Optimal Designs for Individual Prediction in Random Coefficient Regression Models

Maryna Prus*, Rainer Schwabe

Otto-von-Guericke University Magdeburg, Institute for Mathematical Stochastics, PF 4120, D-39016 Magdeburg, Germany

Abstract: In this note we present optimal designs for the estimation of the individual response ("prediction") for the participates in a study within the framework of hierarchical linear mixed models. These optimal designs may differ substantially from those propagated in the literature so far.

Keywords: Linear Mixed Model, Random Coefficient Regression, Optimal Design, Prediction.

1 Introduction

Random coefficient regression models attract an increasing popularity in many fields of applications starting from animal breeding over population pharmacokinetics to the development of individualized medicine. In random coefficient regression models interest may be either in a "typical" response described by some averaged characteristics of the population (estimation of the population parameters) or in the individual responses of the subjects involved in the study themselves (estimation of the individual parameters, "prediction"). The statistical analysis of such models has become tractable during the last years by the now available computer facilities, which pushes this field forward. However, less has been done in optimal design for such experiments. In the present note we want to develop optimal designs for the estimation of the individual response.

In their seminal paper Gladitz and Pilz (1982) established that Bayesian optimal designs are optimal for the estimation of individual responses, when the prior covariance is set equal to the covariance of the individual parameters, in the case of known population parameters, i. e. in the case of the knowledge of the "typical" response. Subsequently this last condition of known population parameters has been often overlooked, which resulted in some kind of "folklore" in the design community that all design problems have been solved related to linear mixed models. Some years later Fedorov and Hackl (1997) (section 5.2) considered design criteria for either the estimation of the population parameters or the estimation (prediction) of the individual responses under various assumptions on the knowledge of parts of the parameters. For the situation of prediction, when the population parameters are unknown, they claim that the optimal design in the corresponding model without random effects retains its optimality for the estimation of the individual parameters. This statement seems to be motivated by a conditional approach, in which the individual response is estimated by the observations of the corresponding individual only, neglecting the fact that the individual effects are assumed to be random and stem from a common population. As often the truth lies somewhere in between: As will be developed in this note the correct characterization of the mean squared error for the prediction of the individual response results in a compound criterion, which is a weighted average of the fully Bayesian criterion and the criterion related to the model without random effects.

The present paper is organized as follows: In the second section we specify the model, introduce the predictor for the individual random parameters and develop the corresponding mean squared error matrix. The third section provides some theoretical results for the determination of

*Maryna Prus: maryna.prus@ovgu.de

designs, which are optimal for prediction. In section 4 we illustrate these results by the example of simple straight line regression and conclude with some discussion in section 5.

2 Model Specification and Prediction

In random coefficient regression the observations are assumed to come from a hierarchical model. On the individual level the jth observation Y_{ij} of individual i is given by

$$Y_{ij} = \mathbf{f}(x_{ij})^\top \boldsymbol{\beta}_i + \varepsilon_{ij} \tag{1}$$

for $j = 1, .., m_i$ and $i = 1, .., n$, where m_i is the number of observations at individual i, n is the number of individuals, $\mathbf{f} = (f_1, .., f_p)^\top$ is a set of known regression functions, and the experimental settings x_{ij} may come from the experimental region $\mathcal{X}$. The observational errors ε_{ij} are assumed to be homoscedastic and uncorrelated with mean 0 and common variance $\sigma^2 > 0$.

On the population level the individual parameters $\boldsymbol{\beta}_i = (\beta_{i1}, .., \beta_{ip})^\top$ are assumed to come from common distribution with unknown population mean $\mathrm{E}\,(\boldsymbol{\beta}_i) = \boldsymbol{\beta} = (\beta_1, .., \beta_p)^\top$ and a $p \times p$ population covariance matrix $\mathrm{Cov}\,(\boldsymbol{\beta}_i) = \sigma^2 \mathbf{D}$. Moreover, all individual parameters $\boldsymbol{\beta}_i$ and all observational errors $\varepsilon_{i'j}$ are assumed to be uncorrelated.

To simplify the notations we will only consider random coefficient regression models throughout this note, in which all individuals get the same experimental treatments, i. e. all individuals i have the same number $m_i = m$ of observations at the same experimental settings $x_{ij} = x_j$.

To make use of the theoretical results available for the estimation of individual parameters (prediction) we will identify the above specified model as a special case of the general mixed model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{2}$$

with a particular nice structure of the fixed effects and random effects design matrices $\mathbf{X}$ and $\mathbf{Z}$, respectively. Here $\boldsymbol{\beta}$ denotes again the fixed effect (population parameter), and $\boldsymbol{\gamma}$ are the random effects. These random effects and the observational errors $\boldsymbol{\varepsilon}$ have zero mean and are all uncorrelated with corresponding full rank covariance matrices $\mathrm{Cov}\,(\boldsymbol{\gamma}) = \mathbf{G}$ and $\mathrm{Cov}\,(\boldsymbol{\varepsilon}) = \mathbf{R}$, respectively.

Under the assumption of Gaussian normal distributions Henderson *et al.* (1959) established that the mixed model equations produce the Best Linear Unbiased Estimator $\hat{\boldsymbol{\beta}}$ and the Best Linear Unbiased Predictor $\hat{\boldsymbol{\gamma}}$ for $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$ by

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Y} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Y} \end{pmatrix}, \tag{3}$$

if the fixed effects design matrix $\mathbf{X}$ has full column rank. According to Christensen (2002) the distributional assumptions can be relaxed to the moment conditions stated above.

Denote by

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^\top & \mathbf{C}_{22} \end{pmatrix} = \mathrm{Cov} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{pmatrix} \tag{4}$$

the joint mean squared error matrix for both $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$, which is partitioned according to these components. Then Henderson (1975) has shown that

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{X}^\top\mathbf{R}^{-1}\mathbf{Z} \\ \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{X} & \mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1}. \tag{5}$$

We now adapt our model (1) to the more general model (2). First we introduce the centered random effects $\mathbf{b}_i := \boldsymbol{\beta}_i - \boldsymbol{\beta}$. Then a single observation in model (1) can be written as

$$Y_{ij} = \mathbf{f}(x_j)^\top \boldsymbol{\beta} + \mathbf{f}(x_j)^\top \mathbf{b}_i + \varepsilon_{ij}\,, \tag{6}$$

where the random effects are separated from the population mean. This results in the vector notation

$$\mathbf{Y}_i = \mathbf{F}\boldsymbol{\beta} + \mathbf{F}\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \tag{7}$$

for all observations $\mathbf{Y}_i = (Y_{i1}, ..., Y_{im})^\top$ of individual i, where $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{im})^\top$ denotes the corresponding error vector and $\mathbf{F} = (\mathbf{f}(x_1), ..., \mathbf{f}(x_m))^\top$ is the individual design matrix, which coincides for all individuals under the assumptions made.

Finally, the complete observation vector $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_n^\top)^\top$ for all individuals has the form

$$\mathbf{Y} = (\mathbf{1}_n \otimes \mathbf{F})\boldsymbol{\beta} + (\mathbf{I}_n \otimes \mathbf{F})\mathbf{b} + \boldsymbol{\varepsilon}, \tag{8}$$

where $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, ..., \boldsymbol{\varepsilon}_n^\top)^\top$ is the vector of all observational errors and $\mathbf{b} = (\mathbf{b}_1^\top, ..., \mathbf{b}_n^\top)^\top$ is the common vector of random effects. Here $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{1}_n = (1, .., 1)^\top$ is a n-dimensional vector with all entries equal to 1, and "$\otimes$" denotes the Kronecker product. Hence, model (8) attains the form of the general model (2). However, the covariance matrix $\mathrm{Cov}\,(\mathbf{b}) = \sigma^2 \mathbf{I}_n \otimes \mathbf{D}$ may fail to be of full rank, if the population covariance matrix $\mathbf{D}$ is singular, which is the case for example for models with random intercepts (random block effects).

Therefore we have to rewrite the model (8) by reducing the dimensionality of the random effects such that the corresponding covariance matrix becomes regular: Let q be the rank of the matrix $\mathbf{D}$. Then there exists a $p \times q$ matrix $\mathbf{H}$ with $\mathbf{D} = \mathbf{H}\mathbf{H}^\top$ and $\mathrm{rank}\,(\mathbf{H}) = q$ such that $\mathbf{H}^\top\mathbf{H}$ is regular. Then we introduce the random variables

$$\mathbf{c}_i := (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{b}_i\,, \tag{9}$$

which satisfy $\mathrm{E}\,(\mathbf{c}_i) = 0$, $\mathrm{Cov}\,(\mathbf{c}_i) = \sigma^2\mathbf{I}_q$, and $\mathbf{b}_i = \mathbf{H}\mathbf{c}_i$ almost surely. The model equation (7) can then be written as

$$\mathbf{Y}_i = \mathbf{F}\boldsymbol{\beta} + \mathbf{F}\mathbf{H}\mathbf{c}_i + \boldsymbol{\varepsilon}_i\,, \tag{10}$$

and consequently

$$\mathbf{Y} = (\mathbf{1}_n \otimes \mathbf{F})\boldsymbol{\beta} + (\mathbf{I}_n \otimes (\mathbf{F}\mathbf{H}))\mathbf{c} + \boldsymbol{\varepsilon}\,, \tag{11}$$

where $\mathbf{c} = (\mathbf{c}_1^\top, .., \mathbf{c}_n^\top)^\top$. Now with $\mathbf{X} = \mathbf{1}_n \otimes \mathbf{F}$, $\mathbf{Z} = \mathbf{I}_n \otimes (\mathbf{F}\mathbf{H})$ and $\boldsymbol{\gamma} = \mathbf{c}$ our model (11) has the form of the general model (2) with regular covariance matrices $\mathbf{R} = \sigma^2\mathbf{I}_{nm}$ and $\mathbf{G} = \sigma^2\mathbf{I}_{nq}$.

In this model the blocks of the mean squared error matrix $\mathbf{C}$ can be readily calculated as

$$\mathbf{C}_{11} = \mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{n}\left((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{H}\mathbf{H}^\top\right), \tag{12}$$

$$\mathbf{C}_{22} = \mathrm{Cov}(\hat{\mathbf{c}} - \mathbf{c}) = \sigma^2\left(\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{I}_q + (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{H}^\top\mathbf{F}^\top\mathbf{F}\mathbf{H} + \mathbf{I}_q)^{-1}\right), \tag{13}$$

$$\mathbf{C}_{12} = \mathrm{Cov}(\hat{\boldsymbol{\beta}}, \hat{\mathbf{c}} - \mathbf{c}) = -\frac{\sigma^2}{n}\mathbf{1}_n^\top \otimes \mathbf{H}\,. \tag{14}$$

Note that in the present case of identical design matrices $\mathbf{F}$ for all individuals the estimator $\hat{\boldsymbol{\beta}}$ for the population parameter $\boldsymbol{\beta}$ can be calculated as the ordinary least squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\bar{\mathbf{Y}}$, where $\bar{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^n \mathbf{Y}_i$ is the averaged response across the individuals. The estimator $\hat{\boldsymbol{\beta}}$ for the population parameter $\boldsymbol{\beta}$ can also be rewritten as the average $\hat{\boldsymbol{\beta}} = \frac{1}{n}\sum_{i=1}^n \hat{\boldsymbol{\beta}}_{i;\mathrm{ind}}$ of the individualized estimates $\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{Y}_i$ of the individual parameters $\boldsymbol{\beta}_i$ based on the observations $\mathbf{Y}_i$ of subject i only (see Entholzner *et al.* (2005)).

Also the predictor $\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i = \hat{\boldsymbol{\beta}} + \mathbf{H}\hat{\mathbf{c}}_i$ for the individual parameter $\boldsymbol{\beta}_i$ simplifies in this situation.

Theorem 1. *In the case of identical design matrices* $\mathbf{F}$ *for all individuals the predictor*

$$\hat{\boldsymbol{\beta}}_i = \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} + (\mathbf{F}^\top\mathbf{F})^{-1}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}, \tag{15}$$

is a weighted average of the individualized estimate $\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}}$ *based on the observations of subject* i *only and the estimator* $\hat{\boldsymbol{\beta}}$ *for the population parameter.*

Consequently the corresponding covariance matrix of

$$\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n \end{pmatrix} = (\mathbf{1}_n \otimes \mathbf{I}_p \mid \mathbf{I}_n \otimes \mathbf{H}) \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{c}} - \mathbf{c} \end{pmatrix}, \tag{16}$$

is independent of the choice of $\mathbf{H}$.

Theorem 2.

$$\begin{aligned} &\mathrm{Cov}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n \end{pmatrix} \\ &= \sigma^2 \left(\left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \right) \otimes \left(\mathbf{D} - \mathbf{D}((\mathbf{F}^\top \mathbf{F})^{-1} + \mathbf{D})^{-1} \mathbf{D} \right) + \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top \otimes (\mathbf{F}^\top \mathbf{F})^{-1} \right). \end{aligned} \tag{17}$$

If the dispersion matrix $\mathbf{D}$ is regular, the mean squared error matrix for the predictors $(\hat{\boldsymbol{\beta}}_1^\top, ..., \hat{\boldsymbol{\beta}}_n^\top)^\top$ simplifies to a weighted average of the covariance matrix $\sigma^2(\mathbf{F}^\top\mathbf{F})^{-1}$ for the fixed effects in the model without random effects ($\mathbf{D} = \mathbf{0}$) and the Bayesian covariance matrix $\sigma^2(\mathbf{F}^\top\mathbf{F} + \mathbf{D}^{-1})^{-1}$ propagated by Gladitz and Pilz (1982).

Corollary 1. *If* $\mathbf{D}$ *is regular, then*

$$\mathrm{Cov}\begin{pmatrix} \hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1 \\ \vdots \\ \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_n \end{pmatrix} = \sigma^2 \left(\left(\mathbf{I}_\mathrm{n} - \frac{1}{\mathrm{n}} \mathbf{1}_\mathrm{n} \mathbf{1}_\mathrm{n}^\top \right) \otimes (\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1})^{-1} + \frac{1}{\mathrm{n}} \left(\mathbf{1}_\mathrm{n} \mathbf{1}_\mathrm{n}^\top \right) \otimes (\mathbf{F}^\top \mathbf{F})^{-1} \right). \tag{18}$$

3 Optimal Design

The performance of the prediction may be measured in terms of the mean squared error matrix derived in Theorem 2 and Corollary 1 and its quality depends on the design of the experiment, i. e. on the choice of the experimental settings $x_1, ..., x_m$ for each individual. These experimental settings need not be distinct. As the performance does not depend on the order of the observations within the individuals, we may rewrite the mean squared error matrix of the predictor in terms of distinct settings $x_1, ..., x_k$, say, and their respective numbers $m_1, ..., m_k$ of replications ($\sum_{j=1}^k m_j = m$). The standardized individual design $\xi = \begin{pmatrix} x_1, ..., x_k \\ w_1, ..., w_k \end{pmatrix}$ is then introduced, where x_j are the distinct settings and $w_j = \frac{m_j}{m}$ the corresponding relative frequencies of replications, respectively, satisfying $\sum_{j=1}^k w_j = 1$. For such a standardized design ξ we denote by

$$\mathbf{M}(\xi) := \sum_{j=1}^k w_j \mathbf{f}(x_j)\mathbf{f}(x_j)^\top = \frac{1}{m}\mathbf{F}^\top\mathbf{F} \tag{19}$$

the standardized information matrix in the model without random effects. Further let $\boldsymbol{\Delta} = m\mathbf{D}$ be a standardized version of the dispersion matrix for the random effects.

In the present paper we will lay emphasis on the Integrated Mean Squared Error ("IMSE") criterion

$$\mathrm{IMSE}_{\mathrm{pred}}(\xi) := \int \mathrm{E}_{\xi}\left(\sum_{i=1}^{n}(\hat{\mu}_i(x)-\mu_i(x))^2\right)\nu(\mathrm{d}x) = \sum_{i=1}^{n}\int \mathrm{Cov}_{\xi}\left(\hat{\mu}_i(x)-\mu_i(x)\right)\nu(\mathrm{d}x) \tag{20}$$

for prediction, which measures the average distance of the predicted individual response $\hat{\mu}_i(x) = \mathbf{f}(x)^\top\hat{\boldsymbol{\beta}}_i$ from the true individual response $\mu_i(x) = \mathbf{f}(x)^\top\boldsymbol{\beta}_i$. E_ξ and Cov_ξ denote the expectation and the covariance matrix, when the design ξ is used and ν is a weight distribution on the design region $\mathcal{X}$, which is typically uniform on $\mathcal{X}$ with $\nu(\mathcal{X}) = 1$.

After some rearrangements it can be seen that the IMSE-criterion for prediction is linear, and by Theorem 2 it follows that

$$\mathrm{IMSE}_{\mathrm{pred}}(\xi) = \frac{\sigma^2}{m}\,\mathrm{tr}\left(\left(\mathbf{M}(\xi)^{-1} + (n-1)\left(\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1}+\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\right)\right)\mathbf{V}\right), \tag{21}$$

where $\mathbf{V} := \int \mathbf{f}(x)\mathbf{f}(x)^\top\nu(\mathrm{d}x)$ is the "information matrix" for the weight distribution ν in the model without random effects and "tr" denotes the trace of a matrix.

For a regular dispersion matrix $\mathbf{D}$ the criterion (21) simplifies by Corollary 1 to

$$\mathrm{IMSE}_{\mathrm{pred}}(\xi) = \frac{\sigma^2}{m}\,\mathrm{tr}\left(\left(\mathbf{M}(\xi)^{-1} + (n-1)(\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}\right)\mathbf{V}\right), \tag{22}$$

which is proportional to a weighted average of the IMSE-criterion $(\mathbf{M}(\xi)^{-1}\mathbf{V})$ in the model without random effects and the corresponding Bayesian IMSE-criterion $\mathrm{tr}\left((\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}\mathbf{V}\right)$. Hence the IMSE-criterion for prediction can be interpreted as a compound criterion, which could equivalently be identified with a certain constrained criterion according to Cook and Wong (1994).

If we allow for approximate designs in the sense of Kiefer (1974), then a standard equivalence theorem can be obtained, which characterizes the IMSE-optimal design, which minimizes the IMSE-criterion.

Theorem 3. *The approximate design ξ^* is IMSE-optimal for prediction, if and only if*

$$\begin{aligned}
&(n-1)\,\mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\left(\mathbf{M}(\xi^*)^{-1}+\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}\mathbf{V}\boldsymbol{\Delta}\left(\mathbf{M}(\xi^*)^{-1}+\boldsymbol{\Delta}\right)^{-1}\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x)\\
&\qquad + \mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\mathbf{V}\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x)\\
&\leq\; (n-1)\,\mathrm{tr}\left(\boldsymbol{\Delta}\left(\mathbf{M}(\xi^*)^{-1}+\boldsymbol{\Delta}\right)^{-1}\mathbf{M}(\xi^*)^{-1}\left(\mathbf{M}(\xi^*)^{-1}+\boldsymbol{\Delta}\right)^{-1}\boldsymbol{\Delta}\mathbf{V}\right) + \mathrm{tr}\left(\mathbf{M}(\xi^*)^{-1}\mathbf{V}\right)
\end{aligned} \tag{23}$$

for all $x \in \mathcal{X}$. Moreover, for any support point x_j of ξ^* with positive weight ($w_j = \xi^*(x_j) > 0$) equality holds in (23).

For regular dispersion matrices $\mathbf{D}$ condition (23) of Theorem 3 simplifies.

Corollary 2. *If $\mathbf{D}$ is regular, the approximate design ξ^* is IMSE-optimal for prediction if and only if*

$$\begin{aligned}
&\mathbf{f}(x)^\top\left((n-1)(\mathbf{M}(\xi^*)+\boldsymbol{\Delta}^{-1})^{-1}\mathbf{V}(\mathbf{M}(\xi^*)+\boldsymbol{\Delta}^{-1})^{-1} + \mathbf{M}(\xi^*)^{-1}\mathbf{V}\mathbf{M}(\xi^*)^{-1}\right)\mathbf{f}(x)\\
&\leq\; \mathrm{tr}\left(\left((n-1)(\mathbf{M}(\xi^*)+\boldsymbol{\Delta}^{-1})^{-1}\mathbf{M}(\xi^*)(\mathbf{M}(\xi^*)+\boldsymbol{\Delta}^{-1})^{-1} + \mathbf{M}(\xi^*)^{-1}\right)\mathbf{V}\right)
\end{aligned} \tag{24}$$

for all $x \in \mathcal{X}$. Moreover, for any support point x_j of ξ^* with positive weight ($w_j = \xi^*(x_j) > 0$) equality holds in (24).

Remark 1. Consider the special case of random intercepts (random block effects). There an explicit individual specific constant term is included, $f_1(x) \equiv 1$, say. The dispersion matrix $\mathbf{D} = d_1\mathbf{e}_1\mathbf{e}_1^\top$ has rank one, where $\mathbf{e}_1^\top = (1, 0, ..., 0)$ is the first unit vector of length p. In this situation the IMSE-criterion

$$\mathrm{IMSE}_{\mathrm{pred}}(\xi) = \sigma^2 \left(\frac{1}{m} \operatorname{tr} \left(\mathbf{M}(\xi)^{-1}\mathbf{V} \right) + \frac{(n-1)d_1}{1 + m\, d_1} \nu(\mathcal{X}) \right), \tag{25}$$

depends on the dispersion matrix only through an additive constant. Hence, the IMSE-optimal design in the fixed effects model remains IMSE-optimal for prediction in the random intercepts model.

4 Example: Straight Line Regression

For illustrative purposes we consider here the model

$$Y_{ij} = \beta_{i1} + \beta_{i2}x_j + \varepsilon_{ij}\,, \tag{26}$$

of a straight line regression on the experimental region $\mathcal{X} = [0, 1]$. The experimental setting x_j may be considered as a dosage and $x_j = 0$ yields an observation at baseline.

In the case of no random effects it is well-known that the IMSE-optimal design assigns half of the observations to the baseline and half of the observations to the maximal dose. According to Remark 1 this design retains its optimality for prediction in the case of random intercepts β_{i1}. However, this does not remain true, if we allow for random slopes β_{i2}. To keep calculations simple we assume that only the slope is random, i. e. $\beta_{i1} \equiv \beta_1$ and $\mathbf{D} = d_2\mathbf{e}_2\mathbf{e}_2^\top$. In view of the equivalence theorem (Theorem 1) the IMSE-optimal design ξ^* takes only observations at the baseline ($x = 0$) and at the highest dose ($x = 1$) and has hence the form $\xi^* = \begin{pmatrix} 0 & 1 \\ 1 - w^* & w^* \end{pmatrix}$, where only the optimal weight w^* at $x = 1$ has to be determined.

For numerical calculations we assume a sample size of $n = 100$ individuals and an intra-individual sample size of $m = 10$ observations per subject. For these numbers the optimal weight w^* is depicted in Figure 4 in dependence on the "intra-class correlation" $\rho = d_2/(1 + d_2)$ at $x = 1$. Here and in the following figure ρ is chosen instead of d_2 on the horizontal axis to cover the whole range of variances d_2 in a finite interval ($0 \le \rho \le 1$). The optimal weight w^* increases in d_2 from $w^* = 0.50$ for $d_2 = 0$ ($\rho = 0$, no random effects) to $w^* \approx 0.91$ for $d_2 \to \infty$ ($\rho = 1$).

To see, what may be gained by using the optimal design for prediction, we have plotted the efficiency of the IMSE-optimal design in the fixed effects model, which assigns equal weights $w = 1/2$ to both the baseline $x = 0$ and the maximal dose $x = 1$, for varying dispersions d_2 in Figure 4. Of course, for $d_2 = 0$ ($\rho = 0$, no random effects) the efficiency equals 1. The efficiency decreases, as the dispersion d_2 increases, with a limiting value 0.60 for $d_2 \to \infty$ ($\rho = 1$).

These pictures do not change much, if we introduce additionally a random intercept with small dispersion d_1, which makes $\mathbf{D}$ regular. The Bayes-optimal design in this situation, which minimizes $\operatorname{tr}\left((\mathbf{F}^\top\mathbf{F} + \mathbf{D}^{-1})^{-1}\right)$, assigns all observations to the maximal dose $x = 1$ and does, hence, not allow for prediction, as the fixed effects information $\mathbf{F}^\top\mathbf{F}$ is singular. Consequently in this situation the efficiency of the Bayes-optimal design equals zero.

5 Discussion and Conclusions

In the present note we develop the mean squared error for the estimation (prediction) of the individual responses as a performance characteristic of the design of an experiment. The resulting

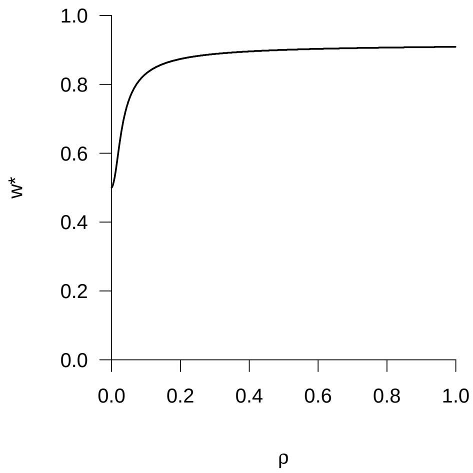

Figure 1: Optimal weight w^* in dependence of the variance parameter $\rho = d_2/(1 + d_2)$

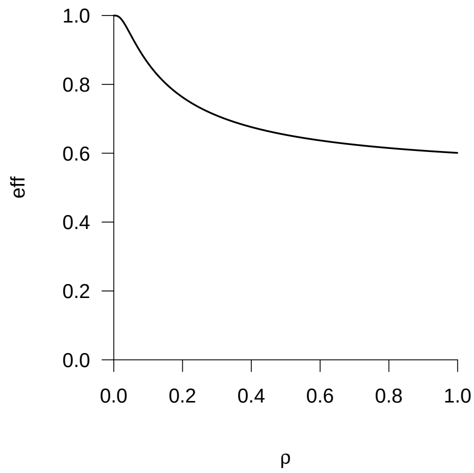

Figure 2: Efficiency of the equi-replicated design ($w = \frac{1}{2}$) in dependence of the variance parameter $\rho = d_2/(1 + d_2)$

objective function is a compromise between the Bayesian and the standard fixed effects objective function proposed so far and accounts for both the information from the observations within an individual as well as the information obtained from the other subjects within the population. In particular, the Bayesian criterion, which in a way neglects the intra-individual information and takes the population information for granted, may lead to useless designs, which do not allow for the estimation of the quantities of interest. As a specific design criterion under investigation we use here the integrated mean squared error for approximating the individual response curves over the whole design region, which is a quite natural choice. For this criterion we derive some characterizations of optimal designs, which may substantially differ from the competing designs obtained by maximizing the corresponding Bayesian or standard criteria. As a by-product we obtain that in the particular case of random intercepts the standard optimal designs retain their optimality. For computational ease we only considered approximate individual designs, which might be criticized, as typically sample sizes within individuals will be small. However these optimal approximate designs may serve well as a benchmark for candidate designs, which themselves can be derived from the optimal approximate designs by intelligent rounding. But this goes beyond the scope of the present paper. Finally we illustrate the dependence of the optimal design on the dispersion parameters as well as the efficiency of the competing designs by a simple example. This gives rise to the problem that the obtained optimal designs are only locally optimal for a specific dispersion. Robustness and sensitivity with respect to the dispersion should be investigated in the future and potentially more robust designs should be developed, which are maximin efficient or optimal with respect to some weighted (“pseudo Bayesian”) criterion. Furthermore, it is to be expected that the qualitative results obtained here will carry over to other design criteria, which will be object of further investigations.

Acknowledgment

This research was partially supported by grant SKAVOE 03SCPAB3 of the German Federal Ministry of Education and Research (BMBF). Part of the work was done, while the second author was visiting the International Newton Institute in Cambridge.

References

Christensen, R. (2002). *Plane Answers to Complex Questions: The Theory of Linear Models.* New York, Springer.

Cook, R. D. and Wong, W. K. (1994). On the equivalence of constrainer and compound designs. *Journal of the American Statistical Association*, **89**, 687–692.

Entholzner, M., Benda, N., Schmelter, T., and Schwabe, R. (2005). A note on designs for estimating population parameters. *Biometrical Letters - Listy Biometryczne*, **42**, 25–41.

Fedorov, V. and Hackl, P. (1997). *Model-Oriented Design of Experiments.* New York, Springer.

Gladitz, J. and Pilz, J. (1982). Construction of optimal designs in random oefficient regression models. *Mathematische Operationsforschung und Statistik, Series Statistics*, **13**, 371–385.

Henderson, C. R. (1975). Best linear unbiased estimation and prediction under a selection model. *Biometrics*, **31**, 423–477.

Henderson, C. R., Kempthorne, O., Searle, S. R., and von Krosigk, C. M. (1959). The estimation of environmental and genetic trends from records subject to culling. *Biometrics*, **15**, 192–218.

Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Annals of Statistics*, **2**, 849–879.