**Evaluation of the parameter estimators behavior in a nonlinear model with misspecified random effects distribution**

María del Carmen Garcia
*Institute of Theoretical and Applied Research, School of Statistics. National University of Rosario. Argentina–* mgarcia@fcecon.unr.edu.ar

Cecilia Rapelli
*Institute of Theoretical and Applied Research, School of Statistics. National University of Rosario. Argentina–* cecirapelli@hotmail.com

Mara Lis Catalano*
*Institute of Theoretical and Applied Research, School of Statistics. National University of Rosario. Argentina–* catalano@fceia.unr.edu.ar

Noelia Castellana
*Institute of Theoretical and Applied Research, School of Statistics. National University of Rosario. Argentina–* noecastellana@hotmail.com

**Abstract**

Nonlinear mixed models provide a framework for analysis longitudinal data within a large number of applications. To characterize the population variation, these models express the individual specific parameters in terms of fixed and random effects, and to consider the correlation between repeated measurements introduce intra unit errors. A basic assumption for the random effects is the normal distribution. This assumption is often not met and its compliance is difficult to verify with standard statistical tools. The methods for estimating these models are based on the normality assumption. In this paper, through simulations, we present a preliminary investigation of the impact of misspecifying the random effects distribution on the fixed effects estimation and the random effects prediction. Data are generated according to a specified model in which the random effects have not Gaussian distribution and analyzed assuming that the normality assumption is met. For the random effects are considered three distributions, normal, t-Student and multivariate Gamma. Results allow a better understanding of the consequences of violating the estimation procedure assumptions, but cannot draw general conclusions.

**Keywords:** nonlinear mixed models; random effects; no Normality.

## 1. Introduction

The mixed nonlinear models are used, frequently, to characterize longitudinally biological processes
In these models the specific parameters of the units are expressed in terms of fixed and random effects, and to consider the repeated measurements correlation introduced intra unit errors..
A standard assumption is the normal distribution for the errors and the random effects. The assumption on the latter usually is not met and its compliance could be difficult to check with standard statistical tools.
Several authors have evaluated the non-normality of the random effects on the fixed effects estimates in mixed nonlinear models, but have not investigated the effect of the noncompliance of this assumption on the random effects prediction.
In this paper, through simulations, it is investigated the impact of the incorrect specification of the random effects distribution on the fixed effects estimation and random effects prediction, as well as, how these last effects recover the true underlying distribution.

## 2. Nonlinear mixed model

The general form of the nonlinear mixed model is given by,

$$\mathbf{Y}_i = f(\mathbf{X}_i, \boldsymbol{\beta}_i) + \mathbf{e}_i, \qquad (2.1)$$

$$\boldsymbol{\beta}_i = \mathbf{A}_i\boldsymbol{\beta} + \mathbf{B}_i\mathbf{b}_i, \qquad (2.2)$$

where, $\mathbf{Y}_i = [Y_{i1}, \ldots, Y_{in}]'$, i = 1, ... ,N, is the vector (nx1) composed of repeated measurements of the unit i, and $Y_{ij}$ is the observation to the unit i at the time $t_j$, $j = 1, \ldots, n$,

$\boldsymbol{f}(\mathbf{X}_i, \boldsymbol{\beta}_i) = [f(\mathbf{x}_{i1}, \boldsymbol{\beta}_i), \ldots, f(\mathbf{x}_{in}, \boldsymbol{\beta}_i)]'$ being, $f$ a known nonlinear function that relates the response vector with time and with other intra-unit possible covariates ($\mathbf{X}_i$) and $\boldsymbol{\beta}_i$ is a unit-specific vector containing the parameters of the nonlinear function.

$\boldsymbol{\beta}$: fixed effects vector (sx1), $\mathbf{b}_i$: random effects vector (qx1), $\mathbf{A}_i$: design matrix for fixed effects (rxs), $\mathbf{B}_i$: design matrix for random effects (rxq).

It is supposed that $\mathbf{b}_i$ and $\mathbf{e}_i$ are independent with distribution, $\mathbf{b}_i \overset{\text{i.i.d.}}{\sim} N_q(\mathbf{0}, \mathbf{D})$ and $\mathbf{e}_i \overset{\text{i.i.d.}}{\sim} N_n(\mathbf{0}, \boldsymbol{\Sigma})$, where, $\mathbf{D}$ is the random effects covariance matrix and $\boldsymbol{\Sigma}$, the intra-unit covariance matrix which have the same structure for all units.

The model can be estimated by the maximum likelihood method. There are various alternatives for estimating the complete likelihood of these models, which are based on the first order Taylor expansion of the model function $f$. The main distinction between these methods, known as "of linearization", lies in the point where the expansion is done: around the expected value of the random effects vector (**0**) (Sheiner y Beal, 1980), or around any estimate of the random effects vector, usually called the best linear unbiased predictor (EBLUP) (Lindstrom y Bates, 1990).

## 3. Simulation study

In order to evaluate the effect of misspecifying the random-effects distribution is performed a simulation study. The data is generated from a model that considers the nonlinear function of Wood.

The choice of parameters for the simulation is based on the results of the fit of a model to assess the evolution of lactation in "Holando" cows (Garcia et. al., 2009).

The data represent measurements of milk production in 15 times to 120 cows, recorded according to parity (1, 2, 3 and more) that corresponds that breastfeeding. It is assumed that the parameters of the Woods curve are a linear function of three fixed effects and the two first have random effects.

Function expression ((2.1) and (2.2)) and parameters values used in the simulation are as follows,

$$Y_{ij} = \beta_{0i}\, t_{ij}^{\beta_{1i}} \exp(-\beta_{2i}\, t_{ij}) + e_{ij} \qquad (3.1)$$

$$\mathbf{e}_i \sim N(\mathbf{0}, \sigma^2\mathbf{I})$$

$$\beta_{0i} = 24.01 - 6.40\,P_1 - 1.88\,P_2 + b_{0i}, \quad \beta_{1i} = 0.35 - 0.10\,P_1 - 0.06\,P_2 + b_{1i}, \quad \beta_{2i} = 0.10 - 0.04\,P_1 - 0.01\,P_2$$

being,

$P_1$=1 first birth and 0 otherwise. $P_2$=1 second birth and 0 otherwise.

$\beta_{0i}$ the initial value of production, $\beta_{1i}$ ascent rate, $\beta_{2i}$ descent rate.

$\mathbf{b}_i$ random effects vector (kx1), $\mathbf{e}_i$ intra cow errors vector (nx1).

The intra cow variance is $\sigma^2 = 7.254$.

The vector $\mathbf{b}_i = \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix}$ has known distribution with zero mean and covariance matrix $\mathbf{D} = \begin{pmatrix} 33.291 & -0.297 \\ -0.297 & 0.011 \end{pmatrix}$.

To generate the random effects are considered three distributions,

1.- normal $\mathbf{b}_i \overset{iid}{\sim} N(\mathbf{0},\mathbf{D})$,

2.- t Student with 3 degree of freedom $\mathbf{b}_i \overset{iid}{\sim} t_2(3,\mathbf{0},\mathbf{D})$ y

3.- Gamma multivariate with covariance matrix **D**.

**3.1. Estimating fixed effects and covariance parameters**

Three datasets of size 120 are generated, assuming for each a different distribution for the random effects. For each set, the proposed model assuming normality of the random effects is fitted and parameter estimates and their variances are recorded. The procedure is repeated 200 times. The average of the parameter estimates and their variances are calculated.

**3.2. Prediction of random effects**

Six sets of 1200 data units are generated, 400 for each number of lactation considering different magnitudes for intra cows variance ($\sigma^2$= 30, 5 y 0.5). Each set corresponds to a combination of distributions of random effects and $\sigma^2$.

For each cow estimated parameters and random effects are calculated assuming normality.

IML procedure and macro NLINMIX from SAS statistical software was used for both schemes.

## 4. Results

The values generated are used to fit the model

$$Y_{ij} = \beta_{0i}\, t_{ij}^{\beta_{1i}} \exp(-\beta_{2i}\, t_{ij}) + e_{ij}$$

$$\beta_{0i} = \beta_0 + \beta_{01}\ P_1 + \beta_{02}\ P_2 + b_{0i},\quad \beta_{1i} = \beta_1 + \beta_{11}\ P_1 + \beta_{12}\ P_2 + b_{1i},\quad \beta_{2i} = \beta_2 + \beta_{21}\ P_1 + \beta_{22}\ P_2\ ,$$

and calculate the prediction of random effects.

The following table summarizes the behavior of the estimators of the parameters of: fixed effects, covariance and variance of the error in terms of bias and standard error of the estimates.

Table 1: Simulation results for the fixed effects and covariance parameters

| Parameters | True values | Normal distribution | | | t distribution with 3 df | | | Gamma distribution | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Bias | Std Error | Mean | Bias | Std Error | Mean | Bias | Std Error |
| $\beta_0$ | 24.01 | 23.92 | -0.09 | 0.9277 | 23.80 | -0.21 | 0.9229 | 29.67 | 5.66 | 0.9067 |
| $\beta_{01}$ | -6.40 | -6.44 | -0.04 | 1.4410 | -6.36 | 0.04 | 1.2566 | -6.28 | 0.12 | 1.1578 |
| $\beta_{02}$ | -1.88 | -1.94 | -0.06 | 1.4657 | -1.80 | 0.08 | 1.2503 | -1.89 | -0.01 | 1.4169 |
| $\beta_1$ | 0.34 | 0.35 | 0.01 | 0.0267 | 0.35 | 0.01 | 0.0236 | 0.45 | 0.11 | 0.0207 |
| $\beta_{11}$ | -0.10 | -0.09 | 0.01 | 0.0437 | -0.09 | 0.01 | 0.0366 | -0.10 | 0.00 | 0.0300 |
| $\beta_{12}$ | -0.05 | -0.06 | -0.01 | 0.0406 | -0.05 | 0.00 | 0.0361 | -0.05 | 0.00 | 0.0325 |
| $\beta_2$ | 0.10 | 0.10 | 0.00 | 0.0034 | 0.10 | 0.00 | 0.0038 | 0.10 | 0.00 | 0.0024 |
| $\beta_{21}$ | -0.03 | -0.03 | 0.00 | 0.0060 | -0.03 | 0.00 | 0.0059 | -0.03 | 0.00 | 0.0040 |
| $\beta_{22}$ | -0.01 | -0.01 | 0.00 | 0.0057 | -0.01 | 0.00 | 0.0058 | -0.01 | 0.00 | 0.0039 |
| $D_{00}$ | 33.29 | 32.98 | -0.31 | 4.8676 | 27.44 | -5.85 | 8.5643 | 32.84 | -0.45 | 9.1883 |
| $D_{01}$ | -0.30 | -0.30 | 0.00 | 0.0793 | -0.27 | 0.03 | 0.1331 | -0.23 | 0.07 | 0.0596 |
| $D_{11}$ | 0.01 | 0.01 | 0.00 | 0.0018 | 0.01 | 0.00 | 0.0144 | 0.01 | 0.00 | 0.0029 |
| $\sigma^2$ | 7.25 | 7.24 | -0.01 | 0.2414 | 7.29 | 0.04 | 0.5145 | 7.24 | -0.01 | 0.2390 |

The estimation of fixed effects is fairly robust to moderate departures from normality of random effects distribution. The reason for this, as discussed in Hartford et al. (2000), may be due to the fact that the results are deleted when the procedure does not converge and perhaps the non removed data sets were "more normal" than the omitted data sets. Data does not provide the same results for some of the covariance parameters.

Figure 1: "Box plots" of the estimates of the covariance parameters and $\hat{\sigma}^2 = S^2$

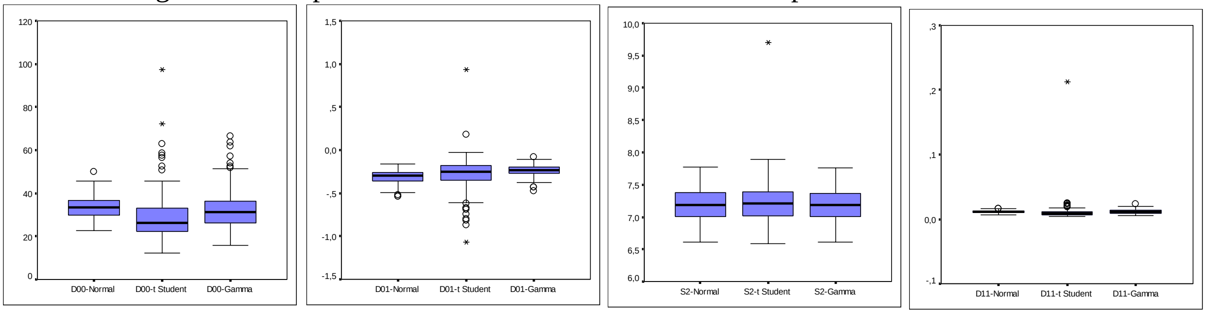



Figure 1 shows that the sampling distribution of the estimators of **D** is highly skewed in some cases where the random effects are not normal. It also notes the presence of outliers.

A general conclusion drawn from the above graphs is that the estimate of the mean of the parameters is little affected by deviations from normality of the random effects, while the variability of the latter is more affected.

The graph presented below allows visualizing how the prediction of random effects behaves when assuming a Gaussian model when the true distribution thereof is not.

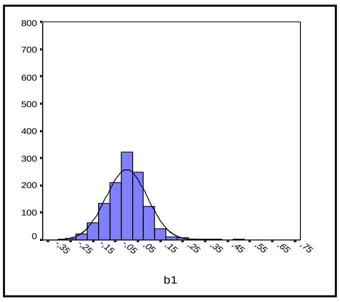
This graph corresponds to the sampling distribution of the prediction of random effects associated to $\hat{\beta}_1$, when Gaussian distribution is assumed, while the true distribution is t-Student (left) and gamma (right).

The histogram of prediction of random effects shows, for large values of $\sigma^2$, that the normality assumption would force values $\hat{b}_i$ satisfy this assumption, not reflecting the distribution from which they come.

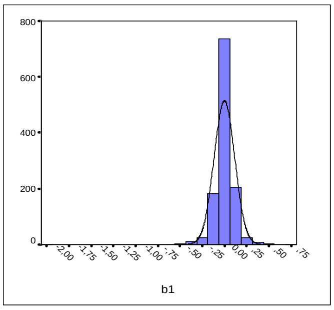
Figure 2: Distribution of the prediction of the random effect of the ascent rate for different variances

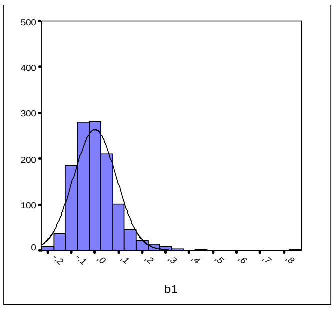
| Distribution prediction when the true distribution is t-Student | True distribution of the random effects | Distribution prediction when the true distribution is Gamma |
|---|---|---|

$\sigma^2 = 30$

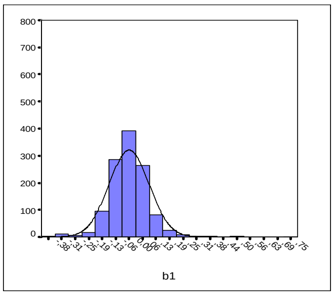



t-Student

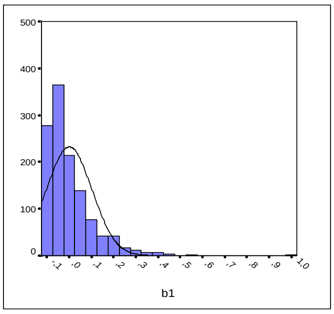


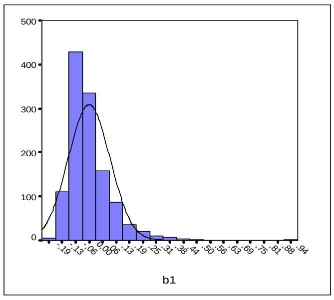




$\sigma^2 = 5$

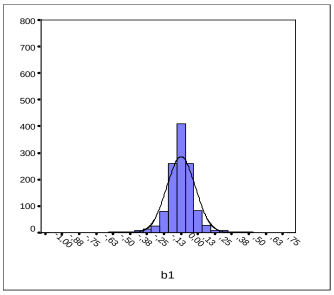


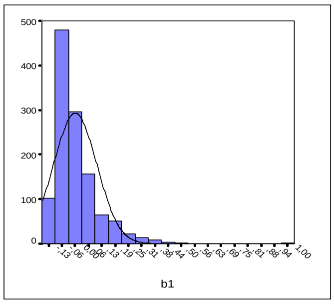

Gamma




$\sigma^2 = 0.5$




## 5. Conclusions

In this paper it is considered the problem that could arise when the random-effects distribution is not Gaussian, which is the distribution that is assumed by the method used to estimate the parameters of nonlinear mixed models. The compliance of this assumption could be difficult to verify with standard statistical tools, because the random-effects prediction depends on both random-effects and random-errors.

Through simulations, it is studied the impact of misspecifying the random-effects distribution on fixed-effects estimation and random-effects prediction, as well as, how these last effects recover the true underlying distribution.

It is observed that:

- √ The fixed-effects estimation is not compromised by the lack of normality, however bias is observed in the estimation of some parameters of covariance.
- √ The random effect prediction seems to depend on the assumed normality assumption and the predictions distribution, in this study, does not reflect the distribution of the population from which they come.
- √ The error variance appears to play an important role in the shape of the distribution of the random effects prediction.

On summary, the estimation of the marginal model parameters in this particular nonlinear mixed model is slightly influenced by the normality assumption for the random effects. However, this conclusion it is not entirely true for the prediction of random effects.

**References**


- Garcia, M. del C., Rapelli, C., Cuatrín, A., 2010 Multilevel nonlinear mixed model for modeling and choosing a lactaction curve BIOCELL vol. 34 ISSN 0327-9545 (abstract)
- Hartford, A. y Davidian, M. 2000 Consequences of misspecifying assumptions in nonlinear mixed effects models. Computational Statistics & Data Analysis 34, 139-164
- Lindstrom, M., Bates, D., 1990. Nonlinear mixed effects models for repeated measures data. Biometrics 46, 673-687.
- Sheiner, L., Beal, S., 1980. Evaluation of methods for estimating population pharmacokinetic parameters. I. Michaelis-Menten model: routine clinical pharmacokinetic data. J. Pharmacokin. Biopharm. 8, 553-571.
- Verbecke, G. and Lesaffre, E. 1996. A linear mixed-effects model with heterogeneity in the random-effects population. Journal of the American Statistical Association 91, 217–221.
- Verbecke, G. and Lesaffre, E. 1997. The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. Computational Statistics and Data Analysis 23, 541–556.