# Optimal Designs for the Prediction of Individual Effects in Random Coefficient Regression

Maryna Prus and Rainer Schwabe

**Abstract** In this note we propose optimal designs for the prediction of the individual responses as well as for the individual deviations from the population mean response in random coefficient models. In this situation the mean population parameters are assumed to be unknown such that the performance measures of the prediction do not coincide for both objectives and, hence, the design optimization lead to substantially diverse results. For simplicity we consider the case, where all individuals are treated in the same way. If the population parameters were known, Bayesian optimal designs would be optimal (Gladitz and Pilz,1982). While the optimal design for the prediction of the individual responses differ from the Bayesian optimal design propagated in the literature (Prus and Schwabe, 2011), the latter designs remain their optimality, if only the individual deviations from the mean response are of interest.

## 1 Introduction

Random coefficient regression models, which incorporate variations between individuals, are getting more and more popular in many fields of application, especially in biosciences. The problem of optimal designs for estimation of the mean population parameters in these models has been widely considered and many theoretical and practical solutions are available in the literature. More recently prediction of the individual response as well as of the individual deviations from the population mean response attracts larger interest in order to create individualized medication and individualized medical diagnostics or to provide information for individual se-

Maryna Prus
Institute for Mathematical Stochastics, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany, e-mail: maryna.prus@ovgu.de

Rainer Schwabe
Institute for Mathematical Stochastics, Otto-von-Guericke University, Universitätsplatz 2, 39106 Magdeburg, Germany, e-mail: rainer.schwabe@ovgu.de

lection in animal breeding, respectively. The frequently applied theory developed by Gladitz and Pilz (1982) for determining designs, which are optimal for prediction, requires the prior knowledge of the population parameters, which can be useful if pilot experiments are available. In this note we consider the practically more relevant situation, where the population parameters are unknown.

The paper is organized as follows: In the second section the model will be specified and the prediction of individual effects will be introduced. Section three provides some theoretical results for the determination of optimal designs, which will be illustrated in section 4 by a simple example. The final section presents some discussion and conclusions.

## 2 Model Specification and Prediction

In the general case of random coefficient regression models the observations are assumed to result from a hierarchical (linear) model: At the individual level the $j$th observation of individual $i$ is given by

$$Y_{ij} = \mathbf{f}(x_{ij})^\top \beta_i + \varepsilon_{ij}, \quad x_{ij} \in \mathscr{X}, \quad j = 1,..,m_i, \quad i = 1,..,n, \tag{1}$$

where $n$ denotes the number of individuals, $m_i$ is the number of observations at individual $i$, $\mathbf{f} = (f_1,..,f_p)^\top$ is the vector of known regression functions, and $\beta_i = (\beta_{i1},..,\beta_{ip})^\top$ is the individual parameter vector specifying the individual response. The experimental settings $x_{ij}$ may be chosen from a given experimental region $\mathscr{X}$. Within an individual the observations are assumed to be uncorrelated given the individual parameters. The observational errors $\varepsilon_{ij}$ have zero mean $\mathrm{E}(\varepsilon_{ij}) = 0$ and are homoscedastic with common variance $\mathrm{Var}(\varepsilon_{ij}) = \sigma^2$.

At the population level the individual parameters $\beta_i$ are assumed to have an unknown population mean $\mathrm{E}(\beta_i) = \beta$ and a given covariance matrix $\mathrm{Cov}(\beta_i) = \sigma^2\mathbf{D}$. All individual parameters and all observational errors are assumed to be uncorrelated.

The model can be represented alternatively in the following form

$$Y_{ij} = \mathbf{f}(x_j)^\top \beta + \mathbf{f}(x_j)^\top \mathbf{b}_i + \varepsilon_{ij} \tag{2}$$

by separation of the random individual deviations $\mathbf{b}_i = \beta_i - \beta$ from the mean response $\beta$. Here these individual deviations $\mathbf{b}_i$ have zero mean $\mathrm{E}(\mathbf{b}_i) = 0$ and the same covariance matrix $\mathrm{Cov}(\mathbf{b}_i) = \sigma^2\mathbf{D}$ as the individual parameters.

We consider the particular case that the number of observations as well as the experimental settings are the same for all individuals ($m_i = m$ and $x_{ij} = x_j$). Moreover, we assume for simplicity that the covariance matrix $\mathbf{D}$ is regular. The singular case will be addressed shortly in the discussion.

In the following we investigate both the predictors of the individual parameters $\beta_1,...,\beta_n$ and of the individual deviations $\mathbf{b}_1,...,\mathbf{b}_n$.

As shown in Prus and Schwabe (2011), the best linear unbiased predictor $\hat{\beta}_i$ of the individual parameter $\beta_i$ is a weighted average of the individualized estimate $\hat{\beta}_{i;\mathrm{ind}} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{Y}_i$, based on the observations at individual $i$, and the estimator of the population mean $\hat{\beta} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\bar{\mathbf{Y}}$,

$$\hat{\beta}_i = \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1}+\mathbf{D})^{-1}\hat{\beta}_{i;\mathrm{ind}} + (\mathbf{F}^\top\mathbf{F})^{-1}((\mathbf{F}^\top\mathbf{F})^{-1}+\mathbf{D})^{-1}\hat{\beta}. \tag{3}$$

Here $\mathbf{F} = (\mathbf{f}(x_1),...,\mathbf{f}(x_m))^\top$ denotes the individual design matrix, which is equal for all individuals, $\mathbf{Y}_i = (\mathbf{Y}_{i1},...,\mathbf{Y}_{im})^\top$ is the observation vector for individual $i$, and $\bar{\mathbf{Y}} = \frac{1}{n}\sum_{i=1}^n \mathbf{Y}_i$ is the average response across all individuals.

It is worth-while mentioning that the estimator of the population mean may be represented as the average $\hat{\beta} = \frac{1}{n}\sum_{i=1}^n \hat{\beta}_{i;\mathrm{ind}}$ of the individualized estimates and does, hence, not require the knowledge of the dispersion matrix $\mathbf{D}$, whereas the predictor of the individual parameter $\beta_i$ does.

The performance of the prediction (3) may be measured in terms of the mean squared error matrix of $(\hat{\beta}_1^\top,...,\hat{\beta}_n^\top)^\top$. Using results of Henderson (1975) it can be shown that this mean squared error matrix is a weighted average of the corresponding covariance matrix in the fixed effects model and the Bayesian one,

$$\mathrm{MSE}_\beta = \sigma^2((\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{F}^\top\mathbf{F})^{-1}), \tag{4}$$

where $\mathbf{I}_n$ is the $n\times n$ identity matrix, $\mathbf{1}_n$ is a $n$-dimensional vector of ones and "$\otimes$" denotes the Kronecker product of matrices as usual. Note that this representation slightly differs from that given in Fedorov and Hackl (1997, section 5.2).

Similarly the best linear unbiased predictor $\hat{\mathbf{b}}_i = \hat{\beta}_i - \hat{\beta}$ of the individual deviation $\mathbf{b}_i$ can be alternatively represented as a scaled difference

$$\hat{\mathbf{b}}_i = \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1}+\mathbf{D})^{-1}(\hat{\beta}_{i;\mathrm{ind}} - \hat{\beta}) \tag{5}$$

of the individualized estimate $\hat{\beta}_{i;\mathrm{ind}}$ from the estimated population mean $\hat{\beta}$. The corresponding mean squared error matrix of the prediction of individual deviations $(\hat{\mathbf{b}}_1^\top,...,\hat{\mathbf{b}}_n^\top)^\top$ can be written as a weighted average of the covariance matrix of the prediction in the Bayesian model and the dispersion matrix $\mathbf{D}$ of the individual effects

$$\mathrm{MSE}_b = \sigma^2((\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes\mathbf{D}). \tag{6}$$

Note that in the case of a known population mean $\beta$, which was considered by Gladitz and Pilz (1982), the mean squared error matrix for the prediction of individual parameters coincides with that for the prediction of individual deviations, which equals $\sigma^2\,\mathbf{I}_n\otimes(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1})^{-1}$.

## 3 Optimal Design

The mean squared error matrix of a prediction depends crucially on the choice of the observational settings $x_1,\ldots,x_m$, which constitute a design and can be chosen by the experimenter to minimize the mean squared error matrix in a certain sense. Typically the optimal settings will be not necessarily all distinct. Then a design

$$\xi = \begin{pmatrix} x_1 & ,\ldots, & x_k \\ w_1 & ,\ldots, & w_k \end{pmatrix} \tag{7}$$

can be specified by its distinct settings $x_1,\ldots,x_k$, $k \leq m$, say, and the corresponding numbers of replications $m_1,\ldots,m_k$ or the corresponding proportions $w_j = m_j/m$.

For analytical purposes we make use of approximate designs in the sense of Kiefer (see e. g. Kiefer, 1974), for which the integer condition on $mw_j$ is dropped and the weights $w_j \geq 0$ may be any real numbers satisfying $\sum_{j=1}^k m_j = m$. For these approximated designs the standardized information matrix for the model without individual effects ($\beta_i \equiv \beta$, i. e. $\mathbf{D} = \mathbf{0}$) is defined as

$$\mathbf{M}(\xi) = \textstyle\sum_{j=1}^k w_j \mathbf{f}(x_j)\mathbf{f}(x_j)^\top = \frac{1}{m}\mathbf{F}^\top\mathbf{F}. \tag{8}$$

Further we introduce the standardized covariance matrix of the random effects $\Delta = m\mathbf{D}$ for notational ease. With these notations we may define the standardized mean squared error matrices as

$$\mathrm{MSE}_\beta(\xi) = (\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{M}(\xi)+\Delta^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes\mathbf{M}(\xi)^{-1} \tag{9}$$

for the prediction of the individual parameters and

$$\mathrm{MSE}_b(\xi) = (\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{M}(\xi)+\Delta^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes\Delta \tag{10}$$

for the prediction of the individual deviations. For any exact design $\xi$ the matrices $\mathrm{MSE}_\beta(\xi)$ and $\mathrm{MSE}_b(\xi)$ coincide with the mean squared error matrices (4) and (6), respectively, up to a multiplicative factor $\sigma^2/m$.

In this paper we focus on the integrated mean squared error (IMSE) criterion, which is defined, in general, as

$$\mathrm{IMSE}_\beta = \textstyle\int_{\mathscr{X}} \mathrm{E}(\sum_{i=1}^n(\hat{\mu}_i(x)-\mu_i(x))^2)\nu(\mathrm{d}x) \tag{11}$$

for prediction of individual parameters, where $\hat{\mu}_i(x) = \mathbf{f}(x)^\top\hat{\beta}_i$ and $\mu_i(x) = \mathbf{f}(x)^\top\beta_i$ denote the predicted and the true individual response, and the integration is with respect to a given weight distribution $\nu$ on the design region $\mathscr{X}$, which is typically uniform. The standardized IMSE-criterion $\Phi_\beta = \frac{m}{\sigma^2}\mathrm{IMSE}_\beta$ can be represented as

$$\Phi_\beta(\xi) = (n-1)\,\mathrm{tr}\,((\mathbf{M}(\xi)+\Delta^{-1})^{-1}\mathbf{V}) + \mathrm{tr}\,(\mathbf{M}(\xi)^{-1}\mathbf{V}), \tag{12}$$

which is a weighted sum of the IMSE-criterion in the fixed effects model and the Bayesian IMSE-criterion, where $\mathbf{V} = \int_{\mathscr{X}} \mathbf{f}(x)\mathbf{f}(x)^\top \nu(\mathrm{d}x)$ is the "information" of the weight distribution $\nu$ and "tr" denotes the trace of a matrix.

With the general equivalence theorem (see e. g. Silvey, 1980) we may obtain the following characterization of an optimal design.

**Theorem 1.** *The approximate design $\xi^*$ is IMSE-optimal for the prediction of individual parameters, if and only if*

$$\begin{aligned}\mathbf{f}(x)^\top((n-1)(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}\mathbf{V}(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}+\mathbf{M}(\xi^*)^{-1}\mathbf{V}\mathbf{M}(\xi^*)^{-1})\,\mathbf{f}(x)\\ \leq \mathrm{tr}\,(((n-1)(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}\mathbf{M}(\xi^*)(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}+\mathbf{M}(\xi^*)^{-1})\mathbf{V}) \quad (13)\end{aligned}$$

*for all $x \in \mathscr{X}$.*

*For any experimental setting $x_j$ of $\xi^*$ with $w_j > 0$ equality holds in* (13).

The IMSE-criterion for the prediction of individual deviations is given by

$$\mathrm{IMSE}_b(\xi) = \int_{\mathscr{X}} \mathrm{E}(\textstyle\sum_{i=1}^{n}(\hat{\mu}_i^b(x) - \mu_i^b(x))^2)\nu(\mathrm{d}x), \quad (14)$$

where $\hat{\mu}_i^b(x) = \mathbf{f}(x)^\top\hat{\mathbf{b}}_i$ and $\mu_i^b(x) = \mathbf{f}(x)^\top\mathbf{b}_i$ denote the predicted and the true individual response deviation form the population mean, respectively. The standardized IMSE-criterion $\Phi_b = \frac{m}{\sigma^2}\mathrm{IMSE}_b$ can again be written as

$$\Phi_b(\xi) = (n-1)\,\mathrm{tr}\,((\mathbf{M}(\xi)+\Delta^{-1})^{-1}\mathbf{V}) + \mathrm{tr}\,(\Delta\mathbf{V}). \quad (15)$$

The first term in (15) coincides with the criterion function of the Bayesian IMSE-criterion and the second term is constant. Hence, Bayesian IMSE-optimal designs are also IMSE-optimal for the prediction of individual deviations. The characterization of IMSE-optimal designs is given by the corresponding equivalence theorem for Bayes optimality.

**Theorem 2.** *The approximate design $\xi^*$ is IMSE-optimal for the prediction of individual deviations, if and only if*

$$\begin{aligned}\mathbf{f}(x)^\top(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}\mathbf{V}(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}\mathbf{f}(x)\\ \leq \mathrm{tr}\,((\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}\mathbf{M}(\xi^*)(\mathbf{M}(\xi^*)+\Delta^{-1})^{-1}\mathbf{V}) \quad (16)\end{aligned}$$

*for all $x \in \mathscr{X}$.*

*For any experimental setting $x_j$ of $\xi^*$ with $w_j > 0$ equality holds in* (16).

## 4 Example

To illustrate our results we consider the model $Y_{ij} = \beta_{i1} + \beta_{i2}x_j + \varepsilon_{ij}$ of a straight line regression on the experimental region $\mathscr{X} = [0, 1]$, where the settings $x_j$ can be interpreted as time or dosage. We assume uncorrelated components such that the covariance matrix $\mathbf{D} = \mathrm{diag}\,(d_1, d_2)$ of the random effects is diagonal with diagonal entries $d_1$ and $d_2$ for the variance of the intercept and the slope, respectively. To exhibit the differences in the design criteria the variance of the intercept is assumed to be small, $d_1 < 1/m$.

According to Theorems 1 and 2, the IMSE-optimal designs only take observations at the endpoints $x = 0$ and $x = 1$ of the design region, as the sensitivity functions, which are the left hand sides in the conditions (13) and (16), are polynomials in $x$ of degree 2. Hence, the optimal design $\xi^*$ is of the form

$$\xi_w = \begin{pmatrix} 0 & 1 \\ 1-w & w \end{pmatrix}, \tag{17}$$

and only the optimal weight $w^*$ has to be determined. For designs $\xi_w$ the criterion functions (12) and (15) are calculated with $\delta_k = m\, d_k$ to

$$\Phi_\beta(\xi_w) = \tfrac{1}{3}\left(\tfrac{(n-1)(3\delta_1+\delta_2+\delta_1\delta_2)}{(\delta_1+1)(w\delta_2+1)-w^2\delta_1\delta_2} + \tfrac{1}{w(1-w)}\right), \tag{18}$$

$$\Phi_b(\xi_w) = \tfrac{1}{3}\left(\tfrac{(n-1)(3\delta_1+\delta_2+\delta_1\delta_2)}{(\delta_1+1)(w\delta_2+1)-w^2\delta_1\delta_2} + 3\delta_1 + \delta_2\right). \tag{19}$$

To obtain numerical results the number of individuals and the number of observations at each individual are fixed to $n = 100$ and $m = 10$. For the variance $d_1$ of the intercept we use the value 0.001. Figure 1 illustrates the dependence of the optimal weight $w^*$ and the rescaled variance parameter $\rho = d_2/(1+d_2)$, which mimics in a way the intraclass correlation and has the advantage to be bounded such that the whole range of slope variances $d_2$ can be shown. The optimal weight for the prediction of individual parameters increases in the slope variance $d_2$ from 0.5 for $d_2 \to 0$ to about 0.91 for $d_2 \to \infty$. For $d_1 < 1/m$ the Bayesian optimal design, which is also optimal for the prediction of individual deviations, has optimal weight $w^* = 1$ for all positive values of $d_2$, which results in a singular design.

In Figure 2 the efficiencies $\mathrm{eff}(\xi) = \Phi(\xi_{w^*})/\Phi(\xi)$ are plotted for the optimal design $\xi_{0.5}$ in the fixed effects model ignoring the individual effects and for the naive equidistant design $\bar{\xi}$, which assigns weights $1/m$ to the $m$ settings $x_j = (j-1)/(m-1)$. For the prediction of individual parameters the efficiency of the design $\xi_{0.5}$ decreases from 1 for $d_2 \to 0$ to approximately 0.60 for $d_2 \to \infty$, whereas $\bar{\xi}$ shows an overall lower performance going down to 0.42 for large $d_2$.

For the prediction of individual deviations the efficiency of both designs show a bathtub shaped behavior with limiting efficiency of 1 for $d_2 \to 0$ or $d_2 \to \infty$. This is due to the fact that all regular designs are equally good for small $d_2$ and equally bad for large $d_2$, since the criterion function (15) behaves like $\mathrm{tr}(\Delta\mathbf{V})$ for $d_2 \to \infty$ independently of $\xi$. The minimal efficiencies are 0.57 for $\xi_{0.5}$ and 0.43 for $\bar{\xi}$.

For the sake of completeness also the efficiency is plotted with respect to the Bayes criterion, which equals the first term in both criteria (12) and (15) to show that the efficiency may differ although the design optimization seems to be the same. This difference is due to the second (constant) term in (15). It should also be noted

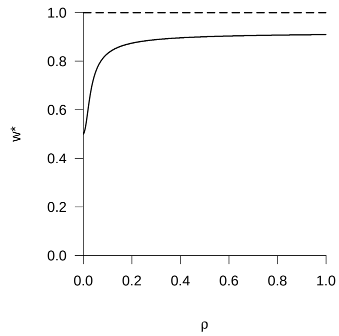


**Fig. 1** Optimal weights $w^*$ for the prediction of individual parameters (solid line) and for the prediction of individual deviations (dashed line)

that the present efficiencies cannot be interpreted as savings or additional needs in terms of sample sizes as in fixed effect models.

## 5 Discussion and conclusions

In this paper we point out similarities and differences in the theory of optimal designs for the prediction of individual parameters and individual deviations compared to Bayesian designs. The objective function of the IMSE-criterion is a weighted av-

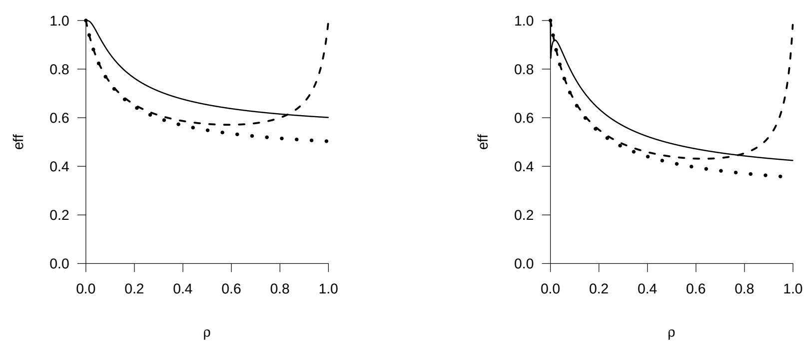


**Fig. 2** Efficiency of $\xi_{0.5}$ (left panel) and $\bar{\xi}$ (right panel) for the prediction of individual parameters (solid line), individual deviations (dashed line) and the Bayes criterion (dotted line)

erage of the Bayesian and the "standard" counterparts in the case of prediction of individual parameters and defines, hence, a compound criterion. For the prediction of individual deviations the Bayesian optimal designs remain optimal, while the criteria differ by an additive constant.

A generalization of the present results to singular dispersion matrices $\mathbf{D}$ is straightforward, although there is no Bayesian counterpart in that case and the formulae become less appealing. Such singular dispersion matrices naturally occur, if only parts of the parameter vector are random and the remaining linear combinations are constant across the population. In particular, in the case of a random intercept model, when all other parameters are fixed, the optimal design for the prediction of the individual parameters can be obtained as the optimal one in the corresponding model without individual effects (Prus and Schwabe, 2011), while for prediction of the individual deviations any meaningful design will be optimal.

The method proposed may be directly extended to other linear design criteria as well as to the class of $\Phi_q$-criteria based on the eigenvalues of the mean squared error matrix. Although the design optimality presented here is formulated for approximate designs, which generally may not be exactly realized. These optimal approximate designs can serve as a benchmark for candidates of exact designs, which for example are obtained by appropriate rounding of the optimal weights. Optimal designs for situations, which allows for different individual designs, will be subject of future research, in particular, in the case of sparse sampling, where the number of observations per individual is less than the number of parameters.

**Acknowledgements** 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 during a visit at the International Newton Institute in Cambridge.

## References

1. Fedorov, V., Hackl, P.: Model-Oriented Design of Experiments. Springer, New York (1997)
2. Gladitz, J., Pilz, J.: Construction of optimal designs in random coefficient regression models. Statistics **13**, 371–385 (1982)
3. Henderson, C. R.: Best linear unbiased estimation and prediction under a selection model. Biometrics **31**, 423–477 (1975)
4. Kiefer, J.: General equivalence theory for optimum designs (approximate theory). Annals of Statistics **2**, 849–879 (1974)
5. Prus, M., Schwabe, R.: Optimal Designs for Individual Prediction in Random Coefficient Regression Models. In: V. Melas, G. Nachtmann and D. Rasch (Eds.): Optimal Design of Experiments - Theory and Application: Proceedings of the International Conference in Honor of the late Jagdish Srivastava, pp. 122–129. Center of Experimental Design, University of Natural Resources and Life Sciences, Vienna (2011)
6. Silvey, S. D.: Optimal Design. Chapman & Hall, London (1980)