1).*

1. *No REML estimators of σ^2 and ρ exist in each of the following cases.*

    (a) $\boldsymbol{1}n \in V$.

    (b) $\mathbb{1}_n \notin V$ *and* $\dim(V) = n - 1$.

2 . *REML estimators exist in all other cases, i.e.,*

    (c) $\mathbb{1}_n \notin V$ *and* $\dim(V) < n - 1$.

In the framework of Theorem 3, we have the following result.

Theorem 5 *Consider the problem of REML estimation in Model (1) with the dispersion matrix of* $\boldsymbol{Y}$ *given by* $\sigma^2((1-\rho)\boldsymbol{I}_n+\rho\xi\xi^\top)$.

1. *No REML estimators of* σ^2 *and ρ exist in each of the following cases.*

   (a) $\xi \in V$.

   (b) $\xi \notin V$ *and* $\dim(V) = n-1$.

2. *REML estimators exist in all other cases, i.e.,*

   (c) $\xi \notin V$ *and* $\dim(V) < n-1$.

Note that the existence of REML estimates in Theorems 4 and 5 is in the almost sure sense. However, REML estimates in cases (a) and (b) do not exist for every data set.

4. A Comparison

We will now present an example for which ML estimates do not exist in the framework of model (1), but ML estimates do exist if model (1) is modified to conform to the model of Demidenko & Massam (1999, pp. 431).

Consider model (1) with $n = 3$ and $\boldsymbol{\mu}$ an unknown vector belonging to the subspace $V = \langle (1, 1, 1)^\top \rangle$ of $\mathbb{R}^3$, spanned by the vector $\boldsymbol{1}_3 = (1, 1, 1)^\top$. Write $\boldsymbol{\mu} = 1_3\beta, \beta \in \mathbb{R}$. Obviously, there are three parameters $\beta \in \mathbb{R}, -1/2 < \rho < 1$, and $\sigma^2 > 0$ to be estimated in this example. Note that $\boldsymbol{1}_3 \in V$. Therefore, no matter what data are available, ML estimates of the parameters do not exist. The same conclusion was also obtained by Arnold (1981), Section 14.9.

To be more specific, let us work with the data $\boldsymbol{y} = (1, -1, 0)^\top$. Then under the transformation

$$\boldsymbol{C}_3 = \begin{pmatrix} 1/\sqrt{3} & 1/\sqrt{3} & 1/\sqrt{3} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \end{pmatrix},$$

we observe that in the transformed model (4), $\boldsymbol{z} = (z_1, z_2, z_3)^\top = \boldsymbol{C}_3\boldsymbol{y} = (0, \sqrt{2}, 0)^\top$, $\boldsymbol{\nu} = \boldsymbol{C}_3\boldsymbol{\mu} = (\sqrt{3}, 0, 0)\beta$ with $\beta \in \mathbb{R}, \boldsymbol{\varepsilon}^* \sim N_3(\boldsymbol{0}, \mathrm{diag}(\sigma^2_1, \sigma^2_2\boldsymbol{I}_2))$ with $\sigma^2_1 = \sigma^2(1 + 2\rho)$ and $\sigma^2_2 = \sigma^2(1-\rho)$, and the log-likelihood of the data $\boldsymbol{z}$ is given by

$$\ln L = -\frac{1}{2}\ln\sigma_1^2 - \frac{3-1}{2}\ln\sigma_2^2 - \frac{(z_1-\sqrt{3}\beta)^2}{2\sigma_1^2} - \frac{(z_2-0)^2+(z_3-0)^2}{2\sigma_2^2}$$
$$= -\frac{1}{2}\ln\sigma_1^2 - \ln\sigma_2^2 - \frac{3\beta^2}{2\sigma_1^2} - \frac{1}{\sigma_2^2}.$$

Note that if $-1/2 < \rho < 1$, the mapping between $(\sigma^2, \rho) \in \{(\sigma^2, \rho); \sigma^2 > 0, -1/2 < \rho < 1\}$ in model (1) and $(\sigma^2_1, \sigma^2_2) \in \{(\sigma^2_1, \sigma^2_2); \sigma^2_1 > 0, \sigma^2_2 > 0\}$ in the transformed model (4) is one-to-one, and all parameters are free to choose without restriction. Obviously, there is no ML estimate.

Now we consider the existence theorem (Theorem 3.1) in Demidenko & Massam (1999) as applied to the data $\boldsymbol{y}= (1, -1, 0)^{\mathsf{T}}$. If we assume $0 \le \rho < 1$, ML estimates of β, ρ, and σ^2 do exist and they are given by $\hat{\beta} = 0, \hat{\rho} = 0$, and $\hat{\sigma}^2 = 2/3$.

It seems that there is a contradiction between our Theorem 1 and Theorem 3.1 of Demidenko & Massam (1999). However, notice that Theorem 3.1 of Demidenko & Massam (1999) can only be applied to our model (1) when $0 \le \rho < 1$. Under the condition $0 \le \rho < 1$, the parameter space $\{(\sigma^2, \rho); \sigma^2 > 0, 0 \le \rho < 1\}$ in model (1) will be one-to-one mapped into $\{(\sigma^2_1, \sigma^2_2); \sigma^2_1 > 0, \sigma^2_2 > 0, \sigma^2_1 \ge \sigma^2_2\}$ in the transformed model (4). Therefore, in the example we considered, if an additional constraint $\sigma^2_1 \ge \sigma^2_2$ is added when ML estimates are calculated, ML estimates for β, σ^2_1, and σ^2_2 will be given by $\hat{\beta} = 0, \hat{\sigma}_1^2 = \hat{\sigma}_2^2 = 2/3$, which implies $\hat{\beta} = 0, \hat{\rho} = 0$, and $\hat{\sigma}^2 = 2/3$. In summary, the conditions $-1/(n-1) < \rho < 1$ (ours) and $0 \le \rho < 1$ (Demidenko & Massam 1999) lead to different conclusions on the existence of ML estimates.

We now consider the work of Birkes & Wulff (2003). To begin with, we put model (1) in the framework of the variance components model ofBirkes & Wulff (2003). Following the notation from Birkes & Wulff (2003) and with the spectral decomposition of $\boldsymbol{J}n$, (3.1) of Birkes & Wulff(2003, pp. 38) can be written as

$$V_\psi = \sigma^2 A(\rho) = \sum_{j=1}^{2}\left(\mathbf{h}_j^{\mathsf{T}}\psi\right)\boldsymbol{E}_j,$$

where $\mathbf{h}_1 = (n, 1)^{\mathsf{T}}, \mathbf{h}_2 = (0, 1)^{\mathsf{T}}, \psi = \sigma^2(\rho, 1-\rho)^{\mathsf{T}}, \boldsymbol{E}_1 = (1/n)\boldsymbol{J}_n$, and $\boldsymbol{E}_2 = \boldsymbol{I}n - (1/n)\, \boldsymbol{J}n$. Therefore, we have

$$\Psi_{pd} = \left\{\psi \in R^2 : \mathbf{h}_j^{\mathsf{T}}\psi > 0, \text{ for } j = 1, 2\right\}$$
$$= \left\{\psi = \sigma^2(\rho, 1-\rho)^{\mathsf{T}}; -1/(n-1) < \rho < 1, \sigma^2 > 0\right\}.$$

We now apply Theorem 5.1 of Birkes & Wulff (2003, pp. 42) to model (1).

Theorem 6 (*A direct translation of Birkes & Wulff*, 2003). *Consider the problem of ML estimation in model*(1).

1. *If* $\mathbf{1}_n \in V^{\perp}$, *and* $\dim(V) < n - 1$, *then with Model*(1)*constrained to a relatively closed subset of* Ψ_{pd}, *ML estimates exist with probability* 1.

2. *If* $\mathbf{1}_n \in V$ *with* $\dim(V) \leq n - 1$, *or neither* $\mathbf{1}_n \in V$ *nor* $\mathbf{1}_n \in V^{\perp}$ *with* $\dim(V) < n - 1$, *ML estimates do not exist for the entire* Ψ_{pd}.

Parallel result holds for the existence of REML estimates.

Theorem 7 (*A direct translation of Birkes & Wulff* (*2003, Corollary 7.3*)) *Consider the problem of REML estimation in Model*(1). *If* $\mathbf{1}_n \notin V$ *and* $\dim(V) < n - 1$, *then with the model*(1)*constrained to a relatively closed subset of* Ψ_{pd}, *REML estimates exist with probability* 1.

Theorems 1 and 4 of this paper cover every possible scenario, whereas Theorems 6 and 7 above cover only a subset of all possible scenarios. However, it should be pointed out that Theorems 1 and 4 work for the entire parameter space Ψ_{pd}, whereas Theorems 6 and 7 are operational for any relatively closed subset of Ψ_{pd}.

Acknowledgments

Acknowledgments. The referee had a considerable impact on the revision of the paper. The identification of two critical papers by the referee has led us to re-examine the issues involved thoroughly and hence to a better revision.

References

Arnold, S. F. (1979). Linear models with exchangeably distributed errors, *J. Amer. Statist. Assoc.* **74**, 194–199.

Arnold, S. F. (1981). *The theory of linear models and multivariate analysis.* New York : Wiley.

Birkes, D. & Wulff, S. (2003). Existence of maximum likelihood estimates in normal variance-components models. *J. Statist. Plann. Inference*. **113**, 35–47.

Chaubey, Y. P. (1980). Application of the method of MINQUE for estimation in regression with intraclass covariance matrix. *SankhyB*.**42**, 28–32.

Demidenko, E. & Massam, H. (1999). On the existence of the maximum likelihood estimate in variance components models. *SankhyA*.**61**, 431–443.

Hartley, H. O. & Jayatillake, K. S. E. (1973). Estimation for linear models with unequal variances. *J. Amer. Statist. Assoc.* **68**, 189–192.

Hartley, H. O. & Rao, J. N. K. (1967). Maximum-likelihood estimation for the mixed analysis of variance model. *Biometrika*. **54**, 93–108.

Hartley, H. O., Rao, J. N. K. & Kiefer, G. (1969). Variance estimation with one unit per stratum. *J. Amer. Statist. Assoc.* **64**, 841–851.

Horn, S. D. & Horn, R. A. (1975). Comparison of estimators of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **70**,872–879.

Horn, S. D., Horn, R. A. & Duncan, D. B. (1975). Estimating heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **70**,380–385.

Neyman, J. & Scott, E. L. (1948). Consistent estimates based on partially consistent observations. *Econometrica*. **16**, 1022–1036.

Press, S. J. (1982). *Applied Multivariate Analysis*, 2nd edn. Huntington : R.E. Krieger Publishing Company.

Putter, J. (1967). Orthogonal bases of error spaces and their use for investigating the normality and variances of residuals. *J. Amer. Statist. Assoc.* **62**, 1022–1036.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **65**, 161–172.

Rao, C. R. (1972). Estimation of variance and covariance components in linear models. *J. Amer. Statist. Assoc.* **67**, 112–115.

Rao, C. R. & Kleffe, K. (1988). *Estimation of variance components and applications.* New York : North-Holland.

Rao, C. R. & Rao, M. B. (1998). *Matrix algebra and its applications to statistics and econometrics.* River Edge , New Jersey : World Scientific.

Rao, P. S. R. S. (1997). *Variance components estimation: Mixed models, methodologies and applications*. London : Chapman & Hall.

Wiorkowski, J. J. (1975). Unbalanced regression analysis with residuals having a covariance structure of intra-class form. *Biometrics* **31**,611–618.