# Highly efficient designs to handle the incorrect specification of linear mixed models

Shirley Ortega Azurduy, Frans E.S. Tan, Martijn P.F. Berger

## ▶ To cite this version:

HAL Id: hal-00514332

https://hal.science/hal-00514332

Submitted on 2 Sep 2010

# Highly efficient designs to handle the incorrect specification of linear mixed models

Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.

LSSP-2007-0223_SAOA_8jul2008.zip

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

# Highly efficient designs to handle the incorrect specification of linear mixed models

**S.A. Ortega-Azurduy,**[1,*]　**F.E.S. Tan**[1] **and M.P.F. Berger**[1]

[1]University of Maastricht,Department of Methodology and Statistics,

P.O. Box 616, 6200 MD Maastricht, The Netherlands

SUMMARY. We apply a maximin criterion to examine the relative efficiency of several $D_q$-optimal designs for a family of linear mixed models. Incorrect specifications of the order of the polynomial, size of the autocorrelation parameter, number of random parameters, and the correlation between random intercept and random slope are investigated. The results of our study allow us to draw the following conclusions: 1) the maximin $D_q$-optimal design encountered appears to be highly efficient;2) the variation of the minimum relative efficiencies of $D_q$-optimal designs of the family of linear mixed models that were studied, decreases as the order of the polynomial increases; 3) the effect of the autocorrelation parameter on the relative efficiencies of $D_q$-optimal designs is the largest for first-degree polynomials; and 4) the relative efficiency of the equidistant design is lower than that of the maximin value and also lower than the reference value 0.85.

KEY WORDS: $D$-optimality; $D_q$-optimality; first-order autocorrelation; linear mixed models; maximin criterion; relative efficiency; robust designs

$^*$*email:* shirley.ortega@STAT.unimaas.nl

1

## 1.  Introduction

Longitudinal and repeated measure designs are widely used by practitioners for experimental and observational studies. These designs target to investigate changes over time in subjects and are usually analyzed by linear mixed models (LMM). See e.g. Laird and Ware(1982); Verbeke and Molenberghs (2000).

An optimal design for longitudinal and repeated measurements depends on the assumed model. Practitioners, however, do not know beforehand the underlying model of a studied phenomenon and data analysts usually choose designs with sufficient number of time points at convenient places. This often results in loss of efficiency. The estimated coefficients of the underlying model are often not estimated efficiently because a chosen model with covariance matrix may differ from the true model and covariance matrix. In this paper, we will focus on the problem of finding efficient designs for linear mixed models that remain highly efficient even if the true model is not known.

The importance of studying optimal designs for linear mixed models has been advocated by many authors, Abt et.al. (1988, 1997). These authors searched for optimal designs for linear and quadratic growth models with random intercept and autocorrelated structures.

The robustness of designs against incorrect specifications of the polynomial model has been studied in recent years by Wong(1994); Dette(1997) and Moerbeek (2004). Few authors, however, have provided evidence of the drawbacks of specifying incorrectly random-effects models. Tan and Berger(1999); Berger et.al(2002) and Berger and Tan(2004) compared some $D$-optimal designs for polynomial regression models where either the random intercept

2

variance or the random slope variance was fixed given a first-order auto-correlation structure, AR(1). They showed that regardless the underlying polynomial regression model, the number of repeated measures should be chosen as close as possible to the number of regression parameters.

Berger et.al(2002), Ouwens et.al. (2002) and Berger and Tan(2004) studied maximin designs for mixed models. Their papers focuss on $D-$optimal cohort designs for first- and second-degree polynomial models with a (fixed) value for either the random intercept or the random slope variance and within specific intervals for the serial correlation parameter and the error variance.

In this paper we will expand their results and search systematically for efficient designs for a larger range of linear mixed models. The issues of concern will be 1. expand the class of models to a third-degree polynomial model, 2. extend the parameter values of both: the random intercept and the random slope variances, and 3. show that a discrete approximation of the standardized maximin method proposed by Dette and Neugebauer(1997) and Dette et.al.(2006) will lead to highly efficient designs for models and combinations of parameter values which often occur in practice.

The remainder of this paper is as follows. In section 2, we will include a motivating example. In section 3, we will introduce the linear mixed model framework and the formulation of the maximin criterion for these models. In section 4, we will explain our numerical approach. Thereafter, we will present the obtained results in section 5. We will discuss some special cases in section 6. Finally, in section 7 we will summarize our conclusions.

3

## 2.  Motivating example

In longitudinal studies of psychiatric disorders, uncovering relationships between pharmacological treatments and its therapeutic effects in patients is decisive. An experiment involving an antidepressive drug treatment for depressive inpatients was conducted to examine the relationship between antidepressants plasma levels and severity of depression. See for details Gram et.al.(1976) and Reisby et.al(1977). The inpatients were followed during an one-week baseline placebo period. Thereafter, a fixed drug's dose was administered during four weeks. Plasma levels of the drug in blood samples were measured and the clinical response of patients was rated using the Hamilton Rating Scales for Depression (HAM-D) at a total of six equally spaced time points.

Hedeker et.al(1989) and Hedeker and Gibbons(1996) used linear mixed models to analyze the experimental data and regress HAM-D scores on weeks by polynomial models with random intercepts and random slopes. Their results indicated a significant effect of the treatment on the response across the time points and evidence of a first-order autoregressive process in the residual errors.

The design of this study is characterized by the number of time points, the location of the time points and by the number of patients measured at the time points. In general, the design of such studies can be improved if more specific information is available about the best fitting linear mixed model. Information about the degree of the polynomial describing the trend over time, the random parameters and the autocorrelation errors may enable the researcher to choose a locally optimal design for such studies. The prob-

4

lem is, however, that such information is rarely available in the design stage of a study. Sometimes, previous research may provide reasonable guesses about these parameter values, but usually the researcher does not know the information required to find an optimal design for his/her study. The problem of finding an optimal design for linear mixed models is hampered by the fact that the information about the parameter may be incorrectly specified. A robust design for these studies should be capable to deal with the incorrect specification of the degree of the polynomial involved, the random variance-covariance structure and the size of the serial correlation. The use of a maximin criterion to identify maximin designs would be an alternative to overcome model and covariance structure uncertainty and reduce experimental costs.

## 3. Linear mixed models and optimality of designs

Let $\mathbf{y}_i$ be a continuous response of subjects $i = 1, \ldots, n$ with three components: the overall-population effects, the random subject-specific effects and the first level within-subjects error terms. For each subject, the measured response is given by:

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \mathbf{e}_i, \tag{1}$$

where $\boldsymbol{\beta}$ denotes a $p \times 1$ vector of unknown population parameters, $\mathbf{X}_i$ is a known $q \times p$ design matrix linking $\boldsymbol{\beta}$ to $\mathbf{y}_i$. The number of the fixed parameters of the polynomial regression and the number of design points (or time points) are given by $p$ and $q$, respectively. Let $\mathbf{b}_i$ denote a $k \times 1$ vector of unknown subject-specific effects and $\mathbf{Z}_i$ be a known $q \times k$ design

5

matrix linking $\mathbf{b}_i$ to $\mathbf{y}_i$. The random-effects parameter vector $\mathbf{b}_i$ represents how the $i^{th}$ subject deviates from the average population and it is normally distributed with mean $\mathbf{0}$ and $k \times k$ covariance matrix $D$, i.e., $N(\mathbf{0}, D)$. We will restrict ourselves to a $2 \times 2$ covariance matrix $D$, where $d_{11}$ stands for the variance of the random intercepts, $d_{12}$ is the covariance between the random intercepts and random slopes and $d_{22}$ corresponds to the variance of the random slopes. In practice, the most often encountered mixed models have only random intercept and random slope terms (Hedeker et.al., 1989; Hedeker and Gibbons, 1996). Thus, the quadratic and cubic random effects are assumed to be fixed, throughout.

The errors $\mathbf{e}_i$ are distributed as $N(\mathbf{0}, \sigma^2 \mathbf{\Psi}_i)$. Here $\sigma^2 \mathbf{\Psi}_i$ is a positive-definite covariance matrix, and depends on the correlation parameter $\rho$ of the time lag between the $j^{th}$ and $j'^{th}$ time points $x_j$ and $x_{j'}$, i.e., $\rho^{\mathrm{abs}(x_j - x_{j'})}$. These elements of the within-subject correlation matrix $\mathbf{\Psi}_i$ correspond to an exponentially decreasing function known as autoregressive correlation structure in which the time points $x_j$ are not necessarily equally spaced. We will symbolize this structure by AR1. Although other structures for $\mathbf{\Psi}_i$ exist, we will restrict ourselves to this structure because it is the most often encountered structure in longitudinal designs. For a list of other possible variance-covariance structures, see Verbeke and Molenberghs (2000, p.99-100). Notice that the random variables $\mathbf{b}_i$ and $\mathbf{e}_i$ are independent, identically and normally distributed.

In what follows we will assume that design matrices $\mathbf{X}_i$ and $\mathbf{Z}_i$ and covariance matrix $\mathbf{V}_i$ are the same for each subject. So, if we will consider that measurements are taken at the same design points for all subjects, the model (1)

6

can be written as a family of linear mixed models: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \mathbf{e}$, whereof $\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta}, \mathbf{Z}, \mathbf{b}$, and $\mathbf{e}$ are all matrices concatenating $n$ subject-specific regression models, i.e. $\mathbf{Y} = (\mathbf{y}'_1, \ldots, \mathbf{y}'_n)'$, $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_n)'$, $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_p)'$, $\mathbf{Z} = \mathrm{Diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_n)$, $\mathbf{b} = (\mathbf{b}'_1, \ldots, \mathbf{b}'_n)'$ and $\mathbf{e} = (\mathbf{e}'_1, \ldots, \mathbf{e}'_n)'$. This linear mixed model implies a marginal model $\mathbf{Y}$ distributed normally with mean $\mathbf{X}\boldsymbol{\beta}$ and covariance matrix:

$$\mathbf{V} = \mathbf{Z}\boldsymbol{D}\mathbf{Z}' + \sigma^2 \boldsymbol{\Psi}, \tag{2}$$

where $\boldsymbol{D} = \mathrm{Diag}(D, \ldots, D)$ and $\boldsymbol{\Psi} = \mathrm{Diag}(\boldsymbol{\Psi}_1, \ldots, \boldsymbol{\Psi}_n)$. The matrix product $\mathbf{Z}\boldsymbol{D}\mathbf{Z}'$ and matrix $\boldsymbol{\Psi}$ have size $(nq \times nq)$.

The best linear unbiased estimator of $\boldsymbol{\beta}$ has a variance-covariance matrix

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}) = \left(\mathbf{X}'\mathbf{V}^{-1}\mathbf{X}\right)^{-1}, \tag{3}$$

which is equal to the inverse Fisher information matrix. Small valued asymptotic covariance matrices of the parameter estimates correspond to large valued information matrices. Fisher information matrix represents the amount of information contained in a specific design and expresses the amount of precision by which the model parameters $\boldsymbol{\beta}$ are estimated by $\hat{\boldsymbol{\beta}}$. Good estimators with high precision have small valued estimated (asymptotic) variances, $\mathrm{Var}(\hat{\boldsymbol{\beta}})$.

### 3.1 *Optimal designs*

In this section, we introduce a design space $\boldsymbol{\Xi}$ and a parameter space $\boldsymbol{\Omega}$ as the building blocks required to handle and apply optimality and maximin criteria.

7

### 3.1.1  Design space, $\boldsymbol{\Xi}$

Let us introduce a design $\xi$ in the design space $\boldsymbol{\Xi}$. $\xi$ is a probability measure with finite support in $x_j$, where $j = 1, \ldots, q$ and without loss of generality $-1 \leq x_j \leq 1$; that is, $\xi$ represents the relative proportion of the total observations taken at $x_j$'s time points Kiefer(1959). The design $\xi$ is defined as:

$$\xi = \left\{ \begin{matrix} x_1 & x_2 & \cdots & x_j & \cdots & x_q \\ 1/q & 1/q & \cdots & 1/q & \cdots & 1/q \end{matrix} \right\}, \text{where } (1 \leq j \leq q). \quad (4)$$

where $\int_{-1}^{1} \xi(dx) = 1$ and the weights are all equal to $1/q$. In our study we have considered data having equal weights, meaning that no missing patterns arise. Unequal weights may arise in practice and are known as unbalanced designs.

### 3.1.2  Parameter space $\boldsymbol{\Omega}$

Let us define the parameter space $\boldsymbol{\Omega}$ as the parameters connected to the family of marginal linear mixed models in (1):

$$\boldsymbol{\Omega} = \big\{ \omega = (q, p, \boldsymbol{\beta}, D, \rho, \sigma^2) : 2 \leq p \leq q, \boldsymbol{\beta} \in \mathcal{R}^p,$$
$$D \geq \boldsymbol{0}, 0 \leq \rho < 1; \sigma^2 > 0 \big\} \quad (5)$$

$\boldsymbol{\Omega}$ contains the mean parameters $\boldsymbol{\beta}$, the variance and covariance parameters of $\boldsymbol{D}$ and $\boldsymbol{\Psi}$, and the variance of the errors, $\sigma^2$.

We consider the space $\boldsymbol{\Omega}$ to have a total of $W$ different combinations of parameters. Let us denote $w = 1, \ldots, W$ to identify each possible combination of parameters in $\boldsymbol{\Omega}$. There is a one-to-one correspondence between each element $\omega$ of the parameter space $\boldsymbol{\Omega}$ and the model $M_w$. Each combination

8

of parameter values $\omega$ determines a linear mixed model of the form (1) and for each model $M_w$ we can derive an optimal design.

Notice that quadratic and cubic-effects are kept as fixed. In practice, one often only specify deviations from the average that change linearly in time. Quadratic- and cubic -effects could also be considered, but its existence and significance might be much harder to show because of lack of power. See Verbeke and Molenberghs, p.70-71, 2000.

### 3.1.3   Optimality criterion

Different optimality criteria are proposed in the literature to compare competing designs (Kiefer, 1959; Kiefer and Wolfowitz, 1961; Fedorov, 1972; Atkinson and Donev, 1996). In this paper, we will focus on the determinant criterion or $D$-optimality criterion. The idea is to minimize the determinant function of the asymptotic covariance matrix for the estimated parameter $\hat{\boldsymbol{\beta}}$ for a given model $M_\omega$, i.e.,

$$\min_{\xi \in \boldsymbol{\Xi}} \Phi\{\underset{M_w}{\operatorname{Var}}(\hat{\boldsymbol{\beta}})\}. \tag{6}$$

This criterion is particularly interesting since it has a natural and intuitive interpretation among practitioners as the minimum volume of the confidence ellipsoid about $\boldsymbol{\beta}$. Another important advantage is that the $D-$optimality criterion is invariant with respect to reparametrization (Ouwens et.al., 2002).

The determinant of the $\operatorname{Var}(\hat{\boldsymbol{\beta}})$ can be written as,

$$\begin{aligned}
\operatorname{Det}\{\underset{M_w}{\operatorname{Var}}(\hat{\boldsymbol{\beta}}|\xi)\} &= \operatorname{Det}\left(\mathbf{X}_\xi' \mathbf{V}_\xi^{-1} \mathbf{X}_\xi\right)^{-1}, \\
&= \operatorname{Det}\{\mathbf{X}_\xi' \left(\mathbf{Z}_\xi \boldsymbol{D} \mathbf{Z}_\xi' + \sigma^2 \boldsymbol{\Psi}_\xi\right)^{-1} \mathbf{X}_\xi\}^{-1},
\end{aligned} \tag{7}$$

where $\xi$ refers to the design that is used for the estimation of $\operatorname{Var}(\hat{\boldsymbol{\beta}})$ under model $M_w$.

9

By applying relation (6) to (7), we obtain that the $D$-optimal design $\xi^*$ is the design that satisfies:

$$\text{Det}\{\underset{M_w}{\text{Var}}(\hat{\boldsymbol{\beta}}|\xi^*)\} \leq \text{Det}\{\underset{M_w}{\text{Var}}(\hat{\boldsymbol{\beta}}|\xi)\} \qquad \forall \ \xi \in \boldsymbol{\Xi}, \tag{8}$$

where $\hat{\boldsymbol{\beta}}$ is the vector of the estimated parameters for designs $\xi^*$ and $\xi$ under model $M_w$, respectively.

A complete description about approximate theory on the general equivalence theory for optimum designs is provided in Fedorov(1972); Atkinson and Donev(1996); Kiefer(1974) and Pukelsheim(1993).

### 3.2   *Maximin criterion*

Different approaches can be used to investigate efficiency loss when a model is incorrectly specified; among them are the sequential approach (Wynn, 1970) and the Bayesian approach (Chaloner and Verdinelli, 1995). For a review of these methods, the reader is referred to Atkinson and Donev(1996). In this paper we will focus on the maximin approach.

The maximin approach was used to overcome the problem of local optimality (Dette and Neugebauer, 1997; Muller, 1995; Bischoff, 1996). In this approach practitioners and analysts select a region of interest for each parameter, such that maximin designs do not depend on the individual parameter values. Thus, the estimation of the variance random components is not covered in this manuscript but a set of values is considered instead.

The maximin designs in this paper are based on $D$-optimal designs. The space of all $D$-optimal designs $\xi_\omega^*(w = 1, \ldots, W)$ for all linear mixed models is $\boldsymbol{\Xi}^*$; all $\xi_\omega^*$ are elements of $\boldsymbol{\Xi}^*$ and $\boldsymbol{\Xi}^*$ is a subset of $\boldsymbol{\Xi}$, i.e., $\xi_\omega^* \in \boldsymbol{\Xi}^* \subset \boldsymbol{\Xi}$.

Two types of $D$-optimal designs $\xi_\omega^*$ can be distinguished:

$$\xi_S^* := D\text{-optimal design for a } \textit{selected model } M_S,$$

$$\xi_T^* := D\text{-optimal design for a } \textit{true model } M_T.$$

Notice that the $D$-optimal design space $\Xi^*$ includes the $D$-optimal designs $\xi_S^*$ and $\xi_T^*$ for the selected model $M_S$ and true model $M_T$, respectively.

The only parameters that are fixed in models $M_S$ and $M_T$ are the number of points $q = 4$, the residual variance $\sigma^2 = 1$ and the interval $[-1, 1]$. The polynomial degree $(p - 1)$, the variance of the random intercept $d_{11}$, the variance of the random slope $d_{22}$, the autocorrelation parameter $\rho$ and the covariance $d_{12}$ between random parameters under the selected model $M_S$ and the true model $M_T$ are not necessarily the same.

The maximin procedure is implemented in three steps. First, we compare the standardized relative efficiencies of an optimal design $\xi_S^*$ for a selected model with respect to the efficiencies of the optimal designs $\xi_T^*$ for a set of assumed true models, i.e., for each assumed true model $M_T$, the generalized variance of the estimated regression coefficients $\underset{M_T}{\mathrm{Var}}(\hat{\boldsymbol{\beta}}|\xi_T^*)$ given the optimal design for the true model is computed, and compared to the generalized variance of the estimated regression coefficients $\underset{M_T}{\mathrm{Var}}(\hat{\boldsymbol{\beta}}|\xi_S^*)$ given the optimal design of a selected model.

The relative efficiency of the selected $D-$optimal design compared to the

11

true optimal design is:

$$
\begin{aligned}
RE(\xi_S^*|\xi_T^*) &= \left[ \frac{\mathrm{Det}\{\underset{M_T}{\mathrm{Var}}(\hat{\boldsymbol{\beta}}|\xi_T^*)\}}{\mathrm{Det}\{\underset{M_T}{\mathrm{Var}}(\hat{\boldsymbol{\beta}}|\xi_S^*)\}} \right]^{\frac{1}{p_T}} \\
&= \left\{ \frac{\mathrm{Det}(\mathbf{X}'_{\boldsymbol{\xi}_T^*}\hat{\mathbf{V}}_{\boldsymbol{\xi}_T^*}^{-1}\mathbf{X}_{\boldsymbol{\xi}_T^*})^{-1}}{\mathrm{Det}(\mathbf{X}'_{\boldsymbol{\xi}_S^*}\hat{\mathbf{V}}_{\boldsymbol{\xi}_S^*}^{-1}\mathbf{X}_{\boldsymbol{\xi}_S^*})^{-1}} \right\}^{\frac{1}{p_T}},
\end{aligned} \tag{9}
$$

where $\hat{\mathbf{V}}_{\boldsymbol{\xi}_S^*}$ and $\hat{\mathbf{V}}_{\boldsymbol{\xi}_T^*}$ are the estimated variance-covariance matrix of the selected optimal design $\boldsymbol{\xi}_S^*$ and true optimal design $\boldsymbol{\xi}_T^*$ both under model $M_T$. Notice that the $D$-optimal design $\xi_S^*$ is optimal for the selected model $M_S$, but not for the true model $M_T$ and that the numerator of (9) has a smaller variance than the denominator when the optimal design $\xi_S^*$ for the selected model is not equal to $\xi_T^*$ for the true model. The number of fixed parameters of the true model is equal to $p_T$. Taking the ratio of the determinant in (9) to the $1/p_T$ power results in an efficiency measure which is proportional to design size, irrespective of the dimension of the model (Pukelsheim, 1993; Atkinson and Donev, 1996). The interpretation of the relative efficiency in (9) is straightforward. Two replicates of a selected optimal design $\xi_S^*$ for which $RE = 0.5$ would be as efficient as one replicate of the true optimal design $\xi_T^*$. Relative efficiencies of 0.85 or higher are usually preferred and the relative efficiency of an optimal design with respect to itself is clearly equal to 1. Notice that equation (9) can be also used to compare designs with different number of repeated measures (Ouwens et.al., 2002, p.736).

Second, we seek for the minimum of all these relative efficiencies over all true (assumed) models, i.e., $M_T = M_w \in \{M_1, \ldots, M_W\}$. Finally, the maximum of all the minimum relative efficiencies is selected over all possible $D$-optimal designs for the selectable models, i.e., $M_S = M_w \in \{M_1, \ldots, M_W\}$.

12

The maximin value $MMV$, in terms of relative efficiency, is defined as:

$$MMV = \max_{\xi_S^* \in \boldsymbol{\Xi}^*} [\min_{\xi_T^* \in \boldsymbol{\Xi}^*} RE(\xi_S^*|\xi_T^*)]. \tag{10}$$

For each $\xi_S^*$ in the design space $\boldsymbol{\Xi}^*$, the minimum of relative efficiencies $RE(\xi_S^*|\xi_T^*)$ is selected over all $\xi_T^*$ designs. Then the maximum of all minima is selected over all $\xi_S^*$ designs. The design that has this maximin value $MMV$ is referred to as the maximin $D_q$-optimal design $\xi_{MMV}$. Notice that this paper focus on model misspecification, thus the number of design points are the same for the selected and the true model and we will only compare design with the same number of time points.

### 3.2.1 $D_q$-optimal designs

'$D_q$-optimal' design refers to $D$-optimal designs where the number of time points is fixed at $q \geq p$, the number of regression parameters. A design $\xi_q^*$ is $D_q$-optimal if $\mathrm{Det}\{\underset{M_w}{\mathrm{Var}}(\hat{\boldsymbol{\beta}}|\xi_q^*)\} \leq \mathrm{Det}\{\underset{M_w}{\mathrm{Var}}(\hat{\boldsymbol{\beta}}|\xi_q)\}$, for all $\xi_q \in \boldsymbol{\Xi}$. For further details on these optimal designs, the reader is referred to Tan and Berger(1999).

In general, notice that, we will refer to $D_q$ designs as the local optimum design if Model $M_w$ where $w = 1, \ldots, W$ is used. Later on, the reader will notice that the subindex $w$ is used to point to a true model $M_T$ with optimal design $\xi_T^*$ or to the selected model $M_S$ with optimal design $\xi_S^*$.

## 4. Numerical Analysis

In this section, we tackle numerically the problem of finding efficient designs that remain efficient even if the model is misspecified.

It is well-known that the correct specification of the linear mixed models

13

depends upon four sources of uncertainty, namely the order of the polynomial $(p-1)$, the size of the autocorrelation coefficient $\rho$, the number $k$ of random effects, and the correlation $\rho^*$ between the random intercept and random slope.

To deal with uncertainty about the order of the polynomial, the models $M_w$ will be restricted to the set $S_p = \{2, 3, 4\}$, i.e., the first-degree ($p = 2$, linear), the second-degree ($p = 3$, quadratic) and the third-degree ($p = 4$, cubic). These polynomial models are mostly encountered in longitudinal studies in health sciences and medicine (Hedeker et.al., 1989; Hedeker and Gibbons, 1996; Keller et.al., 2003). The number of time points chosen is $q = 4$, it enables us to test the goodness-of-fit. Higher (fourth or fifth) order polynomials are not considered because they may not be always meaningfully interpreted in any biological or physical sense.

To deal with the uncertainty of the size of the autocorrelation coefficient $\rho$, a convenient grid is used that is $S_\rho = \{0.0001, 0.1, 0.2, 0.3, 0.6, 0.9\}$. Since $\rho$ is chosen from the set $S_\rho$, six different within-subject residual variance matrices $\mathbf{\Psi}$ can be specified.

To handle uncertainty of the random effects, four classes of linear mixed models are considered. These models are based on the number of random effects $k \in S_k = \{0, 1, 2\}$:

1. The case $k = 0$ leads to fixed effect models with AR1 structure, i.e there is no random term in the variance covariance matrix $\mathbf{V}$. The random variance-covariance matrix $D$ is $\mathbf{0}$ in block matrix $\mathbf{D}$ of equation (2) and the matrix $\mathbf{V}$ depends only on matrix $\mathbf{\Psi}$. This class of fixed models will be denoted by **AR1**.

14

2. The case $k = 1$ enhances random intercept models with AR1 structure. This class of models is denoted by **RI-AR1**. Because, besides the autocorrelated structure for the residuals, there is also a random intercept (RI) variance in matrix $D = (d_{11})$. Random intercept models with AR1 structure are obtained by choosing a value $d_{11}$ from the set of random intercept variances $S_{d_{11}} = \{1, 2, 3, 4, 10\}$ and a value for $\rho \in S_{\rho}$.

3. The case $k = 2$ allows random intercept and random slope models with AR1 structure. This class of models is denoted by **RI-RS-AR1**. The random intercept (RI) variance and random slope (RS) variance of matrix $D$ are both non-zero, i.e. $D = \text{Diag}(d_{11}, d_{22})$ and residuals are autocorrelated. This model is constructed by choosing a value for the random intercept variance $d_{11} \in S_{d_{11}}$, a value for the random slope variance $d_{22}$ from the set of random slope variances $S_{d_{22}} = \{1, 3\}$ and a value for $\rho \in S_{\rho}$.

4. In this case, $k = 2$ and the covariance between random intercept and random slope is non-zero. This covariance depends on the correlation $\rho^*$ between the two random variance parameters. The main diagonal elements of $D = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}$ are filled by values $d_{11} \in S_{d_{11}}$ and $d_{22} \in S_{d_{22}}$. Then, using the relation $d_{12} = \rho^* \sqrt{d_{11} d_{22}}$ the off-diagonal elements of the symmetric matrix $D$ are fixed. Notice that $D$−optimal designs are invariant with respect to the sign of $d_{12}$ (Ouwens et.al., 2002). The correlation $\rho^*$ is chosen from a set $S_{\rho^*} = \{0.5, 0.8\}$ and the parameter $\rho$ is chosen from $S_{\rho}$. This class of models is denoted by **RI-RS-$\rho^*$-AR1**.

15

In this paper, we have neglected the fact that the parameters of the $\mathbf{V}$ matrix need to be estimated. The optimal designs are calculated conditional on the parameters of the $\mathbf{V}$ matrix, which is often done in literature (Näther, 1985, p.9).

Given the sets of parameter values $S_p, S_\rho, S_k, S_{d_{11}}, S_{d_{22}}$ and $S_{\rho^*}$, the parameter space of our numerical analysis is defined as:

$$\mathbf{\Omega} = \{\omega = (q, p, \boldsymbol{\beta}, D, \rho, \sigma^2) : q = 4, p \in S_p, \boldsymbol{\beta} \in \mathcal{R}^p, \tag{11}$$

$$d_{11} \in S_{d_{11}}, d_{22} \in S_{d_{22}}; \rho \in S_\rho, \rho^* \in S_{\rho^*}, \sigma^2 = 1; k \in S_k\},$$

Each combination of parameters $\omega$ in the parameter space $\mathbf{\Omega}$ leads to a model $M_w$, where $w = 1, \ldots, W$. These set of values are chosen to cover the range of estimates obtained in most longitudinal studies (Verbeke and Molenberghs, 2000).

[Table 1 about here.]

The ordering of the models for different combinations of parameters is illustrated in Table 1. In this table, the fastest varying index is the index for the autocorrelation parameter $\rho \in S_\rho$, whereas the slowest varying index is always the index for the random intercept variance $d_{11} \in S_{d_{11}}$. Notice that model complexity increases with the value of $\omega$ and that in the linear case ($p = 2$), the quadratic ($p = 3$) and cubic ($p = 4$) cases, $\omega$ belongs to the interval $[1, 216]$, $[217, 432]$ and $[433, 648]$, respectively. The ordering of models $M_w$ starts with model class **AR1** and ends with **RI-RS-$\rho^*$-AR1**. The parameter combination of $M_w$ can be illustrated by describing the easiest and most complex model classes as examples.

The first model class **AR1** ($M_1 - M_6$) is obtained when $k = 0$ and the $\rho$ values are chosen one-by-one in ascending order from set $S_\rho$.

16

The last class of models **RI-RS-$\rho^*$-AR1** $(M_{97} - M_{216})$ is more complicated because the combination of parameters $d_{11}$, $d_{22}$ and $\rho^*$ is required while $k = 2$. So, let $d_{11}$, $d_{22}$ and $\rho^*$ be fixed at their lowest values, i.e., 1, 1 and 0.5, respectively. Next, choose one-by-one an autocorrelation parameter $\rho$ from the set $S_\rho$, in ascending order. Afterwards, keep $d_{11}$ and $d_{22}$ unchanged and vary $\rho^*$ to its higher level, i.e., 0.8. Then again, vary the autocorrelation parameter $\rho \in S_\rho$, in ascending order. Thereafter, let $d_{11}$ remain at its lowest level, while $d_{22}$ takes the value 3 and put the value of $\rho^*$ back to its lowest level. Vary the autocorrelation parameter $\rho \in S_\rho$, in ascending order. Subsequently, let $d_{11}$ and $d_{22}$ be unchanged and increase $\rho^*$ to 0.8. And, so again choose $\rho$ from the set $S_\rho$. Repeat this procedure with a next $d_{11}$ and so forth.

4.1  *Algorithm*

We numerically derive $D_q$-optimal designs $\xi_q^*$ for all models $M_w$ by optimizing the generalized variance of their estimated parameters. A program using function FMINSEARCH was implemented in MATLAB version 7.0.1(R14). The program uses a multidimensional unconstrained nonlinear minimization algorithm developed by Nelder and Mead (Lagarias et.al., 1998; Matlab, 2004). The program requires an initial design value $\xi_0$. We used equally-spaced designs as initial values.

Thereafter, we compute the $RE$ in (9) for all $D_q$-optimal designs $\xi_q^*$'s. We evaluate $W \times W$ pairs of $D_q$-optimal true $\xi_T^*$ and $D_q$-optimal selected $\xi_S^*$ designs. The application of the criterion (10) leads to the maximin value $MMV$ and the corresponding maximin design $\xi_{MMV}$. The total number of paired comparisons of designs and models is $648^2 - 648 = 419\,256$.

17

The MATLAB programs used to derive $D_q$−optimal designs and to find maximin designs in the parameter space $\boldsymbol{\Omega}$ are available upon request to the authors as well as the list with the most highly efficient designs. The computational time required to obtain our results in a desktop computer with dubble INTEL processor $x86$, Family 6, Model 15 having a speed of 1595 MHz is approx. 35 minutes.

## 5. Results

A total of 648 minimum relative efficiencies are displayed in Figure 1, where the upper, middle and lower subplots show the results for (a) linear, (b) quadratic and (c) cubic models, respectively. On the $x$−axis of each subplot, we distinguished the four model classes: **AR1**, **RI-AR1**, **RI-RS-AR1**, and **RI-RS-$\rho^*$-AR1** represented by their optimal designs $\xi_S^*$ where $S = 1, \ldots, 648$.

[Figure 1 about here.]

In all subplots of Figure 1, a horizontal line is drawn at min $RE = 0.85$ as reference line. This reference line leads to 97 optimal designs with min $RE \geq$ 0.85. Notice that all quadratic models and related optimal designs $\xi_S^*$ turn out to have min $RE$ values smaller than $RE = 0.85$ (Figure 1b).

The following results are based on Figure 1:

1. **The efficiency of the encountered maximin design is very high**

   The maximin $D_q$-optimal design $\xi_{MMV}$ for the parameter space $\boldsymbol{\Omega}$ has a relative efficiency $RE = 0.8727$, see Figure 1a. The design $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ is $D_q$-optimal for a first-degree polynomial with $p = 2$, random intercept $d_{11} = 1$ and with autocorrelation parameter $\rho = 0.2$.

18

2. **The highly efficient designs are located around $\boldsymbol{\xi}^*$ for $q = 4$**

   In Table 2, a list of the most highly efficient $D_q$-optimal designs is presented. All $D_q$ designs have an autocorrelation coefficient $\rho$ smaller than 0.2. The models involved are mostly linear and pertain to either the class model **RI-AR1** or class model **RI-RS-AR1**.

   [Table 2 about here.]

3. **Variation of minimum $RE$ values decreases as the order of $p$ increases**

   Figures 1a-1c show that the variation of the min $RE$ values for the linear, quadratic and cubic models decreases as the degree of the polynomial increases.

   The min $RE$s of locally optimal designs $D_q$ shows steadily less variation on their min $RE$s values as the number of fixed parameters $p$ increases from 2 to 4, i.e., as $p$ approaches to number of time points $q$. We observe that under these conditions the locally optimal design points depend less on the covariance-variance structure and hence, are less affected by its incorrect specification.

4. **Effect of autocorrelation parameter $\rho$ on relative efficiencies**

   The effect of the autocorrelation parameter $\rho$ in the relative efficiencies min $RE$ of $D_q$-optimal designs corresponding to linear and quadratic models is large. For the linear model in Figure 1a, an increase of the value of $\rho$ first leads to a rapid increase of the relative efficiency in the interval $[0.0001, 0.2]$. Once a maximum is reached in the interval $[0.1, 0.2]$, the relative efficiency of the $D_q$-optimal designs decreases

   19

slowly, but steadily, as correlation increases. All linear models studied show this non-linear relation. For the quadratic model, the relation of $\min RE$ with the autocorrelation parameter $\rho$ is positive. These $\min RE$'s reach their maximum values for $\rho = 0.9$, see Figure 1b.

In contrast to linear and quadratic models, the relation in cubic models between the autocorrelation parameter $\rho$ and the relative efficiency $\min RE$ is relatively small and negative (see Figure 1c). High $\min RE$'s are usually encountered for the case where $\rho = 0.0001$ and low $\min RE$'s are found for high serial correlation values, i.e., $\rho = 0.9$.

## 6.    Special cases

### Effect of fixing design points to $q = 4$

The effect of fixing the number of design points to $q = 4$ in the maximin procedure is that the $MMV$ of the maximin design will be very high. However, Wong(1994) and Moerbeek (2004) reported $RE$'s for uncorrelated fixed effects and fixed effects with AR1 structure, which are much lower. In their papers, the $RE$ in equation (9) is evaluated for the $q = 2$ points $D$-optimal design $\xi^* = \{-1 \quad 1\}$ and the $q = 3$ points $D$-optimal design $\xi^* = \{-1 \quad 0 \quad 1\}$ of a linear and quadratic model, respectively. Our results are based on $D_q$-optimal designs with fixed time points, $q = 4$. We will illustrate the consequences of such a choice.

Our maximin design is $\xi_{MMV} = \{-1 \quad -0.4568 \quad 0.4568 \quad 1\}$ with $MMV = 0.8727$. If we recalculate the $RE$ of this design related to the efficiency of the $q = 2$ points $\xi^* = \{-1 \quad 1\}$ then the relative efficiency value is $RE = 0.4963$. This value is almost two times smaller than our current reported $MMV$. This difference is to be expected, because Wong(1994) and Moerbeek (2004)

20

are comparing a $D$-optimal design with $q = 2$ points to a design with $q = 4$ design points, whereas we fix the design points to $q = 4$ points. We choose for the $D_q$-optimality criterion to account for the uncertainty about the order of the polynomial. Because, the effect of this additional uncertainty about the true order of the polynomial $p_T$ adds extra loss of efficiency.

**Effect of the presence of autocorrelation**

In longitudinal and repeated measure designs, repeated measures could be correlated and in practice the uncorrelated condition in our numerical analysis might be unrealistic. So, let us introduce two sets $S_{\rho(1)}$, such that $S_{\rho(1)} \equiv S_\rho$ and $S_{\rho(2)} = \{0.1, 0.2, 0.3, 0.6, 0.9\}$.

In Table 3, the two cases are summarized. For $\rho \in S_{\rho(1)}$, we obtain a maximin design $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ corresponding to a linear model with an AR1 structure having $\rho = 0.2$ and a random intercept variance $d_{11} = 1$. Its $MMV$ is $RE = 0.8727$. For the $\rho \in S_{\rho(2)}$ condition, the maximin design becomes $\xi_{MMV} = \{-1 \ -0.4247 \ 0.4247 \ 1\}$, corresponding to a linear model with random intercept variance $d_{11} = 1$, $\rho = 0.3$ and its $MMV$ has a $RE = 0.9734$. So, if we assume that the correlation $\rho \in S_{\rho(2)}$, our maximin procedure will lead to a maximin design with a much higher $MMV$.

The purpose of studying the effect of the autocorrelation on the chosen maximin design is to answer the question: what-if one has serial correlation. Do we need to know the value of the serial correlation parameter $\rho$ to apply a maximin design? The answer to this question is: No. Once the practitioner knows if there is serial correlation, it does not seem to matter anymore what the value of $\rho$ is. So, the choice of the maximin design only depends on whether or not there is serial correlation.

21

[Table 3 about here.]

## Effect of using of equidistant time points

[Table 4 about here.]

In practice designs with equally-spaced time points are often used. The $q = 4$ equally-spaced design is $\xi_{equal} = \{-1 \quad -0.3333 \ 0.3333 \ 1\}$. In Table 4, the $RE$'s for this equally-spaced design is 0.8373 for $\rho \in S_{\rho(1)}$ and 0.9236 for $\rho \in S_{\rho(2)}$, respectively. Here, again, it can be seen that the $RE$ increases when the autocorrelation parameter corresponding to the uncorrelated case is not considered in our parameter space.

### 6.1 Discussion on Hamilton Depression Scores'example

Hedeker and Gibbons reported the following estimates for a linear mixed model of first-degree $(p = 2)$ with $q = 6$ design points: random intercept variance $\hat{d}_{11} = 11.64$, random slope variance $\hat{d}_{22} = 2.08$, random covariance $\hat{d}_{12} = -1.40$ with $\hat{\sigma}^2 = 12.22$ and autocorrelation coefficient $\hat{\rho} = 0.37$ (Hedeker and Gibbons, 1996).

The generalized variance of the equally spaced designs $\xi_{equal} = \{-1 \quad -0.6 \ -0.2 \ 0.2 \ 0.6 \ 1\}$ can be compared with that of the maximin design $\xi_{MMV} = \{-1 \ -0.4568 \ 0.4568 \ 1\}$ using the relative efficiency measure given in (9). Further, we will assume that the model proposed in Hedeker and Gibbons(1996) as $M_T$. After adjusting the result for the number of repeated measures in designs $\xi_{MMV}$ and $\xi_{equal}$, i.e., four and six, respectively, we find that the relative efficiency of $\xi_{MMV}$ with respect to the $\xi_{equal}$ under the given model parameters is 0.663, i.e., the equally spaced design should be replicated 1.5 times to

22

have the same efficiency as our maximin design. Thus, the maximin design proposed for the Hamilton's depression example is $\xi_{MMV} = \{0\ 9\ 26\ 35\}$ days and the sample size connected to this design has 44 patients place at four time points instead of the 66 patients being measured at six time points.

## 7. Conclusions

In this paper, we have studied the optimality of designs with four types of incorrect specifications in the linear mixed models, namely the order of the polynomial $(p-1)$, the size of the autocorrelation coefficient $\rho$, the number of random effects $k$, and the correlation $\rho^*$ between the random intercept and random slope. Our results are limited to four-points designs for first-, second- and third-degree polynomials. However, the same methodology to find maximin designs can be applied to other misspecifications.

The application of the maximin criterion and the constraint $RE \geq 0.85$ has rendered 97 highly efficient designs. Our main conclusions are that, the maximin $D_q$-optimal design is highly efficient. The maximin $D_q$-optimal design $\xi_{MMV} = \{-1\ -0.4568\ 0.4568\ 1\}$ has a relative efficiency $RE = 0.8727$ in the chosen parameter and model spaces. This maximin design is a $D_q$-optimal design for a first-degree polynomial having covariance structure given by a random intercept and an AR1 structure with autocorrelation $\rho = 0.2$. A list of top-five highest efficient designs that resembles the maximin design very much, is also provided. Moreover, it is shown that the variation of $\min RE$ values decreases as the order of $(p-1)$ increases. Although equally-spaced designs are often used in practice, they are less efficient than the top-five highest efficient designs of our study. It means that when practitioners have budget restrictions or simply a small sample,

23

the maximin designs proposed here could play a deciding role on choosing a design over another and will perform better than equally-spaced designs. Equally-spaced designs will generally be the second best choice under model uncertainties, (cf. Tan and Verger, 1999).

Notice that due to computational constraints, we have chosen few parameter values (2 or 3) for the correlation parameter between the random intercept and random slope variances ($\rho^*$) and the random slope variance ($d_{22}$). From Figure 1(a) one can notice that the optimal designs for $p = 2$ depend strongly on the variance-covariance structure of their models. However, not all the parameters of the variance-covariance structure seem to affect the optimal designs and the relative efficiencies in the same manner. The largest changes on the optimal designs and $RE$s are due to the value of the parameter $\rho$ and the presence/absence of $\rho^*$ and $d_{22}$ rather than the values of the parameters themselves. So, we expect that the use of a finer mesh to discretise the variance-covariance parameters will not influence the results drastically.

Another topic of further research may be to construct highly efficient designs for polynomial models with covariates. Covariates may vary or not over time. It might also be matter of further research to study the effect of using other structures than the AR(1) on the obtained maximin design. Finally, other extensions may be to consider unbalanced data, missing values or drop-out processes.

## Acknowledgements

24

## References

Abt, M., Liski, E.P., Mandal, N.K., and Sinha, B.K. (1997), "Optimal designs in growth curve models: Part I correlated model for linear growth: optimal designs for slope parameter estimation and growth prediction," *Journal of Statistical Planning and Inference*, **64**:141–150.

Abt, M., Gaffke, N., Liski, E.P., and Sinha, B.K. (1988), "Optimal designs in growth curve models: Part II correlated model for quadratic growth: optimal designs for slope parameter estimation and growth prediction," *Journal of Statistical Planning and Inference*, **67**:287–296.

Atkinson A.C., and Donev, A.N. (1996), *Optimum Experimental Designs*. Oxford Science Publications, Oxford.

Chaloner, K., and Verdinelli, I. (1995), "Bayesian experimental design: a review," *Statistical Science*, **10**:273–304.

Berger, M.P.F., Ouwens, M.J.N.M., and Tan, F.E.S. (2002) "Robust designs for longitudinal mixed effects models," In H.Yanai et.al., editor, *Advances in Psychometric Research*, 339–346. Springer Verlag.

Berger, M.P.F., and Tan, F.E.S. (2004), "Robust designs for linear mixed effects models," *Journal of the Royal Statistical Society*, series C, **53**(4):569–581.

Bischoff, W. (1996), "On maximin designs for correlated observations," *Statistics and Probability Letters*, **26**:357–363.

Dette, H., and Neugebauer, H.M. (1997), "Bayesian d-optimal designs for exponential regression models," *Journal of Statistical Planning and Inference*, **60**(2):331–349.

Dette, H. (1997), "Designing experiments with respect to standarized opti-

25

mality criteria," *Journal Royal Statistical Society B*, **59**(1):97–110.

Dette, H., Martinez,I., Ortiz, I.M., and Pepelyshev, A. (2006), "Maximin efficient design of experiment for exponential regression models," *Journal of Statistical Planning and Inference*, **136**(12):4397–4418.

Fedorov, V.V. (1972), *Theory of Optimal Experiments*, Academic Press, New York.

Gram, L.F., Reisby, N., Ibsen, I., Nagy, A., Dencker, S.J., Bech,P., Petersen, G.O., and Christiansen, J. (1976),"Plasma levels and antidepressive effects of imipramine," *Clinical Pharmacological Therapy*, **19**:318–324.

Hedeker, D., Gibbons, R.D. Waternaux, C. and Davis, J.M. (1989), "Investigating drug plasma levels and clinical response using random regression models." *Psychopharmacology Bulletin*, **25**:227–231.

Hedeker, D. and Gibbons, R.D. (1996), "Mixreg: a computer program for mixed-effects regression analysis with autocorrelated errors," *Computer Methods and Programs in Biomedicine*, **49**(3):229-52.

Keller, A., Castellanos, F.X., Vaituzis, A.C., Jeffries, N.O., Giedd, J.N. and Rapport, J.L. (2003),"Progressive loss of cerebellar volume in childhood-onset schizophrenia." *The American Journal of Psychiatry*, **160**:128–133.

Kiefer, J. (1959), "Optimal experimental designs (with discussion)." *Journal of the Royal Statistical Society B*, **21**:272–319.

Kiefer, J. and Wolfowitz, J. (1960), "The equivalence of two extremum problems." *Canadian Journal of Mathematics*, **12**:363–366.

Kiefer, J. and Wolfowitz, J.(1961), "Optimum designs in regression problems." *Annals of Mathematical Statistics*, **30**:271–294.

Kiefer, J. (1974), "General equivalence theory for optimum designs (approx-

imate theory)." *The Annals of Statistics*, **2**(5):849–879.

Lagarias, J.C., Reeds, J.A., Wright, M.H., and Wright, P.E. (1998), "Convergence properties of the nelder-mead simplex method in low dimensions." *SIAM Journal of Optimization*, **9**:112-147.

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

MATLAB (2004), *Handbook MATLAB, version 7.0.1. (R14)*. MathWorks Inc, MA, USA.

Moerbeek, M. (2004), "Robustness properties of A-, D-, and E-optimal designs for polynomial growth models with autocorrelated errors." *Computational Statistics and Data Analysis*, **48**:765–778.

Müller, Ch. H. (1995), "Maximin efficient designs for estimating nonlinear aspects in linear models." *Journal of Statistical Planning and Inference*, **44**:117–132.

Naether, W. (1985), *Effective Observation of Random Fields*, Leipzig: Teubner-Texte.

Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F. (2002), "Maximin D-optimal designs for longitudinal mixed effects models." *Biometrics*, **58**:735–741.

Ouwens, M.J.N.M., Tan, F.E.S. and Berger, M.P.F. (2006), "A maximin criterion for the logistic random intercept model with covariates." *Journal of Statistical Planning and Inference*, **136**(3):962–981.

Pukelsheim, F. (1993), *Optimal Designs of Experiments*, John Wiley & Sons, New York.

Reisby, N. Gram, L.F., Bech, P., Nagy, A., Petersen, G.O., Ortmann,J.,

27

Ibsen, I., Dencker, S.J., Jacobsen,O., Krautwald, O., Sondergaard, I., and Christiansen J. (1977), "Imipramine: Clinical effects and pharmacokinetic variability." *Psychopharmacology*, **54**:263–272.

Tan, F.E.S., and Berger M.P.F. (1999), "Optimal allocation of time points for the random effects models." *Communications in Statistics, Simulations and Computations*, **28**(2):517–540.

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

Wong, W.K. (1994), "Comparing robust properties of A, D, E and G-optimal designs." *Computational Statistics and Data Analysis*, **18**:441–448.

Wynn, H.P. (1970), "The sequential generation of D-optimum experimental designs." *The Annals of Mathematical Statistic*, **41**(5):1655–1664.

28

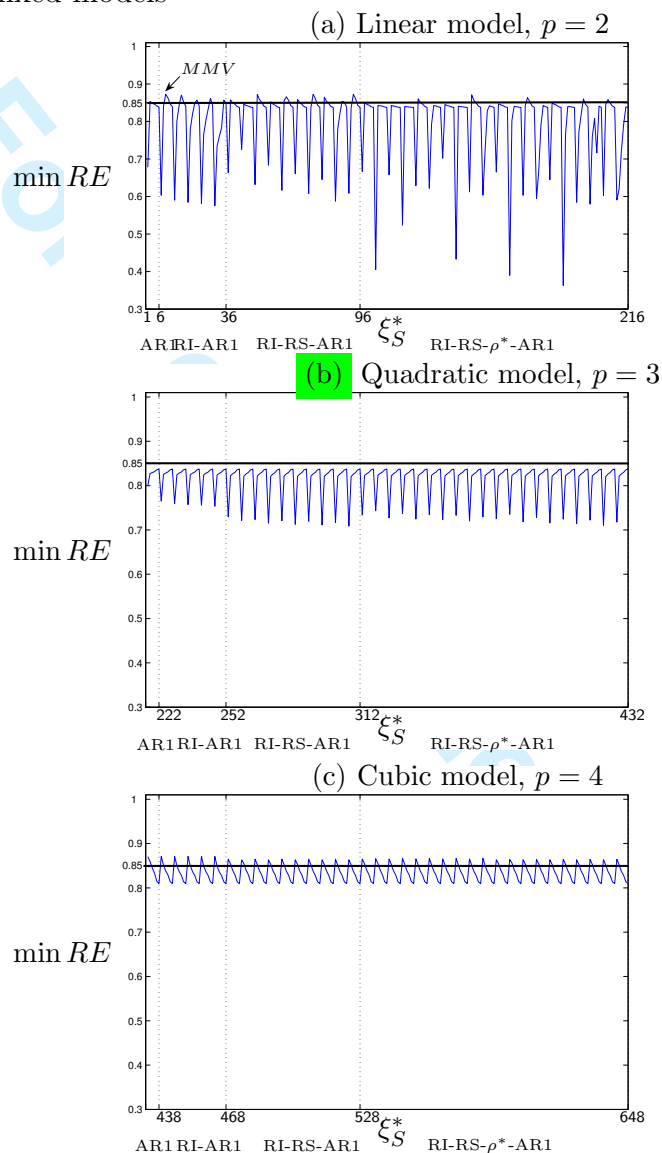**Figure 1.** Minima of relative efficiencies for $D_q$-optimal designs of families of linear mixed models



(a) Linear model, $p = 2$



(b) Quadratic model, $p = 3$



(c) Cubic model, $p = 4$

29

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Table 1**

*Ordering of linear mixed models $M_w$ for the first-order polynomial model,*
*$p = 2$, $q = 4$, $\sigma^2 = 1$, $w = 1, \ldots, 216$.*

| $M_w$ | Model Class | $S_k$ | $S_{d_{11}}$ | $S_{d_{22}}$ | $S_{\rho^*}$ | $S_\rho$ |
|---|---|---|---|---|---|---|
| $M_1$-$M_6$ | AR1 | k=0: $D = \mathbf{0}$ | — | — | — | $\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$ |
| $M_7$-$M_{36}$ | RI-AR1 | k=1: $D = (d_{11})$ | $\{1, 2, 3, 4, 10\}$ | — | — | $\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$ |
| $M_{37}$-$M_{96}$ | RI-RS-AR1 | k=2: $D = \text{Diag}(d_{11}, d_{22})$ | $\{1, 2, 3, 4, 10\}$ | $\left\{ \begin{array}{c} 1 \\ 3 \end{array} \right\}$ | — | $\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$ |
| $M_{97}$-$M_{216}$ | RI-RS-$\rho^*$-AR1 | k=2: $D = \begin{pmatrix} d_{11} & d_{12} \\ d_{12} & d_{22} \end{pmatrix}$ $d_{12} = \rho^*\sqrt{d_{11}d_{22}}$ | $\{1, 2, 3, 4, 10\}$ | $\left\{ \begin{array}{c} 1 \\ 3 \end{array} \right\}$ | $\left\{ \begin{array}{c} 0.5 \\ 0.8 \end{array} \right\}$ | $\left\{ \begin{array}{c} 0.0001 \\ 0.1 \\ 0.2 \\ 0.3 \\ 0.6 \\ 0.9 \end{array} \right\}$ |

30

**Table 2**

*Top 5: $D_q$-optimal designs with highest* $\min RE$'s

| | $D_q$-designs | | | $\min RE$ | $p$ | $\rho$ | $D$ |
|---|---|---|---|---|---|---|---|
| {-1 | -0.4568 | 0.4568 | 1} | 0.8727 | 2 | 0.2 | $\begin{bmatrix} 1 \end{bmatrix}$ |
| {-1 | -0.4567 | 0.4568 | 1} | 0.8726 | 2 | 0.2 | $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ |
| {-1 | -0.4567 | 0.4568 | 1} | 0.8726 | 2 | 0.2 | $\begin{bmatrix} 10 & 0 \\ 0 & 3 \end{bmatrix}$ |
| {-1 | -0.4538 | 0.4538 | 1} | 0.8718 | 2 | 0.1 | $\begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix}$ |
| {-1 | -0.4494 | 0.4494 | 1} | 0.8710 | 4 | 0.0001 | $\begin{bmatrix} 4 \end{bmatrix}$ |

31

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

**Table 3**

*Maximin designs for uncorrelated and correlated cases*

| | $\xi_{MMV}$ | | | | $MMV$ | $p$ | $\rho$ | $D$ |
|---|---|---|---|---|---|---|---|---|
| $S_{\rho(1)}$ | {-1 | -0.4568 | 0.4568 | 1} | 0.8727 | 2 | 0.2 | $\begin{bmatrix} 1 \end{bmatrix}$ |
| $S_{\rho(2)}$ | {-1 | -0.4247 | 0.4247 | 1} | 0.9734 | 2 | 0.3 | $\begin{bmatrix} 1 \end{bmatrix}$ |

32

**Table 4**

*Relative efficiencies RE for equidistant designs $\xi_{equal}$ given uncorrelated and correlated cases*

| | $\xi_{equal}$ | | | | $RE$ | $p$ | $\rho$ | $D$ |
|---|---|---|---|---|---|---|---|---|
| $S_{\rho(1)}$ | {-1 | -0.3333 | 0.3333 | 1} | 0.8373 | 3 | 0.9 | 1 |
| $S_{\rho(2)}$ | {-1 | -0.3333 | 0.3333 | 1} | 0.9236 | 3 | 0.9 | 1 |

33

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

The authors are thankful to the Editor and reviewers for their valuable comments and the time spent on thoroughly reading our paper. Hereby please find enclosed our reactions.

**Responses to Reviewer Comments:**                                    **<Reviewer #1>**

*1.     Please state explicitly what the differences to your former papers (I mean all papers with authors Tan and/or Berger cited in the references) are (with respect to models, covariance structure, …). The paragraph on page 3 is not sufficient to show how much of the work in this manuscript is novel.*

   The reviewer raises a good point. We added a few sentences to meet this concern. On page 3 (new version), lines 11-17:

   "The issues of concern will be
   1. expand the class of models to a third-degree polynomial model,
   2. extend the parameter values of both: the random intercept and the random slope variances, and
   3. show that a discrete approximation of the standardized maximin method proposed by Dette and Neugebauer(1997) and Dette et.al.(2006) will lead to highly efficient designs for models and combinations of parameter values which often occur in practice."

*2. It seems the estimation of variance components is not covered in this manuscript. Could you please comment on this?*

   This crucial assumption was introduced by Näther (1985, p.9).  On page 16 (new version), lines 1-3, we address this issue:

   "In this paper, we have neglected the fact that the parameters of the **V** matrix need to be estimated in practice. The optimal designs are calculated conditional on the parameters of the **V** matrix, which is often done in literature (Nather, 1985, p.9)."

*3. The criterion is not a true maximin criterion in the sense of e.g. Dette and Neugebauer (1997) but a discrete approximation. This should be made clear in the paper.*

   We agree with the reviewer. On page 3, line 14 we added:
   "...show that a discrete approximation of the standardized maximin method ..."

*Also the designs do no longer depend only on the choice of the parameter domains (as is usual for maximin designs) but also on the discretisation. The discretisation seems quite crude in places with only 2 or 3 values from the parameter domain representing the entire set. Do the optimal designs change considerably if finer grids are used?*

   We have addressed this issue on page 24, lines 4-13:

   "Notice that due to computational constraints, we have chosen few parameter values (2 or 3) for the correlation parameter between the random intercept and random slope variances

($\rho^*$) and the random slope variance ($d_{22}$). From Figure 1(a) one can notice that the optimal designs for *p=2* depend strongly on the variance-covariance structure of their models. However, not all the parameters of the variance-covariance structure seem to affect the optimal designs and the relative efficiencies in the same manner. The largest changes on the optimal designs and *RE*s are due to the value of the parameter $\rho$ and the presence/absence $\rho^*$ and $d_{22}$ rather than the value of the parameter itself. So, we expect that the use of a finer mesh to discretise the variance-covariance parameters will not influence the results drastically".

*4. Moreover, the set of eligible designs is highly restricted to locally optimal (4 point) designs with respect to one of the models. What is the reason (apart from computational issues) to admit only locally optimal designs? Why can the designs not have more than 4 support points? These restrictions seem unnecessary and quite restrictive from a design point of view. There could well be designs with more than 4 points that are not locally optimal for any of the models but still outperform all the locally optimal designs in the maximin sense. It is stressed quite often in the manuscript that the maximin optimal designs are highly efficient, but this is of course due to the fact that only this restricted set of designs is admitted into the competition. Please comment on this.*

We agree with the reviewer that the number of design points *(q=4)* looks restrictive. If more time points would be considered (e.g. *q=5*), one may think that there might be designs (connected to first-, second- third- and fourth-degree models, i.e. *p=2, 3, 4* and *5*) that might outperform locally optimal designs from a maximin viewpoint. However, based on our empirical results and those of Tan and Berger (1999), it is expected that the encountered maximin value (RE=0.87) will not change strikingly. The reasoning is as follows. We observed that when *q=4*, the variation of the *min RE* for the first-degree models (*p=2*) is the highest and this variation decreases steadily as *p* approaches or equals *q(=4).* See Figure 1(a-c). Tan and Berger (1999, p. 532-533) have shown that using more designs points than the necessary for a model $M_w$ (i.e. *q>p*) is less efficient for all values of the serial correlation. So, the maximin value, when *q=5, p≤5,* is smaller than *0.85.* The last but not least argument for choosing q=4 has to do with computational time.

*5. Only the AR1 correlation structure is considered. Does it (in most cases) give the best fit to the data in longitudinal studies or is it just the most popular with practitioners?*

A number of approaches have been proposed to model the covariance structure of serial measurements. A commonly used approach for structuring the covariance is the first-order autoregressive AR(1) model (Verbeke and Molenbeghs, p. 152, 2000). This is the main reason of choosing AR(1) in this paper. Other correlation structures could be considered in future research.

We account for this issue on page 24, lines 16-17:

"It might also be matter of further research to study the effect of using other structures than the AR(1) on the obtained maximin design."

*6. Have you thought about using different designs for different subjects?*

We consider briefly this option, but we did not implement it because the complexity that it adds does not pay off in practical terms. In fact, as in question 5, it is not customary to model longitudinal data using different designs for different subjects. Perhaps, clustering subjects having similar designs points was a possibility, but again which criteria will be used to distribute subjects into these clusters with different designs points remained being a problem.

*7. Is there an intuitive explanation why the quadratic and cubic effects are usually assumed as fixed?*

There is mainly a practical reason rather than an intuitive argument for not specifying quadratic and/or cubic random effects. On page 9, lines 3-6, we add:

"Notice that quadratic and cubic-effects are kept as fixed for one reason. In practice, one often only specifies deviations of the average that change linearly in time. Quadratic and cubic effects could also be considered, but its existence and significance might be much harder to show because of lack of power. See Verbeke and Molenberghs, p.70-71, 2000."

*8. Is it restrictive to fix sigma^2=1 (page 10)? It seems that the value of sigma^2 balances the 2 components of V. Would a different choice e.g. sigma^2=10 or 0.1 result in different designs?*

We understand the concern of the reviewer. However, without loss of generality, if a different $\sigma^2$ is used, the maximin design is expected to remain the same. Notice that D-optimal designs are invariant w.r.t. the scale of the designs interval.

*9. How long does it take to run the algorithm on a standard PC?*

We added the following sentences in page 18, lines 3-5:

"The computational time required to obtain our results in a desktop computer with double INTEL processor *x86*, Family *6*, Model *15* having a speed of *1595* MHz is approx. *35* minutes."

*10. (p. 18) "Variation of min RE decreases with increasing p". Is there an intuitive explanation for this? Are locally optimal designs for larger values of p more similar in some sense?*

Unfortunately, we have no intuitive explanation for this phenomenon. On page 19, lines 12-16, we add some sentences to improve this statement:

"The *min RE*s of locally optimal designs $D_q$ show steadily less variation on their *min RE*s values as the number of fixed parameters *p* increases from *2* to *4,* i.e. *p* approaches to the

number of time points *q.* We observed that, under these conditions, the locally optimal design points depend less on the variance-covariance structure and hence, seems to be less affected by its incorrect specification.*"*

Please, see also answer of question 4.

*11. Is there an intuitive explanation for the effect of the correlation parameter rho?*

We lack of an intuitive explanation. We only can state that the effect of the serial correlation parameter on the *RE*s is highly non-linear. As we have mentioned in question 14, once the practitioner knows if there is serial correlation, it does not seem to matter anymore what the value of ρ is.

*12. Is there an intuitive explanation why the D-optimal design for the cubic fixed effects model is close to the maximin design?*

We want to thank the reviewer for making us notice the misleading statement on page 18, lines 33-41 (old version). This sentence is dropped in the new version.

*13. (p.19/20): The MMV is not only high because 2 or 3 point designs are left out of the comparison, but also because the model space {M1, …, MW} is discretised and only a finite number W of designs is admissible.*

The reviewer raises an interesting point because despite the relatively small number of considered parameter values, one can see some trends about the different variance-covariance structures and their corresponding $D_q$ optimal designs. We have tried 419 256 combinations of parameter values and it looks as if the discretisation will not influence the result. Notice that the class of models indicated by the model space is a choice of models that is often encountered in practice. The alternative here would be to find an analytical proof. However, that is extremely difficult.

*14. (p. 20): Leaving out the uncorrelated case leads to a different optimal design and a considerably different efficiency. This shows that the specification of the parameter space (here for rho) is crucial. The maximin designs do not seem to be very robust. Would a finer discretisation help?*

We understand the concern of the reviewer. On page 21, lines 19-25, we address this issue:

"The purpose of studying the effect of the autocorrelation on the chosen maximin design is to answer the question: what-if one has serial correlation. Do we need to know the value of the serial correlation parameter ρ to apply a maximin design? The answer to this question is: No. Once the practitioner knows if there is serial correlation, it does not seem to matter anymore what the value of ρ is. So, the choice of the maximin design only depends on whether or not there is serial correlation."

*(What is the efficiency of the optimal design within S_rho(2) (excluding the uncorrelated case) in the set S_rho(1)?)*

We found that the loss of efficiency of using the maximin design $\xi_{MMV} = \{-1\ -0.4568\ 0.4568\ 1\}$ of $S_{\rho(2)}$ instead of $\xi_{MMV} = \{-1\ -0.4247\ 0.4247\ 1\}$ of $S_{\rho(1)}$ is marginal, i.e. RE=0.9859.

*Is it (this efficiency) better than the efficiency of the equidistant design?*

Yes, this value is higher than those obtained in Table 4.

*15. (p. 21): The efficiencies for the equidistant design are not too bad. How would you convince a practitioner to use maximin designs?*

In this paper we advocate the use of maximin designs instead of the equidistant designs. The loss of efficiency seems indeed not too large, therefore we added a few sentences to support our argument (on the pages 23-24, lines 25 and 1-3, respectively):

"... It means that when practitioners have budget restrictions or simply a small sample, the maximin designs proposed here could play a deciding role on choosing a design over another and will perform better than equally-spaced designs. Equally-spaced designs will generally be the second best choice under model uncertainties (cf. Tan and Berger, 1999)."

*16. (p. 22, line 26 ff): Statement not clear. Of course, the optimal MMV design must have equal weights, since all admissible designs (locally optimal designs) have equal weights. For the AR1 correlation structure, repeated measurements do not make sense (at least in fixed effects models). Or would this be different in mixed effects models due to the fact that the covariance matrix consists of 2 components instead of one, and could so be dominated by the ZDZ' component?*

We agree with the reviewer. We dropped this statement.

-------------------------------------------------------------------------------------------

**<Reviewer #2>**

*2.1. The page 6, the line 31, the number of year,*
(Verbeke and Molenberghs, 2002,
*and in references*
Verbeke, G., and Molenberghs G. (2000),
*See also the page 15, the line 30.*

We agree with the reviewer. On page 6, line 20 (new version), the year is rectified to *2000*. Similarly is done with page 16, line 11.

*2.2. The page 6, the line 47, the corrigendum*
$X = (X_1|\ldots|X_n)'$,
On the page 7, line 3 (new version) is corrected as follows:
$X = (X_1,\ldots,X_n)'$

*2.3. On the page 8 (the lines 26 and 36) identical names (labels) for*

$$\omega = (q, p, \boldsymbol{\beta}, D, \rho, \sigma^2)$$

*and*

$$w = 1, \ldots, W$$ *for is used but the sense of this names (labels) is various. See also the page 10, the line 22, the page 15, the lines 20 and 26.*

We corrected this ambiguity. On the page 8, line 20 (and onwards, i.e. all over the manuscript), the variable

$$w = 1, \ldots, W$$

is used instead of $\omega$. Same holds for $M_w$ .

Notice also that the earlier definition

$$w = 1/q$$

is omitted and the expression (4) of page 8 turns into

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \ldots & x_j & \ldots & x_q \\ 1/q & 1/q & \ldots & 1/q & \ldots & 1/q \end{array} \right\}, \text{where } (1 \leq j \leq q). \quad (4)$$

*2.3. The Page 9, the line 13, a point after the formula is absent*

$$\min_{\xi \in \Xi} \Phi[\underset{M_\omega}{\mathrm{Var}}(\hat{\beta})],$$

On the page 9, a point replaces this misused comma.

*2.4. The Page 9, the line 26, more in detail to present expression for*

$$X'_\xi V^{-1}_\xi X_\xi$$

On the page 9, the equation (7) is improved as follows:

$$\mathrm{Det}\{\underset{M_w}{\mathrm{Var}}(\hat{\beta}|\xi)\} = \mathrm{Det}\left(X'_\xi V^{-1}_\xi X_\xi\right)^{-1},$$
$$= \mathrm{Det}\{X'_\xi \left(Z_\xi D Z'_\xi + \sigma^2 \Psi_\xi\right)^{-1} X_\xi\}^{-1}, \quad (7)$$

*2.5. The Page 9, the line 40, see the text*

$$\hat{\beta} \text{ is the parameter estimator}$$

On the page 10, line 4: we adjusted the text to

$$\hat{\beta} \text{ is the vector of the estimated parameters}$$

*2.6. The Page 11, the line 22, the formula (9), to unify labels in*

$$\left[ \frac{\mathrm{Det}\{\underset{M_T}{\mathrm{Var}}(\hat{\beta}|\xi^*_T)\}}{\mathrm{Det}\{\underset{M_T}{\mathrm{Var}}(\hat{\beta}|\xi^*_S)\}} \right]$$

*and*

$$\left\{ \frac{\mathrm{Det}(\mathrm{X}'_{\xi_T^*} \hat{\mathrm{V}}^{-1}_{\xi_T^*} \mathrm{X}_{\xi_T^*})^{-1}}{\mathrm{Det}(\mathrm{X}'_{\xi_S^*} \hat{\mathrm{V}}^{-1}_{\xi_S^*} \mathrm{X}_{\xi_S^*})^{-1}} \right\}$$

*See also the formula (8) and the page 12, the line 45*

$$\mathrm{Det}\{\underset{M_w}{\mathrm{Var}}(\hat{\beta}|\xi_q^*)\} \le \mathrm{Det}\{\underset{M_w}{\mathrm{Var}}(\hat{\beta}|\xi_q)\},$$

The reviewer raises a good point. We added a few sentences to meet this concern, see the page 13, lines 16-19:

In general, notice that, we will refer to $D_q$ designs as the local optimum design if Model $M_w$ where $w = 1, \ldots, W$ is used. Later on, the reader will notice that the subindex $w$ is used to point to a true model $M_T$ with optimal design $\xi_T^*$ or to the selected model $M_S$ with optimal design $\xi_S^*$.

*2.7. The Page 15, the lines 4 and 6, the parentheses*

$D-$optimal designs are invariant with respect to the sign of $d_{12}$ Ouwens et.al. (2002). The correlation $\rho^*$ is chosen from a set $S_{\rho^*} = \{0.5, 0.8\}$

On the page 15, lines 21-22, the reference is been rectified.

*2.8. The Page 19, the line 16, to unify labels for intervals*

$[0.1 - 0.2],$

We agree with the reviewer. In the manuscript, intervals have been unified, on page 19, line 23, e.g.

$[0.1, 0.2],$

*2.9. The page 19, the lines 50, 52 (and other places of the manuscript), to unify labels of the designs of experiments*

$\xi^* = (-1 \quad 1)$

*Compare to the label on the page 8, the formula (4)*

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \ldots & x_j & \ldots & x_q \\ w & w & \ldots & w & \ldots & w \end{array} \right\}$$

The reviewer is very right. All over the manuscript, designs of experiments are represented by ξ and *curly brackets,* i.e.

$$\xi = \left\{ \begin{array}{cccccc} x_1 & x_2 & \ldots & x_j & \ldots & x_q \\ 1/q & 1/q & \ldots & 1/q & \ldots & 1/q \end{array} \right\}, \text{where } (1 \le j \le q). \quad (4)$$

The optimal design is denoted by ξ*, curly brackets and weights are omitted, since they are all the same, e.g. on page 20, line 17:

$\xi^* = \{-1 \quad 1\}$

*2.10. The page 21, the lines 34 and 36, the parentheses*

$\hat{d}_{12} = -1.40$ with $\hat{\sigma}^2 = 12.22$ and autocorrelation coefficient $\hat{\rho} = 0.37$ Hedeker

and Gibbons (1996).

We agree with the reviewer. The text of page 22, line 15 should reads as follows:
(Hedeker and Gibbons, 1996)

*2.11. The page 24, the line 17*

growth curve models: Part ii

*On t*he page 25, the line 7 is corrected

growth curve models:  Part II

*2.12. The page 24, the lines 12, 21, 42, and others, to unify labels in the list of the literature (references)*

64:141–150.

10, 273–304.

We agree with the reviewer. In the manuscript, the list of references has been unified, i.e.

64:141–150.

10:273–304.

*2.13. The page 25, the line 43, there are no numbers of the pages*

Methods and Programs in Biomedicine, 49.

The reviewer points out an important gap. The reference is corrected on page 26, line 15 as follows:

Methods and Programs in Biomedicine, 49(3):229-52.

*2.14. The page 28, the line 22, to replace "a" to "b"*

(a) Quadratic model, $p = 3$

On the page 29, the correct label is

(b) Quadratic model, $p = 3$

-----------------------------------------------

We do hope that the amendments made in the paper honor the valuable comments done by the reviewers.