Extending the Box–Cox transformation to the linear mixed model

Matthew J. Gurka

University of Virginia, Charlottesville, USA

and Lloyd J. Edwards, Keith E. Muller and Lawrence L. Kupper

University of North Carolina, Chapel Hill, USA

Summary. For a univariate linear model, the Box–Cox method helps to choose a response transformation to ensure the validity of a Gaussian distribution and related assumptions. The desire to extend the method to a linear mixed model raises many vexing questions. Most importantly, how do the distributions of the *two* sources of randomness (pure error and random effects) interact in determining the validity of assumptions? For an otherwise valid model, we prove that the success of a transformation may be judged solely in terms of how closely the *total error* follows a Gaussian distribution. Hence the approach avoids the complexity of separately evaluating pure errors and random effects. The extension of the transformation to the mixed model requires an exploration of its potential effect on estimation and inference of the model parameters. Analysis of longitudinal pulmonary function data and Monte Carlo simulations illustrate the methodology discussed.

Keywords: Linear mixed model; Longitudinal data; Lung function; Normality; Random effects; Transformation

1. Introduction

1.1. Motivation

In a univariate linear model, satisfying the Gaussian distribution and related assumptions, even approximately, often requires transforming the response variable. The work of Box and Cox (1964), among others, has led to the development of systematic methods for transforming the response in linear models with independent and identically distributed error terms. Although often used in practice in longitudinal data analysis settings owing to the familiarity of the Box–Cox transformation for univariate linear models, little is known about such transformations when applied to linear mixed models.

It could prove insightful to assess the utility of a transformation in a mixed model setting for an area of research that has not traditionally used transformations in the past, particularly when analysing repeated measurements. One such area is the analysis of pulmonary function. An abundance of research exists studying patient characteristics and risk factors that influence various measures of lung function. Longitudinal models to analyse these pulmonary function variables have become increasingly popular. Some of these measures include forced expiratory volume in 1 s (FEV_1), forced vital capacity and maximum mid-expiratory flow (MMEF).

0964–1998/06/169273

Even though it is not commonplace, natural logarithmic transformations have been applied to some of these pulmonary function studies, both for cross-sectional as well as longitudinal designs (e.g. Manzke *et al.* (2001) and Berhane *et al.* (2000)). However, the examples cited do not document the reasons for choosing the log-transformation. In the case when it is applied to longitudinal mixed models, we have yet to find research that describes how that transformation strengthened the multiple assumptions of the mixed model. A methodological examination of the use of transformations in the context of longitudinal pulmonary function data would greatly add to this particular area of research.

On the basis of its intermittent use in previous pulmonary studies, we wish to assess the utility of a transformation when modelling longitudinal pulmonary function in cystic fibrosis (CF) patients. CF is a chronic obstructive pulmonary disease which is both studied and treated within a longitudinal framework. Linear mixed models have become increasingly popular in this area of research in determining characteristics that affect pulmonary function (Edwards, 2000). Specifically, age and gender are often assessed as predictors of various measures of lung function, such as FEV_1 and MMEF. From Edwards (2000), longitudinal pulmonary function data were available from clinical follow-up of 47 adult CF patients (23 female; 24 male) seen at the University of North Carolina pulmonary clinic. We wished to explore these data further by analysing the percentage predicted MMEF (%PredMMEF) in a similar fashion to the models of percentage predicted FEV_1 in Edwards (2000). MMEF, also known as $FEF_{25\%-75\%}$, is the mean forced expiratory flow during the middle half of the forced vital capacity. MMEF is a measure of function of the small airways of the lungs (Berhane *et al.*, 2000). We specifically wanted to assess the effect of gender and age on %PredMMEF in adult CF patients. To demonstrate the methods that are discussed later in a straightforward fashion, the following model is of interest:

$$\%\text{PredMMEF}_{ij} = \beta_0 + \beta_1 \, \text{GENDER}_i + \beta_2 \, \text{AGE}_{ij} + b_{0i} + b_{1i} \, \text{AGE}_{ij} + e_{ij}, \tag{1}$$

where $\%\text{PredMMEF}_{ij}$ is an $n_i \times 1$ vector of n_i measures of percentage predicted MMEF over time on the ith CF subject. There were a total of 1463 observations from the 47 subjects. The model expresses $\%\text{PredMMEF}_{ij}$ as a function of gender ($\text{GENDER}_i = 0$ if person i is female; $\text{GENDER}_i = 1$ if male) and age (AGE_{ij}, centred at 20 years). Preliminary analyses of the data indicated no significant gender $\times$ age interaction in the set of fixed effects. As is standard in the linear mixed model, we assume that the random effects and within-subject error term independently follow normal distributions: $b_{0i} \sim N(0, \sigma_{b0}^2)$, $b_{1i} \sim N(0, \sigma_{b1}^2)$ and $e_{ij} \sim N(0, \sigma_e^2)$. Also, preliminary covariance model selection procedures indicated that it is suitable to assume that the random intercept and random slope are independent of one another.

On the basis of the literature, there is no clear consensus on the need for a transformation of %PredMMEF, or any other lung function measure. It would be helpful to determine whether or not a transformation for model (1) is needed, and how that potential transformation helps us to attain validity of the assumptions of the model. Rather than choosing a particular transformation *a priori*, we wish to estimate an appropriate transformation. Likewise, we aim to examine the effect of estimating a transformation on inference about the gender and age fixed effects.

Throughout, we evaluate methods that are intended to achieve Gaussian distributions of the random effects and the pure error of the linear mixed model by transforming the response. We begin with careful definitions of Box–Cox transformation methods in the linear mixed model. Next we discuss potential pitfalls that result from estimating the transformation and propose solutions when interest lies primarily in the fixed effects. In turn, we derive results that not only characterize the properties of the method but also provide guidance for practical use. Free software that is available from the author implements the methods proposed.

1.2.   *Literature review*

Seminal work by Harville (1977) and Laird and Ware (1982) helped to popularize the use of linear mixed models. Many theoretical and applied papers followed (e.g. Kenward and Roger (1997)), and several texts now include discussions of linear and non-linear mixed models (e.g. Vonesh and Chinchilli (1997)). Verbeke and Lesaffre (1997) showed that the maximum likelihood estimators of the fixed effect parameters are consistent and asymptotically normally distributed regardless of the distribution of the random effects; however, they provided a sandwich-type correction for the estimates of the standard errors when the random effects are not normal. Additionally, Verbeke and Lesaffre (1996) demonstrated that the random effects may be poorly estimated when their distribution is a mixture of Gaussian distributions. Lange and Ryan (1989) presented a method to assess graphically the assumption of normality of the random effects in mixed models. Techniques to assess and improve normality of both the random effects and pure error simultaneously have received little or no attention in the literature.

Transformation of the response has become a very simple and popular remedy in model fitting when the validity of the assumptions of the model are called into question. The sandwich-type correction that was proposed by Verbeke and Lesaffre (1997) allows for improved inference on the fixed effects when the random effects are non-normal. However, transformation of the response in a mixed model setting may be an alternative that not only serves the purpose of improving inference about the fixed effects but also may lead to valid estimation and inference of the random effects. Additionally, in fields where transformations are utilized even sparingly, such as with pulmonary function research, a thorough examination of the transformation approach for the linear mixed model would prove especially useful.

The use of parametric transformations in univariate linear models has been studied extensively (e.g. Box and Cox (1964)). In the univariate setting, making inferences about the model parameters in the presence of a transformation has generated much debate (Hinkley and Runger, 1984). Some argue that treating the transformation parameter as non-stochastic when making inferences on the other parameters is valid. However, many contend that we must take into account the estimation of the unknown transformation parameter in this case (Bickel and Doksum, 1981). Though debated from a methodological standpoint, the application of transformations in univariate models is commonplace to utilize a parametric model that best meets the assumptions that are made on its error term.

The discussion of transformations and their use in practice naturally leads to the issue of interpretability. Most of the time researchers are not interested in making decisions based on a model on a different scale. For instance, the analysis of cost data often leads to the use of the log-transformation but, as Manning (1998) pointed out, 'First Bank will not cash a check for log dollars'. The idea of 'retransformation' to the original scale of the data is not straightforward and should be used with caution. Hence, retransformation methods have been proposed for univariate linear models (e.g. Duan (1983)) and may need to be studied further. However, retransformation methods for the linear mixed model is beyond the scope of this paper.

Additionally, as the popularity of generalized linear models has increased, their use in this setting has also been advocated and discussed in the literature (e.g. Manning and Mullahy (2001)). A comparison of the generalized linear model approach to the Box–Cox method of transformation led Manning and Mullahy (2001) to detail the pros and cons of each of the approaches in different settings. Manning and Mullahy (2001) concluded that no single technique is best under all the conditions that they examined. Whether their findings translate to the linear mixed model setting is unknown. Generalized linear mixed models would be one approach in resolving normality questions of the data to be analysed. However, the proper link function is often unknown. Estimating a transformation in such a situation is very beneficial and at the very least can serve

as a basis for future application of the generalized linear model approach. Research that has not completely ruled out the use of the Box–Cox transformation, such as the literature cited, serves as motivation for continued exploration of the transformation approach in the linear mixed model.

Andrews *et al.* (1971) extended the transformation theory to multivariate models by introducing a vector of parameters, one for each of the response variables. Sakia (1990), Gianola *et al.* (1990) and Solomon and Taylor (1999) studied approaches to transformations for simple variance component models. Oberg and Davidian (2000) extended the methods for estimating transformations to non-linear mixed effects models for repeated measurement data, employing the transform-both-sides model that was proposed by Carroll and Ruppert (1984). Lipsitz *et al.* (2000) analysed longitudinal CD4 cell count data by applying a Box–Cox transformation on the response of a marginal mean linear model (a repeated measures model with only fixed effects that is often referred to as a population-averaged model). Since the model did not explicitly contain random effects, the authors assumed that the transformation achieved normality of the overall error term only. We know of no research for the linear mixed model which comprehensively examines the process and implications of transformations that are chosen to help to meet the multiple distributional assumptions. Additionally, an examination of the effect of the addition of the transformation parameter on the estimation and inference of the mixed model parameters is warranted.

2.   A likelihood formulation of the Box–Cox transformation for the linear mixed model

2.1.   *Notation*

The linear mixed model with a parametrically transformed response for $i \in \{1, \ldots, m\}$, m the number of independent sampling units (subjects), is given by

$$\mathbf{y}_i^{(\lambda)} = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i. \tag{2}$$

Here, $\mathbf{y}_i$ is an $n_i \times 1$ vector of observations on subject i, $\mathbf{X}_i$ is an $n_i \times p$ known, constant design matrix for the ith subject with rank p and $\boldsymbol{\beta}$ is a $p \times 1$ vector of unknown, constant population parameters. Also, $\mathbf{Z}_i$ is an $n_i \times q$ known, constant design matrix for the ith subject with rank q corresponding to $\mathbf{b}_i$, a $q \times 1$ vector of unknown, random individual-specific parameters, and $\mathbf{e}_i$ is an $n_i \times 1$ vector of random within-subject, or pure, error terms. Additionally, let $\varepsilon_i = \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i$ be the 'total' error term of model (2).

When utilizing linear mixed models in general, we assume that the following distributional assumptions are reasonably valid: $\mathbf{b}_i$ and $\mathbf{e}_i$ independently follow normal distributions with mean vectors $\mathbf{0}$ and covariance matrices $\mathbf{D}$ and $\mathbf{R}_i$ respectively. We wish to examine whether and how a simple extension of the Box–Cox transformation targets the validity of both assumptions of normality in the mixed model. The covariance matrices $\mathbf{D}$ and $\mathbf{R}_i$ are characterized by unique parameters that are contained in the $k \times 1$ vector $\boldsymbol{\theta}$. Thus, the total variance for the response vector in model (2) is

$$\mathcal{V}(\varepsilon_i) = \Sigma_i = \mathbf{Z}_i \mathbf{D} \mathbf{Z}_i' + \mathbf{R}_i,$$

where $\mathcal{V}(\cdot)$ is the variance–covariance operator. Since the model pertains to a single outcome measured repeatedly, $\mathbf{y}_i$ is transformed by using the form that was proposed by Box and Cox (1964) with a single transformation parameter λ, for all subjects $i \in \{1, \ldots, m\}$ and measurement occasions $j \in \{1, \ldots, n_i\}$, $y_{ij} > 0$:

$$y_{ij}^{(\lambda)} = \begin{cases} (y_{ij}^{\lambda} - 1)/\lambda & \lambda \neq 0, \\ \log(y_{ij}) & \lambda = 0. \end{cases} \tag{3}$$

The standard linear mixed model with response y_{ij} is obtained when $\lambda = 1$. Using the transformation (3) requires $y_{ij} > 0$ for all i and j. Box and Cox (1964) also studied a shifted power transformation to avoid the restriction. A second limitation arises from the fact that the range of the observed values of $y_{ij}^{(\lambda)}$ depends on the sign of λ. Thus, the method can achieve only approximate normality, which seems an acceptable goal.

To obtain the most accurate parameter estimators overall of model (2), a residual maximum likelihood (REML) approach is employed. For fixed λ, the residual log-likelihood for all m individuals in terms of the original observations, $\mathbf{y}$, can be written as (Verbeke and Molenberghs, 2000)

$$L_{\mathrm{R}}(\mathbf{y}, \lambda | \boldsymbol{\theta}) = -\frac{N-p}{2}\log(2\pi) + \frac{1}{2}\log\left|\sum_{i=1}^{m}\mathbf{X}_i'\mathbf{X}_i\right| - \frac{1}{2}\sum_{i=1}^{m}\log|\boldsymbol{\Sigma}_i| - \frac{1}{2}\log\left|\sum_{i=1}^{m}\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right|$$
$$- \frac{1}{2}\sum_{i=1}^{m}(\mathbf{y}_i^{(\lambda)} - \mathbf{X}_i\hat{\boldsymbol{\beta}})'\boldsymbol{\Sigma}_i^{-1}(\mathbf{y}_i^{(\lambda)} - \mathbf{X}_i\hat{\boldsymbol{\beta}}) + N(\lambda - 1)\log(\tilde{y}), \tag{4}$$

in which

$$\tilde{y} = \left(\prod_{i=1}^{m}\prod_{j=1}^{n_i} y_{ij}\right)^{1/N}$$

is the geometric mean of the $N = \Sigma_{i=1}^{m} n_i$ observations, and

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}(\lambda, \boldsymbol{\theta}) = \left(\sum_{i=1}^{m}\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{m}\mathbf{X}_i'\boldsymbol{\Sigma}_i^{-1}\mathbf{y}_i^{(\lambda)}.$$

The exploration of the following theory utilizes the REML approach to estimation. As noted in Verbeke and Molenberghs (2000), pages 44–46, the residual likelihood differs from its maximum likelihood counterpart by additional terms that are not dependent on $\boldsymbol{\beta}$. In the case of the transformed linear mixed model (2), these additional terms in the REML function (4) also do not depend on λ. Thus, the methods that are presented here for REML estimation apply to maximum likelihood estimation as well.

In the interest of developing methods that are accessible to the widest possible audience, we wish to take advantage of existing computational procedures to estimate λ. To do this, we must replace expression (3) with a scaled transformation (Box and Cox, 1964):

$$w_{ij}^{(\lambda)} = \begin{cases} (y_{ij}^{\lambda} - 1)/\lambda\tilde{y}^{\lambda-1} & \lambda \neq 0, \\ \tilde{y} \cdot \log(y_{ij}) & \lambda = 0. \end{cases} \tag{5}$$

The scaled transformation leads to the model

$$\mathbf{w}_i^{(\lambda)} = \mathbf{X}_i\boldsymbol{\beta}_* + \mathbf{Z}_i\mathbf{b}_{*i} + \mathbf{e}_{*i}, \tag{6}$$

with the same assumptions as model (2). Conditioned on the geometric mean, the Jacobian of expression (5) is equal to 1. Thus, the residual log-likelihood in terms of the original observations $\mathbf{y}$, ignoring the constant terms, is that of a standard linear mixed model:

$$L_{\mathrm{R}}(\mathbf{w}, \lambda | \boldsymbol{\theta}_*) = -\frac{1}{2}\sum_{i=1}^{m}\log|\boldsymbol{\Sigma}_{*i}| - \frac{1}{2}\log\left|\sum_{i=1}^{m}\mathbf{X}_i'\boldsymbol{\Sigma}_{*i}^{-1}\mathbf{X}_i\right| - \frac{1}{2}\sum_{i=1}^{m}(\mathbf{w}_i^{(\lambda)} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_*)'\boldsymbol{\Sigma}_{*i}^{-1}(\mathbf{w}_i^{(\lambda)} - \mathbf{X}_i\hat{\boldsymbol{\beta}}_*),$$
$$\tag{7}$$

where $\mathcal{V}(\mathbf{b}_{*i}) = \mathbf{D}_*$, $\mathcal{V}(\mathbf{e}_{*i}) = \mathbf{R}_{*i}$, $\boldsymbol{\theta}_*$ constitutes the set of variance parameters and

$$\mathcal{V}(\mathbf{w}_i^{(\lambda)}) = \boldsymbol{\Sigma}_{*i} = \mathbf{Z}_i\mathbf{D}_*\mathbf{Z}_i' + \mathbf{R}_{*i}.$$

Thus, by using the scaled transformation (5), we can employ existing linear mixed model estimation procedures to obtain the REML estimates of the model parameters for fixed λ. With respect to λ, maximizing the likelihood (7) by using the scaled transformation, $\mathbf{w}_i^{(\lambda)}$, is equivalent to maximizing the likelihood (4) by using $\mathbf{y}_i^{(\lambda)}$ (Spitzer, 1982). Since we assume that model (2) is the model of interest, we need to convert the model parameter estimates that maximize log-likelihood (7) back to those based on log-likelihood (4). Since $w_{ij}^{(\lambda)} = y_{ij}^{(\lambda)}/\tilde{y}^{\lambda-1}$, we can state that

$$
\left.
\begin{aligned}
\boldsymbol{\beta} &= \tilde{y}^{\lambda-1} \cdot \boldsymbol{\beta}_*, \\
\mathbf{b}_i &= \tilde{y}^{\lambda-1} \cdot \mathbf{b}_{*i}, \\
\mathbf{e}_i &= \tilde{y}^{\lambda-1} \cdot \mathbf{e}_{*i},
\end{aligned}
\right\}
\tag{8}
$$

where $\tilde{y}$ is considered fixed. As discussed previously, interpretability is always an issue when a transformation of the response is utilized. Interpretability becomes even more convoluted if we were to use the model of the scaled transformation (5) due to the inclusion of the geometric mean of the response data. Hence, estimation and inference of the parameters of model (2) need to be discussed further, taking advantage of the fact that $\hat{\lambda}$ obtained from the scaled model (6) holds for the preferred transformation model (2).

For the linear mixed model, a grid search provides one simple approach for finding $\hat{\lambda}$, i.e., for a large set of values for λ, we can fit the model with the scaled transformed response (6) by using standard linear mixed model software. Calculating and plotting the residual likelihood values for the fitted models of the scaled response against the set of values for λ will locate $\hat{\lambda}$. Lipsitz *et al.* (2000) implemented a similar grid search utilizing the profile likelihood of the original transformed response (3). However, as Lipsitz *et al.* (2000) noted, application of their method to an example led to a non-smooth graph. Wade and Cole (2001) raised concerns about the nature of the problem. In contrast, practical experience with our grid search has yet to produce such a non-smooth figure.

2.2. Gaussian distribution of both error terms in the linear mixed model

As in the univariate linear model, we transform the response of the linear mixed model in the hopes of meeting (at least approximately) the assumptions that are inherent in the model. We are primarily interested in examining whether or not transforming the response does lead to a linear mixed model in which both the random effects and the pure error have approximate Gaussian distributions. By maximizing the residual likelihood of the scaled response (7) with respect to λ, we have described a technique to transform the response variable in the linear mixed model that finds the optimum value of λ for the marginal model

$$
\mathbf{y}_i^{(\lambda)} = \mathbf{X}_i \boldsymbol{\beta} + \boldsymbol{\varepsilon}_i.
$$

In other words, estimating λ directly targets the validity of the assumption that $\boldsymbol{\varepsilon}_i = \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_i)$. We can assess the validity of the assumption of multivariate normality of $\mathbf{b}_i$ and $\mathbf{e}_i$, rather than solely $\boldsymbol{\varepsilon}_i$, because of the following lemma and corollary.

Lemma 1. If $\mathbf{x}_1$ and $\mathbf{x}_2$ are independent vectors and $\mathbf{x}_1 + \mathbf{x}_2$ follows a multivariate Gaussian distribution, then both $\mathbf{x}_1$ and $\mathbf{x}_2$ each individually follow a multivariate Gaussian distribution (Cramer, 1970).

Corollary 1. The assumptions underlying the mixed model dictate that $\mathbf{b}_i$ and $\mathbf{e}_i$ are statistically independent. Hence, if $\boldsymbol{\varepsilon}_i$ is assumed to be multivariate normal, then $\mathbf{Z}_i \mathbf{b}_i$ and $\mathbf{e}_i$ each follow multivariate normal distributions. It directly follows that $\mathbf{b}_i$ is multivariate normal.

Thus, parametrically transforming the response to achieve normality of the total error term ε_i will target both the normality of the random effects $\mathbf{b}_i$ and the within-unit error term $\mathbf{e}_i$. We can make this conclusion because, in the linear mixed model, $\mathbf{b}_i$ and $\mathbf{e}_i$ are assumed to be independent. We recognize that the assumptions of independence and normality for the transformed model are so strongly related that it is almost impossible to distinguish non-normality from non-independence by using the lemma. With no transformation of the response, the standard assumption for linear mixed models is that $\mathbf{b}_i$ and $\mathbf{e}_i$ are independent of each other as well as normally distributed. As one reviewer noted, since only the outcomes are observed, there may be little evidence in the data to justify decomposing the sources of variability into independent random components $\mathbf{b}_i$ and $\mathbf{e}_i$. However, the study design and sampling scheme can help to dictate the legitimacy of the assumption of independence. In adding a transformation parameter to the model, we now assume that for an optimum value of the transformation the assumptions of normality and independence remain. We are simply extending this assumption without concerning ourselves with the assessment of its validity, stating that this independence is present for the optimal transformation, if such a transformation exists.

3. Estimation and inference: complications and solutions

The extension of the Box–Cox transformation to the linear mixed model would not be complete without discussion of how the estimation of λ affects inference about the other model parameters. As stated previously, this topic has been extensively debated for the univariate linear model. Simulations will provide insight into the effect of the estimation of λ on the mixed model parameters and hence may add to this debate.

We wish to estimate and make inference about the parameters $(\boldsymbol{\beta}, \boldsymbol{\theta})$ of the original model (2). The estimate of λ is invariant to the use of the scaled transformed model (6) or the preferred model (2) involving the original transformation $\mathbf{y}_i^{(\lambda)}$. However, problems arise when estimating the parameters of model (2) and their variances. Using equations (8), it would be natural to use $\hat{\boldsymbol{\beta}} = \tilde{y}^{\hat{\lambda}-1} \cdot \hat{\boldsymbol{\beta}}_*$ in deriving an estimate of $\boldsymbol{\beta}$. This is what occurs when we fit model (2) by using $\hat{\lambda}$ obtained from model (6). Unfortunately, $\hat{\boldsymbol{\beta}}$ is a slightly biased estimator of $\boldsymbol{\beta}$. Details found in Appendix A quantify this bias and provide a bias-corrected estimator of $\boldsymbol{\beta}$, $\hat{\boldsymbol{\beta}}_{\text{adj.}}$

We take the position that one should attempt to adjust for the estimation of λ when estimating the variances of the other model parameters. Simulations will show that this notion is justified. Lipsitz *et al.* (2000) also made this assessment and proposed a jackknife method to estimate the variances of $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}, \hat{\lambda})$ given their dependence. This approach is sound for the most general case. Nevertheless, we present a likelihood-based method for estimating $\mathcal{V}(\hat{\boldsymbol{\beta}})$ when inferential interest lies primarily on the fixed effects. This method is a computationally less intensive alternative that is valid for large sample sizes and is potentially easier to implement. Starting from the scaled model (6) and assuming that $\boldsymbol{\theta}_*$ is known, we initially compute the observed information matrix for $\{\boldsymbol{\beta}_*, \lambda\}$ in the likelihood (7); see Appendix A for details. The inverse of the observed information matrix evaluated at $\{\hat{\boldsymbol{\beta}}_*, \hat{\lambda}\}$ will provide an estimate of the large sample covariance matrix of $\{\hat{\boldsymbol{\beta}}_*, \hat{\lambda}\}$. Standard practice in linear mixed models with $\boldsymbol{\theta}_*$ unknown is to use a large sample approximation: replace $\boldsymbol{\Sigma}_{*i}$ with its REML estimate $\hat{\boldsymbol{\Sigma}}_{*i}$, and compute $\hat{\boldsymbol{\beta}}_*$ and $\mathcal{V}(\hat{\boldsymbol{\beta}}_*)$ (Harville, 1977; Laird and Ware, 1982).

It is worth mentioning again that the observed information matrix and the resulting variance estimates are based on the model using the scaled transformed response. Following the lead of Spitzer (1982), again treating $\boldsymbol{\theta}$ as known, we can approximate $\mathcal{V}(\hat{\boldsymbol{\beta}})$ by use of the delta method. The details of the problems as well as the approximation and formulae for the variance estimates can be found in Appendix A. This proposed variance estimate of $\hat{\boldsymbol{\beta}}$ 'adjusts' for the estimation of

λ, and hence any resulting correlation between $\hat{\beta}$ and $\hat{\lambda}$. Unfortunately, this corrected variance estimate relates to the biased estimator $\hat{\beta}$ and not to its corrected counterpart, $\hat{\beta}_{\text{adj}}$. A similarly corrected variance estimate for $\hat{\beta}_{\text{adj}}$ is beyond the scope of this paper. Hence, formulae that are found in Appendix A provide for separate corrections for the bias and the increased variance stemming from the estimation of λ.

When focusing on the fixed effects, we have treated θ as fixed and known, substituting its estimate $\hat{\theta}$ when θ is unknown. The calculations that are summarized here and detailed in Appendix A are simply needed to obtain an improved estimate of $\mathcal{V}(\hat{\beta})$ that accounts for the estimation of λ. Ideally, we would prefer to estimate the observed information matrix for $(\hat{\beta}, \hat{\theta}, \hat{\lambda})$. However, this would require extremely difficult derivations of second derivatives of the residual log-likelihood, as pointed out by Lipsitz *et al.* (2000). We could solve these derivatives numerically by using sophisticated mathematical tools, but our alternative allows for straightforward generalization and application. When interest lies in estimation and inference on θ, it would be natural to recommend the use of the jackknife approach that is described in Lipsitz *et al.* (2000), or a bootstrap method. But, as alluded to in Appendix A as well as displayed through simulations, the estimators of the variance components themselves are biased owing to the estimation of λ. Hence we must caution the analyst in using the transformation approach when interest lies in the variance components of the mixed model. Lipsitz *et al.* (2000) did not demonstrate the utility of their approach via simulations, so this research helps to serve as evidence of the effect of the transformation method on the estimation and inference of the mixed model parameters.

Even though estimating β does not depend on the normality of $\mathbf{b}_i$, inference about β does rely on the Gaussian distribution of $\mathbf{b}_i$. The employment of a transformation is one way to resolve the question of the validity of the Gaussian distribution in the mixed model setting and is used often in practice. Hence, in using our transformation approach as a tool to achieve normality of $\mathbf{b}_i$, we have proposed a variance estimate of $\hat{\beta}$ that takes into account the estimation of λ when interest also lies in inference about β. Simulations that are presented indicate that such a correction is necessary. The quality of the estimates of the random effects is dependent on the validity of the assumption of their normality (Verbeke and Molenberghs, 2000). In targeting the normality of the random effects of the model, the Box–Cox transformation would hopefully provide improved estimation of the random effects. Future research should target improved inference on the random effects by taking into account the estimation of λ.

Finally, for interpretability a 'convenient' value of a transformation is often chosen after estimating λ. For instance, we may wish to use the log-transformation instead of a value of $\hat{\lambda} = 0.17$. It is reasonable to assume that the use of transformations in this manner will affect inference on the model parameters as well. However, the likelihood approach that is presented is no longer valid when $\hat{\lambda}$ is not used. We recommend in this instance the use of a jackknife or bootstrap approach. As Lipsitz *et al.* (2000) so astutely noted, examination of the optimal transformation (even if we do not use it) may be an appropriate sensitivity analysis (in comparing inferences across the fitted models, including untransformed, transformed and transformed by a convenient value). Ultimately, this may be the best use of such transformation techniques for those cases when inferences on the optimal transformed scale are not preferred.

4. Simulation results

We performed simulation studies to demonstrate the effect of estimation of λ on the other model parameters in a large sample setting due to the asymptotic basis of the derivations. For each $\lambda \in \{0, 0.5, 1\}$, we generated 10 000 Monte Carlo data sets, each with 100 subjects and five

observations per subject at five time points equally spaced between (0, 1). We employed a missing data mechanism to reflect data missing completely at random, randomly removing on average 20% of the observations. For the fixed effects, we used $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2)'$, containing a gender parameter β_1 and a common time parameter β_2. For simplicity, only a random intercept was included in the model, $b_{0i} \sim \mathcal{N}(0, 0.5)$, and $e_i \sim \mathcal{N}(0, 0.5)$. We set the intercept parameter $\beta_0 = 5$ to assure $y_{ij}^{(\lambda)} > 0$. The following model was then generated:

$$y_{ij}^{(\lambda)} = 5 + 2.0x_i + 1.0t_{ij} + b_{0i} + e_{ij}, \tag{9}$$

with $x_i = 0$ if subject i is in the first group and $x_i = 1$ if in the second group. We fitted the scaled transformation model to use existing estimation procedures:

$$w_{ij}^{(\lambda)} = \beta_{*0} + \beta_{*1}x_i + \beta_{*2}t_{ij} + b_{*0i} + e_{*ij}. \tag{10}$$

Parameter estimates and estimated standard errors for the model of interest (9) are displayed in Table 1. We wished to replicate the 'naïve' approach in finding an estimate of λ and then fitting

Table 1. Simulation results for 10 000 Monte Carlo replications for the transformed mixed model (9): means and standard errors of parameter estimates†

Estimate‡	$\lambda = 0$				$\lambda = 0.5$				$\lambda = 1$			
	CS		IID		CS		IID		CS		IID	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE	Mean	SE
$\hat{\lambda}$	−0.00	0.02	−0.00	0.03	0.50	0.10	0.49	0.13	0.99	0.18	0.99	0.22
$\widehat{SE}_1(\hat{\lambda})$	0.00	—	0.00	—	0.01	—	0.01	—	0.02	—	0.01	—
$\widehat{SE}_2(\hat{\lambda})$	0.02	—	0.03	—	0.10	—	0.11	—	0.18	—	0.19	—
$\hat{\beta}_0$ ($\beta_0 = 5$)	5.01	0.33	5.02	0.42	5.04	0.80	5.07	1.01	5.07	1.07	5.13	1.37
$\hat{\beta}_{0,\text{adj}}$	5.03	0.33	5.04	0.42	5.03	0.79	5.06	1.00	5.02	1.05	5.08	1.33
$\widehat{SE}_1(\hat{\beta}_0)$	0.12	—	0.10	—	0.13	—	0.10	—	0.13	—	0.11	—
$\widehat{SE}_2(\hat{\beta}_0)$	0.33	—	0.34	—	0.79	—	0.83	—	1.05	—	1.12	—
$\hat{\beta}_1$ ($\beta_1 = 2$)	2.02	0.36	2.04	0.45	2.06	0.63	2.10	0.80	2.09	0.78	2.16	1.01
$\hat{\beta}_{1,\text{adj}}$	2.00	0.35	2.01	0.44	1.97	0.61	2.00	0.77	1.96	0.73	2.01	0.94
$\widehat{SE}_1(\hat{\beta}_1)$	0.16	—	0.10	—	0.16	—	0.10	—	0.17	—	0.11	—
$\widehat{SE}_2(\hat{\beta}_1)$	0.36	—	0.36	—	0.62	—	0.65	—	0.76	—	0.81	—
$\hat{\beta}_2$ ($\beta_2 = 1$)	1.01	0.19	1.02	0.24	1.03	0.32	1.05	0.41	1.04	0.40	1.08	0.51
$\hat{\beta}_{2,\text{adj}}$	1.00	0.19	1.01	0.24	0.98	0.31	1.00	0.39	0.98	0.37	1.00	0.48
$\widehat{SE}_1(\hat{\beta}_2)$	0.10	—	0.14	—	0.11	—	0.15	—	0.11	—	0.15	—
$\widehat{SE}_2(\hat{\beta}_2)$	0.19	—	0.22	—	0.32	—	0.35	—	0.38	—	0.43	—
$\hat{\sigma}_{b0}^2$ ($\sigma_{b0}^2 = 0.5$)	0.52	0.19	—	—	0.57	0.37	—	—	0.62	0.50	—	—
$\hat{\sigma}_e^2$ ($\sigma_e^2 = 0.5$)	0.52	0.18	1.07	0.46	0.57	0.36	1.25	1.03	0.62	0.49	1.39	1.48

†Column values of λ are considered the true values from which the data sets were generated. Results are displayed for when both the covariance structure is correctly assumed (compound symmetry, CS) and incorrectly specified (independent and identically distributed observations, IID).
‡$\widehat{SE}_1(\cdot)$, estimated unadjusted standard error of the parameter estimates; $\widehat{SE}_2(\cdot)$, estimated adjusted standard error of the parameter estimates.
§SE, Monte Carlo standard error of the model parameter estimates.

model (9), transforming the response by $\hat{\lambda}$. For $k \in \{0, 1, 2\}$, we report the average unadjusted standard error estimate of $\hat{\beta}_k$, $\widehat{SE}_1(\hat{\beta}_k)$, the standard error estimate resulting from the fit of model (9) using $\hat{\lambda}$ found from model (10). This naïve approach treats λ as fixed and known when making inferences about β. In other words, the unadjusted standard error estimates are based on the assumption that there is no correlation between the estimates of λ and β. We also report the average value of $\widehat{SE}_2(\hat{\beta}_k)$, the adjusted standard error estimate of $\hat{\beta}_k$ that is obtained by using the methods of Section 3. In a similar fashion, $\widehat{SE}_2(\hat{\lambda})$ is the 'adjusted' standard error approximation calculated while taking into account the estimation of β. Additionally, the average value of the bias-corrected estimator of β, $\hat{\beta}_{adj}$, is displayed, but with no corresponding standard error estimates.

We first assume the correct covariance structure from which the data were simulated, namely from a model with a random intercept and an independent and identically distributed pure error term (i.e. compound symmetry). It also would prove interesting to determine the effect of misspecification of the covariance structure on the estimation and inference about (β, λ). Thus, the estimates that were described above were calculated when assuming that all observations were independent; i.e. parameter estimators were examined from a univariate linear model of correlated data.

The results assuming the correct covariance structure are first described. Estimation of λ by using the methods described is proven reliable. Estimation of β is shown to be more problematic. As described and quantified in Appendix A, an increasing bias of $\hat{\beta}$ is observed for increasing values of λ. Such a phenomenon has been pointed out in previous simulation studies involving the transformed univariate linear model (Spitzer, 1978). To our knowledge the discrepancy has never been thoroughly explained. The proposed correction, $\hat{\beta}_{adj}$, does not exhibit such a trend.

The simulation results demonstrate that ignoring the estimation of λ leads to poor estimation of $\mathcal{V}(\beta)$. Adjusting for the estimation of λ by using the methods of Section 3 leads to an improved estimate of $\mathcal{V}(\hat{\beta})$. Likewise, a distinct advantage exists in taking into account the estimation of β when approximating $\mathcal{V}(\hat{\lambda})$. The corrected standard error estimates are still slightly different from the Monte Carlo standard errors. Part of the disparity is most probably explained by the fact that we are replacing θ with its estimate and treating $\hat{\theta}$ as known in deriving the variance estimates of $\hat{\beta}$ and $\hat{\lambda}$. Even though we recommend further investigation of $\mathcal{V}(\hat{\beta}_{adj})$, the average adjusted standard error of its uncorrected counterpart, $\hat{\beta}$, is seen in this situation to be relatively accurate when compared with the Monte Carlo standard error of $\hat{\beta}_{adj}$.

The estimates of σ_{b0}^2 and σ_e^2 are biased and can be seen as evidence of correlation between estimates of λ and θ that the implemented methods are not taking into account. The increase in bias of the estimators of the fixed effects and variance component parameters as the true value of λ increases is probably due to the greater variability that is seen in $\hat{\lambda}$ for larger true values of λ and should not come as a surprise after reading the details of Appendix A. Use of the 'corrected' estimates of β show promise as a remedy for the observed bias of the fixed effects. Similar corrections for the variance component estimators would be extremely beneficial. However, such corrections are difficult to formulate for the components of a general covariance model and should be studied further.

Finally, we observe the same results when the incorrect covariance structure was assumed (independent and identically distributed errors and no random effects). As expected, a slight bias of the estimators of (β, λ) is seen, as well as a more substantial increase in the variability of these estimates. This loss of precision is not accurately estimated with the estimators proposed. However, the corrected variance estimates are still superior to their uncorrected counterparts. These results can serve as evidence that we must proceed very carefully when employing a transformed linear mixed model if any doubt exists over the correct covariance structure of the data.

Further research is warranted that addresses inference for the transformed linear mixed model when the covariance is misspecified (e.g. a quasi-likelihood sandwich-type correction).

5. Example results

We explore the lung function data by using our proposed approach, not only to validate the methods further, but also to examine the necessity of a transformation for the model of interest (1). The transformed linear mixed model for the data is a generalization of model (1):

$$\%\text{PredMMEF}_{ij}^{(\lambda)} = \beta_0 + \beta_1 \, \text{GENDER}_i + \beta_2 \, \text{AGE}_{ij} + b_{0i} + b_{1i} \, \text{AGE}_{ij} + e_{ij}. \qquad (11)$$

Methods that were described in Sections 2 and 3 are used to estimate λ (using the scaled transformed model) and to obtain parameter estimates of model (11), as well as accurate variance estimates.

Fig. 1 provides a graphical representation of the grid search method that was described previously which found $\hat{\lambda} = 0.008$. Table 2 displays the model parameter estimates for $\lambda = 1$ (no transformation) and $\hat{\lambda}$. Using the log-transformation seems reasonable in this context since the value of $\hat{\lambda}$ is close to 0. This example proves interesting when comparing inference across the models. When ignoring the estimation of λ, significance of the fixed effects does not change relative to the untransformed model. However, when using the methods that were described in Section 3 and detailed in Appendix A, the corrected standard error for gender is drastically different from its uncorrected counterpart, leading to a significant gender effect. This change in significance is further evidence that estimation of λ does indeed impact inference about the fixed effects. We would conclude on the basis of the transformed model that males have a significantly lower percentage predicted MMEF, and that this index of small airway function decreases over time for both genders.
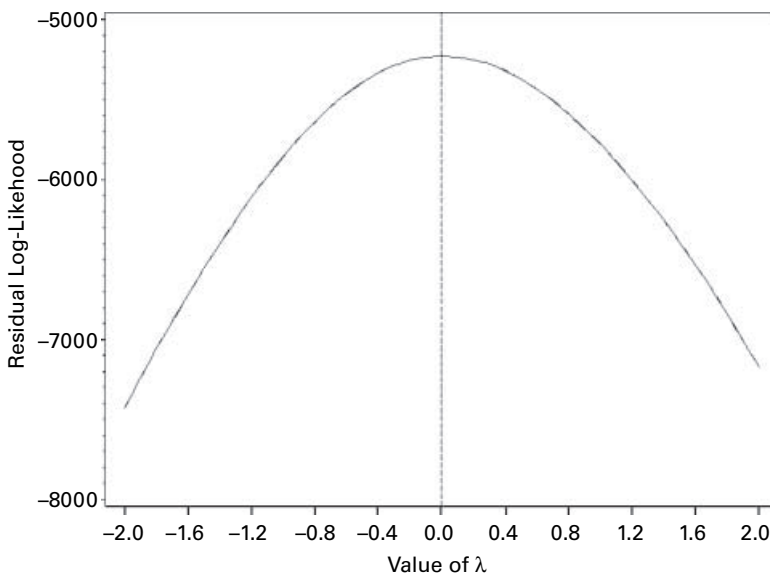
Fig. 1.   Graphical representation of the grid search estimation method for the transformed linear mixed model applied to the MMEF data: residual log-likelihoods for values of the Box–Cox transformation parameter λ ($\hat{\lambda} = 0.008$)

Table 2. MMEF model parameter estimates and standard errors for the untransformed model ($\lambda = 1$) and the transformed model ($\hat{\lambda} = 0.008$)†

Parameter	Results for the untransformed model			Results for the transformed model					
	Estimate	*SE*	*p*	*Estimate*	$\widehat{SE}_1$	*p*	$\widehat{SE}_2$	*p*	*Bias-corrected estimate*
λ	—	—	—	0.008	0.017	—	0.026	—	—
Intercept	39.844	4.521	<0.001	3.558	0.156	<0.001	0.206	<0.001	3.556
GENDER	−6.473	6.295	0.310	−0.197	0.217	0.370	0.061	<0.001	−0.196
AGE	−2.020	0.294	<0.001	−0.072	0.008	<0.001	0.010	<0.001	−0.072
σ^2_{b0}	384.290	93.356	—	0.483	0.114	—	‡	—	‡
σ^2_{b1}	2.247	0.750	—	0.002	0.001	—	‡	—	‡
σ^2_e	133.310	5.102	—	0.105	0.004	—	‡	—	‡

†SE, observed standard error of the parameter estimate when λ is treated as fixed; $\widehat{SE}_1$, calculated unadjusted standard error of the parameter estimate; $\widehat{SE}_2$, calculated adjusted standard error of the parameter estimate; *p*, *p*-value computed from the *t*-statistic and using the Kenward–Roger degrees of freedom.
‡Not applicable.

Additionally, the bias-corrected estimates of the fixed effects for the transformed model are provided to compare with the uncorrected estimates. In this case, there is very little difference between the two sets of estimates. Even though the simulations indicate that the bias-corrected estimators are not perfect, this example acts as guidance on how they can be employed in their current form. Specifically, computation of both forms of estimators of β and a comparison of the two values will help to quantify the bias for the analyst, and in the case where there is no difference we can proceed with the uncorrected estimators in which reliable variance estimates exist.

To assess the distributional assumptions for this particular example, we examined the normal quantile–quantile (*q*–*q*-) plots on the resulting residuals, $\hat{b}_{0i}$, $\hat{b}_{1i}$ and $\hat{e}_{ij}$, using the fit of model (11) after calculating $\hat{\lambda}$ by using the scaled transformation. Simple *q*–*q*-plots could be used rather than the weighted *q*–*q*-plots of Lange and Ryan (1989) since we assume that the random intercept and random slope are independent. Examination of covariances for both transformed and untransformed models indicated that this independence assumption was valid for the data. The comparison of the *q*–*q*-plots for the untransformed model with those of the transformed model in Fig. 2 demonstrates the utility of the corollary. To simplify, we rounded to $\lambda = 0$, or the log-transformation. The transformation method, in targeting the normality of the total error, results in stronger assumptions of normality for all error sources of the model.

6. Discussion

Many areas of application utilize transformations, most often in the context of cross-sectional data. The increasing popularity of longitudinal data analyses requires a thorough examination of all the available tools to improve on the assumptions of the models. It is vital in practice to build a valid model of the data in which sound inferences can be made, and the use of transformations is one simple tool to help to reach that goal.

We have presented a general methodological overview and exploration of the linear mixed model for longitudinal data when the response is transformed parametrically. We have examined the effect of such a transformation on the distributional assumptions that are inherent
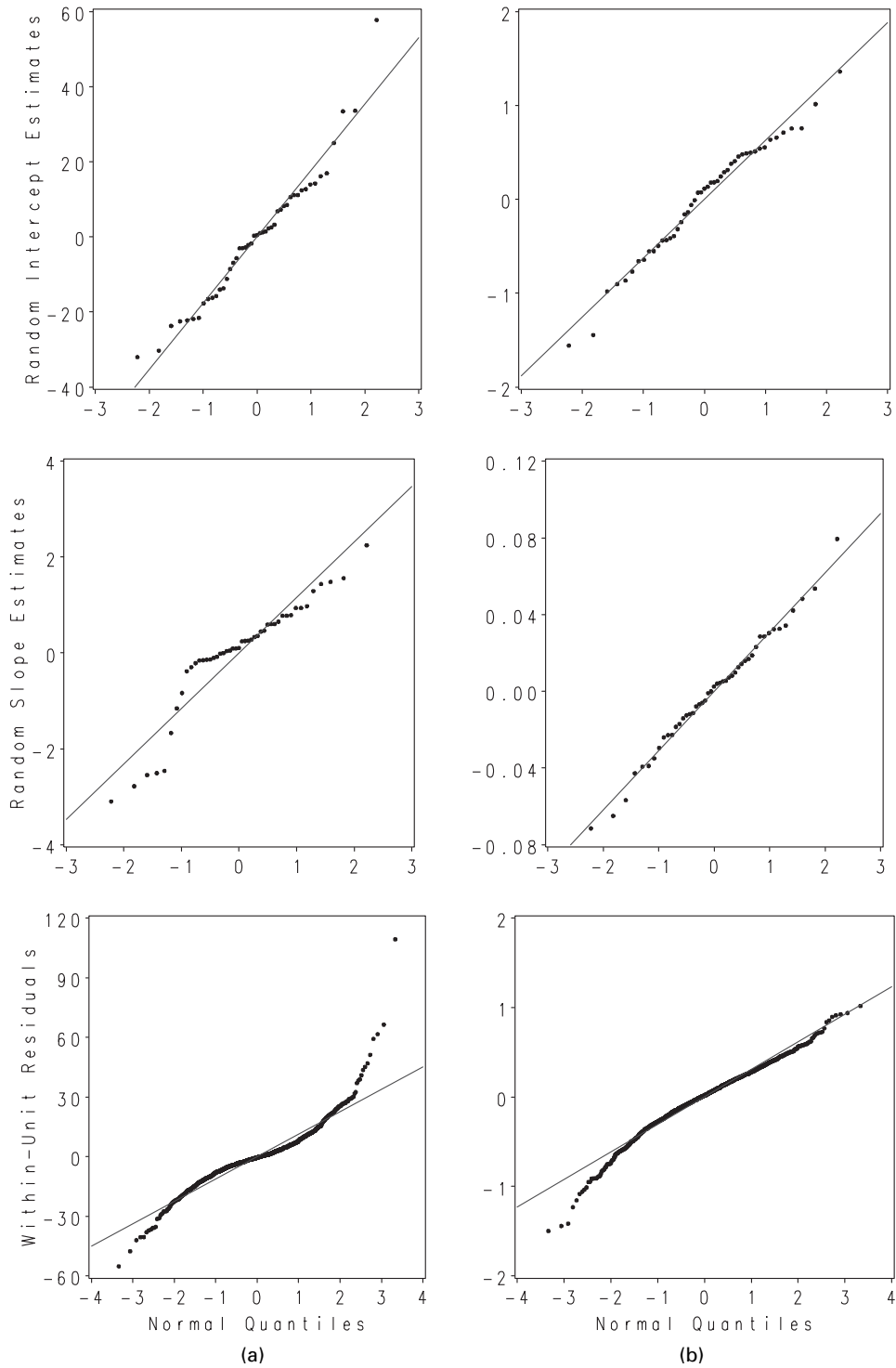
Fig. 2. Normal quantile–quantile plots for fitted MMEF data models: (a) no transformation ($\lambda = 1$); (b) log-transformation ($\lambda = 0$)

in the linear mixed model to understand exactly how the transformation helps to meet those assumptions. We have shown that, if the extension of the standard Box–Cox transformation to the linear mixed model results in near normality of the total error term, then the random effects and the pure error term will each have approximate Gaussian distributions. We can then avoid separately evaluating the normality of the pure error and the random effects. The accuracy of the assessment of multivariate normality of the total error term should be examined in more detail.

As is the case with univariate linear models, the Box–Cox transformation procedure does not allow us to state conclusively that the assumption of normality of the total error term in the linear mixed model is valid after applying the transformation. However, it does provide a model in which the assumption of normality is more reasonable than if we did not transform the response at all. Consequently, the interpretation is that the best parametric model of a given form is fitted relative to the assumption of approximate normality of both the random effects and the within-unit error term in the model. Of course, such conclusions depend on the valid assumption that both sources of randomness are independent of one another for the optimally transformed model.

The observed effect of outliers has led to interest in influence diagnostics for the linear mixed model. The presented corollary allows us to propose that, when the focus of a particular analysis is on estimation and inference of the fixed effects of the linear mixed model, we may aggregate the random effects and pure error term and hence only examine diagnostics on the level of the independent sampling unit. Further examination of this conclusion is necessary to help to simplify diagnostic analyses in the linear mixed model.

Even though the proposed transformation approach is promising in regard to addressing the normality assumptions of the mixed model, we clearly demonstrate that one must use caution when applying a transformation to the mixed model. The simulation results indicate that we should take into account the estimation of λ when focusing on the other parameters of the model. We have provided straightforward large sample methods that do this for the fixed effect parameter estimators. When the particular area of research suggests the use of a transformation, the methods are an alternative to the sandwich-type correction of the standard errors of the fixed effect estimators that are available when the random effects do not exhibit a Gaussian distribution.

This paper is one component of a general exploration of the idea of transformations in the linear mixed model for longitudinal data. Further areas of research include examination of influential data points and robust estimation. Most importantly, since previous research has indicated that accurate estimation and inference on the random effects is dependent on the assumption of normality (Verbeke and Molenberghs, 2000), it would be useful to study in more detail estimation and inference on the random effects after application of a transformation. We demonstrate that simple estimation of a transformation in the context of a linear mixed model leads to biased estimates of the variance components themselves. Thus, this work gives a clear warning about the naïve implementation of a transformation to the linear mixed model when interest lies in the variance components and/or the random effects. Future work that focuses on improving these estimators would be beneficial.

The methodological debates on the topic of the Box–Cox transformation for the univariate linear model carry over to the linear mixed model as well. We do not attempt to propose an answer to the controversy when extending transformation theory to the linear mixed model. We simply wish to assist in the understanding of a procedure that is sometimes used in practice without exact knowledge of its methodological implications. The transformation technique is a commonly employed tool for the univariate linear model. As a result of this research, its value as well as its limitations are evident when extended to the linear mixed model.

Acknowledgements

Appendix A: Theory and details from Section 3

The observed information matrix that is described in Section 3 and used in estimating the variance of $\hat{\lambda}$ and the fixed effects estimators in the scaled model is

$$\boldsymbol{\mathcal{I}}(\hat{\boldsymbol{\beta}}_{*}, \hat{\lambda}) = \begin{pmatrix} \sum\limits_{i=1}^{m} \mathbf{X}_i' \boldsymbol{\Sigma}_{*i}^{-1} \mathbf{X}_i & \boldsymbol{\mathcal{I}}_{12} \\ \boldsymbol{\mathcal{I}}_{12}' & \boldsymbol{\mathcal{I}}_{22} \end{pmatrix},$$

where

$$\boldsymbol{\mathcal{I}}_{12} = -\sum_{i=1}^{m} \mathbf{X}_i' \boldsymbol{\Sigma}_{*i}^{-1} \left(\frac{\partial \mathbf{w}_i^{(\lambda)}}{\partial \lambda} \right)_{\lambda = \hat{\lambda}},$$

$$\boldsymbol{\mathcal{I}}_{22} = \left\{ \sum_{i=1}^{m} \left(\frac{\partial \mathbf{w}_i^{(\lambda)}}{\partial \lambda} \right)' \boldsymbol{\Sigma}_{*i}^{-1} \left(\frac{\partial \mathbf{w}_i^{(\lambda)}}{\partial \lambda} \right) + \sum_{i=1}^{m} \left(\frac{\partial^2 \mathbf{w}_i^{(\lambda)}}{\partial \lambda^2} \right)' \boldsymbol{\Sigma}_{*i}^{-1} (\mathbf{w}_i^{(\lambda)} - \mathbf{X}_i \hat{\boldsymbol{\beta}}_*) \right\}_{\lambda = \hat{\lambda}},$$

$$\frac{\partial \mathbf{w}_i^{(\lambda)}}{\partial \lambda} = \lambda^{-1} \tilde{y}^{1-\lambda} \mathbf{s}_i^{(\lambda)} - \{\lambda^{-1} + \log(\tilde{y})\} \mathbf{w}_i^{(\lambda)},$$

$$\frac{\partial^2 \mathbf{w}_i^{(\lambda)}}{\partial \lambda^2} = \lambda^{-1} \tilde{y}^{1-\lambda} [\mathbf{t}_i^{(\lambda)} - 2\{\lambda^{-1} + \log(\tilde{y})\} \mathbf{s}_i^{(\lambda)}] + [\{\lambda^{-1} + \log(\tilde{y})\}^2 + \lambda^{-2}] \mathbf{w}_i^{(\lambda)},$$

and $\mathbf{s}_i^{(\lambda)}$ and $\mathbf{t}_i^{(\lambda)}$ are $n_i \times 1$ vectors such that the jth element is $y_{ij}^{\lambda} \log(y_{ij})$ and $y_{ij}^{\lambda} \log(y_{ij})^2$ respectively. Inverting the observed information matrix for $(\hat{\boldsymbol{\beta}}_*, \hat{\lambda})$ given above leads to estimates of $\mathcal{V}(\hat{\boldsymbol{\beta}}_*)$, $\mathcal{V}(\hat{\lambda})$ and $\mathrm{cov}(\hat{\boldsymbol{\beta}}_*, \hat{\lambda})$, denoted by $\boldsymbol{\Omega}_*$, σ_{λ}^2 and $\boldsymbol{\gamma}_*$ ($p \times 1$) respectively.

We now wish to obtain estimates for the preferred, original transformed model (2). Since $w_{ij}^{(\lambda)} = y_{ij}^{(\lambda)} / \tilde{y}^{\lambda-1}$, we can state $\boldsymbol{\beta} = \tilde{y}^{\lambda-1} \cdot \boldsymbol{\beta}_*$, $\mathbf{b}_i = \tilde{y}^{\lambda-1} \cdot \mathbf{b}_{*i}$ and $\mathbf{e}_i = \tilde{y}^{\lambda-1} \cdot \mathbf{e}_{*i}$, where $\tilde{y}$ is considered fixed. Focusing on the fixed effects, it is natural to use the estimator $\hat{\boldsymbol{\beta}} = \tilde{y}^{\hat{\lambda}-1} \cdot \hat{\boldsymbol{\beta}}_*$ to obtain estimators from the original transformed model. However, $\hat{\boldsymbol{\beta}}$ is a biased estimator of $\tilde{y}^{\lambda-1} \cdot \boldsymbol{\beta}_*$. Using a Taylor series expansion, the bias that is incurred in using $\hat{\boldsymbol{\beta}} = \tilde{y}^{\hat{\lambda}-1} \cdot \hat{\boldsymbol{\beta}}_*$, which is displayed in the simulations presented and in Spitzer (1978), can be characterized by the form

$$E(\hat{\boldsymbol{\beta}}) = \tilde{y}^{\lambda-1} \cdot \boldsymbol{\beta}_* + \tfrac{1}{2} \tilde{y}^{\lambda-1} \log^2(\tilde{y}) \boldsymbol{\beta}_* \sigma_{\lambda}^2 + \tilde{y}^{\lambda-1} \log(\tilde{y}) \boldsymbol{\gamma}_*,$$

since $E(\hat{\boldsymbol{\beta}}_*) = \boldsymbol{\beta}_*$ and $E(\hat{\lambda}) = \lambda$. Using the delta method, we obtain the variance of $\hat{\boldsymbol{\beta}}$:

$$\mathcal{V}(\hat{\boldsymbol{\beta}}) = \tilde{y}^{2(\lambda-1)} \log^2(\tilde{y}) \cdot \boldsymbol{\beta}_* \boldsymbol{\beta}_*' \sigma_{\lambda}^2 + \tilde{y}^{2(\lambda-1)} \boldsymbol{\Omega}_* + \tilde{y}^{2(\lambda-1)} \log(\tilde{y}) (\boldsymbol{\beta}_* \boldsymbol{\gamma}_*' + \boldsymbol{\gamma}_* \boldsymbol{\beta}_*').$$

Replacing the parameters with their estimates leads to a variance estimate of $\hat{\boldsymbol{\beta}}$ that 'adjusts' for the estimation of λ. Separately, an estimator of $\boldsymbol{\beta}$ that corrects for the incurred bias follows:

$$\hat{\boldsymbol{\beta}}_{\mathrm{adj}} = \tilde{y}^{\hat{\lambda}-1} \cdot \hat{\boldsymbol{\beta}}_* - \tfrac{1}{2} \tilde{y}^{\hat{\lambda}-1} \log^2(\tilde{y}) \hat{\boldsymbol{\beta}}_* \hat{\sigma}_{\lambda}^2 - \tilde{y}^{\hat{\lambda}-1} \log(\tilde{y}) \hat{\boldsymbol{\gamma}}_*.$$

However, $\mathcal{V}(\hat{\boldsymbol{\beta}}_{\mathrm{adj}})$ should be estimated to provide a usable alternative estimator. Such estimation is not straightforward, though, and further research is required. Thus, the theory that is presented here provides a bias-corrected estimator of the fixed effects parameters, taking into account the estimation of λ, as well as a separate adjusted variance estimate of the uncorrected fixed effects estimator.

Similar theory can be applied to the variance components of the mixed model, namely $\boldsymbol{\theta}$. The bias of the estimators and their variances can be explained by similar extensions of the above formulae. However, further complications arise in trying to extend the above theory to provide 'corrections' to the variance component estimators.

References

Andrews, D. F., Gnanadesikan, G. and Warner, J. L. (1971) Transformations of multivariate data. *Biometrics*, **27**, 825–840.

Berhane, K., McConnell, R., Gilliland, F., Islam, T., Gauderman, W. J., Avol, E., London, S. J., Rappaport, E., Margolis, H. G. and Peters, J. M. (2000) Sex-specific effects of asthma on pulmonary function in children. *Am. J. Resp. Crit. Care Med.*, **162**, 1723–1730.

Bickel, P. J. and Doksum, K. A. (1981) An analysis of transformations revisited. *J. Am. Statist. Ass.*, **76**, 296–311.

Box, G. E. P. and Cox, D. R. (1964) An analysis of transformations (with discussion). *J. R. Statist. Soc.* B, **26**, 211–252.

Carroll, R. J. and Ruppert, D. (1984) Power transformations when fitting theoretical models to data. *J. Am. Statist. Ass.*, **79**, 321–328.

Cramer, H. (1970) *Random Variables and Probability Distributions*, 3rd edn, pp. 53–56. London: Cambridge University Press.

Duan, N. (1983) Smearing estimate: a nonparametric retransformation method. *J. Am. Statist. Ass.*, **78**, 605–610.

Edwards, L. J. (2000) Modern statistical techniques for the analysis of longitudinal data in biomedical research. *Ped. Pulmonol.*, **30**, 330–344.

Gianola, D., Im, S., Fernando, R. L. and Foulley, J. L. (1990) Mixed model methodology and the Box-Cox theory of transformations: a Bayesian approach. In *Advances in Statistical Methods for Genetic Improvement of Livestock*, pp. 15–40. New York: Springer.

Harville, D. A. (1977) Maximum likelihood approaches to variance component estimation and to related problems. *J. Am. Statist. Ass.*, **72**, 320–338.

Hinkley, D. V. and Runger, G. (1984) The analysis of transformed data (with discussion). *J. Am. Statist. Ass.*, **79**, 302–320.

Kenward, M. G. and Roger, J. H. (1997) Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics*, **53**, 983–997.

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

Lange, N. and Ryan, L. (1989) Assessing normality in random effects models. *Ann. Statist.*, **17**, 624–642.

Lipsitz, S. R., Ibrahim, J. and Molenberghs, G. (2000) Using a Box–Cox transformation in the analysis of longitudinal data with incomplete responses. *Appl. Statist.*, **49**, 287–296.

Manning, W. G. (1998) The logged dependent variable, heteroscedasticity, and the retransformation problem. *J. Hlth Econ.*, **17**, 283–295.

Manning, W. G. and Mullahy, J. (2001) Estimating log models: to transform or not to transform? *J. Hlth Econ.*, **20**, 461–494.

Manzke, H., Stadlober, E. and Schellauf, H.-P. (2001) Combined body plethysmographic, spirometric and flow volume reference values for male and female children aged 6 to 16 years obtained from "hospital normals". *Eur. J. Ped.*, **160**, 300–306.

Oberg, A. and Davidian, M. (2000) Estimating data transformations in nonlinear mixed effects models. *Biometrics*, **56**, 65–72.

Sakia, R. M. (1990) Retransformation bias: a look at the Box-Cox transformation to linear balanced mixed ANOVA models. *Metrika*, **37**, 345–351.

Solomon, P. J. and Taylor, J. M. G. (1999) Orthogonality and transformations in variance components models. *Biometrika*, **86**, 289–300.

Spitzer, J. J. (1978) A Monte Carlo investigation of the Box-Cox transformation in small samples. *J. Am. Statist. Ass.*, **73**, 488–495.

Spitzer, J. J. (1982) A primer on Box-Cox estimation. *Rev. Econ. Statist.*, **64**, 307–313.

Verbeke, G. and Lesaffre, E. (1996) A linear mixed-effects model with heterogeneity in the random-effects population. *J. Am. Statist. Ass.*, **91**, 217–221.

Verbeke, G. and Lesaffre, E. (1997) The effect of misspecifying the random-effects distribution in linear mixed models for longitudinal data. *Computnl Statist. Data Anal.*, **23**, 541–556.

Verbeke, G. and Molenberghs, G. (2000) *Linear Mixed Models for Longitudinal Data*. New York: Springer.

Vonesh, E. F. and Chinchilli, V. M. (1997) *Linear and Nonlinear Models for the Analysis of Repeated Measurements*. New York: Dekker.

Wade, A. and Cole, T. J. (2001) Use of Box–Cox transformations. *Appl. Statist.*, **50**, 403.