Robust Estimation in Linear Mixed-Effects Models Using the Multivariate t-Distribution

José C. Pinheiro, Chuanhai Liu and Yingnian Wu

Bell Laboratories
Lucent Technologies
Murray Hill, NJ 07974

Department of Statistics
University of Michigan
Ann Arbor, MI 48109

Abstract

Linear mixed-effects models are frequently used to analyze repeated measures data, because they model flexibly the within-subject correlation often present in this type of data. The most popular linear mixed-effects model for a continuous response assumes normal distributions for the random effects and the within-subject errors, making it sensitive to outliers. Such outliers are more problematic for mixed-effects models than for fixed-effects models, because they may occur in the random effects, in the within-subject errors, or in both, making them harder to be detected. We propose a robust linear mixed-effects model in which the random effects and the within-subject errors have multivariate t-distributions, with known or unknown degrees-of-freedom, which are allowed to vary with subject. We show how the EM algorithm and its extensions can be used for efficient maximum likelihood estimation in our multivariate t model. A real dataset is used to motivate the extension of the linear mixed-effects model and to illustrate the use of the proposed algorithms. We compare the performance of the Gaussian and the multivariate t models under different patterns of outliers. Simulation results indicate that the multivariate t substantially outperforms the Gaussian model when outliers are present in the data, even in moderate amounts.

Key words: EM; ECM; ECME; Random effects; Repeated measures; Longitudinal data; Outliers; Goodness-of-fit.

1 Introduction

Linear mixed-effects models (Hartley and Rao, 1967) have become a popular tool for analyzing repeated measures data which arise in many areas as diverse as agriculture, biology, economics, and geophysics. The increasing popularity of these models is explained by the flexibility they offer in modeling the within-subject correlation often present in repeated measures data, by the handling of both balanced and unbalanced data, and by the availability of reliable and efficient software for fitting them (Wolfinger, Tobias and Sall, 1991; MathSoft, 1997). The most commonly used linear mixed-effects model for a continuous response was proposed by Laird and Ware (1982) and is expressed as

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i, \quad i = 1, \ldots, m, \tag{1}$$

where i is the subject index, $\boldsymbol{y}_i$ is an n_i-dimensional vector of observed responses, $\boldsymbol{X}_i$ and $\boldsymbol{Z}_i$ are known $n_i \times p$ and $n_i \times q$ design matrices corresponding to the p-dimensional fixed effects vector $\boldsymbol{\beta}$ and the q-dimensional random effects vector respectively, and $\boldsymbol{e}_i$ is an n_i-dimensional vector of within-subject errors. The $\boldsymbol{b}_i$ are assumed to be independent with distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and the $\boldsymbol{e}_i$ are assumed to be independent with distribution $\mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_i)$, independent of the $\boldsymbol{b}_i$. The $\boldsymbol{\Psi}$ covariance matrix may be unstructured or structured – e.g. diagonal (Jennrich and Schluchter, 1986). The $\boldsymbol{\Lambda}_i$ matrices are typically assumed to depend on i only through their dimensions, being parametrized by a fixed, generally small, set of parameters $\boldsymbol{\rho}$ – e.g. an AR(1) covariance structure (Box, Jenkins and Reinsel, 1994). The most popular estimation methods for the parameters in model (1) are maximum likelihood and restricted maximum likelihood (Lindstrom and Bates, 1988). Confidence intervals and hypothesis tests for the parameters are generally based on asymptotic results (Miller, 1977).

Though model (1) offers great flexibility for modeling the within-subject correlation frequently present in repeated measures data, it suffers from the same lack of robustness against outlying observations as other statistical models based on the Gaussian distribution. An interesting feature of mixed-effects models is that outliers may occur either at the level of the within-subject error $\boldsymbol{e}_i$, called $\boldsymbol{e}-$outliers, or at the level of the random effects $\boldsymbol{b}_i$, called $\boldsymbol{b}-$outliers. In the first case, some unusual within-subject values are observed, whereas in the second case some unusual subjects are observed. Depending on the percentage of $\boldsymbol{e}$-outliers and the number of observations per subject, it may not be possible to distinguish between the two cases.

A vast statistical literature exists on robust modeling methods, with some authors concentrating more on methods for outlier identification (Barnett and Lewis, 1994) and others on methods for outlier accommodation (Huber, 1981; Hampel, Ronchetti, Rousseeuw and Stahel, 1986). We follow here the robust statistical modeling approach described in Lange, Little and Taylor (1989) and consider a version of model (1) in which the multivariate normal distributions for the $\boldsymbol{b}_i$ and the $\boldsymbol{e}_i$ are replaced by multivariate t-distributions, with known or unknown degrees-of-freedom, which are allowed to vary with subject. This approach can be regarded as outlier-accommodating, though it also provides useful information for outlier identification. A similar approach, but restricted to the distribution of the $\boldsymbol{b}_i$, has been considered by Wakefield, Smith, Racine-Poon and Gelfand (1994) and Racine-Poon (1992), within a Bayesian framework. Pendergast and Broffitt (1986) also have mentioned the multivariate t-distribution in connection with M-estimation for growth curve models. Robust estimation in mixed-effects models with variance components only (i.e. without covariance among random effects) using bounded influence estimators has been considered by Richardson and Welsh (1995) and Richardson (1997).

In Section 2, we describe growth curve data in which both $\boldsymbol{b}$- and $\boldsymbol{e}$-outliers seem to be present.

The multivariate t version of model (1) is described in Section 3. In Section 4 we show how EM-type algorithms can be used for maximum likelihood estimation in the multivariate t linear mixed-effects model. We compare the robust maximum likelihood estimators obtained under the multivariate t-distribution to the Gaussian maximum likelihood estimators corresponding to model (1) in Section 5. Our conclusions and suggestions for further research are presented in Section 6.

2 An example: orthodontic distance growth in boys and girls

Our data come from an orthodontic study of 16 boys and 11 girls between the ages of 8 and 14 years and were originally reported in Potthoff and Roy (1964). The response variable is the distance (in millimeters) between the pituitary and the pterygomaxillary fissure, which was measured at 8, 10, 12, and 14 years for each boy and girl. Figure 1 presents a Trellis display (Becker, Cleveland and Shyu, 1996) of the data, along with individual least-squares fits of the simple linear regression model.

Figure 1 about here

Figure 1 reveals that the estimated slope for subject `M13` is considerably larger than the remaining estimated slopes and that the responses for subject `M09` are more variable around the fitted line. Overall, the responses for the boys vary more around the least squares lines, than do those for the girls. These features are more evident in the residuals plots by gender, displayed in Figure 2 and in the normal plots of the individual coefficients estimates, displayed in Figure 3. These plots suggest that two of the observations on subject `M09` are $\boldsymbol{e}$-outliers and that subject `M13` is a $\boldsymbol{b}$-outlier. Subject `M10` is also identified in Figure 3 because he is indicated as a possible $\boldsymbol{b}$-outlier later in Section 5.1.

Figures 2 and 3 about here

Because both intercept and slope seem to vary with subject and the within-subject variation is larger among boys than girls, the following linear mixed-effects model can be used to describe the orthodontic distance growth with age.

$$y_{ij} = \beta_0 + \delta_0 I_i(F) + (\beta_1 + \delta_1 I_i(F))\, t_j + b_{0i} + b_{1i} t_j + e_{ij}, i = 1, \ldots, 27 \text{ and } j = 1, \ldots, 4, \qquad (2)$$

where y_{ij} denotes the orthodontic distance for the ith subject at age t_j, β_0 and β_1 denote respectively the intercept and the slope fixed effects for boys, δ_0 and δ_1 denote respectively the difference in intercept and slope fixed effects between girls and boys, $I_i(F)$ denotes an indicator variable for females, $\boldsymbol{b}_i = (b_{0i}, b_{1i})$ is the random effects vector for the ith subject, and e_{ij} is the within-subject error.

In Section 5.1, we compare the maximum likelihood estimates (MLEs) under the Gaussian version of the linear mixed-effects model (2) to the MLEs obtained under the multivariate t model described in Section 3.

3 A multivariate t linear mixed-effects model

The Gaussian linear mixed-effects model (1) can alternatively be written as:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} \overset{\text{ind}}{\sim} \mathcal{N}_{n_i+q}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Psi}Z_i' + \boldsymbol{\Lambda}_i & \boldsymbol{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\boldsymbol{Z}_i' & \boldsymbol{\Psi} \end{bmatrix} \right), \quad i = 1, \ldots, m,$$

with $\boldsymbol{\Lambda}_i = \boldsymbol{\Lambda}_i(\boldsymbol{\rho})$. For robust estimation of $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\rho}$, we proceed as in Lange et al. (1989) and replace the multivariate normal distribution in (3) with the multivariate t-distribution:

$$\begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} \overset{\text{ind}}{\sim} t_{n_i+q}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i' + \boldsymbol{\Lambda}_i & \boldsymbol{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\boldsymbol{Z}_i' & \boldsymbol{\Psi} \end{bmatrix}, \nu_i \right), \quad i = 1, \ldots, m, \tag{3}$$

where ν_i represents the multivariate t-distribution degrees of freedom (d.f.) for the i-th subject. It follows from (3) that the $\boldsymbol{y}_i$ are independent and marginally distributed as $t_{n_i}\left(\boldsymbol{X}_i\boldsymbol{\beta}, \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i' + \boldsymbol{\Lambda}_i, \nu_i\right)$, which provides yet another characterization of the multivariate t linear mixed-effects model.

From standard properties of the multivariate t-distribution (Johnson and Kotz, 1972), it follows that, for $\nu_i > 2$,

$$\operatorname{var}(\boldsymbol{b}_i) = \frac{\nu_i}{\nu_i - 2}\boldsymbol{\Psi} \quad \text{and} \quad \operatorname{var}(\boldsymbol{e}_i) = \frac{\nu_i}{\nu_i - 2}\boldsymbol{\Lambda}_i, \quad i = 1, \ldots, m.$$

Therefore, the interpretation of $\boldsymbol{\Psi}$ and $\boldsymbol{\Lambda}_i$ is different in the Gaussian model (1) and in the multivariate t model (3). Note, in particular, that $\operatorname{var}(\boldsymbol{b}_i)$ is allowed to change with i in the multivariate t model, while it is independent of i in the Gaussian model. Provided $\nu_i > 1$ in (3), both models have $\mathrm{E}(\boldsymbol{y}_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, so that the fixed effects have the same interpretation.

Generally, some constraints are needed on the ν_i when these are to be estimated from the data. Common constraints are $\nu_i = \nu$ for all $i = 1, \ldots, m$, or, more generally,

$$\nu_i = \lambda_{h(i)}, \quad i = 1, \ldots, m, \tag{4}$$

where $h(i) \in \{1, \ldots, l\}$ denotes the group to which the ith subject belongs and $\lambda_1, \ldots, \lambda_l$ are l distinct positive scalar parameters, which can be treated as known, or unknown. We shall focus here on the t linear mixed-effects model (3) with the constraints (4).

The t model (3) can be expressed as the marginal distribution of $[\boldsymbol{y}_i', \boldsymbol{b}_i']'$ in the following hierarchical models:

$$\begin{aligned} \begin{bmatrix} \boldsymbol{y}_i \\ \boldsymbol{b}_i \end{bmatrix} \Bigg| \tau_i &\overset{\text{ind}}{\sim} \mathcal{N}_{n_i+q}\left(\begin{bmatrix} \boldsymbol{X}_i\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \frac{1}{\tau_i}\begin{bmatrix} \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i' + \Lambda_i & \boldsymbol{Z}_i\boldsymbol{\Psi} \\ \boldsymbol{\Psi}\boldsymbol{Z}_i' & \boldsymbol{\Psi} \end{bmatrix} \right) \quad \text{and} \\ \tau_i &\overset{\text{ind}}{\sim} \text{Gamma}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right), \quad i = 1, \ldots, m, \end{aligned} \tag{5}$$

or

$$\begin{aligned} \boldsymbol{y}_i | \boldsymbol{b}_i, \tau_i &\overset{\text{ind}}{\sim} \mathcal{N}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i, \frac{1}{\tau_i}\boldsymbol{\Lambda}_i), \quad \boldsymbol{b}_i|\tau_i \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \frac{1}{\tau_i}\boldsymbol{\Psi}), \quad \text{and} \\ \tau_i &\overset{\text{ind}}{\sim} \text{Gamma}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right), \quad i = 1, \ldots, m. \end{aligned} \tag{6}$$

Integrating out the $\boldsymbol{b}_i$, we can express the distribution of $\boldsymbol{y}_i$ as the marginal distribution of the following hierarchical model.

$$\boldsymbol{y}_i|\tau_i \overset{\text{ind}}{\sim} \mathcal{N}\left(\boldsymbol{X}_i\boldsymbol{\beta}, \frac{1}{\tau_i}\left(\boldsymbol{\Lambda}_i + \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i'\right)\right) \quad \text{and} \quad \tau_i \overset{\text{ind}}{\sim} \text{Gamma}\left(\frac{\nu_i}{2}, \frac{\nu_i}{2}\right), \quad i = 1, \ldots, m. \tag{7}$$

A useful consequence of (7) is that

$$\tau_i|\boldsymbol{y}_i \overset{\text{ind}}{\sim} \text{Gamma}\left(\frac{\nu_i + n_i}{2}, \frac{\nu_i + \delta_i^2\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}\right)}{2}\right),$$

where

$$\delta_i^2(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}) = (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\left(\boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i' + \boldsymbol{\Lambda}_i\right)^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}). \tag{8}$$

Note that, in particular,

$$\text{E}(\tau_i|\boldsymbol{y}_i) = \frac{\nu_i + n_i}{\nu_i + \delta_i^2(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho})}.$$

4 Maximum likelihood estimation using the EM algorithms

In this section, we consider the maximum likelihood (ML) estimation of the parameters in the multivariate t linear mixed-effects model (3). We describe two EM-type algorithms for ML estimation with known and unknown degrees of freedom, based on two types of missing data structures. The first uses the hierarchical model (6) with both the $\boldsymbol{b}_i$ and the τ_i treated as missing. The second is based on the hierarchical model (7) which, by integrating out the $\boldsymbol{b}_i$, has just the τ_i as missing data. The first algorithm is computationally simpler, with closed-form expressions for the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\rho}$ in the maximization step, but requires additional assumptions about the structure of the $\boldsymbol{\Lambda}_i$ matrices. The second algorithm has a more computationally intensive maximization step, but allows more generality in the model specification and only requires minor modifications to existing software for fitting the Gaussian linear mixed-effects model (1). It should be noted that both algorithms lead to the same MLEs (up to numerical round-off error).

Letting $\boldsymbol{\psi}$ denote a minimal set of parameters to determine $\boldsymbol{\Psi}$ (e.g. the upper triangular elements of $\boldsymbol{\Psi}$ in the unstructured case), we define the population parameters vector $\boldsymbol{\theta} = \left[\boldsymbol{\beta}', \boldsymbol{\psi}', \boldsymbol{\rho}', \boldsymbol{\lambda}'\right]'$.

Compared to the Gaussian linear mixed-effects model (1), the multivariate t model (3) allows each subject to have its own scale τ_i, which is unobserved and needs to be *imputed* from the data. The different individual scales result in different *weights* for estimating the population parameters $\boldsymbol{\theta}$. For example, conditional on $\boldsymbol{\Psi}$, $\boldsymbol{\rho}$, and the τ_i, the ML estimate of $\boldsymbol{\beta}$ minimizes $\sum_{i=1}^m \tau_i\delta_i^2\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}\right)$, with δ_i^2 as defined as (8). Because $\text{E}\left(\tau_i \mid \boldsymbol{y}_i\right)$ decreases with δ_i^2, subjects with larger *residual sum of squares* δ_i^2 will have less weight in the determination of the ML estimates. The influence of δ_i^2 on the τ_i scales is controlled by the individual degrees of freedom ν_i – the smaller ν_i the larger the influence of δ_i^2 on τ_i.

4.1 The EM algorithm

The EM algorithm (Dempster, Laird and Rubin, 1977) is a popular iterative algorithm for ML estimation in models with incomplete data. More specifically, let $\boldsymbol{y}_{\text{obs}}$ denote the observed data and $\boldsymbol{y}_{\text{mis}}$ denote the missing data. The complete data $\boldsymbol{y}_{\text{com}} = (\boldsymbol{y}_{\text{obs}}, \boldsymbol{y}_{\text{mis}})$ is $\boldsymbol{y}_{\text{obs}}$ augmented with $\boldsymbol{y}_{\text{mis}}$. We denote by $f(\boldsymbol{y}_{\text{com}}|\boldsymbol{\theta})$ the complete-data likelihood function of a parameter vector $\boldsymbol{\theta} \in \Theta$, by $L(\boldsymbol{\theta}) = f(\boldsymbol{y}_{\text{obs}}|\boldsymbol{\theta})$ the log-likelihood function and by $Q(\boldsymbol{\theta}|\boldsymbol{\theta}')$ the expected complete-data log-likelihood

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}') = \mathrm{E}\left\{\ln\left[f\left(\boldsymbol{y}_{\text{com}}|\boldsymbol{\theta}\right)\right]|\boldsymbol{y}_{\text{obs}}, \boldsymbol{\theta}'\right\}.$$

Each iteration of the EM algorithm consists of two steps, the *Expectation* step and the *Maximization* step:

E-step: Compute $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ as a function of $\boldsymbol{\theta}$;

M-step: Find $\boldsymbol{\theta}^{(t+1)}$ such that $Q(\boldsymbol{\theta}^{(t+1)}|\boldsymbol{\theta}^{(t)}) = \max_{\boldsymbol{\theta}\in\Theta} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$.

Each iteration of the EM algorithm increases the likelihood function $L(\boldsymbol{\theta})$ and, under mild conditions, the EM algorithm converges to a local or global maximum of $L(\boldsymbol{\theta})$ (Dempster et al., 1977; Wu, 1983).

When the M-step in the EM algorithm is difficult to implement, it is often useful to replace it with a sequence of constrained maximization (CM) steps, each of which maximizes $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(t)})$ over $\boldsymbol{\theta}$ with some function of $\boldsymbol{\theta}$ held fixed. The sequence of CM-steps is such that the overall maximization is over the full parameter space. This leads to a simple extension of the EM algorithm, called the ECM algorithm (Meng and Rubin, 1993). A further extension of the EM algorithm is the ECME algorithm (Liu and Rubin, 1994). This algorithm replaces each CM-step of ECM with a CM-step that maximizes either the constrained Q function, as in ECM, or the correspondingly constrained L function.

4.2 EM algorithm with $\boldsymbol{b}_i$ and τ_i as missing data

First consider the hierarchical multivariate t model (6) with both the $\boldsymbol{b}_i$ and the τ_i as missing data. For simplicity, assume that

$$\boldsymbol{\Lambda}_i = \sigma_i^2 \boldsymbol{R}_i, \quad \sigma_i^2 = \sigma_{g(i)}^2, \quad i = 1, \ldots, m, \tag{9}$$

with $g(i) \in \{1, \ldots, k\}$ representing the group to which the ith subject belongs. The $\boldsymbol{R}_i$ are known matrices, usually equal to the identity. We denote by $\boldsymbol{\sigma}^2$ the unique elements in $\{\sigma_1^2, \ldots, \sigma_m^2\}$.

The within-subject covariance structure (9) allows for variance heterogeneity among different groups of subjects, but does not include serial correlation structures such as in ARMA models (Box et al., 1994).

4.2.1 ML estimation with known degrees of freedoms using ECM

Let $\boldsymbol{y} = [\boldsymbol{y}_1', \ldots, \boldsymbol{y}_m']'$, $\boldsymbol{b} = [\boldsymbol{b}_1', \ldots, \boldsymbol{b}_m']'$, and $\boldsymbol{\tau} = [\tau_1, \ldots, \tau_m]$. Under the constraints (9), the log-likelihood for the complete data in the multivariate t linear mixed-effects model (3) is

$$\mathrm{L}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\sigma}^2|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}) = \mathrm{L}_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^2|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}) + \mathrm{L}_2(\Psi|\boldsymbol{b}, \boldsymbol{\tau}) + constant,$$

where

$$\begin{aligned}
&\mathrm{L}_1(\boldsymbol{\beta}, \boldsymbol{\sigma}^2|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau}) \\
&= \sum_{i=1}^m \left[-\frac{n_i}{2} \ln \sigma_i^2 - \frac{\tau_i}{2\sigma_i^2} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i)' \boldsymbol{R}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i) \right] \\
&= -\sum_{i=1}^m \frac{n_i}{2} \ln \sigma_i^2 - \sum_{i=1}^m \frac{\tau_i}{2\sigma_i^2} \mathrm{trace} \left[\boldsymbol{R}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{b}_i)(\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{b}_i)' \right] \\
&\quad + \sum_{i=1}^m \frac{\tau_i}{\sigma_i^2} \boldsymbol{\beta}' \boldsymbol{X}_i' \boldsymbol{R}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{Z}_i\boldsymbol{b}_i) - \sum_{i=1}^m \frac{\tau_i}{2\sigma_i^2} \boldsymbol{\beta}' \boldsymbol{X}_i' \boldsymbol{R}_i^{-1} \boldsymbol{X}_i \boldsymbol{\beta}
\end{aligned}$$

and

$$\mathrm{L}_2(\boldsymbol{\Psi}|\boldsymbol{b}, \boldsymbol{\tau}) = -\frac{m}{2} \ln |\boldsymbol{\Psi}| - \frac{1}{2} \mathrm{trace} \left(\boldsymbol{\Psi}^{-1} \sum_{i=1}^m \tau_i \boldsymbol{b}_i \boldsymbol{b}_i' \right).$$

Letting

$$\widehat{\tau}_i = \mathrm{E}(\tau_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \boldsymbol{y}), \quad \widehat{\boldsymbol{b}}_i = \mathrm{E}(\boldsymbol{b}_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \boldsymbol{y}), \quad \text{and} \quad \widehat{\boldsymbol{\Omega}}_i = \tau_i \mathrm{cov}(\boldsymbol{b}_i|\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}, \boldsymbol{y}),$$

we obtain

$$\begin{aligned}
\widehat{\boldsymbol{\Omega}}_i &= \widehat{\boldsymbol{\Psi}} - \widehat{\boldsymbol{\Psi}} \boldsymbol{Z}_i' (\boldsymbol{Z}_i \widehat{\boldsymbol{\Psi}} \boldsymbol{Z}_i' + \widehat{\sigma}_i^2 \boldsymbol{R}_i)^{-1} \boldsymbol{Z}_i \widehat{\boldsymbol{\Psi}} = \left(\widehat{\boldsymbol{\Psi}}^{-1} + \frac{1}{\widehat{\sigma}_i^2} \boldsymbol{Z}_i' \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i \right)^{-1}, \qquad (10) \\
\widehat{\boldsymbol{b}}_i &= \widehat{\boldsymbol{\Psi}} \boldsymbol{Z}_i' (\boldsymbol{Z}_i \widehat{\boldsymbol{\Psi}} \boldsymbol{Z}_i' + \widehat{\sigma}_i^2 \boldsymbol{R}_i)^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}) \\
&= \left(\widehat{\boldsymbol{\Psi}}^{-1} + \frac{1}{\widehat{\sigma}_i^2} \boldsymbol{Z}_i' \boldsymbol{R}_i^{-1} \boldsymbol{Z}_i \right)^{-1} \frac{1}{\widehat{\sigma}_i^2} \boldsymbol{Z}_i' \boldsymbol{R}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}), \quad \text{and} \\
\widehat{\tau}_i &= \frac{\nu_i + n_i}{\nu_i + \delta_i^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\sigma}}^2)}.
\end{aligned}$$

From standard multivariate analysis results (Fang and Zhang, 1990, p. 4) we have

$$\begin{aligned}
&[(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})', \boldsymbol{b}_i'] \begin{bmatrix} \boldsymbol{Z}_i \boldsymbol{\Psi} \boldsymbol{Z}_i' + \sigma_i^2 \boldsymbol{R}_i & \boldsymbol{Z}_i \boldsymbol{\Psi} \\ \boldsymbol{\Psi} \boldsymbol{Z}_i' & \boldsymbol{\Psi} \end{bmatrix}^{-1} \begin{bmatrix} \boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} \\ \boldsymbol{b}_i \end{bmatrix} \\
&= \boldsymbol{b}_i' \boldsymbol{\Psi}^{-1} \boldsymbol{b}_i + \frac{1}{\sigma_i^2} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i)' \boldsymbol{R}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i).
\end{aligned}$$

Replacing $\boldsymbol{\theta}$ and $\boldsymbol{b}_i$ with their current estimates, we obtain the following useful decomposition:

$$\begin{aligned} \delta_i^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\sigma}}^2) &= (\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}})' \left(\boldsymbol{Z}_i\widehat{\boldsymbol{\Psi}}\boldsymbol{Z}_i' + \widehat{\sigma}_i^2\boldsymbol{R}_i\right)^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}}) \qquad (11) \\ &= \widehat{\boldsymbol{b}}_i'\widehat{\boldsymbol{\Psi}}^{-1}\widehat{\boldsymbol{b}}_i + \frac{1}{\widehat{\sigma}_i^2}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i)'\boldsymbol{R}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i) = \widehat{\delta}^2_{\boldsymbol{b}_i} + \widehat{\delta}^2_{\boldsymbol{e}_i}. \end{aligned}$$

Equation (11) provides a simple way to compute $\delta_i^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\sigma}^2)$ as well as the weights $\widehat{\tau}_i$. It also gives some insight on how the estimated random effects $\widehat{\boldsymbol{b}}_i$ and the estimated residuals $\widehat{\boldsymbol{e}}_i = \boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i$ affect the individual weights $\widehat{\tau}_i$.

Using simple algebra we get

$$\begin{aligned} &\mathrm{E}\left[\mathrm{L}_1(\boldsymbol{\beta}, \sigma^2|\boldsymbol{y}, \boldsymbol{b}, \tau)|\boldsymbol{y}, \widehat{\boldsymbol{\theta}})\right] \\ &= -\sum_{i=1}^m \frac{n_i}{2}\ln\sigma_i^2 - \sum_{i=1}^m \frac{1}{2\sigma_i^2}\mathrm{trace}\left[\boldsymbol{R}_i^{-1}\left(\widehat{\tau}_i(\boldsymbol{y}_i - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i)(\boldsymbol{y}_i - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i)' + \boldsymbol{Z}_i\widehat{\boldsymbol{\Omega}}_i\boldsymbol{Z}_i'\right)\right] \\ &\quad + \sum_{i=1}^m \frac{\widehat{\tau}_i}{\sigma_i^2}\boldsymbol{\beta}'\boldsymbol{X}_i'\boldsymbol{R}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i) - \sum_{i=1}^m \frac{\widehat{\tau}_i}{2\sigma_i^2}\boldsymbol{\beta}'\boldsymbol{X}_i'\boldsymbol{R}_i^{-1}\boldsymbol{X}_i\boldsymbol{\beta} \end{aligned}$$

and

$$\mathrm{E}\left[\mathrm{L}_2(\boldsymbol{\Psi}|\boldsymbol{b}, \tau)|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right] = -\frac{m}{2}\ln|\boldsymbol{\Psi}| - \frac{1}{2}\mathrm{trace}\left[\boldsymbol{\Psi}^{-1}\sum_{i=1}^m\left(\widehat{\tau}_i\widehat{\boldsymbol{b}}_i\widehat{\boldsymbol{b}}_i' + \widehat{\boldsymbol{\Omega}}_i\right)\right].$$

We then have the following ECM algorithm:

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{\boldsymbol{b}}_i$, $\widehat{\tau}_i$, and $\widehat{\boldsymbol{\Omega}}_i$ for $i = 1, \ldots, m$, using (10).

CM-step 1: Fix $\sigma_i^2 = \widehat{\sigma}_i^2$ for $i = 1, \ldots, m$ and update $\widehat{\boldsymbol{\beta}}$ by maximizing $\mathrm{E}\left[\mathrm{L}_1(\boldsymbol{\beta}, \widehat{\sigma}^2|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau})|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{\beta}$, which leads to

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^m \frac{\widehat{\tau}_i}{\widehat{\sigma}_i^2}\boldsymbol{X}_i'\boldsymbol{R}_i^{-1}\boldsymbol{X}_i\right)^{-1}\sum_{i=1}^m \frac{\widehat{\tau}_i}{\widehat{\sigma}_i^2}\boldsymbol{X}_i'\boldsymbol{R}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i\right).$$

CM-step 2: Fix $\boldsymbol{\beta} = \widehat{\boldsymbol{\beta}}$ and update $\widehat{\sigma}_i^2$ for $i = 1, \ldots, m$ by maximizing $\mathrm{E}\left[\mathrm{L}_1(\widehat{\boldsymbol{\beta}}, \sigma^2|\boldsymbol{y}, \boldsymbol{b}, \boldsymbol{\tau})|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over σ_i^2, which gives

$$\widehat{\sigma}_j^2 = \sum_{i:g(i)=j}\left[\widehat{\tau}_i(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i)'\boldsymbol{R}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i\widehat{\boldsymbol{b}}_i) + \mathrm{trace}(\widehat{\boldsymbol{\Omega}}_i\boldsymbol{Z}_i'\boldsymbol{R}_i^{-1}\boldsymbol{Z}_i)\right] \Big/ \sum_{i:g(i)=j} n_i,$$

for $j = 1, \ldots, k$.

CM-step 3: Update $\widehat{\boldsymbol{\Psi}}$ by maximizing $\mathrm{E}\left[\mathrm{L}_2(\boldsymbol{\Psi}|b, \boldsymbol{\tau})|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right]$ over $\boldsymbol{\Psi}$, that is,

$$\widehat{\boldsymbol{\Psi}} = \frac{1}{m}\sum_{i=1}^m\left(\widehat{\tau}_i\widehat{\boldsymbol{b}}_i\widehat{\boldsymbol{b}}_i' + \widehat{\boldsymbol{\Omega}}_i\right).$$

4.2.2 ML estimation with unknown degrees of freedom using ECME

When some, or all, of the degrees of freedom $\nu_1, \ldots, \nu_m$ are unknown, we can use the ECME algorithm that has the same E and CM steps as the ECM algorithm described in 4.2.1 for updating the estimates of $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\sigma}^2$ and an additional CML step that maximizes the constrained likelihood over the degrees of freedom with $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\sigma}^2$ fixed at their current estimates. When λ_j is unknown, the constrained likelihood is computed using

$$\boldsymbol{y}_i \overset{\text{ind}}{\sim} \mathbf{t}_{n_i}\left(\boldsymbol{X}_i\widehat{\boldsymbol{\beta}}, \boldsymbol{Z}_i\widehat{\boldsymbol{\Psi}}\boldsymbol{Z}_i' + \widehat{\sigma}_i^2\boldsymbol{R}_i, \lambda_j\right), \quad \text{for } i \in \{i : g(i) = j\}.$$

More specifically, we have

CML-step: Update each unknown λ_j $(j = 1, \ldots, l)$ by maximizing

$$\mathrm{L}_3(\lambda|\boldsymbol{y}, \boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\sigma}^2) = \sum_{i:h(i)=j}\left\{\ln\left[\Gamma\left(\frac{\lambda+n_i}{2}\right)\right] - \ln\left[\Gamma\left(\frac{\lambda}{2}\right)\right] + \frac{\lambda}{2}\ln(\lambda) - \frac{\lambda+n_i}{2}\ln\left[\lambda + \delta_i^2\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\sigma}}^2\right)\right]\right\}$$

over λ. This requires only a one-dimensional search and can be obtained, for example, using the Newton-Raphson method (Thisted, 1988, §4.2.2).

4.3 ML estimation integrating out the $\boldsymbol{b}_i$

The EM algorithm described in Section 4.2 provides closed form expressions for updating the estimates of $\boldsymbol{\theta}$, but require that the within-subject covariance matrices $\boldsymbol{\Lambda}_i$ be constrained to the form given in (9). A more flexible formulation, with no constraints on the $\boldsymbol{\Lambda}_i$, can be used when the $\boldsymbol{b}_i$ are integrated out of the complete data likelihood, so that only the τ_i are treated as missing data, at the expense of a more computationally intensive CM-step. We describe here an ECME algorithm for this missing data scheme.

The log-likelihood of the complete data $[\boldsymbol{y}', \boldsymbol{\tau}']'$ in the multivariate t model (3) is

$$\mathrm{L}(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\tau}) = \mathrm{L}_1(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\tau}) + \mathit{constant},$$

where

$$\mathrm{L}_1(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}|\boldsymbol{y}, \boldsymbol{\tau}) = -\frac{1}{2}\sum_{i=1}^{m}\left[n_i\log|\boldsymbol{V}_i| + \tau_i(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right],$$

with $\boldsymbol{V}_i = \boldsymbol{\Lambda}_i + \boldsymbol{Z}_i\boldsymbol{\Psi}\boldsymbol{Z}_i'$. Letting $\widehat{\tau}_i$ be defined as in (10), it follows that

$$\mathrm{E}\left[\mathrm{L}_1\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}|\boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right)\right] = -\frac{1}{2}\sum_{i=1}^{m}\left[n_i\log|\boldsymbol{V}_i| + \widehat{\tau}_i(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})'\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta})\right]$$

and, therefore, the following ECME algorithm can be used to obtain the MLEs of $\boldsymbol{\theta}$.

E-step: Given $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}$, compute $\widehat{\tau}_i = (\nu_i + n_i)\,/\left[\nu_i + \delta_i^2\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\rho}}\right)\right]$, with $\delta_i^2(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho})$ as defined in (8).

CM-step: For fixed $\widehat{\boldsymbol{\tau}}$, update $\widehat{\boldsymbol{\beta}}$, $\widehat{\boldsymbol{\Psi}}$, and $\widehat{\boldsymbol{\rho}}$ by maximizing the function $\mathrm{E}\left[\mathrm{L}_1\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho} | \boldsymbol{y}, \widehat{\boldsymbol{\theta}}\right)\right]$ over $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, and $\boldsymbol{\rho}$.

The CM-step in this ECME algorithm is equivalent to maximum likelihood estimation in the Gaussian linear mixed-effects model $\boldsymbol{y}_i^* = \boldsymbol{X}_i^* \boldsymbol{\beta} + \boldsymbol{Z}_i \boldsymbol{b}_i + \boldsymbol{e}_i$, $i = 1, \ldots, m$, where $\boldsymbol{y}_i^* = \sqrt{\tau_i} \boldsymbol{y}_i$ and $\boldsymbol{X}_i^* = \sqrt{\tau_i} \boldsymbol{X}_i$. Reliable and efficient implementations of Newton-Raphson algorithms for obtaining the MLEs in the general Gaussian linear mixed-effects model (1) are available in commercial products such as SAS (PROC MIXED) and S-PLUS (`lme` function). These programs can be used to implement the ECME algorithm described here at low additional cost.

The decomposition of $\delta_i^2\left(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\rho}}\right)$ given in (11) remains valid for general $\boldsymbol{\Lambda}_i$. That is,

$$\widehat{\delta}_i^2 = \delta_i^2(\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\Psi}}, \widehat{\boldsymbol{\rho}}) = \widehat{\boldsymbol{b}}_i' \widehat{\boldsymbol{\Psi}}^{-1} \widehat{\boldsymbol{b}}_i + (\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i)' \widehat{\boldsymbol{\Lambda}}_i^{-1} (\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}} - \boldsymbol{Z}_i \widehat{\boldsymbol{b}}_i) = \widehat{\delta}_{\boldsymbol{b}_i}^2 + \widehat{\delta}_{\boldsymbol{e}_i}^2,$$

where $\widehat{\boldsymbol{b}}_i = \mathrm{E}\left(\boldsymbol{b}_i | \boldsymbol{y}_i, \widehat{\boldsymbol{\theta}}\right) = \widehat{\boldsymbol{\Psi}} \boldsymbol{Z}_i' \widehat{\boldsymbol{V}}_i^{-1} \left(\boldsymbol{y}_i - \boldsymbol{X}_i \widehat{\boldsymbol{\beta}}\right)$.

When the degrees of freedom ν_i are unknown, an additional CML-step, identical to the one described in Section 4.2.2, can be used to estimate the λ_j, $j = 1, \ldots, l$. We have used the `lme` function to implement the ECME algorithm described here in S-PLUS. This implementation allows the degrees of freedom to be fixed in advance, or estimated from the data.

4.4 Inference based on the maximum likelihood estimates

One is generally interested in using MLEs to obtain confidence intervals and test hypotheses about the parameters. Because the distribution of the MLEs cannot be explicitly derived, approximate inference methods must be employed. The most common method uses the asymptotic normal approximation to the distribution of the MLEs (Miller, 1977; Lange et al., 1989). Other methods include the bootstrap (Efron and Tibshirani, 1993) and likelihood profiling (Bates and Watts, 1988). These last two methods usually give more accurate approximations, but are computationally intensive for the multivariate t model (3). This paper considers only confidence intervals and tests based on the normal approximation, concentrating on methods for the fixed effects $\boldsymbol{\beta}$.

Asymptotic confidence intervals and tests based on the MLEs can be obtained using either the observed or the expected Fisher information matrix. For the multivariate t model, these can be derived using the results in Appendix B of Lange et al. (1989). Let $\boldsymbol{J}$ denote the expected Fisher information matrix for the marginal log-likelihood L of the multivariate t model and $\boldsymbol{\omega}$ denote the set of parameters excluding the fixed effects, so that $\boldsymbol{\theta} = (\boldsymbol{\beta}', \boldsymbol{\omega}')'$. It can be shown that

$$\boldsymbol{J}_{\boldsymbol{\beta\beta}} = \mathrm{E}\frac{\partial^2 \mathrm{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} = \sum_{i=1}^m \frac{\nu_i + n_i}{\nu_i + n_i + 2} \boldsymbol{X}_i' \boldsymbol{V}_i^{-1} \boldsymbol{X}_i \quad \text{and} \quad \boldsymbol{J}_{\boldsymbol{\beta\omega}} = \mathrm{E}\frac{\partial^2 \mathrm{L}}{\partial \boldsymbol{\beta} \partial \boldsymbol{\omega}'} = \mathbf{0}.$$

It follows that the expected Fisher information matrix is block diagonal and, in particular, $\left[\boldsymbol{J}^{-1}\right]_{\boldsymbol{\beta\beta}} = \boldsymbol{J}_{\boldsymbol{\beta\beta}}^{-1}$. Asymptotic confidence intervals and hypothesis tests for the fixed effects are obtained assuming that the MLE $\widehat{\boldsymbol{\beta}}$ has approximately a $\mathcal{N}_p\left(\boldsymbol{\beta}, \boldsymbol{J}_{\boldsymbol{\beta\beta}}^{-1}\right)$ distribution. In practice, $\boldsymbol{J}_{\boldsymbol{\beta\beta}}$ is usually unknown and has to be replaced by its MLE $\widehat{\boldsymbol{J}}_{\boldsymbol{\beta\beta}}$.

5 Comparing the Gaussian and the multivariate t MLEs

In this section we compare the MLEs under the Gaussian model (1) to the MLEs obtained under the multivariate t model (3). Firstly, we compare the Gaussian MLEs and the multivariate t MLEs for the orthodontic growth example of Section 2. The performance of the two sets of estimators are then compared under different outlier patterns, using results of a simulation.

5.1 The orthodontic growth example revisited

The distributional assumptions for the Gaussian version of the orthodontic growth model (2) are: $\boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and $e_{ij} \overset{\text{ind}}{\sim} \mathcal{N}\left(0, \sigma^2_{g(i)}\right)$, with the $\boldsymbol{b}_i$ independent of the e_{ij}. $g(i) = I_i(F) + 1$ denotes the gender group for the ith subject. The corresponding MLEs are given below.

$$\begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\delta}_0 \\ \widehat{\beta}_1 \\ \widehat{\delta}_1 \end{bmatrix} = \begin{bmatrix} 16.34 \\ 1.03 \\ 0.78 \\ -0.31 \end{bmatrix}, \quad \widehat{\boldsymbol{\Psi}} = \begin{bmatrix} 3.20 & -0.11 \\ -0.11 & 0.02 \end{bmatrix}, \quad \begin{bmatrix} \widehat{\sigma}^2_1 \\ \widehat{\sigma}^2_2 \end{bmatrix} = \begin{bmatrix} 2.63 \\ 0.45 \end{bmatrix}. \tag{12}$$

The multivariate t version of model (2) has the following distributional assumptions: $\boldsymbol{b}_i|\tau_i \overset{\text{ind}}{\sim} \mathcal{N}\left(\mathbf{0}, \tau_i^{-1}\boldsymbol{\Psi}\right)$, $e_{ij}|\tau_i \overset{\text{ind}}{\sim} \mathcal{N}\left(0, \tau_i^{-1}\sigma^2_{g(i)}\right)$, and $\tau_i \overset{\text{ind}}{\sim} \text{Gamma}\left(\lambda_{g(i)}/2, \lambda_{g(i)}/2\right)$.

As mentioned in Section 3, the parameters $\boldsymbol{\Psi}$, σ^2_1, and σ^2_2 in the Gaussian model do not have the same interpretation as in the multivariate t model. To make the MLEs comparable, we consider the parameters $\text{var}(\boldsymbol{b}_i) = \boldsymbol{\Psi}_{g(i)}(t) = \left[\lambda_{g(i)}/\left(\lambda_{g(i)} - 2\right)\right]\boldsymbol{\Psi}$ and $\text{var}(e_{ij}) = \sigma^2_{g(i)}(t) = \left[\lambda_{g(i)}/\left(\lambda_{g(i)} - 2\right)\right]\sigma^2_{g(i)}$. The MLEs for the multivariate t model are shown below.

$$\begin{gathered} \begin{bmatrix} \widehat{\beta}_0 \\ \widehat{\delta}_0 \\ \widehat{\beta}_1 \\ \widehat{\delta}_1 \end{bmatrix} = \begin{bmatrix} 16.83 \\ 0.54 \\ 0.73 \\ -0.25 \end{bmatrix}, \quad \widehat{\boldsymbol{\Psi}}_1(t) = \begin{bmatrix} 4.79 & -0.16 \\ -0.16 & 0.03 \end{bmatrix}, \quad \widehat{\boldsymbol{\Psi}}_2(t) = \begin{bmatrix} 3.13 & -0.11 \\ -0.11 & 0.02 \end{bmatrix} \\ \begin{bmatrix} \widehat{\sigma}^2_1(t) \\ \widehat{\sigma}^2_2(t) \end{bmatrix} = \begin{bmatrix} 2.43 \\ 0.45 \end{bmatrix}, \quad \begin{bmatrix} \widehat{\lambda}_1 \\ \widehat{\lambda}_2 \end{bmatrix} = \begin{bmatrix} 5.78 \\ 6 \times 10^6 \end{bmatrix}. \end{gathered} \tag{13}$$

Comparing these estimates to the Gaussian MLEs in (12), we see that the estimates of the incremental parameters (δ_0 and δ_1) fixed effects and the boys' random effects covariance matrix $\boldsymbol{\Psi}_1(t)$ are considerable different. The multivariate t MLEs of δ_0 and δ_1 are respectively 50% smaller and 20% larger than the corresponding Gaussian MLEs. The boys' random effects variances multivariate t MLEs are 50% larger than the Gaussian MLEs. The MLEs of the girls' parameters are essentially unchanged. Using

$$t\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \nu\right) \overset{\nu \to \infty}{\Longrightarrow} \mathcal{N}\left(\boldsymbol{\mu}, \boldsymbol{\sigma}^2\right), \tag{14}$$

it is clear that the estimated degrees of freedom $\widehat{\lambda}_2$ indicate that a Gaussian model is adequate for the girls' orthodontic growth. The multivariate t linear mixed-effects model (3) can be easily modified into a hybrid model in which some of the subjects have a multivariate t-distribution, while others follow a Gaussian distribution (by setting $\tau_i = 1$ for these subjects).

To better understand the differences between the MLEs under the Gaussian and the multivariate t models, we consider the approximate distributions of the fixed effects estimators (corresponding to the asymptotic distributions evaluated at the MLEs) for each model, presented in Figure 4.

Figure 4 about here

The incremental parameters δ_0 and δ_1 have estimates closer to zero in the multivariate t model and the slope for the girls β_1 appears to be overestimated under the Gaussian model. The estimated variability for the MLEs is smaller in the multivariate t fit (the 95% confidence intervals are between 12 and 16% smaller than in the Gaussian model), suggesting that the parameters are estimated with greater precision.

Because of (14), the Gaussian linear mixed-effects model (1) can be viewed as a particular case of the multivariate t model (3). In the orthodontic growth example, the maximum log-likelihood for the Gaussian model is -203.021 and for the multivariate t model the maximum log-likelihood is -184.555, corresponding to likelihood ratio statistic of 36.932 (p-value of 10^{-8}). This indicates that the multivariate t model fits the data substantially better than the Gaussian model.

The estimated average distances δ_i^2, $\delta_{\boldsymbol{b}_i}^2$, and $\delta_{\boldsymbol{\epsilon}_i}^2$, defined in (11), provide useful diagnostic statistics for identifying subjects with outlying observations. Note that, under the Gaussian model (1), $\mathrm{E}\left(\delta_{\boldsymbol{b}_i}\right) = \mathrm{E}\left(\boldsymbol{b}_i' \boldsymbol{\Psi}^{-1} \boldsymbol{b}_i\right) = q$, $\mathrm{E}\left(\delta_{\boldsymbol{\epsilon}_i}\right) = \mathrm{E}\left[\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\right)' \boldsymbol{\Lambda}_i^{-1}\left(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta} - \boldsymbol{Z}_i\boldsymbol{b}_i\right)\right] = n_i$, and $\mathrm{E}\left[\delta_i^2\left(\boldsymbol{\beta}, \boldsymbol{\Psi}, \boldsymbol{\rho}\right)\right] = n_i$. Therefore, $\widehat{\delta}_i^2/n_i$, $\widehat{\delta}_{\boldsymbol{b}_i}^2/q$, and $\widehat{\delta}_{\boldsymbol{\epsilon}_i}^2/n_i$ are expected to be close to 1 under the Gaussian model, and can be used as diagnostics statistics for identifying subjects with outliers (under this Gaussian model). Figure 5 presents these diagnostic statistics for the boys (because of the large value of $\widehat{\lambda}_2$, the girls' estimated weights $\widehat{\tau}_i$ are all essentially equal to 1). Subjects `M09` and `M13` present large values of $\widehat{\delta}_i^2$ and $\widehat{\delta}_{\boldsymbol{e}_i}^2$, suggesting outlying observations at the within-subject level. This is consistent with the preliminary plot of the data, included in Figure 1, which suggests that both subjects have unusual growth patterns. The $\widehat{\delta}_{\boldsymbol{b}_i}^2$ plot gives some indication that subject `M10` is possibly a $\boldsymbol{b}$-outlier, which can not be concluded from Figure 3. Inspection of Figure 1 reveals that this subject has an unusually high orthodontic distance at the time of the first measurement.

Figure 5 about here

5.1.1 Influence of a single outlier

The robustness of the multivariate t MLEs with respect to the Gaussian MLEs can also be assessed through the influence of a single outlying observation (corresponding to a single $\boldsymbol{e}$-outlier) on the estimated parameters. To simplify, we consider only the model for the girls, which can be represented as

$$y_{ij} = \beta_0 + \beta_1 t_j + b_{0i} + b_{1i} t_j + e_{ij}, \quad i = 1, \ldots, 11 \; j = 1, \ldots, 4$$

with $\boldsymbol{b}_i \overset{\text{ind}}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{\Psi})$ and $e_{ij} \overset{\text{ind}}{\sim} \mathcal{N}(0, \sigma^2)$ in the Gaussian model and $\boldsymbol{b}_i | \tau_i \overset{\text{ind}}{\sim} \mathcal{N}\left(\mathbf{0}, \tau_i^{-1}\boldsymbol{\Psi}\right)$, $e_{ij}|\tau_i \overset{\text{ind}}{\sim} \mathcal{N}\left(0, \tau_i^{-1}\sigma^2_{g(i)}\right)$, and $\tau_i \overset{\text{ind}}{\sim} \text{Gamma}(\lambda/2, \lambda/2)$ in the multivariate t model

We consider the influence of a change of Δ units in a single measurement on the estimated parameters. That is, we replace a single data point y_{ij} by the contaminated value $y_{ij}(\Delta) = y_{ij} + \Delta$, re-estimate the parameters, and record the relative change in the estimates $\left(\widehat{\theta}(\Delta) - \widehat{\theta}\right)/\widehat{\theta}$, where $\widehat{\theta}$ denotes the original estimate and $\widehat{\theta}(\Delta)$ the estimate for the contaminated data. In this example, we contaminated a *typical* value, the fourth observation (age = 14 years) on subject `F01`, and varied Δ between -20mm and 20mm by increments of 2mm. The Gaussian and the multivariate t fits were identical for the uncontaminated data in this case. Because $\boldsymbol{\Psi}$ and σ^2 have different interpretations under the Gaussian model (1) and the multivariate t model (3), and even within the multivariate t model for different degrees of freedom, we concentrate here on the estimation of the fixed effects $\boldsymbol{\beta}$, which have the same interpretation under both models and for different degrees of freedom within the multivariate t model. We study the influence of the single outlier $y_{ij}(\Delta)$ on the estimation of $\widehat{\boldsymbol{\beta}}$ and of its estimated covariance matrix $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$.

Figure 6 presents the percent change curves for $\widehat{\boldsymbol{\beta}}$ and the upper-triangular elements of $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ for different values of Δ.

Figure 6 about here

The influence of the single outlier is unbounded in the case of the Gaussian model, but clearly bounded in the multivariate t model. In the Gaussian model, the outlying observation has considerable more impact on the estimates of $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ (changes between -2000% and 1800%), than on the fixed effects $\widehat{\boldsymbol{\beta}}$ (changes up to $\pm 60\%$). This has a direct impact on inferences drawn from the fit: confidence intervals increase unboundedly and test statistics go to zero. In the multivariate t fit, the influence of the single outlier for the fixed effects estimates remains bounded between -10% and 6% and for the estimates of $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ it remains between -107% and 86%.

For closer contamination values ($|\Delta| \leq 2$), the multivariate t fit and the Gaussian fit are essentially identical and therefore have the same influence curves. This occurs because the contaminated observation is not distant enough from the *typical* data to be identified as an outlier, resulting in $\widehat{\lambda} = \infty$. Therefore, the two estimation methods will have about the same efficiency for no or close contamination cases.

5.2 Comparing the MLEs under different outlier patterns

To compare the performance of the maximum likelihood estimators under the Gaussian model (1) and the multivariate t model (3), we conducted a simulation study involving different patterns of $\boldsymbol{b}$- and $\boldsymbol{e}$-outliers.

The linear mixed-effects model used to simulate the data is

$$\boldsymbol{y}_i = \boldsymbol{X}\left(\boldsymbol{\beta} + \boldsymbol{b}_i\right) + \boldsymbol{e}_i,\ i = 1,\ldots,27, \quad \boldsymbol{X} = \begin{bmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{bmatrix}, \tag{15}$$

with the following mixture of normals models being used to *contaminate* the distributions of the $\boldsymbol{b}_i$ and the $\boldsymbol{e}_i$.

$$\begin{aligned} \boldsymbol{b}_i &\overset{\text{ind}}{\sim} \left(1 - p_{\boldsymbol{b}}\right) \cdot \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Psi}\right) + p_{\boldsymbol{b}} f \cdot \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{\Psi}\right), \\ e_{ij} &\overset{\text{ind}}{\sim} \left(1 - p_{\boldsymbol{e}}\right) \cdot \mathcal{N}\left(0, \sigma^2\right) + p_{\boldsymbol{e}} f \cdot \mathcal{N}\left(0, \sigma^2\right),\ i = 1,\ldots,27, \quad j = 1,\ldots,4, \end{aligned} \tag{16}$$

where $p_{\boldsymbol{b}}$ and $p_{\boldsymbol{e}}$ denote, respectively, the expected percentage of $\boldsymbol{b}$- and $\boldsymbol{e}$-outliers in the data and f denotes the contamination factor. This model is a simplified version of the orthodontic growth model (2), with no gender differences. The parameters in the uncontaminated distributions are similar to the MLEs (13). It follows from (16) that $\operatorname{var}\left(\boldsymbol{b}_i\right) = \left[1 + \left(f^2 - 1\right) p_{\boldsymbol{b}}\right] \boldsymbol{\Psi}$ and $\operatorname{var}\left(e_{ij}\right) = \left[1 + \left(f^2 - 1\right) p_{\boldsymbol{e}}\right] \sigma^2$.

All thirty-two combinations of $p_{\boldsymbol{b}},\ p_{\boldsymbol{e}} = 0, 0.05, 0.1, 0.25$, and $f = 2, 4$ were used in the simulation study. The $f = 2$ case corresponds to a close contamination pattern, while $f = 4$ illustrates a more distant contamination pattern. A total of 500 Monte Carlo replications were obtained for each $\left(p_{\boldsymbol{b}}, p_{\boldsymbol{e}}, f\right)$ combination.

An S-PLUS implementation of the ECME algorithm of Section 4.3 was used to obtain the multivariate t MLEs at each replication. For the Gaussian MLEs, the `lme` function in S-PLUS (MathSoft, 1997) was used. To enhance the comparability of the results, the same data set was used to obtain the multivariate t estimates and the Gaussian estimates, at each replication. The degrees of freedom for the multivariate t-distribution were assumed unknown, being estimated in the ECME algorithm.

As mentioned in Section 5.1, $\boldsymbol{\Psi}$ and σ^2 have different interpretations under the Gaussian model (1) and the multivariate t model (3) and their corresponding MLEs under the two models cannot be directly compared. As before, we concentrate on the estimation of the fixed effects $\boldsymbol{\beta}$, which have the same interpretation under both models.

For the simulation model (15), under estimation method E, the approximate covariance matrix $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ of the fixed effects estimates has the form

$$\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = \left[\widehat{\sigma}_E^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} + \widehat{\boldsymbol{\Psi}}_E\right] / m.$$

Under the Gaussian model, $\widehat{\sigma}_G^2 = \widehat{\sigma}^2$ and $\widehat{\boldsymbol{\Psi}}_G = \widehat{\boldsymbol{\Psi}}$, while under the multivariate t model $\widehat{\sigma}_T^2 = m\left[\sum_{i=1}^m \left(\widehat{\nu}_i + n_i\right) / \left(\widehat{\nu}_i + n_i + 2\right)\right]^{-1} \widehat{\sigma}^2$ and $\widehat{\boldsymbol{\Psi}}_T = m\left[\sum_{i=1}^m \left(\widehat{\nu}_i + n_i\right) / \left(\widehat{\nu}_i + n_i + 2\right)\right]^{-1} \widehat{\boldsymbol{\Psi}}$, with $\widehat{\theta}$ denoting the MLE of θ under the appropriate estimation method. For the purpose of the simulation study, robustness is determined by how close the estimated values are to the parameters of the uncontaminated distribution. The asymptotic covariance matrix for the MLE of $\boldsymbol{\beta}$ based on the uncontaminated data only is $\left[\sigma^2 \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} + \boldsymbol{\Psi}\right]/m$. Therefore, we define σ^2 as the target value for $\widehat{\sigma}^2_G$ and $\widehat{\sigma}^2_T$ σ^2 and $\boldsymbol{\Psi}$ as the target value for $\widehat{\boldsymbol{\Psi}}_G$ and $\widehat{\boldsymbol{\Psi}}_T$. These estimators can then be used to compare the performance of the two estimation methods with respect to the variance-covariance components $\boldsymbol{\Psi}$ and σ^2.

The following parameters, with respective *target* values, are used in the comparison of the two estimation methods:

$$\beta_0 = 17, \quad \beta_1 = 0.8, \quad \Psi_{11} = 4, \quad \Psi_{22} = 0.0225, \quad \Psi_{12} = 0, \quad \text{and} \quad \sigma^2 = 1.$$

For the Gaussian model, the MLEs are considered and for the multivariate t model the MLEs of the fixed effects and the modified estimators $\widehat{\boldsymbol{\Psi}}_T$ and $\widehat{\sigma}^2_T$ of the variance-covariance components are considered.

Let θ denote a parameter of interest, with target value $\theta_0 \neq 0$, estimated by $\widehat{\theta}$. The efficiency of the Gaussian estimator $\widehat{\theta}_G$ relative to the multivariate t estimator $\widehat{\theta}_T$ is defined as the ratio of the respective mean square errors, $\mathrm{E}\left(\widehat{\theta}_G - \theta_0\right)^2 / \mathrm{E}\left(\widehat{\theta}_T - \theta_0\right)^2$. Expectations are taken with respect to the simulation distribution, that is, $\mathrm{E}\left(\widehat{\theta} - \theta_0\right)^2 = \sum_{i=1}^{500}\left(\widehat{\theta}_i - \theta_o\right)^2/500$.

Figures 7 and 8 present the relative efficiency of the multivariate t estimators with respect to the Gaussian estimators. There are substantial gains in efficiency for all parameters under the more distant contamination patterns ($f = 4$) and moderate gains under the close contamination patterns ($f = 2$). The efficiency gains are bigger for the variance-covariance components than for the fixed effects. The two methods have about the same efficiency under the no-contamination case. For the close contamination patterns (Figure 7), the efficiency increases with the percentage of $\boldsymbol{b}-$ and $\boldsymbol{e}-$outliers (except for the Ψ_{11} parameter, for which there is a slight efficiency decrease when the percentage of $\boldsymbol{e}-$outliers increases from 10% to 25%). In the case of distant contamination (Figure 8, the efficiency shows a non-monotone behavior with respect to the percentage of $\boldsymbol{b}-$ and $\boldsymbol{e}-$outliers. This pattern suggests that the multivariate t model is more robust than the Gaussian model especially for moderate percentages (5-10%) of outliers.

Figures 7 and 8 about here

The simulation results for the mean square error (not shown here) indicate that outliers affect the variance-covariance components estimates more than they affect the fixed effects estimates. The precision of the estimator of σ^2 seems to be affected only by the percentage of $\boldsymbol{e}-$outliers, while the fixed effects and random effects variance-covariance components estimators are affected by both types of outliers.

The MLEs of the fixed effects are nearly unbiased (relative bias $\leq 0.7\%$) for both estimation methods under all contamination patterns. The bias for the variance-covariance components follows the same basic pattern as the mean square error: it increases with the percentage of $\boldsymbol{e}-$outliers, is insensitive to the percentage of $\boldsymbol{b}-$outliers for σ^2, and increases in absolute value with the percentage of both types of outliers for the random effects variance-covariance components.

The coverage probabilities of the approximate 95% confidence intervals for the fixed effects, not included here, are generally close to the nominal level for both estimation methods, with the smallest coverage probability 90.4% and the largest 97%. The coverage probabilities tend to increase with the percentage of outliers, because the fixed effects estimators remain unbiased and the confidence intervals get larger. The average length of the 95% confidence intervals is about the same under Gaussian and multivariate t estimation for the close contamination patterns, but 10% to 25% larger in the Gaussian model for the more distant contamination patterns.

6 Conclusion

This article describes a robust version of the linear mixed-effects model of Laird and Ware (1982) in which the Gaussian distributions for the random effects and the within-subject errors are replaced by multivariate t-distributions. Analysis of examples and simulation results indicate that the multivariate t linear mixed-effects model substantially outperforms the Gaussian model when outliers are present in the data, even in moderate amounts. Gains in efficiency for the multivariate t MLEs relative to the Gaussian MLEs, under outlier contamination, are observed for all parameters, being particularly high in the estimation of variance-covariance components, ranging from 20%–30% in the case of close contamination (two standard deviations) to 200%–400% in the case of distant contamination (four standard deviations). This has a direct impact on confidence intervals and test statistics obtained from the fit, which determine all inferences drawn from the estimated model. The influence function is bounded for the multivariate t model and unbounded for the Gaussian model. The multivariate t model also provides diagnostics tools for graphically identifying subjects with outlying observations.

We describe EM-type algorithms for efficient maximum likelihood estimation under two missing data structures: with both the random effects and the individual weights treated as missing and with only the individual weights treated as missing. The former leads to an algorithm with closed form expressions for both the E- and the M-step, but imposes some restrictions on the correlation structure of the within-subject errors. The algorithm corresponding to the latter missing data structure, which allows general correlation structures for the within subject errors, involves a more computationally intensive M-step, but can be implemented using existing, reliable software.

Further extensions of the multivariate t model we propose are still possible. Our model formulation assumes that a single weight τ_i is associated with both the random effects and the within-subject error, for each subject. A more general formulation would allow different weights for the random effects and for the within-subject errors, with possibly different degrees of freedom. The main difficulty with this approach is that the marginal distribution of $\boldsymbol{y}_i$ and the conditional distributions of $\tau_i|\boldsymbol{y}_i$ and $\phi_i|\boldsymbol{y}_i$ become intractable and, as a result, the E-step of the EM algorithms becomes a far more complex numerical problem. Because we believe that the cases in which this more general multivariate t model could perform better than the simpler *single weight* model are not likely to be observed in practice, we feel that the latter is more useful in applications. The robust estimation approach described in this article can also be extended to nonlinear mixed-effects models (Lindstrom and Bates, 1990). The computations become considerably more complex, but algorithms based on linear approximations to the marginal distribution of the $\boldsymbol{y}_i$ can, in principle, be used in conjunction with the methods described here.

References

Barnett, V. and Lewis, T. (1994). *Outliers in statistical data (Third edition)*, John Wiley & Sons.

Bates, D. M. and Watts, D. G. (1988). *Nonlinear Regression Analysis and Its Applications*, Wiley, New York.

Becker, R. A., Cleveland, W. S. and Shyu, M.-J. (1996). The visual design and control of trellis graphics displays, *J. of Computational and Graphical Statistics* **5**(2): 123–156.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, 3rd edn, Holden-Day, San Francisco.

Dempster, A. P., Laird, N. M. and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm (c/r: P22-37), *Journal of the Royal Statistical Society, Ser. B* **39**: 1–22.

Efron, B. and Tibshirani, R. J. (1993). *An Introduction to the Bootstrap*, Chapman & Hall, New York.

Fang, K. T. and Zhang, Y. Y. (1990). *Generalized Multivariate Analysis*, Science Press, Beijing and Springer–Verlag, Berlin.

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J. and Stahel, W. A. (1986). *Robust Statistics: The Approach Based on Influence Functions*, John Wiley & Sons, New York.

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

Huber, P. J. (1981). *Robust Statistics*, Wiley, New York.

Jennrich, R. I. and Schluchter, M. D. (1986). Unbalanced repeated measures models with structural covariance matrices, *Biometrics* **42**(4): 805–820.

Johnson, N. L. and Kotz, S. (1972). *Distributions in Statistics: Continuous Multivariate Distributions*, Wiley, New York.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data, *Biometrics* **38**: 963–974.

Lange, K. L., Little, R. J. A. and Taylor, J. M. G. (1989). Robust statistical modeling using the t-distribution, *Journal of the American Statistical Association* **84**: 881–896.

Lindstrom, M. J. and Bates, D. M. (1988). Newton-Raphson and EM algorithms for linear mixed-effects models for repeated-measures data, *Journal of the American Statistical Association* **83**: 1014–1022.

Lindstrom, M. J. and Bates, D. M. (1990). Nonlinear mixed effects models for repeated measures data, *Biometrics* **46**: 673–687.

Liu, C. and Rubin, D. B. (1994). The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence, *Biometrika* **81**: 633–648.

MathSoft (1997). *S-PLUS 4 Guide to Statistics*, Data Analysis Products, MathSoft Inc., Seattle, WA.

Meng, X. L. and Rubin, D. B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework, *Biometrika* **80**: 267–278.

Miller, J. J. (1977). Asymptotic properties of maximum likelihood estimates in the mixed model of the analysis of variance, *Ann. of Statistics* **5**: 746–762.

Pendergast, J. F. and Broffitt, J. D. (1986). Robust estimation in growth curve models, *Communications in Statistics: Theory and Methods* **14**: 1919–1939.

Potthoff, R. F. and Roy, S. N. (1964). A generalized multivariate analysis of variance model useful especially for growth curve problems, *Biometrika* **51**: 313–326.

Racine-Poon, A. (1992). Saga: Samples assisted graphical analysis (disc: P401-404), *Bayesian Statistics 4. Proceedings of the Fourth Valencia International Meeting*, pp. 389–401.

Richardson, A. M. (1997). Bounded influence estimation in the mixed linear model, *Journal of the American Statistical Association* **92**(437): 154–161.

Richardson, A. M. and Welsh, A. H. (1995). Robust estimation in the mixed linear model, *Biometrics* **51**: 1429–1439.

Thisted, R. A. (1988). *Elements of Statistical Computing*, Springer-Verlag, London.

Wakefield, J. C., Smith, A. F. M., Racine-Poon, A. and Gelfand, A. E. (1994). Bayesian analysis of linear and nonlinear population models using the Gibbs sampler, *Applied Statistics*. Accepted for publication.

Wolfinger, R., Tobias, R. and Sall, J. (1991). Mixed models: A future direction, *SUGI* **16**: 1380–1388.

Wu, C. F. J. (1983). On the convergence properties of the EM algorithm, *The Annals of Statistics* **11**: 95–103.

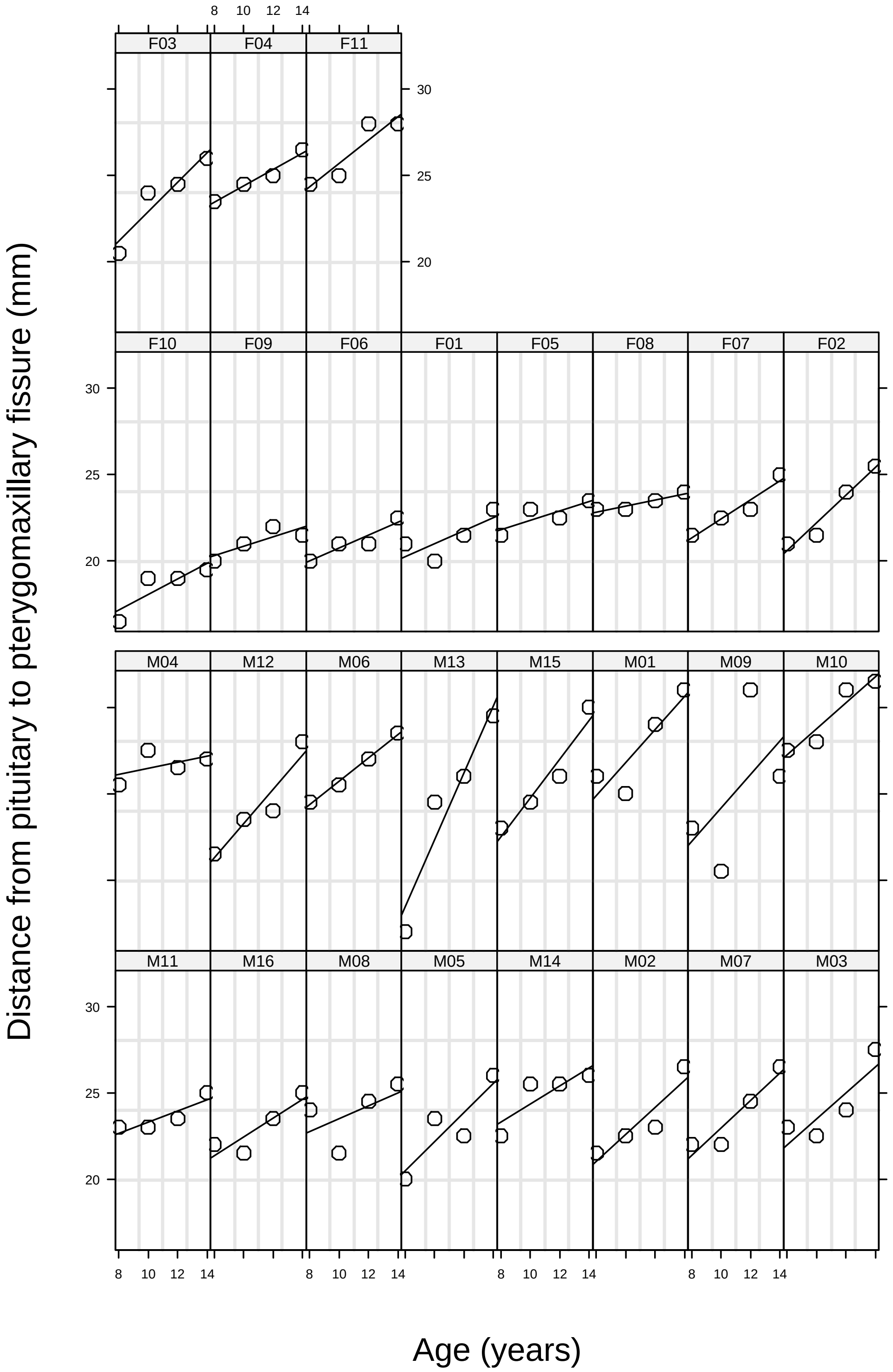

Figure 1: Orthodontic growth patterns in 16 boys(M) and 11 girls(F) between 8 and 14 years of age. Lines represent the individual least squares fits of the simple linear regression model.

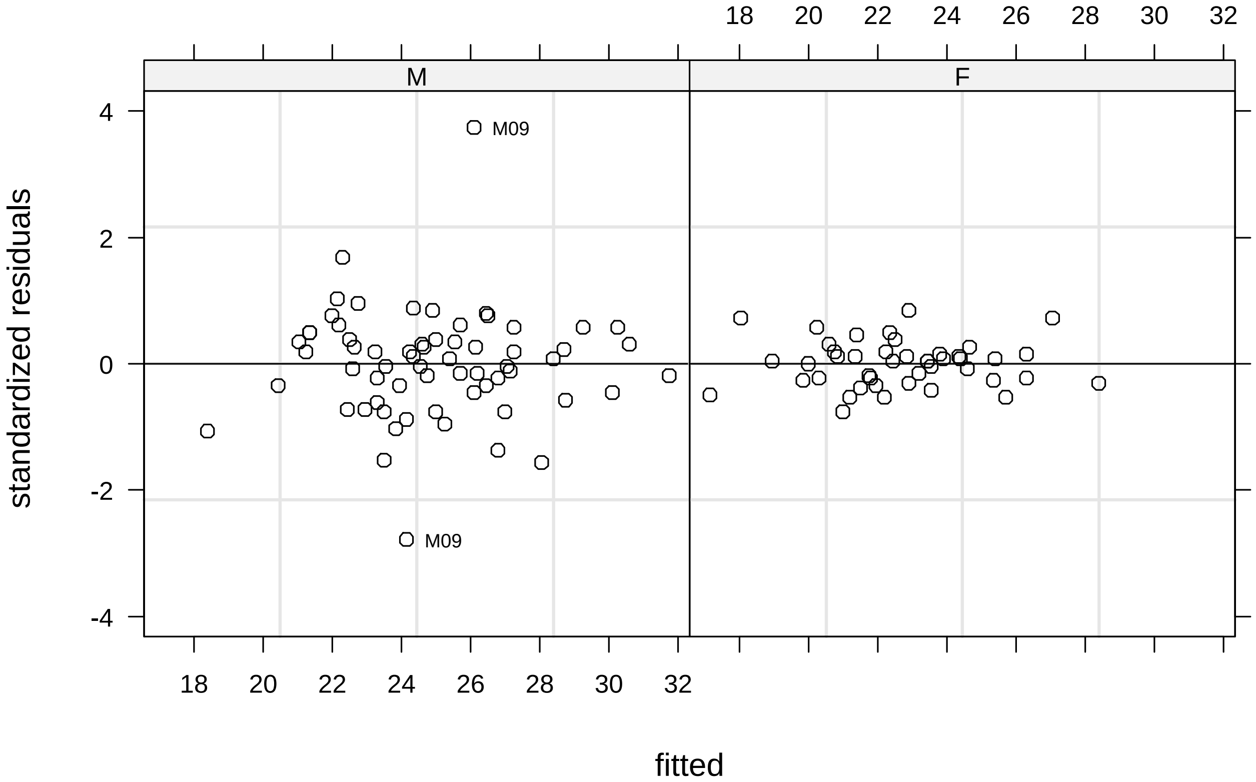

Figure 2: Residuals versus fitted values plots by gender, corresponding to individual least squares fits of the orthodontic growth data.

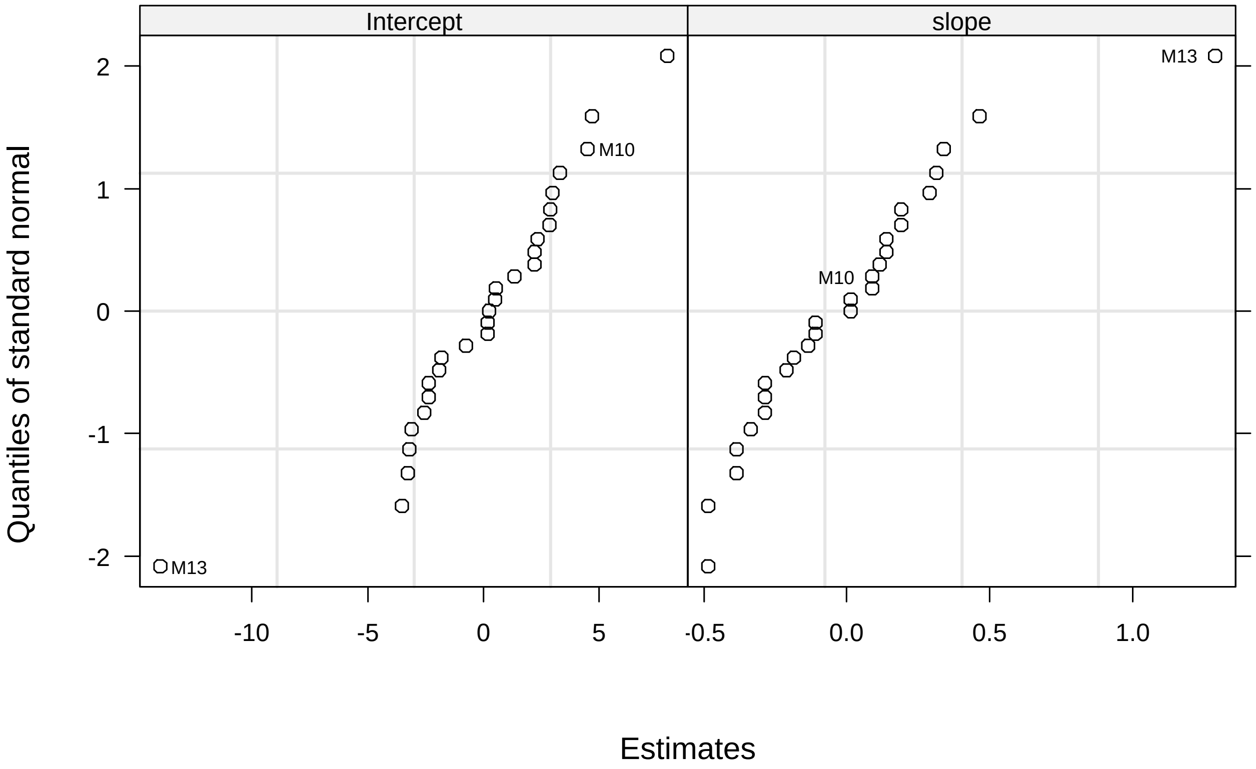

Figure 3: Normal plots of estimated coefficients corresponding to individual least squares fits of the orthodontic growth data.

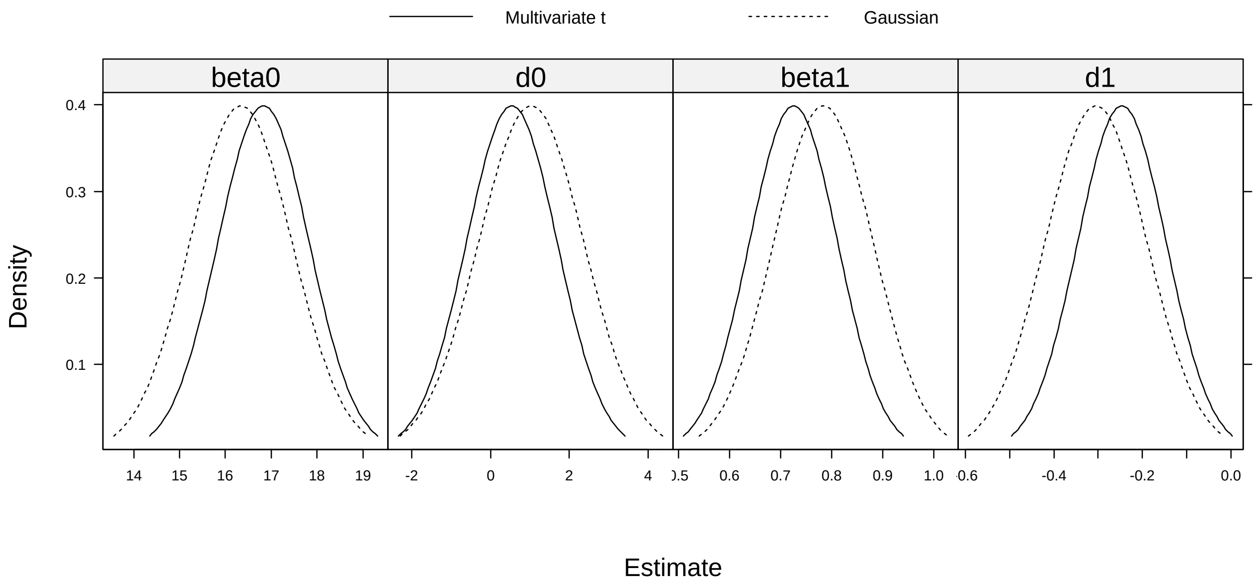

Figure 4: Approximate densities of the fixed effects MLEs in the orthodontic growth model (2) under Gaussian and multivariate t estimation.

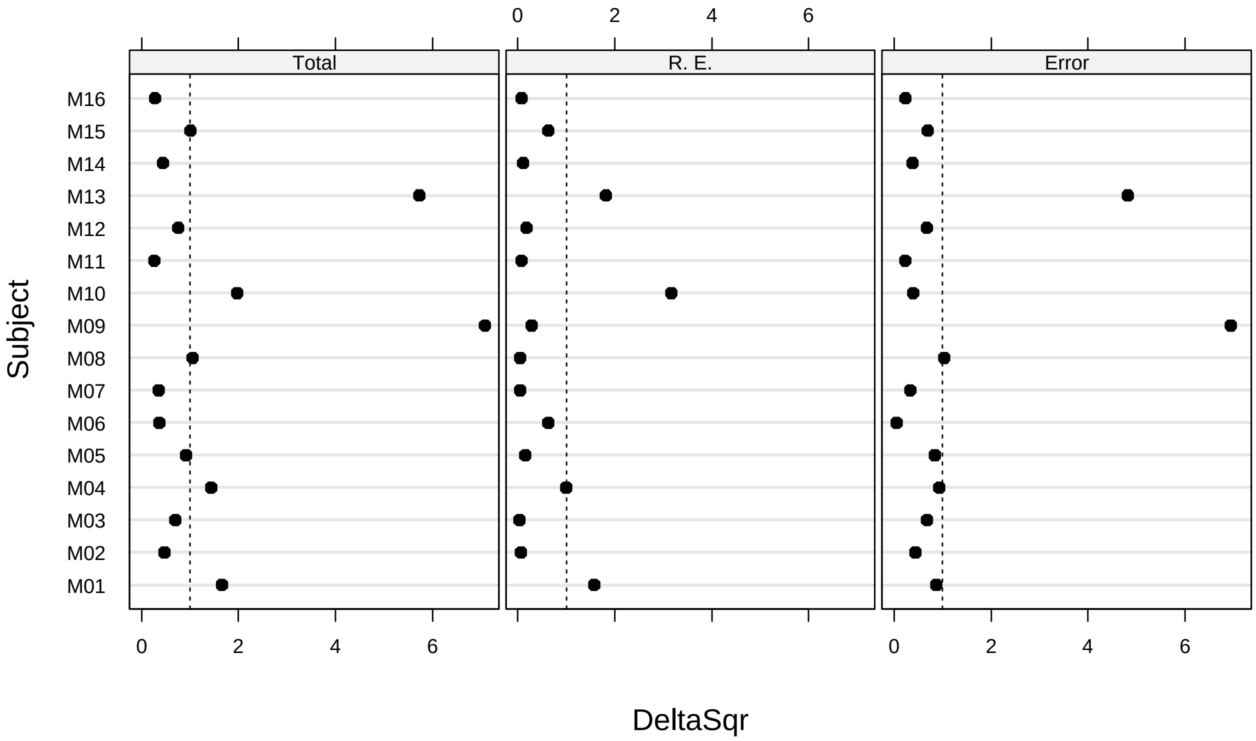

Figure 5: Estimated δ_i^2 (Total), $\delta_{\boldsymbol{b}_i}$ (R.E.), and $\delta_{\boldsymbol{e}_i}$ (Error) for boys in the multivariate t fit of the orthodontic distance data.

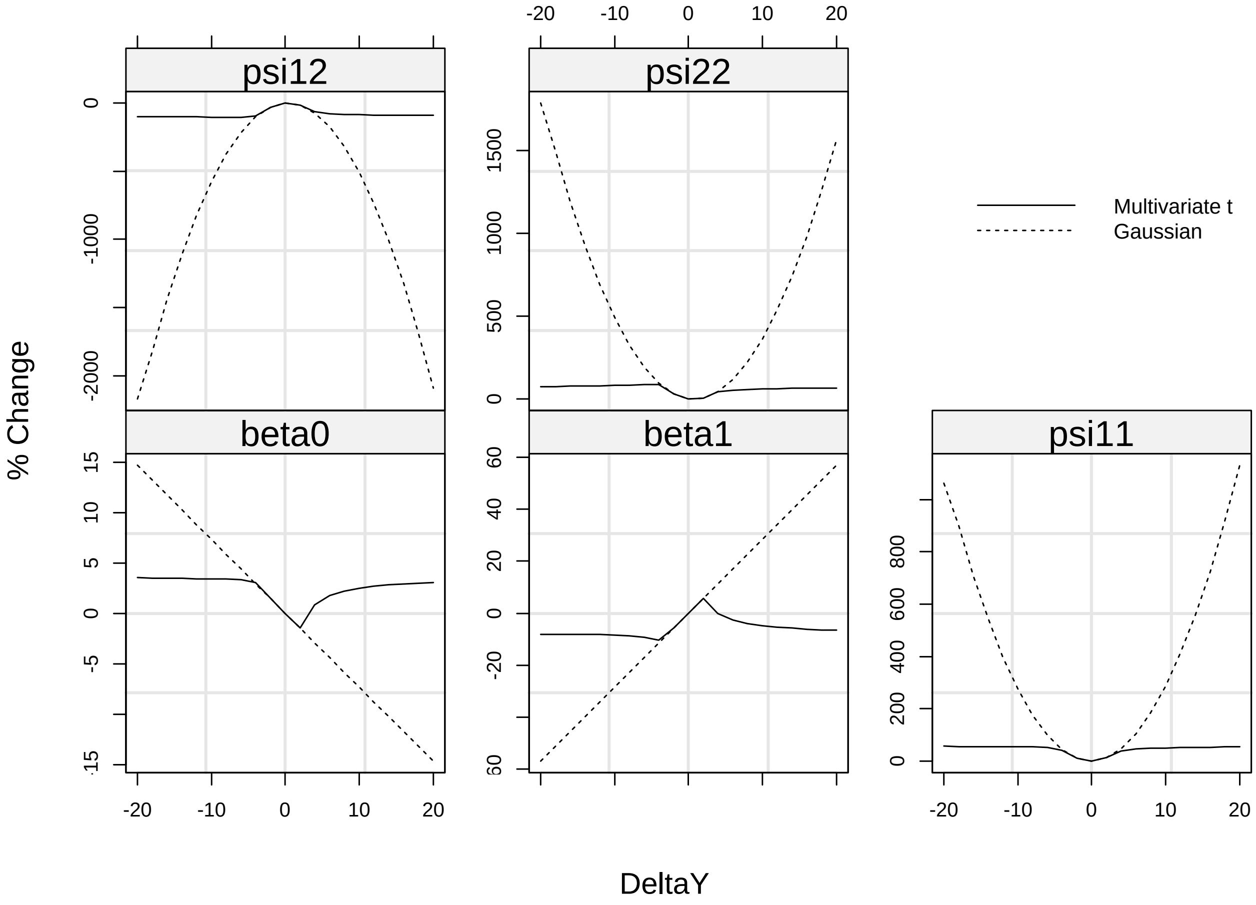

Figure 6: Percent change in maximum likelihood estimates under the Gaussian and multivariate t models for different contaminations Δ of a single observation.

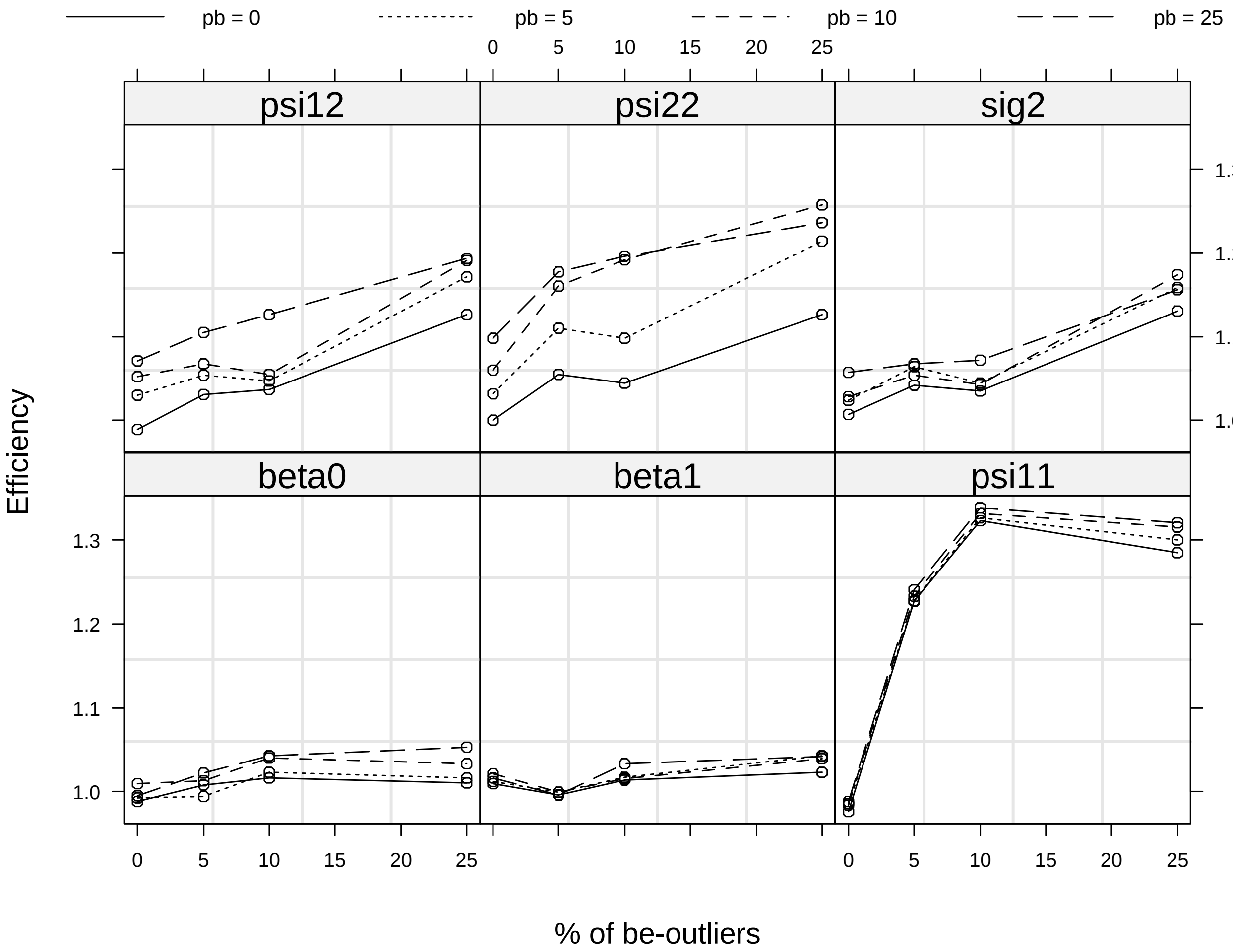

Figure 7: Relative efficiencies of the multivariate t MLEs with respect to the Gaussian MLEs under close outlier contamination patterns.

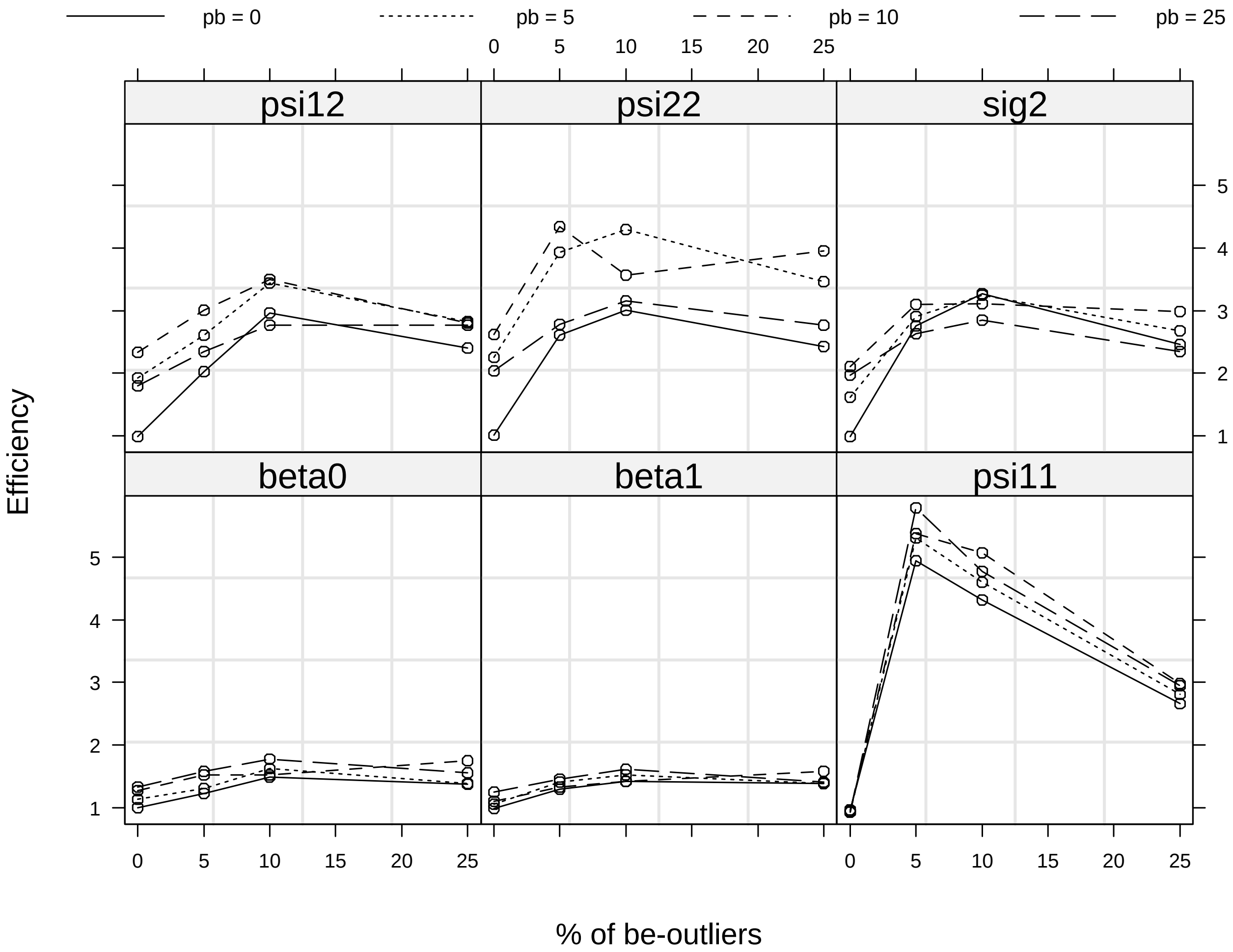

Figure 8: Relative efficiencies of the multivariate t MLEs with respect to the Gaussian MLEs under distant outlier contamination patterns.