Interpolation and Extrapolation in Random Coefficient Regression Models: Optimal Design for Prediction

Maryna Prus and Rainer Schwabe

Abstract The problem of optimal designs for the prediction of individual parameters in random coefficient regression in the particular case of a given population mean was considered by Gladitz and Pilz (1982). In the more general situation, where the population parameter is unknown, D- and L-optimal designs were discussed in Prus and Schwabe (2016). Here we present analytical results for designs, which are optimal for the prediction in the case of interpolation as well as extrapolation of the individual response.

1 Introduction

Hierarchical random coefficient regression models are very popular in many fields of statistical applications. The subject of this paper is the problem of finding optimal designs for interpolation and extrapolation in such models. Optimal designs for interpolation and extrapolation in fixed effects models were considered in detail in Kiefer (1964a, 1964b). Some theory for determining Bayesian optimal designs, which are also optimal for the prediction of the individual deviations from the population mean in random coefficients regression models (see Prus (2015)), was developed by Chaloner (1984) (see also Chaloner (1982)). The problem of optimal designs for the prediction of individual parameters in random coefficient regression models was considered in Prus and Schwabe (2016) for the linear and the generalized D-criteria. Here we deal with the c-criterion for the prediction and formulate the optimality condition for approximate designs. We consider the problem of optimal designs for interpolation and extrapolation as a particular case of the c-criterion for the straight line regression model and illustrate the results by a numerical example.

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

The paper has the following structure: In the second part the hierarchical random coefficient regression model is specified and the best linear unbiased prediction of the individual parameters is introduced. The third part provides analytical results for c-optimal designs. In the fourth part optimal designs for interpolation and extrapolation are considered. The last section presents some discussion and an outlook.

2 Model Specification and Prediction

In the hierarchical random coefficient regression model the j-th observation of individual i is given by

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

at the individual level, where n denotes the number of individuals, m_i is the number of observations at individual i, $\mu_i(x)$ are the response functions of the form $\mu_i(x) = \mathbf{f}(x)^\top \beta_i$, $\mathbf{f} = (f_1,..,f_p)^\top$ 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 ε_{ij} have zero mean $\mathrm{E}(\varepsilon_{ij}) = 0$ and are homoscedastic with common variance $\mathrm{var}(\varepsilon_{ij}) = \sigma^2$.

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

We consider the particular case of the model (1) where the number of observations as well as the experimental settings are the same for all individuals ($m_i = m$ and $x_{ij} = x_j$).

We investigate the predictor of the individual parameters $\beta_1,...,\beta_n$. This predictor is also sometimes called estimator of the random parameters and can be viewed as an empirical Bayes estimator.

As exhibited in Prus and Schwabe (2011), the best linear unbiased predictor $\hat{\beta}_i$ of the individual parameter β_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 $\hat{\beta} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\bar{\mathbf{Y}}$,

$$\hat{\beta}_i = \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\beta}_{i;\mathrm{ind}} + (\mathbf{F}^\top\mathbf{F})^{-1}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\beta}. \tag{2}$$

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.

The performance of the prediction (2) may be measured in terms of the mean squared error matrix of $(\hat{\beta}_1^\top,...,\hat{\beta}_n^\top)^\top$. The latter matrix has the form

$$\mathrm{MSE} = \sigma^2\left((\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{F}^\top\mathbf{F}+\mathbf{D}^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{F}^\top\mathbf{F})^{-1}\right), \qquad (3)$$

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. The mean squared error matrix (3) is a weighted average of the corresponding covariance matrix in the fixed effects model and the Bayesian one.

Note that the response functions $\mu_i(x_0) = \mathbf{f}(x_0)^\top\beta_i$ are predictable for all values of x_0 (not necessarily from the experimental region $\mathscr{X}$) also if the design matrix $\mathbf{F}$ is not of full rank and consequently the matrix $\mathbf{F}^\top\mathbf{F}$ is singular (see Prus (2015)). However, the individual parameters β_i themselves are not predictable if the design matrix is not full rank.

3 c-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 can be chosen by the experimenter to minimize the mean squared error matrix and which constitute an exact design. Typically the optimal settings will be not necessarily all distinct. Then a design

$$\xi = \begin{pmatrix} x_1 & ,\ldots, & x_k \\ w_1 & ,\ldots, & w_k \end{pmatrix} \qquad (4)$$

can be specified by its distinct settings $x_1,\ldots,x_k$, $k \le m$, 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 $w_j \ge 0$ may be any real numbers satisfying $\sum_{j=1}^k w_j = 1$ or equivalently $\sum_{j=1}^k m_j = m$. For these approximate designs the standardized information matrix for the model without individual effects ($\beta_i \equiv \beta$, i. e. $\mathbf{D} = \mathbf{0}$) is defined as

$$\mathbf{M}(\xi) = \textstyle\sum_{j=1}^k w_j \mathbf{f}(x_j)\mathbf{f}(x_j)^\top = \tfrac{1}{m}\mathbf{F}^\top\mathbf{F}. \qquad (5)$$

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

$$\mathrm{MSE}(\xi) = (\mathbf{I}_n - \tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes(\mathbf{M}(\xi)+\Delta^{-1})^{-1} + (\tfrac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top)\otimes\mathbf{M}(\xi)^{-1}. \qquad (6)$$

For any exact design ξ all mw_j are integer. Then the matrix $\mathrm{MSE}(\xi)$ coincides with the mean squared error matrices (3) up to a multiplicative factor σ^2/m.

In this paper we focus on the extended c-criterion for the prediction (see Prus (2015) or Prus and Schwabe (2016)), which is defined as the sum of the variances of $\mathbf{c}^\top\hat{\beta}_i - \mathbf{c}^\top\beta_i$ across all individuals, where $\mathbf{c}$ is a specified vector of dimension p:

$$\mathrm{c}_\beta(\xi) = \sum_{i=1}^{n} \mathrm{var}\,(\mathbf{c}^\top \hat{\beta}_i - \mathbf{c}^\top \beta_i). \tag{7}$$

Using (6), the standardized c-criterion $\Phi_\beta = \frac{m}{\sigma^2}\mathrm{c}_\beta$ can be represented as

$$\Phi_\beta(\xi) = \mathbf{c}^\top \mathbf{M}(\xi)^{-1}\mathbf{c} + (n-1)\,\mathbf{c}^\top(\mathbf{M}(\xi) + \Delta^{-1})^{-1}\mathbf{c}, \tag{8}$$

which is a weighted sum of the c-criterion in the fixed effects model and the Bayesian c-criterion.

With the general equivalence theorem (see e. g. Silvey, 1980, ch. 3) we may obtain the following characterization of an optimal design.

Theorem 1. *The approximate design ξ^* with non-singular information matrix $\mathbf{M}(\xi^*)$ is c-optimal for the prediction of individual parameters if and only if*

$$\begin{aligned}[\mathbf{f}(x)^\top \mathbf{M}(\xi^*)^{-1}\mathbf{c}]^2 &+ (n-1)\,[\mathbf{f}(x)^\top(\mathbf{M}(\xi^*) + \Delta^{-1})^{-1}\mathbf{c}]^2 \\ \leq\ & \mathbf{c}^\top \mathbf{M}(\xi^*)^{-1}\mathbf{c} + (n-1)\,\mathbf{c}^\top(\mathbf{M}(\xi^*) + \Delta^{-1})^{-1}\mathbf{M}(\xi^*)(\mathbf{M}(\xi^*) + \Delta^{-1})^{-1}\mathbf{c}\end{aligned} \tag{9}$$

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

For any experimental setting x_j of ξ^* with $w_j > 0$ equality holds in (9).

Note that optimal designs, which result in singular information matrices, may also exist.

4 Optimal Designs for Interpolation and Extrapolation

For models without random effects the interpolation and extrapolation problem was considered in detail by Kiefer and Wolfowitz (1964a, 1964b). The Bayesian optimal designs were discussed in Caloner (1984) (see also Chaloner (1982)).

In this section we examine the straight line regression

$$Y_{ij} = \beta_{i1} + \beta_{i2} x_j + \varepsilon_{ij} \tag{10}$$

on the experimental region $\mathscr{X} = [0,1]$. The settings x_j can be interpreted as time or dosage and $x = 0$ means a measurement at baseline. We assume uncorrelated components such that the dispersion matrix $\mathbf{D} = \mathrm{diag}\,(d_1, d_2)$ of the random effects is diagonal with entries d_1 and d_2 for the variance of the intercept and slope, respectively. The variance of the intercept is assumed to be small, $d_1 < 1/m$.

The problem of optimal designs for interpolation and extrapolation of the response functions $\mu_i(x_0) = \mathbf{f}(x_0)^\top \beta_i$ at some given point $x_0 \in \mathscr{X} = [0,1]$ and $x_0 \notin \mathscr{X} = [0,1]$, respectively, may be recognized as a special case of the c-criterion (8) for the prediction with $c = \mathbf{f}(x_0) = (1, x_0)^\top$.

It follows from Theorem 1 that for $c = (1, x_0)^\top$ the c-optimal designs (with non-singular information matrices) only take observations at the endpoints $x = 0$ and

$x = 1$ of the design region, since the sensitivity function, given by the left hand side of inequality (9), is then a polynomial in x of degree 2 with positive leading term. Hence, the optimal design ξ^* is of the form

$$\xi_w = \begin{pmatrix} 0 & 1 \\ 1-w & w \end{pmatrix}, \tag{11}$$

and only the optimal weight w^* has to be determined. For designs ξ_w the criterion function (8) is calculated with $\gamma_k = 1/(m d_k)$ for $k = 1,2$ to

$$\Phi_\beta(\xi_w) = \frac{x_0^2 - 2wx_0 + w}{w(1-w)} + (n-1)\frac{x_0^2(1+\gamma_1) - 2wx_0 + w + \gamma_2}{(1+\gamma_1)(w+\gamma_2) - w^2}. \tag{12}$$

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^* on the rescaled variance parameter $\rho = d_2/(1+d_2)$, which in a way mimics the intraclass correlation and has the advantage of being bounded, so that the whole range of slope variances d_2 can be shown. We use the values 0.9, 0.5 and 0.3 for the interpolation point and the values 1.2, 2 and 100 for the extrapolation point x_0. Note that in both cases of interpolation and extrapolation the optimal weights w^* are equal for $\rho = 0$ to that in the model without random effects: $w^* = x_0$ and $w^* = x_0/(2x_0 - 1)$, respectively (see e.g. Schwabe (1996), ch. 2).

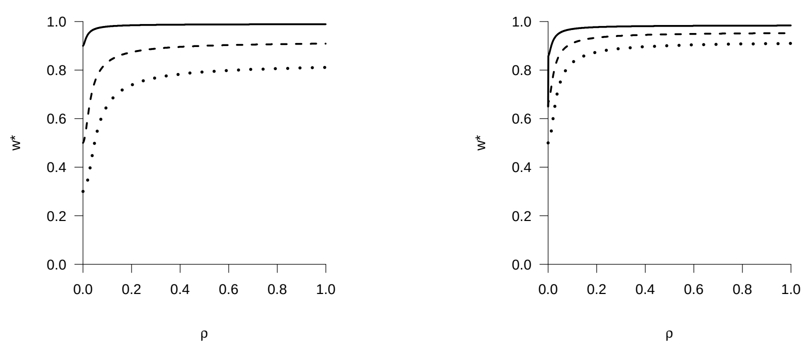

Fig. 1 Optimal weights w^* for interpolation (left panel) in $x_0 = 0.9$ (solid line), $x_0 = 0.5$ (dashed line) and $x_0 = 0.3$ (dotted line) and extrapolation (right panel) in $x_0 = 1.2$ (solid line), $x_0 = 2$ (dashed line) and $x_0 = 100$ (dotted line) as functions of $\rho = d_2/(1+d_2)$

In Figure 2 the efficiency $\mathrm{eff}(\xi) = \Phi_\beta(\xi_{w^*})/\Phi_\beta(\xi_w)$ is plotted for the optimal designs ξ_w with $w = x_0$ and $w = x_0/(2x_0 - 1)$ in the fixed effects model for interpolation and extrapolation, respectively.

Figure 3 represents the efficiency of the Bayesian optimal designs. Note that the latter designs are optimal for the prediction of the individual deviations $\mathbf{b}_i = \beta_i - \beta$ from the population mean in the random coefficient regression model (1) (see Prus (2015)).

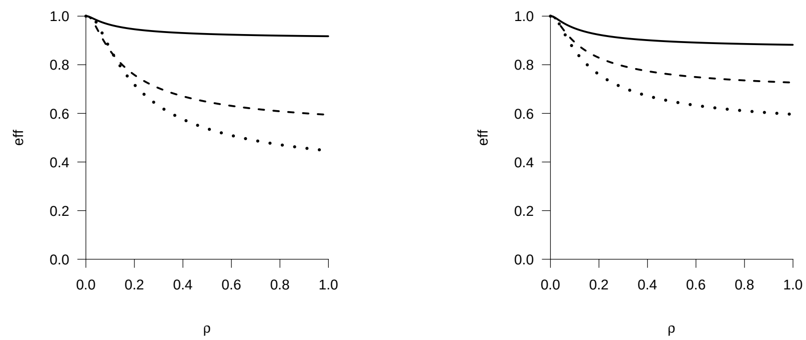

Fig. 2 Efficiency of the optimal designs in fixed effects model for interpolation (left panel) in $x_0 = 0.9$ (solid line), $x_0 = 0.5$ (dashed line) and $x_0 = 0.3$ (dotted line) and extrapolation (right panel) in $x_0 = 1.2$ (solid line), $x_0 = 2$ (dashed line) and $x_0 = 100$ (dotted line) as functions of $\rho = d_2/(1+d_2)$

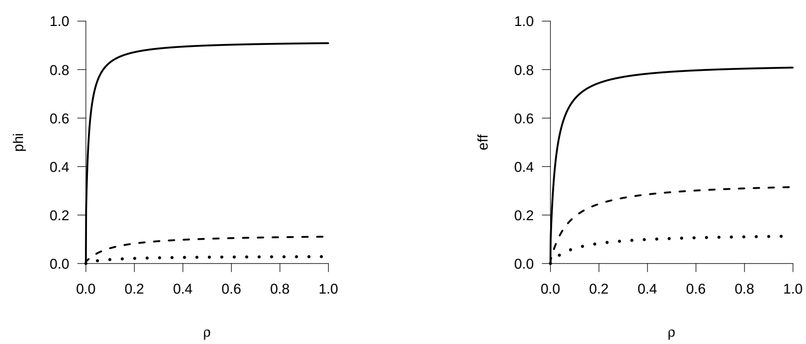

Fig. 3 Efficiency of the Bayesian optimal designs for interpolation (left panel) in $x_0 = 0.9$ (solid line), $x_0 = 0.5$ (dashed line) and $x_0 = 0.3$ (dotted line) and extrapolation (right panel) in $x_0 = 1.2$ (solid line), $x_0 = 2$ (dashed line) and $x_0 = 100$ (dotted line) as functions of $\rho = d_2/(1+d_2)$

5 Discussion and Outlook

We have proposed an analytical method for determining optimal interpolation and extrapolation designs for the prediction of individual parameters in hierarchical random coefficient regression. This problem was considered as a particular case of the c-criterion for the prediction. The criterion function of the c-criterion is a weighted sum of the c-criterion in the fixed effects models and the Bayesian c-criterion and can be recognized as a special case of the compound criterion (see e.g. Atkinson et al. (2007), ch. 21).

It was established that the optimal two-points designs in the fixed effects models are in general not optimal for the prediction. Note that the one-point designs, which take all observations at point x_0 and are also optimal for estimation in models without random effects, result in even larger values of the criterion function than the two-points designs mentioned above and perform hence worse. In the numerical example for the straight line regression with a diagonal dispersion matrix the Bayesian optimal designs show a similar behavior as the optimal designs for the prediction. They lead, however, for $x_0 \to 1$ to a singular information matrix and cannot be used for prediction.

The optimality condition for the c-criterion proposed here was formulated for approximate designs, which are not directly applicable and the optimal weights have to be appropriately rounded. The analytical results presented in this paper are based on the assumption that all individuals (observational units) get the same treatment. In one of the next steps of the research the design optimality problem will be considered for a more general case of the hierarchical random coefficient regression, where different designs are allowed for different individuals.

References

1. Atkinson, A. C., Donev, A. N., Tobias, R. D.: Optimum Experimental Designs, with SAS. Oxford University Press, Oxford (2007).
2. Chaloner, K.: Optimal Bayesian Experimental Design for Linear Models. Department of Statistics Technical Report 238, Carnegie-Mellon University, Pittsburgh (1982)
3. Chaloner, K.: Optimal Bayesian experimental design for linear models. Ann. Statist. **12**, 283–300 (1984)
4. Cook, R. D., Wong, W. K.: On the equivalence of constrained and compound optimal designs. Journal of the American Statistical Association **89**, 687-692 (1994).
5. Gladitz, J., Pilz, J.: Construction of optimal designs in random coefficient regression models. Statistics **13**, 371-385 (1982)
6. Kiefer, J.: General equivalence theory for optimum designs (approximate theory). Annals of Statistics **2**, 849–879 (1974)
7. Kiefer, J., Wolfowitz, J.: Optimum extrapolation and interpolation designs i. Annals of the Institute of Statistical Mathematics, **16**, 79-108 (1964a)
8. Kiefer, J., Wolfowitz, J.: Optimum extrapolation and interpolation designs ii. Annals of the Institute of Statistical Mathematics, **16**, 295-303 (1964b)
9. Prus, M.: Optimal Designs for the Prediction in Hierarchical Random Coefficient Regression Models. Ph.D. thesis, Otto-von-Guericke University Magdeburg (2015)

10. Prus, M., Schwabe, R.: Optimal designs for individual prediction in random coefficient regression models. In: Melas V., Nachtmann G., Rasch D.: 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)
11. Prus, M., Schwabe, R.: Optimal designs for the prediction of individual parameters in hierarchical models. Journal of the Royal Statistical Society: Series B., DOI: 10.1111/rssb.12105 (2016)
12. Schwabe, R.: Optimum Designs for Multi-Factor Models. Springer, New York (1996)
13. Silvey, S. D.: Optimal Design. Chapman & Hall, London (1980)