Estimating parameters in one-way analysis of covariance model with short-tailed symmetric error distributions

Birdal Şenoğlu

Department of Statistics, Eskişehir Osmangazi University, Eskişehir 26480, Turkey

Received 18 March 2005; received in revised form 13 February 2006

Abstract

We consider one-way analysis of covariance (ANCOVA) model with a single covariate when the distribution of error terms are short-tailed symmetric. The maximum likelihood (ML) estimators of the parameters are intractable. We, therefore, employ a simple method known as modified maximum likelihood (MML) to derive the estimators of the model parameters. The method is based on linearization of the intractable terms in likelihood equations. Incorporating these linearizations in the maximum likelihood, we get the modified likelihood equations. Then the MML estimators which are the solutions of these modified equations are obtained. Computer simulations were performed to investigate the efficiencies of the proposed estimators. The simulation results show that the proposed estimators are remarkably efficient compared with the conventional least squares (LS) estimators.

Keywords: Covariance analysis; Modified likelihood; Short-tailed symmetric family; Non-normality; Efficiency

1. Introduction

Analysis of covariance (ANCOVA), in its most general definition, is a combination of regression analysis with an analysis of variance (ANOVA). The prime advantage of using the ANCOVA model is to reduce the variability of the random error that is associated with covariates and to result in more precise estimates and more powerful tests, see, for example, [9,24]. The simplest and most important ANCOVA model in a completely randomized design is

\[ y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}_i) + e_{ij} \quad (i = 1, 2, \ldots, c; \quad j = 1, 2, \ldots, n_i), \]

where \( e_{ij} \) are independently and identically distributed (iid) random errors and \( x_{ij} \) represents the value of the non-stochastic covariate, or supplementary information, corresponding to \( y_{ij} \). It is assumed that there is a linear relationship between the response variable \( y \) and the covariate \( x \).

Traditionally, the distribution of the error terms is assumed to be normal. However, in many applications, populations that are far from being normal are more prevalent, see [10,6,5,8,12–14,18]. The violation of the normality assumption can adversely affect the efficiencies of the least squares (LS) estimators, i.e., the LS estimators have relatively low efficiency when the error term has a non-normal distribution. In fact, the impact of violating normality on the performance of
the estimators is often overlooked. It has, therefore, been of enormous interest to develop efficient estimators for the non-normal error distributions.

Much work has been done about how to obtain efficient parametric estimators in ANCOVA under long-tailed symmetric error distributions, see, for example, [4,1]. However, there is no previous work about short-tailed symmetric error distributions. The originality of this work lies in assuming short-tailed symmetric error distributions for the one-way ANCOVA model. Since the maximum likelihood method does not provide explicit estimators for the parameters.

It is clear that the kurtosis of this family is not defined for \( a < r \). The values of its kurtosis \( a > r \), and the constant

\[
\begin{align*}
C = \frac{1}{\sum_{j=0}^{r} \binom{r}{j} (\lambda/2r)^j (2j)!/2^j (j)!}
\end{align*}
\]

It should also be noted that \( E(y) = \mu \) and \( V(y) = \mu^2 \sigma^2 \). The kurtosis of this family (\( \mu_k / \mu_k^2 \)) assumes values between 1 and 3 for all \( r \) and \( a \) values. Here, \( \mu_k \) is defined as \( E(z_k) \). \( f(z) \) is unimodal for \( a \leq 0 \) and is generally multimodal for \( a > 0 \). The values of its kurtosis (\( \beta_2 \)) are given in the following table; \( a < r (\lambda > 0) \), see [18].

<table>
<thead>
<tr>
<th>( r )</th>
<th>( a = -0.5 )</th>
<th>( 0.0 )</th>
<th>( 0.5 )</th>
<th>( 1.0 )</th>
<th>( 1.5 )</th>
<th>( 2.5 )</th>
<th>( 3.5 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>( \beta_2 = 2.559 )</td>
<td>2.437</td>
<td>2.265</td>
<td>2.026</td>
<td>1.711</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>4</td>
<td>( \beta_2 = 2.464 )</td>
<td>2.370</td>
<td>2.255</td>
<td>2.118</td>
<td>1.957</td>
<td>1.591</td>
<td>1.297</td>
</tr>
</tbody>
</table>

It is clear that the kurtosis of this family is not defined for \( a > r \), and therefore the dashed entries are used when \( r = 2 \) and \( a = 2.5 \) and 3.5.

2. Likelihood equations

Consider the observations \( y_{ij} (1 \leq i \leq c, 1 \leq j \leq n_i) \) in the \( i \)th treatment. Let \( y_{i(1)} \leq y_{i(2)} \leq \cdots \leq y_{i(n_i)} \) be the order statistics obtained by arranging \( y_{ij} \) in ascending order of magnitude. Since complete sums are invariant to ordering, i.e., \( \sum_{i=1}^{n} f(y_i) = \sum_{i=1}^{n} f(y_{ij}) \) where \( f(y) \) is any function of \( y \), the likelihood equations for estimating \( \mu, \tau_i \) (\( i \leq i \leq c \)), \( \beta \) and \( \sigma \) can be expressed in terms of \( z_{i(j)} \) as follows:

\[
\begin{align*}
\hat{\lambda} \ln L \hat{\sigma} \hat{\mu} &= \frac{-\lambda}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} g(z_{i(j)}) + \frac{1}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} z_{i(j)} = 0, \\
\hat{\lambda} \ln L \hat{\sigma} \hat{\tau}_i &= \frac{-\lambda}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} g(z_{i(j)}) + \frac{1}{\sigma} \sum_{j=1}^{n_i} z_{i(j)} = 0, \\
\hat{\lambda} \ln L \hat{\sigma} \hat{\beta} &= \frac{-\lambda}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{i[j]} - \bar{x}_{i[j]}) g(z_{i(j)}) + \frac{1}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{i[j]} - \bar{x}_{i[j]}) z_{i(j)} = 0
\end{align*}
\]
We also write
\[
\frac{\partial \ln L}{\partial \sigma} = -\frac{N}{\sigma} - \frac{\lambda}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} z_{i(j)} g(z_{i(j)}) + \frac{1}{\sigma} \sum_{i=1}^{c} \sum_{j=1}^{n_i} z_{i(j)}^2 = 0,
\]
where
\[
z_{i(j)} = (y_{i(j)} - \mu - \tau_i - \beta(x_{i[j]} - \bar{x}_{[i]}))/\sigma, \quad 1 \leq j \leq n_i \quad \text{(for a given } \beta),
\]
\[
g(z) = z/[1 + (\lambda/2r)z^2].
\]

$z_{i(j)}$ are the ordered variates and $(y_{i[j]}, x_{i[j]})$ is that pair of observations $(y_{ij}, x_{ij})$ which corresponds to $z_{i(j)} (1 \leq j \leq n_i)$; $(y_{i[j]}, x_{i[j]})$ may be called the concomitant of $z_{i(j)}$. Note that $\bar{x}_{[i]}$ in (2.2) is the overall mean of $x_{i[j]}$'s and is equal to $\sum_{i=1}^{c} \sum_{j=1}^{n_i} x_{i[j]}/N (N = \sum_{i=1}^{c} n_i)$.

The maximum likelihood (ML) estimators are the solutions of the equations in (2.1). However, they do not admit explicit solutions. Iteration is the only way to solve these equations, but that is difficult and time consuming indeed, since there are $c + 2$ equations to iterate simultaneously, see, for example, [11,21,22,18]. The ML estimators are, therefore, elusive. In such situations, we obtain modified likelihood equations which have no such difficulties. The solutions of these equations are unique and we shall call them MML estimators.

3. MML estimators

Let $t_{i(j)} = E(z_{i(j)})$ ($1 \leq i \leq c$, 1 $\leq j \leq n_i$) be the expected values of the standardized order statistics $z_{i(j)}$. To obtain the MML estimators, we linearize $g(z_{i(j)})$ as a Taylor series around $t_{i(j)}$ [17,18]. Recalling that a differentiable function is almost linear in a small interval $a < z < b$ and $z_{i(j)}$ converges to its expected value $t_{i(j)}$ as $n$ becomes large. Then, we obtain a linear approximation of $g(z_{i(j)})$ from the first two terms of a Taylor series expansion. That is
\[
g(z_{i(j)}) \approx g(t_{i(j)}) + (z_{i(j)} - t_{i(j)}) \left\{ \frac{d}{dz} g(z) \right\}_{z=t_{i(j)}},
\]
\[
g(z_{i(j)}) \approx \gamma_{ij} + \gamma_{ij} z_{i(j)}, \quad (1 \leq i \leq c, 1 \leq j \leq n_i),
\]
where
\[
\gamma_{ij} = \left\{ \frac{d}{dz} g(z) \right\}_{z=t_{i(j)}} \quad \text{and} \quad \alpha_{ij} = g(t_{i(j)}) - \gamma_{ij} t_{i(j)}.
\]

The exact values of $t_{i(j)}$ are not available, however, for $n \geq 10$, we use their approximate values obtained from the equation
\[
\int_{-\infty}^{t_{i(j)}} f(z) \, dz = \frac{j}{n_i + 1}, \quad 1 \leq i \leq c, \quad 1 \leq j \leq n_i,
\]
see, [19,18]. Approximating validity of this equation stems from the fact that $F(z_{i(j)})$ has a beta distribution $B(j, n - j + 1)$ with expected value $j/(n_i + 1)$ ($1 \leq i \leq c, 1 \leq j \leq n_i$), since $F(z) = \int_{-\infty}^{z} f(z) \, dz$ is the cumulative distribution function and has a uniform (0,1) distribution. Note that as $n_i \to \infty$, $t_{i(j)}$ is exactly equal to $E(z_{i(j)})$. The use of these approximate values does not adversely affect the efficiency of the MML estimators.

It follows that
\[
\alpha_{ij} = (\lambda/r)^3/(1 + (\lambda/2r)^2) \quad \text{and} \quad \gamma_{ij} = (1 - (\lambda/2r)^2)/(1 + (\lambda/2r)^2)^2 \quad (t = t_{i(j)}).
\]
We also write
\[
\beta_{ij} = 1 - \hat{\gamma}_{ij} \quad (1 \leq i \leq c, 1 \leq j \leq n_i).
Remark. For \(a \leq 0(\lambda \leq 1)\), all of the \(\beta_{ij}\) coefficients are greater than or equal to zero.

Incorporating (3.1) in (2.1), we obtain the modified likelihood equations. The solutions of these equations are the following MML estimators (\(i = 1, 2, \ldots, c\)):

\[
\hat{\mu} = \hat{\mu}_{[i]} - \hat{\beta}_{[i]} h_{[i]}, \\
\hat{\tau}_i = \hat{\mu}_{[i]} - \hat{\mu}_{[i]} - \hat{\beta}(\hat{\mu}_{[i]} - \hat{\mu}_{[i]}), \\
\hat{\beta} = K - L \hat{\sigma}
\]

and

\[
\hat{\sigma} \left\{ -B + \sqrt{B^2 + 4AC} \right\} / 2\sqrt{N(N - c - 1)},
\]

where

\[
\hat{\mu}_{[i]} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \beta_{ij} y_i [j] / m, \quad \hat{\mu}_{[i]} = \sum_{i=1}^{n_i} \beta_{ij} y_i [j] / m_i, \quad \hat{\mu}_{[i]} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \beta_{ij} (x_i [j] - \bar{x}_{[i]}) / m,
\]

\[
\hat{\mu}_{x,[i]} = \sum_{i=1}^{n_i} \beta_{ij} (x_i [j] - \bar{x}_{[i]}) / m_i, \quad m_i = \sum_{i=1}^{n_i} \beta_{ij}, \quad m = \sum_{i=1}^{c} m_i, \quad K = \frac{E_{XY}}{E_{XX}},
\]

\[
L = \frac{\lambda \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_i [j] - \bar{x}_{[i]}) \beta_{ij}}{E_{XX}}, \quad A = N,
\]

\[
B = \lambda \sum_{i=1}^{c} \sum_{j=1}^{n_i} \alpha_{ij} (y_i [j] - \hat{\mu}_{i}) + K (\hat{\mu}_{x[i]} - (x_i [j] - \bar{x}_{[i]})],
\]

\[
C = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \beta_{ij} (y_i [j] - \hat{\mu}_{i}) + K (\hat{\mu}_{x[i]} - (x_i [j] - \bar{x}_{[i]})]^2,
\]

\[
S_{XX} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \beta_{ij} (x_i [j] - \bar{x}_{[i]})^2, \quad S_{XY} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \beta_{ij} (x_i [j] - \bar{x}_{[i]}) y_i [j], \quad T_{XX} = \sum_{i=1}^{c} m_i \hat{\mu}_{x,[i]}^2,
\]

\[
T_{XY} = \sum_{i=1}^{c} m_i \hat{\mu}_{x,[i]} \hat{\mu}_{i}, \quad E_{XX} = S_{XX} - T_{XX}, \quad E_{XY} = S_{XY} - T_{XY}.
\]

It is clear that the MML estimators above have all closed form algebraic expressions. For \(a > 0\), the MML estimators are asymptotically equivalent to ML estimators when regularity conditions hold [3,23]. For small \(n\), they are almost fully efficient in terms of the minimum variance bounds (MVBs) and have little or no bias, see, for example, [22,14]. For \(a > 0\), however, MVB do not generally exist, therefore, the variances of the MML estimators and the LS estimators are compared with each other. Moreover, they have the invariance property like the ML estimators. Simulation results show that the MML estimators are more efficient than the traditional LS estimators as expected (see Section 5).

It should also be noted that the MML estimators are robust estimators. The robustness of the MML methodology is due to the decreasing sequence of the \(\beta_{ij}\) coefficients till the middle value and then increasing sequences of \(\beta_{ij}\)’s in a symmetric fashion. Since small weights are assigned to the middle order statistics (inliers), see [19], this depletes the dominant effects of the inliers and makes MML estimators robust.
Remark. For $a > 0$ ($\lambda > 1$), some of the $\beta_{ij}$ coefficients in the middle can be negative [19]. This makes $\hat{\sigma}$ not real or negative. Therefore, the coefficients $\alpha_{ij}$ and $\gamma_{ij}$ are replaced by $\alpha_{ij}^\ast$ and $\gamma_{ij}^\ast$, respectively:

$$
\alpha_{ij}^\ast = ((\lambda/a) t^3 + (1 - 1/\lambda)) t^2 (1 + (\lambda/2a) t^2)^2, \\
\gamma_{ij}^\ast = ((1/\lambda) - (\lambda/2a) t^2) / (1 + (\lambda/2a) t^2)^2 \quad (\lambda > 1 \text{ and } t = t_{i(j)}).
$$

(3.6)

It should be realized that this alternative linear approximation does not alter the asymptotic properties of the estimators since $z_{i(j)} - \hat{z}_{i(j)}$ is asymptotically zero and, consequently, $\alpha_{ij} + \gamma_{ij} z_{i(j)} = \alpha_{ij}^\ast + \gamma_{ij}^\ast z_{i(j)}$ ($1 \leq i \leq c, 1 \leq j \leq n_i$), see [18] for more detailed information. Thus $\hat{\sigma}$ is always real and positive. We also write $\hat{\beta}_{ij} = 1 - \check{\lambda}_{ij}^\ast$, which is greater than or equal to zero for all $i$ and $j$. It should also be noted that

$$
\sum_{j=1}^{n_i} \alpha_{ij} = \sum_{j=1}^{n_i} \alpha_{ij}^\ast = 0, \quad \beta_{ij} = \beta_{i,n_i-j+1}, \quad \beta_{ij}^\ast = \beta_{i,n_i-j+1}^\ast \quad (1 \leq i \leq c, 1 \leq j \leq n_i).
$$

This is because of the symmetry, see also [18].

Remark. The LS estimators $\hat{\mu}$, $\hat{\tau}_i$, $\check{\beta}$ and $\hat{\sigma}$ are obtained from (3.4) simply by equating $\alpha_{ij}$ to zero and $\beta_{ij}$ to 1 ($1 \leq i \leq c, 1 \leq j \leq n_i$). In particular,

$$
\check{\beta} = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \ddot{x}_i) y_{ij} / \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - \ddot{x}_i)^2
$$

and

$$
\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{c} \sum_{j=1}^{n_i} (y_{ij} - \ddot{y}_i - \check{\beta}(x_{ij} - \ddot{x}_i))^2}{(N - c - 1) \mu_2}},
$$

(3.7)

where

$$
\mu_2 = \sum_{j=0}^{r} \left( \frac{r}{2^r} \right) ^j \left( \frac{(2(i + j))!}{2^j + j(i + j)!} \right) / \sum_{j=0}^{r} \left( \frac{r}{2^r} \right) ^j \left( \frac{(2j)!}{2^j(j)!} \right),
$$

(3.8)

$\mu_2$ is the variance of the short-tailed symmetric distribution and is used in the expression for $\hat{\sigma}$ as a bias correction.

Computations. To initiate the ordering, the concomitants $(y_{ij}, x_{ij})$ are obtained from the order statistics $w_{i(j)} = y_{i(j)} - \check{\beta}x_{i(j)}$ ($1 \leq i \leq c, 1 \leq j \leq n_i$) by using the LSE $\check{\beta}$ in (3.7). It should be noted that the ordering of $z_{ij} = (y_{ij} - \mu - \tau_i - \check{\beta}(x_{ij} - \ddot{x}_i))/\sigma$ does not depend on $\mu$, $\tau_i$ and $\sigma$, because $\mu$ and $\tau_i$ are additive constants and $\sigma$ is positive. Then the MML estimators $\hat{\beta}$ is calculated from the concomitants $(y_{ij}, x_{ij})$. We repeat the computations one more time to obtain the revised values of $\hat{\sigma}$ and $\check{\beta}$ by replacing $\hat{\sigma}$ by $\check{\beta}$ and then compute $\hat{\mu}$ and $\check{\tau}_i$ ($1 \leq i \leq c$). Only two iterations are needed for the estimators to stabilize sufficiently enough.

4. Properties of the estimators

The Fisher information matrix of the MLE $\delta_1 = \hat{\mu}_i$ (where $\mu_i = \mu + \tau_i, 1 \leq i \leq c$), $\delta_2 = \check{\beta}$ and $\delta_3 = \hat{\sigma}$ are given by the elements of the symmetric matrix

$$
I_{ij} = -E \left( \frac{\partial^2 \ln L}{\partial \delta_i \partial \delta_j} \right)_{i,j=1,2,3}.
$$

(4.1)

The modified likelihood equations are asymptotically equivalent with the corresponding likelihood equations, see [7]. The asymptotic variances and covariances of the MML estimators are, therefore, equal to the diagonal elements of $I^{-1}$. 


For the short-tailed symmetric distribution, the elements of the Fisher information $I(\mu, \beta, \sigma)$ matrix are given by

$$
I_{11} = \frac{n_i}{\sigma^2} D, \quad I_{12} = \frac{n_i}{\sigma^2} Ds_x, \quad I_{13} = 0, \quad I_{22} = \frac{n_i}{\sigma^2} Ds_{xx}, \quad I_{23} = 0, \quad I_{33} = \frac{n_i}{\sigma^2} N_i D^*,
$$

(4.2)

where

$$
D = 1 - \lambda E_{0,1} + \frac{j^2}{r} E_{1,2}, \quad D^* = 2 - 3\lambda E_{1,1} + \frac{j^2}{r} E_{2,2},
$$

$$
E_{u,v} = C \sum_{j=0}^{r-v} \binom{r-v}{j} \left( \frac{\lambda}{2r} \right)^j \left( \frac{2(u+j)!}{2u+j(u+j)} \right), \quad s_{xx} = \sum_{i=1}^c \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / n_i
$$

and

$$
s_x = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i) / n_i.
$$

It is easy to invert $I$, and we will write

$$
\text{Cov}(\hat{\mu}_i, \hat{\beta}, \hat{\sigma}) = I^{-1} = \frac{\sigma^2}{n_i} \begin{bmatrix} v_{11} & v_{12} & v_{13} \\ v_{21} & v_{22} & v_{23} \\ v_{31} & v_{32} & v_{33} \end{bmatrix}, \quad (1 \leq i \leq c).
$$

(4.3)

Note that $v_{22}, v_{23}$ and $v_{33}$ are same for all $i = 1, 2, \ldots, c$.

In particular, the asymptotic variance of $\hat{\mu}_i$ is

$$
\text{V}(\hat{\mu}_i) \approx \frac{\sigma^2}{n_i D} \frac{s_{xx}}{s_x - s_{x}^2}.
$$

(4.4)

The following results are true asymptotically.

**Lemma 1.** Asymptotically ($n_i$ tends to infinity), the estimators $\hat{\mu}_i$ ($\beta$ known) and $\hat{\beta}$ ($\sigma$ known) are independently distributed of $\hat{\sigma}$.

**Proof.** To prove this we note that asymptotically $\partial^{r+s} \ln L^*/\partial \mu_i^r \partial \sigma^s$ and $\partial^{r+s} \ln L^*/\partial \beta^r \partial \sigma^s$ are equivalent to $\partial^{r+s} \ln L/\partial \mu_i^r \partial \sigma^s$ and $\partial^{r+s} \ln L/\partial \beta^r \partial \sigma^s$ and

$$
E(\partial^{r+s} \ln L/\partial \mu_i^r \partial \sigma^s) = 0 \quad \text{and} \quad E(\partial^{r+s} \ln L/\partial \beta^r \partial \sigma^s) = 0
$$

for all $r \geq 1$ and $s \geq 1$; see [2]. \(\square\)

**Lemma 2.** Asymptotically, the distribution of $\hat{\mu}_i - \mu_i$ and $\hat{\beta} - \beta$ are jointly distributed as bivariate normal with zero means and variance–covariance matrix

$$
\begin{bmatrix} -E(\partial^2 \ln L/\partial \mu_i^2) & -E(\partial^2 \ln L/\partial \mu_i \partial \beta) \\ -E(\partial^2 \ln L/\partial \mu_i \partial \beta) & -E(\partial^2 \ln L/\partial \beta^2) \end{bmatrix}^{-1}.
$$

(4.5)

**Proof.** The result follows from the fact that $\partial \ln L/\partial \mu_i$ and $\partial \ln L/\partial \beta$ are asymptotically jointly distributed as bivariate normal and the expected values of the first and second derivatives of $\ln L^*$ are exactly the same as those of $\ln L$ [23]. \(\square\)

**Lemma 3.** Conditional on $\sigma$ known, the distribution of $\hat{\mu}_i - \mu_i$ is asymptotically normal

$$
N \left( 0, \frac{\sigma^2}{n_i D} \frac{s_{xx}}{s_x - s_{x}^2} \right) \text{ (see (4.4)).}
$$
To prove Lemma 3, we can use, alternatively, 3-moment chi-square approximation for $\hat{\sigma}^2$ and sample size, they are not reported here. Without loss of generality, we choose the following setting in our simulation: error distribution in (1.2). Because the results for the $B/\sqrt{NC}$ and realizing that $B/\sqrt{NC} \approx 0$, it is easy to show that

$$
\frac{\partial \ln L^*}{\partial \sigma} \approx \frac{N}{\sigma^2} \left\{ \frac{1}{N} C - \sigma^2 \right\} \quad (1 \leq i \leq c).
$$

Since the expective values of the successive derivatives of $\ln L^* / \hat{\sigma}$ are exactly the same as those of

$$
\frac{\partial \ln L}{\partial \sigma} = \frac{N}{\sigma^3} \left\{ \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (y_{ij} - \mu)^2 - \sigma^2 \right\},
$$

where $y_{ij}$ ($1 \leq i \leq c$, $1 \leq j \leq n_i$) is a random sample from $N(\mu, \sigma^2)$, the result follows. □

**Lemma 4.** Conditional on $\mu_i$ and $\beta$ known, the distribution of $N \hat{\sigma}^2(\mu_i, \beta)/\sigma^2$ is asymptotically chi-square with $N$ degrees of freedom.

**Proof.** Writing

$$
B = \lambda \sum_{i=1}^{c} \sum_{j=1}^{n_i} \{ y_{ij} - \mu_i - \beta(x_{ij} - \bar{x}_{.[.]}) \}
$$

and

$$
C = \sum_{i=1}^{c} \sum_{j=1}^{n_i} \beta_{ij} \{ y_{ij} - \mu_i - \beta(x_{ij} - \bar{x}_{.[.]}) \}^2
$$

and realizing that $B/\sqrt{NC} \approx 0$, it is easy to show that

$$
\frac{\partial \ln L^*}{\partial \sigma} \approx \frac{N}{\sigma^2} \left\{ \frac{1}{N} C - \sigma^2 \right\} \quad (1 \leq i \leq c).
$$

Since the expective values of the successive derivatives of $\ln L^* / \hat{\sigma}$ are exactly the same as those of

$$
\frac{\partial \ln L}{\partial \sigma} = \frac{N}{\sigma^3} \left\{ \frac{1}{N} \sum_{i=1}^{c} \sum_{j=1}^{n_i} (y_{ij} - \mu)^2 - \sigma^2 \right\},
$$

where $y_{ij}$ ($1 \leq i \leq c$, $1 \leq j \leq n_i$) is a random sample from $N(\mu, \sigma^2)$, the result follows. □

5. Simulation study

In this section, the efficiencies of the MML and the LS estimators are studied by means of simulation (based on [100, 000/n] Monte Carlo runs). The mean and MSE values are used to evaluate estimator performance. The design values $x_{ij}$ ($1 \leq j \leq n_i$) were generated only once to be common to all the [100, 000/n] (integer value) samples $y_{ij}$ ($1 \leq j \leq n_i$) generated to complete a simulation run. This was done for each $i = 1, 2, \ldots, c$. The study design includes three sample sizes ($n = 10, 15$ and $20$) crossed with six different parameter values of short-tailed symmetric error distribution in (1.2). Because the results for the $n = 15$ sample size are intermediate to those for the $n = 10$ and 20 sample size, they are not reported here. Without loss of generality, we choose the following setting in our simulation:

$$
\mu = 0, \quad \tau_i = 0 \quad (1 \leq i \leq c), \quad \beta = 1 \quad \text{and} \quad \sigma = 1.
$$

Table 1 shows that the MML estimators are considerably more efficient than the LS estimators. The efficiency of the LS estimators as compared to the MML estimators decreases as $n$ increases. Simulated means of the estimators of $\mu$ and $\tau_i$ ($1 \leq i \leq c$) are not given since their biases were found to be negligible. The mean and MSE values were similar for each $\tau_i$ ($1 \leq i \leq c$), therefore we just reproduce the values for $\tau_1$.

The poor performance of the LS estimators under non-normal error distributions serves to reaffirm the importance of assessing underlying assumptions as part of any ANCOVA analysis. It should be noted that the methodology and the results of the computer simulations are exactly the same as in [19] for $i = 1$, because the model given in (1.1) becomes the simple linear regression model.
Table 1
Simulated means and n × Mean square errors of the MML and LS estimators; c = 3 and nᵢ = n = 10, 20 (1 ≤ i ≤ 3)

<table>
<thead>
<tr>
<th>n</th>
<th>Mean</th>
<th>n × MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>̂β</td>
<td>̂μ</td>
</tr>
<tr>
<td></td>
<td>̂σ</td>
<td>̂β_n</td>
</tr>
<tr>
<td></td>
<td>̂σ</td>
<td>̂σ_n</td>
</tr>
</tbody>
</table>

| r = 2, a = −0.5 | | | |
| 10 | 1.019 | 1.013 | 0.988 | 0.961 | 0.63 | 0.59 | 1.24 | 1.19 | 5.80 | 5.60 | 0.15 | 0.15 |
| 20 | 1.015 | 1.013 | 0.992 | 0.979 | 0.65 | 0.61 | 1.08 | 1.03 | 5.98 | 5.63 | 0.13 | 0.13 |

| r = 2, a = 0.0 | | | |
| 10 | 1.005 | 0.993 | 0.988 | 0.958 | 0.69 | 0.62 | 1.46 | 1.35 | 8.77 | 8.00 | 0.14 | 0.14 |
| 20 | 0.953 | 0.965 | 0.993 | 0.979 | 0.71 | 0.64 | 1.21 | 1.11 | 8.26 | 7.72 | 0.12 | 0.12 |

| r = 2, a = 0.5 | | | |
| 10 | 1.039 | 1.012 | 0.990 | 0.984 | 0.79 | 0.66 | 1.59 | 1.39 | 13.66 | 11.64 | 0.12 | 0.11 |
| 20 | 1.012 | 1.013 | 0.994 | 0.991 | 0.82 | 0.68 | 1.41 | 1.20 | 8.52 | 7.18 | 0.11 | 0.10 |

| r = 4, a = −0.5 | | | |
| 10 | 1.012 | 1.005 | 0.988 | 0.953 | 0.80 | 0.73 | 1.54 | 1.44 | 13.02 | 12.27 | 0.14 | 0.14 |
| 20 | 0.965 | 0.976 | 0.993 | 0.976 | 0.83 | 0.76 | 1.39 | 1.29 | 10.44 | 9.14 | 0.12 | 0.12 |

| r = 4, a = 1.0 | | | |
| 10 | 1.000 | 1.004 | 0.992 | 0.988 | 1.09 | 0.88 | 2.43 | 2.04 | 21.08 | 17.96 | 0.11 | 0.10 |
| 20 | 1.014 | 0.996 | 0.995 | 0.993 | 1.13 | 0.87 | 1.96 | 1.55 | 13.99 | 10.32 | 0.10 | 0.09 |

| r = 4, a = 2.0 | | | |
| 10 | 1.003 | 1.012 | 0.995 | 1.026 | 1.49 | 1.16 | 2.83 | 2.20 | 13.00 | 9.74 | 0.09 | 0.09 |
| 20 | 0.957 | 0.955 | 0.997 | 1.014 | 1.55 | 1.09 | 2.69 | 1.91 | 14.98 | 7.76 | 0.07 | 0.06 |

6. Conclusions and future work

Our results have shown that the MML estimators are more efficient than the classical LS estimators under short-tailed symmetric error distributions. This is because of the fact that non-normal error distributions substantially impact the efficiency of the LS estimators. In other words, the normal-theory estimators are efficient estimators only if the error distribution is normal or near-normal.

In our future study, we will use the MML estimators to develop test statistics for testing equality of the treatment means and assumed values of linear contrasts in a simple ANCOVA model under short-tailed symmetric distributions (having kurtosis smaller than 3). Power and robustness properties of these tests will be compared with the classical normal-theory tests.

Acknowledgements

The author would like to thank the referees for their many helpful comments and suggestions.

References

[10] E.S. Pearson, The analysis of variance in cases of nonnormal variation, Biometrika 23 (1932) 114–133.