Optimal Designs for the Prediction in Hierarchical Random Coefficient Regression Models

Dissertation

zur Erlangung des akademischen Grades

doctor rerum naturalium
(Dr. rer. nat.)

von Dipl.-Wirt.-Math. Maryna Prus

geb. am 17.08.1982 in Gorlowka, Ukraine

genehmigt durch die Fakultät für Mathematik
der Otto-von-Guericke-Universität Magdeburg

Gutachter: Prof. Dr. Rainer Schwabe
             Prof. Julia Volaufova, Ph.D.

eingereicht am: 28.04.2015

Verteidigung am: 15.07.2015

Summary

Hierarchical random coefficient regression models are popular in many fields of statistical application, for example in pharmacokinetics. In these models the observational units are assumed to come from the same population and differ from each other by individual random parameters with an unknown population mean. The aim of this thesis is to develop an analytical approach for determining designs, which are optimal for the prediction of linear aspects in hierarchical random coefficient regression models.

After the first introductory chapter the main concepts of optimal design for classical linear regression models and Bayesian models (random coefficient regression with a given population parameter) are presented in the second chapter.

The third chapter provides some analytical results for the estimation and prediction of linear aspects in linear mixed models. This theory is used in Chapter 4 for determining best linear unbiased predictions of linear aspects, in particular individual parameters and individual deviations from the population mean, in hierarchical random coefficient regression models.

In Chapter 5 design criteria - linear criteria and a generalized version of the determinant criterion - in the models under investigation are introduced. For prediction of the individual deviations Bayesian optimal designs retain their optimality. In the case of prediction of the individual parameters the design criteria result in compound criteria with weights depending on the number of experimental units.

A more general case of the random coefficient regression models (multi-group models) is considered in Chapter 6. In these models the statistical analysis can be performed in each group separately if the group sizes are fixed.

The thesis is concluded by a discussion and an outlook on further open problems.

Zusammenfassung

Hierarchische Regressionsmodelle mit zufälligen Parametern sind populär in vielen statistischen Anwendungsbereichen, beispielsweise in Pharmakokinetik. In diesen Modellen wird angenommen, dass die Beobachtungseinheiten aus derselben Population kommen und sich voneinander durch individuelle zufällige Parameter mit einem unbekannten Populationsmittelwert unterscheiden. Das Ziel dieser Arbeit ist die Entwicklung eines analytischen Ansatzes zur Ermittlung von Designs, die optimal für die Vorhersage linearer Aspekte in hierarchischen Regressionsmodellen mit zufälligen Parametern sind.

Nach dem ersten einleitenden Kapitel werden im Kapitel 2 die grundliegenden Konzepte der optimalen Designs in klassischen linearen Regressionsmodellen sowie Bayesschen Modellen (Regressionsmodelle mit zufälligen Parametern mit einem bekannten Populationsmittelwert) präsentiert.

Das dritte Kapitel liefert analytische Ergebnisse für die Schätzung und Vorhersage linearer Aspekte in linearen gemischten Modellen. Anhand dieser Theorie werden im vierten Kapitel beste lineare unverfälschte Vorhersagen für lineare Aspekte, insbesondere für individuelle Parameter und individuelle Abweichungen vom Populationsparameter, in hierarchischen Regressionsmodellen ermittelt.

Im fünften Kapitel werden Designkriterien - lineare Kriterien sowie eine verallgemeinerte Version vom Determinantenkriterium - in den untersuchten Modellen eingeführt. Für die Vorhersage individueller Abweichungen sind Bayessche optimale Designs weiterhin optimal. Im Fall der Vorhersage individueller Parametern resultieren die Designkriterien in zusammengesetzten Kriterien, deren Gewichte von der Anzahl der Beobachtungseinheiten abhängig sind.

Im sechsten Kapitel wird eine allgemeinere Klasse von Regressionsmodellen mit zufälligen Parametern (Mehrgruppenmodelle) betrachtet. In diesen Modellen erfolgt die statistische Analyse separat in jeder Gruppe, wenn die Gruppengrößen fixiert sind.

Die Arbeit wird mit einer Diskussion und einem Ausblick auf weitere offene Probleme abgeschlossen.

Contents

V

1 Introduction

Hierarchical random coefficient regression models, which allow for variations between individuals, are popular in many fields of statistical application. In these models observational units are assumed to come from the same population and differ from each other by individual random parameters with an unknown population mean. Originally, hierarchical random coefficient regression was introduced in biosciences and applied for selection purposes, in particular, in plant and animal breeding (see e.g. Henderson (1984)). More recently, this concept appears in sociological and psychological research and is often used in statistical methodology, for instance, for small area estimation (see e.g. Rao (2003)). Some real data examples can be found in Pinheiro and Bates (2000) and Molenberghs and Verbeke (2001).

The problem of estimation of the population parameters (population mean) has been well discussed in the literature (see e.g. Rao (1965), Spjotvoll (1977), Isotalo *et al.* (2011)). If the population mean is assumed to be known, the Bayesian approach (see e.g. Pukelsheim (1993)) is commonly used for the estimation (prediction) of the individual parameters. For the situation, where the population parameters are unknown, under the assumption of normal distribution with a non-singular covariance matrix of random effects the Maximum-Likelihood estimator can be directly derived (see e.g. Fedorov and Jones (2005)). In the general case Henderson's approach (see Henderson (1975)) based on moment assumptions only is the common solution (see e.g. Christensen (2002)). The prediction of the individual parameters was also considered by Fedorov and Leonov (2013). Some related findings can be found in Haslett *et al.* (2014). In this thesis we avoid distributional assumptions and allow for a singular covariance matrix of random parameters.

Optimal designs for the estimation of population mean parameters in hierarchical random coefficient regression models have been considered in detail in many scientific works (see e.g. Fedorov and Hackl (1997), Liski *et al.* (2002), Entholzner *et al.* (2005) or Schmelter (2007)). Gladitz and Pilz (1982) have investigated the models with a known population mean (Bayesian models) and have established that Bayesian optimal designs are optimal for the prediction of the individual parameters. Some results for the prediction of the individual deviations from the population mean in the models with unknown population parameters are briefly presented by Prus and Schwabe (2013). Prus and Schwabe

(2016) (see also Prus and Schwabe (2011)) propose an analytical approach for determining optimal designs for the prediction of the individual parameters. Candel (2009) has considered this problem in the particular case of polynomial growth curves.

In this work we focus on the models with unknown population mean and search for optimal designs for the prediction of various linear aspects. We consider linear design criteria, in particular the commonly used integrated mean squared error criterion, and introduce a generalized version of the determinant criterion, which can also be used when the dispersion matrix of random effects is singular. For the prediction of the individual deviations Bayesian optimal designs retain their optimality. The prediction of the individual parameters leads for the criteria mentioned above to a compound criterion as considered by Cook and Wong (1994) (see also Läuter (1974), Atkinson *et al.* (2007), ch. 10, 21): a weighted sum of the corresponding criterion in the fixed effects model and the Bayesian criterion.

The present thesis has the following structure: The main concepts of optimal design for classical linear regression models and Bayesian models are introduced in the second chapter. The third chapter provides some analytical results for the estimation and prediction of linear aspects in linear mixed models. This theory is used in Chapter 4 for determining best linear unbiased predictions of linear aspects, in particular individual parameters and individual deviations, in hierarchical random coefficient regression models. In Chapter 5 design criteria in the models under investigation are proposed and an analytical approach for determining optimal designs is given. Moreover some tools for construction of optimal designs by in- or equivariance are provided. Theoretical results are illustrated by simple examples. In Chapter 6 the theory developed in the previous chapters is applied to multi-group models. In these models the statistical analysis can be performed in each group separately if the group sizes are fixed. The thesis is concluded by a discussion and an outlook on further research.

2 Optimal Designs in Classical Linear Regression and Bayesian Models

This chapter provides the basic concepts of optimal experimental designs in classical linear regression and Bayesian models. The problem of optimal designs in classical linear regression (or fixed effects) models is well discussed in the literature. The books by Fedorov (1972), Silvey (1980), Pazman (1986) and Atkinson and Donev (1992) build a good introduction. Optimal designs in the general case of Bayesian models has been considered by Pukelsheim (1993). Gladitz and Pilz (1982) propose solutions for the particular case of random coefficient regression with a known population parameter (see also Fedorov and Hackl (1997)). The results presented in this chapter come form the mentioned publications.

2.1 Model Specification and Estimation

In this section the fixed effects linear regression and Bayesian models are described and the best linear unbiased estimators in these models are introduced.

2.1.1 Fixed Effects Models

In fixed effects linear regression models the j-th observation is given by the formula

$$Y_j = \mathbf{f}(x_j)^\top \boldsymbol{\beta} + \varepsilon_j, \quad j = 1, .., m, \quad x_j \in \mathcal{X}, \tag{2.1}$$

where m is the number of observations, Y_j denotes the j-th observation, $\mathbf{f} = (f_1, .., f_p)^\top : \mathcal{X} \to \mathbb{R}^p$ is a vector of known regression functions and $\boldsymbol{\beta} = (\beta_1, .., \beta_p)^\top$ is a vector of unknown fixed parameters. The experimental settings x_j may be chosen from the experimental region $\mathcal{X}$ for which the image $\mathbf{f}(\mathcal{X}) \subset \mathbb{R}^p$ is assumed to be a compact set. The observational errors ε_j are uncorrelated, have zero mean and common variance $\sigma^2 > 0$.

In the vector notation the model (2.1) has the form

$$\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{2.2}$$

3

where $\mathbf{Y} = (Y_1, ..., Y_m)^\top$ is the vector of observations, $\mathbf{F} = (\mathbf{f}(x_1), ..., \mathbf{f}(x_m))^\top$ is called design matrix and $\boldsymbol{\varepsilon} = (\varepsilon_1, ..., \varepsilon_m)^\top$ denotes the vector of observational errors.

Now we consider properties of linear aspects, i.e. aspects of the form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta}$ for a specified $\tau \times p$ matrix $\mathbf{K}$, of the parameter $\boldsymbol{\beta}$ as well as of their estimators.

Definition 1. *An estimator $\hat{\boldsymbol{\Psi}}$ of a linear aspect $\boldsymbol{\Psi}$ is called linear if $\hat{\boldsymbol{\Psi}} = \mathbf{U}\mathbf{Y}$ for a specified $\tau \times m$ matrix $\mathbf{U}$.*

Definition 2. *An estimator $\hat{\boldsymbol{\Psi}}$ of a linear aspect $\boldsymbol{\Psi}$ is called unbiased if $\mathrm{E}(\hat{\boldsymbol{\Psi}}) = \boldsymbol{\Psi}$ for all $\boldsymbol{\beta}$.*

Definition 3. *A linear aspect $\boldsymbol{\Psi}$ is called (linear) estimable if there exists a linear unbiased estimator $\hat{\boldsymbol{\Psi}}$ of $\boldsymbol{\Psi}$.*

Now we search for the best estimator $\hat{\boldsymbol{\Psi}}$ of an estimable linear aspect $\boldsymbol{\Psi}$ in the class of linear unbiased estimators.

Definition 4. *A linear unbiased estimator $\hat{\boldsymbol{\Psi}}$ of an estimable linear aspect $\boldsymbol{\Psi}$ is called the best linear unbiased estimator for $\boldsymbol{\Psi}$ if for any linear unbiased estimator $\tilde{\boldsymbol{\Psi}}$ of $\boldsymbol{\Psi}$ the matrix*

$$\mathrm{Cov}\left(\tilde{\boldsymbol{\Psi}}\right) - \mathrm{Cov}\left(\hat{\boldsymbol{\Psi}}\right) \tag{2.3}$$

is non-negative definite.

Note that the best linear unbiased estimator $\hat{\boldsymbol{\Psi}}$ introduced in Definition 4 is unique.

Definition 5. *A generalized inverse of an $m \times n$ matrix $\mathbf{A}$ is any $n \times m$ matrix $\mathbf{A}^-$ such that*

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}. \tag{2.4}$$

The latter definition may be found for example in Harville (1997), ch. 9.

Due to the Gauss-Markov Theorem (see e.g. Christensen (2002), ch. 2) the best linear unbiased estimator of an estimable aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta}$ is given by

$$\hat{\boldsymbol{\Psi}} = \mathbf{K}(\mathbf{F}^\top\mathbf{F})^-\mathbf{F}^\top\mathbf{Y}, \tag{2.5}$$

where $(\mathbf{F}^\top\mathbf{F})^-$ is a generalized inverse of $\mathbf{F}^\top\mathbf{F}$. The estimator (2.5) has covarince matrix

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}}) = \sigma^2\mathbf{K}(\mathbf{F}^\top\mathbf{F})^-\mathbf{K}^\top. \tag{2.6}$$

If the design matrix $\mathbf{F}$ has full column rank, the vector of unknown parameters $\boldsymbol{\beta}$ is estimable with the best linear unbiased estimator

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \mathbf{Y}. \tag{2.7}$$

The covariance matrix of the estimator (2.7) is given by

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2 (\mathbf{F}^\top \mathbf{F})^{-1}. \tag{2.8}$$

2.1.2 Bayesian Models

Now we consider the Bayesian linear models with moment assumptions (see e.g. Pukelsheim (1993), ch. 11). We investigate the particular case of a constant variance of observational errors.

Bayesian linear models with moment assumptions have the general form (2.1):

$$Y_j = \mathbf{f}(x_j)^\top \boldsymbol{\theta} + \varepsilon_j, \quad j = 1, .., m, \quad x_j \in \mathcal{X}. \tag{2.9}$$

However, in (2.9) the unknown parameter $\boldsymbol{\theta}$ is a random vector with known expectation $\mathrm{E}(\boldsymbol{\theta}) = \boldsymbol{\theta}_0$ and covariance matrix $\mathrm{Cov}(\boldsymbol{\theta}) = \sigma^2 \mathbf{D}$. The dispersion matrix $\mathbf{D}$ is known and assumed to be non-singular. The regression functions $\mathbf{f}$ and the observational errors ε_j have the same properties as in the fixed effects model (2.1).

In the vector notation the model equation (2.9) results in

$$\mathbf{Y} = \mathbf{F}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2.10}$$

with $\mathbf{F}$ and $\boldsymbol{\varepsilon}$ as defined in Section 2.1.1.

Now we consider candidate estimators of a linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\theta}$ for a specified $\tau \times p$ matrix $\mathbf{K}$.

Definition 6. *An estimator $\hat{\boldsymbol{\Psi}}$ of a linear aspect $\boldsymbol{\Psi}$ is called unbiased if $\mathrm{E}(\hat{\boldsymbol{\Psi}}) = \mathrm{E}(\boldsymbol{\Psi})$.*

As we will see later there exists in the model (2.9) an unbiased estimator for every linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\theta}$.

Definition 7. *An estimator $\hat{\boldsymbol{\Psi}}$ of a linear aspect $\boldsymbol{\Psi}$ is called affine linear if $\hat{\boldsymbol{\Psi}} = \mathbf{A}\,\mathbf{Y} + \mathbf{b}$ for a specified $\tau \times m$ matrix $\mathbf{A}$ and a specified τ-dimensional vector $\mathbf{b}$.*

Definition 8. *An affine linear estimator* $\hat{\Psi}$ *of a linear aspect* Ψ *is called a Bayes estimator of* Ψ *if for any affine linear estimator* $\tilde{\Psi}$ *of* Ψ *the matrix*

$$\mathrm{E}\left((\tilde{\Psi} - \Psi)(\tilde{\Psi} - \Psi)^\top\right) - \mathrm{E}\left((\hat{\Psi} - \Psi)(\hat{\Psi} - \Psi)^\top\right) \tag{2.11}$$

is non-negative definite, i.e $\hat{\Psi}$ *has minimum mean squared error matrix in the non-negative definite sense.*

According to Pukelsheim (1993), ch. 11 the unique Bayes estimator of a linear aspect $\Psi = \mathbf{K}\boldsymbol{\theta}$ is given by

$$\hat{\Psi} = \mathbf{K}\left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1}(\mathbf{F}^\top \mathbf{Y} + \mathbf{D}^{-1}\boldsymbol{\theta}_0). \tag{2.12}$$

The Bayes estimator (2.12) is unbiased since $\mathrm{E}(\hat{\Psi}) = \mathbf{K}\boldsymbol{\theta}_0 = \mathrm{E}(\Psi)$. The mean squared error matrix of (2.12) is given by

$$\mathrm{Cov}(\hat{\Psi} - \Psi) = \sigma^2\,\mathbf{K}\left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1}\mathbf{K}^\top. \tag{2.13}$$

In the particular case $\Psi = \boldsymbol{\theta}$ the Bayes estimator simplifies to

$$\hat{\boldsymbol{\theta}} = (\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1})^{-1}(\mathbf{F}^\top \mathbf{Y} + \mathbf{D}^{-1}\boldsymbol{\theta}_0) \tag{2.14}$$

and its mean squared error matrix equals

$$\mathrm{Cov}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \sigma^2(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1})^{-1}. \tag{2.15}$$

2.2 Optimal Designs

In this section we want to characterize designs, i.e. collections of experimental settings $x_1, ..., x_m$, which are optimal for the estimation of the unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ in the models (2.1) and (2.9) respectively. The performance of the estimation may be measured by means of the covariance matrix (2.8) and the mean squared error matrix (2.15), which depend on the choice of the not necessarily distinct experimental settings $x_1, ..., x_m$. Let $x_1, ..., x_k$ be the distinct settings and $m_1, ..., m_k$ denote their numbers of replications with $\sum_{j=1}^{k} m_j = m$. Then the general form of an (exact) design is given by

$$\xi = \begin{pmatrix} x_1 & ,..., & x_k \\ m_1 & ,..., & m_k \end{pmatrix}. \tag{2.16}$$

Let $\Xi_{\mathcal{X}}$ be the set of all designs of the form (2.16) (exact designs) of size m.

Further we focus on the concept of approximate (continuous) designs in the sense of Kiefer (1974). These designs have the general form (2.16); however, the restriction of integer values for the replication numbers m_j is dropped and only the conditions $m_j \geq 0$ and $\sum_{j=1}^{k} m_j = m$ have to be satisfied. Let Ξ denote the set of all approximate designs of size m. Then it is easy to see that $\Xi_{\mathcal{X}} \subseteq \Xi$.

For any approximate design ξ the (standardized) information matrix in the model (2.1) is defined by

$$\mathbf{M}(\xi) = \frac{1}{m} \sum_{j=1}^{k} m_j \mathbf{f}(x_j) \mathbf{f}(x_j)^{\top} \tag{2.17}$$

(see e.g. Silvey (1980), ch. 3), which simplifies to the inverse of the covariance matrix (2.8) of the best linear unbiased estimator $\hat{\boldsymbol{\beta}}$

$$\mathbf{M}(\xi) = \frac{1}{m} \mathbf{F}^{\top} \mathbf{F}$$

in the case of an exact design, when we suppress the constant factor $\frac{\sigma^2}{m}$.

Further let

$$\boldsymbol{\Delta} = m \, \mathbf{D} \tag{2.18}$$

denote an appropriately adjusted version of the dispersion matrix $\mathbf{D}$ in the Bayesian model (2.9). With this notation the Bayesian information matrix of an approximate design is defined by

$$\mathrm{MSE}\,(\xi) = (\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1} \tag{2.19}$$

and coincides with (2.15) divided by the factor $\frac{\sigma^2}{m}$ for an exact design. Note that the mean squared error matrix (2.19) is defined for any information matrix $\mathbf{M}(\xi)$, i.e. for any design matrix $\mathbf{F}$.

2.2.1   *D*-Optimal Designs

The *D*-criterion for the estimation of the unknown parameter vector $\boldsymbol{\beta}$ in the classical linear regression model (2.1) is defined as the logarithm of the determinant of the inverse information matrix:

$$\mathrm{D}_{\beta}(\xi) = \ln \det \mathrm{M}(\xi)^{-1}, \tag{2.20}$$

where the information matrix $\mathbf{M}(\xi)$ is assumed to be non-singular.

For the estimation of the unknown parameter $\boldsymbol{\theta}$ in the Bayesian model (2.9) the Bayesian D-criterion is analogously defined as the logarithm of the determinant of the mean squared error matrix of the Bayes estimator $\hat{\boldsymbol{\theta}}$:

$$D_\theta(\xi) = \ln \det \mathrm{MSE}(\xi), \tag{2.21}$$

which results in the explicit form

$$D_\theta(\xi) = \ln \det(\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}. \tag{2.22}$$

Note that the Bayesian D-criterion (2.22) is defined for every information matrix $\mathbf{M}(\xi)$ because the dispersion matrix $\mathbf{D}$ has been assumed to be non-singular.

Then using the general equivalence theorem (see e.g. Silvey (1980), ch. 3) for approximate designs optimality conditions for the D-criterion in the fixed effects and the Bayesian models can be formulated.

Theorem 1. *The approximate design $\xi^* \in \Xi$ is D-optimal for the estimation in the fixed effects model if and only if*

$$\mathbf{f}(x)^\top \mathbf{M}(\xi^*)^{-1} \mathbf{f}(x) \le p \tag{2.23}$$

for all $x \in \mathcal{X}$. Moreover, for any support point x_j^* of ξ^* with positive weight $(m_j^* > 0)$ equality holds in (2.23).

The above result is also known as the Kiefer-Wolfowitz equivalence theorem (see Kiefer and Wolfowitz (1960)).

Theorem 2. *The approximate design $\xi^* \in \Xi$ is D-optimal for the estimation in the Bayesian model if and only if*

$$\mathbf{f}(x)^\top (\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1} \mathbf{f}(x) \le \mathrm{tr}\left((\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1} \mathbf{M}(\xi^*)\right) \tag{2.24}$$

for all $x \in \mathcal{X}$. Moreover, for any support point x_j^* of ξ^* with positive weight $(m_j^* > 0)$ equality holds in (2.24).

The result of Theorem 2 can be found in Gladitz and Pilz (1982) or Fedorov and Hackl (1997), ch. 5.

2.2.2 *L*-Optimal Designs

The classical linear criterion (*L*-criterion) in the fixed effects model (2.1) is defined for exact designs as the trace of the covariance matrix of the linear combination $\mathbf{K}\hat{\boldsymbol{\beta}}$ for a specified fixed $\tau \times p$ matrix $\mathbf{K}$:

$$\mathrm{L}_\beta(\xi) = \mathrm{tr}\left(\mathrm{Cov}(\mathbf{K}\hat{\boldsymbol{\beta}})\right). \tag{2.25}$$

In the Bayesian model (2.9) the *L*-criterion is similarly defined as the trace of the mean squared error matrix of the linear combination $\mathbf{K}\hat{\boldsymbol{\theta}}$:

$$\mathrm{L}_\theta(\xi) = \mathrm{tr}\left(\mathrm{Cov}\left(\mathbf{K}\hat{\boldsymbol{\theta}} - \mathbf{K}\boldsymbol{\theta}\right)\right). \tag{2.26}$$

For an approximate design ξ the linear criteria (2.25) and (2.26) can be generalized for $\mathcal{A} = \mathbf{K}^\top\mathbf{K}$ to

$$\mathrm{L}_\beta(\xi) = \mathrm{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{A}\right) \tag{2.27}$$

and

$$\mathrm{L}_\theta(\xi) = \mathrm{tr}\left(\mathrm{MSE}(\xi)\,\mathcal{A}\right) = \mathrm{tr}\left((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}\mathcal{A}\right) \tag{2.28}$$

respectively if we suppress the constant factor $\frac{\sigma^2}{m}$.

Due to the general equivalence theorem the following optimality condition for approximate designs in the fixed effects models can be formulated.

Theorem 3. *The approximate design $\xi^* \in \Xi$ is L-optimal with respect to $\mathcal{A}$ for the estimation in the fixed effects model if and only if*

$$\mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\mathcal{A}\,\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x) \leq \mathrm{tr}\left(\mathbf{M}(\xi^*)^{-1}\mathcal{A}\right) \tag{2.29}$$

for all $x \in \mathcal{X}$. Moreover, for any support point x_j^* of ξ^* with positive weight $(m_j^* > 0)$ equality holds in (2.29).

The optimality condition for the Bayesian model is given by the next theorem (see e.g. Gladitz and Pilz (1982)) .

Theorem 4. *The approximate design $\xi^* \in \Xi$ is L-optimal with respect to $\mathcal{A}$ for the estimation in the Bayesian model if and only if*

$$\mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\,\mathcal{A}\,\boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x) \tag{2.30}$$
$$\leq \mathrm{tr}\left(\boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\mathbf{M}(\xi^*)^{-1}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\,\mathcal{A}\right)$$

for all $x \in \mathcal{X}$. Moreover, for any support point x_j^* of ξ^* with positive weight $(m_j^* > 0)$ equality holds in (2.30).

The problem of computation of D- and L-optimal designs in the Bayesian model was considered in detail by Gladitz and Pilz (1982).

A commonly used particular case of the linear criterion is the integrated mean squared error (IMSE) criterion. For exact designs this criterion is defined as the integrated squared difference between the estimated and the real mean response:

$$\text{IMSE}_\beta(\xi) = \int_{\mathcal{X}} \text{Cov} \left(\mathbf{f}(x)^\top \hat{\boldsymbol{\beta}} \right) \nu(\mathrm{d}x) \qquad (2.31)$$

and

$$\text{IMSE}_\theta(\xi) = \int_{\mathcal{X}} \text{Cov} \left(\mathbf{f}(x)^\top \hat{\boldsymbol{\theta}} - \mathbf{f}(x)^\top \boldsymbol{\theta} \right) \nu(\mathrm{d}x) \qquad (2.32)$$

for the fixed effects and the Bayesian models respectively, where ν denotes some weight distribution on the design region $\mathcal{X}$. For approximate designs the criterion functions (2.31) and (2.32) can be generalized and equivalently presented as

$$\text{IMSE}_\beta(\xi) = \text{tr} \left(\mathbf{M}(\xi)^{-1} \mathcal{V} \right) \qquad (2.33)$$

and

$$\text{IMSE}_\theta(\xi) = \text{tr} \left(\text{MSE}(\xi) \mathcal{V} \right) = \text{tr} \left((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1} \mathcal{V} \right), \qquad (2.34)$$

where $\mathcal{V} = \int_{\mathcal{X}} \mathbf{f}(x) \mathbf{f}(x)^\top \nu(\mathrm{d}x)$ (see e.g. Liski *et al.* (2002), ch. 3, Fedorov and Leonov (2013), ch. 2).

3 Estimation and Prediction in Linear Mixed Models

In this chapter we present best linear unbiased estimation and prediction of linear aspects in linear mixed models. For the particular case of a full column rank fixed effects design matrix this problem was basically considered by Henderson *et al.* (1959), Henderson (1963) and Henderson (1975). Some results are also proved in Christensen (2002) (see also Isotalo *et al.* (2011)).

The general form of a linear mixed model is given by the following formula:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{3.1}$$

where $\mathbf{X}$ and $\mathbf{Z}$ are known non-zero matrices, $\boldsymbol{\beta}$ is an unknown vector of fixed effects, and $\boldsymbol{\gamma}$ is a vector of random effects. The random effects and the observational errors $\boldsymbol{\varepsilon}$ have zero mean and are uncorrelated. $\mathbf{G}$ and $\mathbf{R}$ are the corresponding non-singular covariance matrices: $\mathbf{G} = \mathrm{Cov}\,(\boldsymbol{\gamma})$ and $\mathbf{R} = \mathrm{Cov}\,(\boldsymbol{\varepsilon})$. We call the matrices $\mathbf{X}$ and $\mathbf{Z}$ fixed and random effects design matrices.

Further we consider linear aspects $\boldsymbol{\Psi}$ of $\left(\boldsymbol{\beta}^{\top}, \boldsymbol{\gamma}^{\top}\right)^{\top}$ of the form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$. Since $\boldsymbol{\Psi}$ includes the random term $\mathbf{L}\boldsymbol{\gamma}$, we will generally speak about prediction and predictable aspects (instead of estimation and estimable aspects) in this chapter. This notation is often used in the literature to distinguish between the fixed and the random case (see e.g. Christensen (2002), ch. 12). Only for special linear aspects of the form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta}$, which are independent of the random effects, we will use the concepts "estimation" and "estimable" as in Section 2.1.1.

Definition 9. *A predictor $\hat{\boldsymbol{\Psi}}$ of a linear aspect $\boldsymbol{\Psi}$ is called linear if $\hat{\boldsymbol{\Psi}} = \mathbf{U}\,\mathbf{Y}$ for some matrix $\mathbf{U}$.*

Definition 10. *A predictor $\hat{\boldsymbol{\Psi}}$ of a linear aspect $\boldsymbol{\Psi}$ is called unbiased if $\mathrm{E}(\hat{\boldsymbol{\Psi}}) = \mathrm{E}(\boldsymbol{\Psi})$ for all values of the population parameter $\boldsymbol{\beta}$.*

Definition 11. *A linear aspect $\boldsymbol{\Psi}$ is called (linear) predictable if there exists a linear unbiased predictor $\hat{\boldsymbol{\Psi}}$ of $\boldsymbol{\Psi}$.*

Further we search for the best predictor $\hat{\boldsymbol{\Psi}}$ of a predictable linear aspect $\boldsymbol{\Psi}$ in the class of linear unbiased predictors.

Definition 12. *A linear unbiased predictor* $\hat{\boldsymbol{\Psi}}$ *of a predictable linear aspect* $\boldsymbol{\Psi}$ *is called the best linear unbiased predictor for* $\boldsymbol{\Psi}$ *if for any linear unbiased predictor* $\tilde{\boldsymbol{\Psi}}$ *of* $\boldsymbol{\Psi}$ *the matrix*

$$\mathrm{Cov}\left(\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right) - \mathrm{Cov}\left(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right) \tag{3.2}$$

is non-negative definite ($\hat{\boldsymbol{\Psi}}$ *has minimum mean squared error matrix in the non-negative definite sense).*

Theorem 5. *A linear aspect* $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ *is predictable if and only if some matrix* $\mathbf{U}$ *exists with* $\mathbf{U}\mathbf{X} = \mathbf{K}$.

Proof. Since $\mathrm{E}(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{E}(\hat{\boldsymbol{\Psi}}) = \mathbf{K}\boldsymbol{\beta}$, we obtain

$$\begin{aligned}
\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma} \text{ predictable} \quad &\Leftrightarrow \quad \exists\, \hat{\boldsymbol{\Psi}} = \mathbf{U}\mathbf{Y} : \mathrm{E}(\hat{\boldsymbol{\Psi}}) = \mathrm{E}\left(\boldsymbol{\Psi}\right) \\
&\Leftrightarrow \quad \exists\, \mathbf{U} : \mathbf{U}\mathbf{X}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\beta}, \quad \forall\, \boldsymbol{\beta} \\
&\Leftrightarrow \quad \exists\, \mathbf{U} : \mathbf{U}\mathbf{X} = \mathbf{K}.
\end{aligned}$$

$\square$

Corollary 1. *If the fixed effects design matrix* $\mathbf{X}$ *has full column rank, all linear aspects of the form* $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ *are predictable.*

The last result follows directly from Theorem 5 since for $\mathbf{X}$ of full column rank the matrix $\mathbf{X}^\top\mathbf{X}$ is non-singular and $\mathbf{U} = \mathbf{K}\left(\mathbf{X}^\top\mathbf{X}\right)^{-1}\mathbf{X}^\top$ satisfies $\mathbf{U}\mathbf{X} = \mathbf{K}$.

3.1 Prediction in the Case of a Full Column Rank Fixed Effects Design Matrix

In this part the fixed effects design matrix $\mathbf{X}$ is assumed to have full column rank. Then according to Corollary 1 all linear aspects of the form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ are predictable. The fixed effects $\boldsymbol{\beta}$ may be estimated using Gauss-Markov theory: Since $\mathrm{E}\left(\mathbf{Y}\right) = \mathbf{X}\boldsymbol{\beta}$ and $\mathrm{Cov}\left(\mathbf{Y}\right) = \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R}$, we get the best linear unbiased estimator

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R})^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^\top(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R})^{-1}\mathbf{Y}. \tag{3.3}$$

Note that the estimator (3.3) is unique.

For the prediction of the random effects $\boldsymbol{\gamma}$ Henderson's mixed model equation

$$\begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Y} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Y} \end{pmatrix} \tag{3.4}$$

(see e.g. Henderson (1963)) may be used. Note that the matrix

$$\begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix} \tag{3.5}$$

is non-singular since the fixed effects design matrix $\mathbf{X}$ is of full column rank. Then the equation (3.4) may be rewritten as

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \mathbf{C} \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Y} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Y} \end{pmatrix} \tag{3.6}$$

for

$$\mathbf{C} = \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} & \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{pmatrix}^{-1} \tag{3.7}$$

and

$$\mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{12}^\top & \mathbf{C}_{22} \end{pmatrix} \tag{3.8}$$

can be partitioned accordingly, where $\mathbf{C}_{11}$ and $\mathbf{C}_{22}$ have the same dimensions as $\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X}$ and $\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1}$ respectively. By a standard inversion formula for partitioned matrices (see e.g. Harville (1997), ch. 8, p. 99), the components $\mathbf{C}_{11}$, $\mathbf{C}_{22}$ and $\mathbf{C}_{12}$ can be determined as

$$\mathbf{C}_{11} = \left(\mathbf{X}^\top \left(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R} \right)^{-1} \mathbf{X} \right)^{-1}, \tag{3.9}$$

$$\mathbf{C}_{22} = \left(\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \right)^{-1} \tag{3.10}$$

and

$$\begin{aligned} \mathbf{C}_{12} &= -\mathbf{C}_{11} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \right)^{-1} \\ &= -(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \, \mathbf{C}_{22}. \end{aligned} \tag{3.11}$$

Henderson *et al.* (1959) and Henderson (1963) have shown that (3.3) and

$$\hat{\boldsymbol{\gamma}} = \mathbf{G}\mathbf{Z}^\top (\mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R})^{-1} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \tag{3.12}$$

13

solve (3.4). According to Henderson (1975) (3.7) is the mean squared error matrix of $\left(\hat{\boldsymbol{\beta}}^{\top}, \hat{\boldsymbol{\gamma}}^{\top}\right)^{\top}$:

$$\mathbf{C} = \mathrm{Cov}\left(\begin{array}{c} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{array}\right) \tag{3.13}$$

with $\mathbf{C}_{11} = \mathrm{Cov}(\hat{\boldsymbol{\beta}})$ and $\mathbf{C}_{22} = \mathrm{Cov}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right)$.

The next theorem gives the best predictor of a predictable linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ of the fixed and random parameters in the model (3.1).

Theorem 6. *The best linear unbiased predictor for a linear aspect* $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ *is* $\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\boldsymbol{\gamma}}$.

The result of Theorem 6 is presented in Henderson (1963) (see also Christensen (2002), ch. 12).

Corollary 2. *The best linear unbiased estimator of the fixed effects* $\boldsymbol{\beta}$ *is* $\hat{\boldsymbol{\beta}}$ *and* $\hat{\boldsymbol{\gamma}}$ *is the best linear unbiased predictor of the random effects* $\boldsymbol{\gamma}$.

The next lemma gives an explicit form of the mean squared error matrix of the best linear unbiased predictor $\hat{\boldsymbol{\Psi}}$ represented in Theorem 6.

Lemma 1. *The mean squared error matrix of the best linear unbiased predictor* $\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\boldsymbol{\gamma}}$ *of a linear aspect* $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ *is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \mathbf{K}\,\mathbf{C}_{11}\,\mathbf{K}^{\top} + \mathbf{K}\,\mathbf{C}_{12}\,\mathbf{L}^{\top} + \mathbf{L}\,\mathbf{C}_{12}^{\top}\,\mathbf{K}^{\top} + \mathbf{L}\,\mathbf{C}_{22}\,\mathbf{L}^{\top}. \tag{3.14}$$

Proof. The result (3.14) follows from the linearity of the predictor $\hat{\boldsymbol{\Psi}} = (\mathbf{K}, \mathbf{L})\,(\hat{\boldsymbol{\beta}}^{\top}, \hat{\boldsymbol{\gamma}}^{\top})^{\top}$ and formula (3.13). $\square$

3.2 Prediction in the Case of a General Fixed Effects Design Matrix

In this section we consider the model (3.1) with some general (not necessarily full column rank) fixed effects design matrix $\mathbf{X}$. Since the matrix $\mathbf{C}$ from the equation (3.7) will not exist in this case, we define for any generalized inverse $(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}$ of $\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X}$ the matrices $\mathbf{C}'_{22}$, $\mathbf{C}'_{11}$ and $\mathbf{C}'_{12}$ by

$$\mathbf{C}'_{22} = \left(\mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{Z}\right)^{-1}, \tag{3.15}$$

$$\mathbf{C}'_{11} \;=\; (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- + (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z}\, \mathbf{C}'_{22}\, \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- \quad (3.16)$$

and

$$\mathbf{C}'_{12} \;=\; -(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z}\, \mathbf{C}'_{22}. \quad (3.17)$$

The matrices $\mathbf{C}'_{11}$ and $\mathbf{C}'_{12}$ can be represented in the following equivalent forms:

$$\begin{aligned}
\mathbf{C}'_{11} &= \left(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X} - \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \right)^{-1} \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X} \right)^- \\
&= \left(\mathbf{X}^\top \left(\mathbf{Z} \mathbf{G} \mathbf{Z}^\top + \mathbf{R} \right)^{-1} \mathbf{X} \right)^- \quad (3.18)
\end{aligned}$$

and

$$\mathbf{C}'_{12} \;=\; -\mathbf{C}'_{11} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \right)^{-1}. \quad (3.19)$$

Since for every non-singular covariance matrix $\mathbf{R}$ there exists a non-singular symmetric matrix $\mathbf{R}_1$ with $\mathbf{R}_1^\top \mathbf{R}_1 = \mathbf{R}$, the matrix $\mathbf{R}_1^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- \mathbf{X}^\top \mathbf{R}_1^{-1}$ is unique and idempotent as the projection matrix of $\mathbf{R}_1^{-1} \mathbf{X}$. Then also $\mathbf{I} - \mathbf{R}_1^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- \mathbf{X}^\top \mathbf{R}_1^{-1}$, where $\mathbf{I}$ denotes the identity matrix, is idempotent and nonnegative definite. The representation

$$\mathbf{C}'_{22} = \left(\mathbf{G}^{-1} + \mathbf{Z}^\top \mathbf{R}_1^{-1} \left(\mathbf{I} - \mathbf{R}_1^{-1} \mathbf{X} (\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^- \mathbf{X}^\top \mathbf{R}_1^{-1} \right) \mathbf{R}_1^{-1} \mathbf{Z} \right)^{-1}$$

establishes the uniqueness and positive definiteness of $\mathbf{C}'_{22}$. The matrices $\mathbf{C}'_{11}$ and $\mathbf{C}'_{12}$ depend on the particular generalized inverse and are in general not unique.

By the standard formula for generalized inverses of a partitioned matrix (see e.g Harville (1997), ch. 9, p. 121) the matrix

$$\mathbf{C}' = \begin{pmatrix} \mathbf{C}'_{11} & \mathbf{C}'_{12} \\ \mathbf{C}'^\top_{12} & \mathbf{C}'_{22} \end{pmatrix} \quad (3.20)$$

is a generalized inverse of

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

Further we consider instead of (3.6) the following candidate estimations and predictions of the fixed and random effects

$$\begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} \end{pmatrix} = \mathbf{C}' \begin{pmatrix} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Y} \\ \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Y} \end{pmatrix}, \quad (3.21)$$

which generally depend on the choice of $\mathbf{C}'$.

Lemma 2. *Solutions of the equation (3.21) are given by*

$$\hat{\boldsymbol{\beta}} \;=\; \left(\mathbf{X}^\top \left(\mathbf{ZGZ}^\top + \mathbf{R}\right)^{-1} \mathbf{X}\right)^{-} \mathbf{X}^\top \left(\mathbf{ZGZ}^\top + \mathbf{R}\right)^{-1} \mathbf{Y}, \tag{3.22}$$

$$\begin{aligned}
\hat{\boldsymbol{\gamma}} \;=\;& \left(\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X}(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-} \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Z}\right)^{-1} \\
& \cdot \left(\mathbf{Z}^\top \mathbf{R}^{-1} - \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X}(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-} \mathbf{X}^\top \mathbf{R}^{-1}\right) \mathbf{Y}. \tag{3.23}
\end{aligned}$$

Proof. From (3.21) it follows that

$$\hat{\boldsymbol{\beta}} = \mathbf{C}'_{11}\, \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Y} + \mathbf{C}'_{12}\, \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Y} \tag{3.24}$$

and

$$\hat{\boldsymbol{\gamma}} = \mathbf{C}'^{\top}_{12}\, \mathbf{X}^\top \mathbf{R}^{-1} \mathbf{Y} + \mathbf{C}'_{22}\, \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Y}. \tag{3.25}$$

Then using (3.19) it is easy to see that

$$\begin{aligned}
\hat{\boldsymbol{\beta}} \;=\;& \mathbf{C}'_{11}\, \mathbf{X}^\top \left(\mathbf{R}^{-1} - \mathbf{R}^{-1} \mathbf{Z} \left(\mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{Z} + \mathbf{G}^{-1}\right)^{-1} \mathbf{Z}^\top \mathbf{R}^{-1}\right) \mathbf{Y} \\
\;=\;& \mathbf{C}'_{11}\, \mathbf{X}^\top \left(\mathbf{ZGZ}^\top + \mathbf{R}\right)^{-1} \mathbf{Y}.
\end{aligned}$$

Applying (3.18) results in

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \left(\mathbf{ZGZ}^\top + \mathbf{R}\right)^{-1} \mathbf{X}\right)^{-} \mathbf{X}^\top \left(\mathbf{ZGZ}^\top + \mathbf{R}\right)^{-1} \mathbf{Y}.$$

From (3.25) and (3.17) it follows that

$$\hat{\boldsymbol{\gamma}} = \mathbf{C}'_{22} \left(\mathbf{Z}^\top \mathbf{R}^{-1} - \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X}(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-} \mathbf{X}^\top \mathbf{R}^{-1}\right) \mathbf{Y}.$$

After inserting (3.15) in the last equation we obtain (3.23). $\qquad\square$

Note that the prediction $\hat{\boldsymbol{\gamma}}$ of the random effects is the same for all generalized inverses $(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-}$. While the estimator $\hat{\boldsymbol{\beta}}$ of the fixed effects is not unique, it follows from properties of projection matrices and formula (3.22) that $\mathbf{X}\hat{\boldsymbol{\beta}}$ is independent of the choice of $(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-}$.

Lemma 3. *The mean squared error matrix of* $\left((\mathbf{X}\hat{\boldsymbol{\beta}})^\top,\ \hat{\boldsymbol{\gamma}}^\top\right)^\top$ *is given by*

$$\mathrm{Cov} \begin{pmatrix} \mathbf{X}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \mathbf{X}\,\mathbf{C}'_{11}\,\mathbf{X}^\top & \mathbf{X}\,\mathbf{C}'_{12} \\ \mathbf{C}'_{12}{}^\top \mathbf{X}^\top & \mathbf{C}'_{22} \end{pmatrix}. \tag{3.26}$$

Proof. For two random vectors $\mathbf{x} \in \mathbb{R}^k$ and $\mathbf{z} \in \mathbb{R}^l$ we define the covariance $\mathrm{Cov}\,(\mathbf{x}, \mathbf{z})$ of $\mathbf{x}$ and $\mathbf{z}$ as

$$\mathrm{Cov}\,(\mathbf{x}, \mathbf{z}) = \left(\mathrm{cov}\,(x_r,\, z_s)\right)_{r,s}, \quad r = 1, ..., k,\ \ s = 1, ..., l. \tag{3.27}$$

Then it is sufficient to establish that

$$\mathrm{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}}\right) = \mathbf{X}\,\mathbf{C}'_{11}\,\mathbf{X}^\top, \tag{3.28}$$

$$\mathrm{Cov}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \mathbf{C}'_{22} \tag{3.29}$$

and

$$\mathrm{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}},\, \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = \mathbf{X}\mathbf{C}'_{12}. \tag{3.30}$$

To verify the equality (3.28) we use the notation

$$\mathbf{V} := \mathbf{Z}\mathbf{G}\mathbf{Z}^\top + \mathbf{R} = \mathrm{Cov}(\mathbf{Y}). \tag{3.31}$$

Then

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right)^{-} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{Y} \tag{3.32}$$

and

$$\mathrm{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}}\right) = \mathbf{X}\left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right)^{-} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\left(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X}\right)^{-} \mathbf{X}^\top.$$

Using properties of projection matrices we get $\mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X} = \mathbf{X}$ and consequently

$$\mathrm{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}}\right) = \mathbf{X}(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-} \mathbf{X}^\top, \tag{3.33}$$

which results in (3.28) according to (3.18).

To establish the statement (3.29) we use the notation

$$\mathbf{A} := \mathbf{Z}^\top \mathbf{R}^{-1} - \mathbf{Z}^\top \mathbf{R}^{-1} \mathbf{X}(\mathbf{X}^\top \mathbf{R}^{-1} \mathbf{X})^{-} \mathbf{X}^\top \mathbf{R}^{-1}. \tag{3.34}$$

Then according to (3.15) and (3.23) we get $\mathbf{A}\,\mathbf{Z} = {\mathbf{C}'_{22}}^{-1} - \mathbf{G}^{-1}$ and

$$\hat{\boldsymbol{\gamma}} = \left(\mathbf{A}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1} \mathbf{A}\mathbf{Y} = \mathbf{C}'_{22}\mathbf{A}\mathbf{Y}. \tag{3.35}$$

The mean squared error matrix of the prediction $\hat{\boldsymbol{\gamma}}$ of the random effects $\boldsymbol{\gamma}$ may be represented in the form

$$\mathrm{Cov}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = \mathrm{Cov}\,(\hat{\boldsymbol{\gamma}}) - \mathrm{Cov}\,(\hat{\boldsymbol{\gamma}},\, \boldsymbol{\gamma}) - \mathrm{Cov}\,(\boldsymbol{\gamma},\, \hat{\boldsymbol{\gamma}}) + \mathrm{Cov}\,(\boldsymbol{\gamma}). \tag{3.36}$$

From (3.35) and (3.31) we obtain

$$\mathrm{Cov}\left(\hat{\boldsymbol{\gamma}}\right) = \mathbf{C}'_{22}\,\mathbf{A}\,\mathbf{Z}\,\mathbf{G}\,\mathbf{Z}^{\top}\mathbf{A}^{\top}\mathbf{C}'_{22} + \mathbf{C}'_{22}\,\mathbf{A}\,\mathbf{R}\,\mathbf{A}^{\top}\mathbf{C}'_{22}, \tag{3.37}$$

where the first term on the right hand side may be rewritten as

$$\begin{aligned}\mathbf{C}'_{22}\,\mathbf{A}\,\mathbf{Z}\,\mathbf{G}\,\mathbf{Z}^{\top}\mathbf{A}^{\top}\mathbf{C}'_{22} &= \mathbf{C}'_{22}\left(\mathbf{C}'_{22}{}^{-1} - \mathbf{G}^{-1}\right)\mathbf{G}\left(\mathbf{C}'_{22}{}^{-1} - \mathbf{G}^{-1}\right)\mathbf{C}'_{22} \\ &= \mathbf{G} - 2\,\mathbf{C}'_{22} + \mathbf{C}'_{22}\,\mathbf{G}^{-1}\mathbf{C}'_{22}. \end{aligned} \tag{3.38}$$

Using (3.35) and

$$\mathrm{Cov}\left(\mathbf{Y},\boldsymbol{\gamma}\right) = \mathbf{Z}\,\mathbf{G} \tag{3.39}$$

we get

$$\mathrm{Cov}\left(\hat{\boldsymbol{\gamma}},\boldsymbol{\gamma}\right) = \mathbf{C}'_{22}\,\mathbf{A}\,\mathbf{Z}\,\mathbf{G} = \mathbf{G} - \mathbf{C}'_{22}, \tag{3.40}$$

which is symmetric and consequently results in

$$\mathrm{Cov}\left(\boldsymbol{\gamma},\hat{\boldsymbol{\gamma}}\right) = \mathbf{G} - \mathbf{C}'_{22}. \tag{3.41}$$

Inserting $\mathrm{Cov}\left(\boldsymbol{\gamma}\right) = \mathbf{G}$, (3.37), (3.40) and (3.41) in formula (3.36) implies

$$\mathrm{Cov}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) = \mathbf{C}'_{22}\left(\mathbf{G}^{-1} + \mathbf{A}\,\mathbf{R}\,\mathbf{A}^{\top}\right)\mathbf{C}'_{22}. \tag{3.42}$$

Using (3.34) it can be easily shown that

$$\begin{aligned}\mathbf{A}\,\mathbf{R}\,\mathbf{A}^{\top} &= \mathbf{Z}^{\top}\mathbf{R}_1^{-1}\left(\mathbf{R}_1^{-1} - \mathbf{R}_1^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}^{-1}\right)\mathbf{R} \\ &\quad \cdot \left(\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}_1^{-1}\right)\mathbf{R}_1^{-1}\mathbf{Z} \end{aligned}$$

with the same $\mathbf{R}_1$ as in the proof of Lemma 2. Since the matrix $\mathbf{R}_1^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}_1^{-1}$ is idempotent as a projection matrix, we obtain $\mathbf{A}\,\mathbf{R}\,\mathbf{A}^{\top} = \mathbf{A}\,\mathbf{Z}$ and consequently

$$\mathbf{G}^{-1} + \mathbf{A}\,\mathbf{R}\,\mathbf{A}^{\top} = \mathbf{C}'_{22}{}^{-1},$$

which implies the result (3.29) according to (3.42).

Now the result (3.30) will be established. From (3.19) and (3.18) it follows that

$$\mathbf{X}\mathbf{C}'_{12} = -\mathbf{X}\left(\mathbf{X}^{\top}\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1}\right)^{-1}. \tag{3.43}$$

Using (3.32) and (3.35) we obtain

$$\text{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}},\,\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)=\mathbf{X}\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^-\mathbf{X}^\top\mathbf{V}^{-1}\left(\text{Cov}(\mathbf{Y})\mathbf{A}^\top\mathbf{C}_{22}'-\text{Cov}(\mathbf{Y},\boldsymbol{\gamma})\right),$$

which results in

$$\text{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}},\,\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)=\mathbf{X}\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^-\left(\mathbf{X}^\top\mathbf{A}^\top\mathbf{C}_{22}'-\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{Z}\mathbf{G}\right)$$

due to (3.31) and (3.39) and simplifies to

$$\text{Cov}\left(\mathbf{X}\,\hat{\boldsymbol{\beta}},\,\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)=-\mathbf{X}\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^-\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{Z}\mathbf{G}$$

since $\mathbf{A}\mathbf{X}=0$.

Using the standard formulas for an inverse of a sum of matrices (see e.g. Harville (1997), ch 18., p. 424) it can be shown that

$$\left(\mathbf{Z}\mathbf{G}\mathbf{Z}^\top+\mathbf{R}\right)^{-1}\mathbf{Z}\mathbf{G}=\mathbf{R}^{-1}\mathbf{Z}\left(\mathbf{Z}^\top\mathbf{R}^{-1}\mathbf{Z}+\mathbf{G}^{-1}\right)^{-1},$$

which proves the result (3.30) according to (3.43). □

Note that the mean squared error matrix (3.26) is unique since the matrices $\mathbf{X}\mathbf{C}_{11}'\mathbf{X}^\top$, $\mathbf{X}\mathbf{C}_{12}'$ and $\mathbf{C}_{22}'$ are independent of the choice of the generalized inverse $\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^-$.

Using $\text{E}(\hat{\boldsymbol{\beta}})=\boldsymbol{\beta}$, $\text{E}(\hat{\boldsymbol{\gamma}})=\mathbf{0}$, Definition 11 and Theorem 5 we can formulate the next statement:

Remark 1. *All linear aspects of the form* $\boldsymbol{\Psi}=\mathbf{S}\mathbf{X}\boldsymbol{\beta}+\mathbf{L}\boldsymbol{\gamma}$ *for a specified matrix* $\mathbf{S}$ *are predictable.*

Theorem 7. *The best linear unbiased predictor for a predictable aspect* $\boldsymbol{\Psi}=\mathbf{K}\boldsymbol{\beta}+\mathbf{L}\boldsymbol{\gamma}$ *is* $\hat{\boldsymbol{\Psi}}=\mathbf{K}\hat{\boldsymbol{\beta}}+\mathbf{L}\hat{\boldsymbol{\gamma}}$.

Proof. First, it will be shown that the predictor $\hat{\boldsymbol{\Psi}}$ is unbiased. According to Theorem 15 $\hat{\boldsymbol{\Psi}}$ may be rewritten in the form $\hat{\boldsymbol{\Psi}}=\mathbf{U}\mathbf{X}\hat{\boldsymbol{\beta}}+\mathbf{L}\hat{\boldsymbol{\gamma}}$. Then it follows from (3.22) and $\text{E}(\mathbf{Y})=\mathbf{X}\boldsymbol{\beta}$ that

$$\begin{aligned}\text{E}(\mathbf{X}\hat{\boldsymbol{\beta}})&=\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^\top\mathbf{V}^{-1}\text{E}(\mathbf{Y})\\&=\mathbf{X}(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X})^-\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\boldsymbol{\beta},\end{aligned}$$

19

which results in

$$\mathrm{E}(\mathbf{X}\hat{\boldsymbol{\beta}}) = \mathbf{X}\boldsymbol{\beta}$$

according to properties of projection matrices. Analogously we obtain from (3.23) that

$$
\begin{aligned}
\mathrm{E}(\hat{\boldsymbol{\gamma}}) ={}& \left(\mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{Z} + \mathbf{G}^{-1} - \mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{Z}\right)^{-1} \\
&\cdot \left(\mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{X} - \mathbf{Z}^{\top}\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X}\right)\boldsymbol{\beta}.
\end{aligned}
$$

Since $\mathbf{R}^{-1}\mathbf{X}(\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X})^{-}\mathbf{X}^{\top}\mathbf{R}^{-1}\mathbf{X} = \mathbf{R}^{-1}\mathbf{X}$, the expected value of $\hat{\boldsymbol{\gamma}}$ is equal to zero, which implies the unbiasedness of the prediction:

$$\mathrm{E}(\hat{\boldsymbol{\Psi}}) = \mathbf{U}\,\mathbf{X}\boldsymbol{\beta} = \mathbf{K}\boldsymbol{\beta} = \mathrm{E}(\boldsymbol{\Psi}).$$

Then according to Definition 11 it has to be established that $\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\boldsymbol{\gamma}}$ has minimum mean squared error matrix in the class of linear unbiased predictors for $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$. First we prove this statement for one-dimensional aspects: $\psi = \mathbf{k}^{\top}\boldsymbol{\beta} + \mathbf{l}^{\top}\boldsymbol{\gamma}$, where $\mathbf{k}$ and $\mathbf{l}$ are specified vectors ($\mathbf{K} = \mathbf{k}^{\top}$ and $\mathbf{L} = \mathbf{l}^{\top}$), and $\hat{\psi} = \mathbf{k}^{\top}\hat{\boldsymbol{\beta}} + \mathbf{l}^{\top}\hat{\boldsymbol{\gamma}}$. Now let $\tilde{\psi}$ be some linear unbiased predictor of ψ, which means that $\tilde{\psi} = \mathbf{u}^{\top}\mathbf{Y}$ for some vector $\mathbf{u}$. Then it follows from Definition 10 and Theorem 5 that $\mathbf{u}^{\top}\mathbf{X} = \mathbf{k}^{\top}$.

Further the following property has to be verified:

$$\mathrm{var}\left(\tilde{\psi} - \psi\right) - \mathrm{var}\left(\hat{\psi} - \psi\right) \geq 0.$$

Using properties of the variance and the covariance it can be easily proved that

$$\mathrm{var}\left(\tilde{\psi} - \psi\right) - \mathrm{var}\left(\hat{\psi} - \psi\right) = \mathrm{var}\left(\tilde{\psi} - \hat{\psi}\right) + 2\,\mathrm{cov}\left(\tilde{\psi} - \hat{\psi},\, \hat{\psi} - \psi\right). \qquad (3.44)$$

Then it is sufficient to show that

$$\mathrm{cov}\left(\tilde{\psi} - \hat{\psi},\, \hat{\psi} - \psi\right) = 0. \qquad (3.45)$$

For this we introduce the following matrices:

$$\mathbf{S}' = \left(\mathbf{X}^{\top}\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}^{\top}\mathbf{V}^{-1} \qquad (3.46)$$

and

$$\mathbf{T}' = \mathbf{G}\mathbf{Z}^{\top}\mathbf{V}^{-1}(\mathbf{I} - \mathbf{X}\mathbf{S}'), \qquad (3.47)$$

20

where
$$\mathbf{V} = \mathbf{ZGZ}^\top + \mathbf{R}.$$

Then we obtain $\hat{\boldsymbol{\beta}} = \mathbf{S}'\mathbf{Y}$ and $\hat{\boldsymbol{\gamma}} = \mathbf{T}'\mathbf{Y}$ respectively. Since

$$\mathbf{X}\left(\mathbf{I} - \mathbf{S}'\mathbf{X}\right) = \mathbf{X} - \mathbf{X}\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X} = \mathbf{0}, \tag{3.48}$$

where $\mathbf{0}$ denotes a zero matrix, it is easy to see that $\mathbf{T}'\mathbf{X} = \mathbf{0}$.

Since the prediction $\hat{\boldsymbol{\Psi}}$ can be represented as $\hat{\boldsymbol{\Psi}} = (\mathbf{k}^\top\mathbf{S}' + \mathbf{l}^\top\mathbf{T}')\mathbf{Y}$, we obtain the following equation:

$$\mathrm{cov}\left(\tilde{\psi} - \hat{\psi},\, \hat{\psi} - \psi\right) = \mathrm{cov}\left(\left(\mathbf{u}^\top - \mathbf{k}^\top\mathbf{S}' - \mathbf{l}^\top\mathbf{T}'\right)\mathbf{Y},\, \left(\mathbf{k}^\top\mathbf{S}' + \mathbf{l}^\top\mathbf{T}'\right)\mathbf{Y} - \mathbf{l}^\top\boldsymbol{\gamma}\right).$$

Then we get from $\mathrm{Cov}\left(\mathbf{Y}\right) = \mathbf{V}$ and $\mathrm{Cov}\left(\mathbf{Y}, \boldsymbol{\gamma}\right) = \mathbf{Z}\,\mathbf{G}$ the equation

$$\mathrm{cov}\left(\tilde{\psi} - \hat{\psi},\, \hat{\psi} - \psi\right) = \left(\mathbf{u}^\top - \mathbf{k}^\top\mathbf{S}' - \mathbf{l}^\top\mathbf{T}'\right)\left(\mathbf{V}\left(\mathbf{k}^\top\mathbf{S}' + \mathbf{l}^\top\mathbf{T}'\right)^\top - \mathbf{Z}\,\mathbf{G}\mathbf{l}\right). \tag{3.49}$$

From (3.46) and (3.47) it follows that

$$\begin{aligned}
\mathbf{V}\left(\mathbf{k}^\top\mathbf{S}' + \mathbf{l}^\top\mathbf{T}'\right)^\top - \mathbf{Z}\,\mathbf{G}\mathbf{l} &= \mathbf{V}\mathbf{S}'^\top\mathbf{k} + \mathbf{V}\mathbf{T}'^\top\mathbf{l} - \mathbf{Z}\,\mathbf{G}\mathbf{l} \\
&= \mathbf{X}\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^{-}\mathbf{k} + \mathbf{V}\left(\mathbf{I} - \mathbf{S}'^\top\mathbf{X}^\top\right)\mathbf{V}^{-1}\mathbf{Z}\,\mathbf{G}\mathbf{l} - \mathbf{Z}\,\mathbf{G}\mathbf{l}.
\end{aligned}$$

Applying (3.46) again results in

$$\mathbf{V}\left(\mathbf{k}^\top\mathbf{S}' + \mathbf{l}^\top\mathbf{T}'\right)^\top - \mathbf{Z}\,\mathbf{G}\mathbf{l} = \mathbf{X}\left(\mathbf{X}^\top\mathbf{V}^{-1}\mathbf{X}\right)^{-}\left(\mathbf{k} - \mathbf{X}^\top\mathbf{V}^{-1}\mathbf{Z}\mathbf{G}\mathbf{l}\right). \tag{3.50}$$

Using $\mathbf{k}^\top = \mathbf{u}^\top\mathbf{X}$ and $\mathbf{T}'\mathbf{X} = \mathbf{0}$ we obtain

$$\left(\mathbf{u}^\top - \mathbf{k}^\top\mathbf{S}' - \mathbf{l}^\top\mathbf{T}'\right)\mathbf{X} = \mathbf{u}^\top\mathbf{X}\left(\mathbf{I} - \mathbf{S}'\mathbf{X}\right) = 0$$

according to (3.48). Then the condition (3.45) follows from (3.50) and (3.49). Note that if $\hat{\psi} \neq \tilde{\psi}$, $\mathrm{var}\left(\tilde{\psi} - \hat{\psi}\right) > 0$. Then due to (3.44) $\mathrm{var}\left(\tilde{\psi} - \psi\right) - \mathrm{var}\left(\hat{\psi} - \psi\right) > 0$, which shows that $\tilde{\psi}$ is not the best linear unbiased predictor of ψ. Consequently, the best linear unbiased predictor $\hat{\psi}$ is unique.

Now we come back to the general case, where $\mathbf{K}$ and $\mathbf{L}$ are specified matrices with two or more rows. Let $\tilde{\boldsymbol{\Psi}} = \mathbf{U}\mathbf{Y}$ be some linear unbiased predictor of $\boldsymbol{\Psi}$. Note that

$\tilde{\boldsymbol{\Psi}}$ is unbiased, which establishes $\mathbf{UX} = \mathbf{K}$. Then it has to be verified that the matrix $\mathrm{Cov}\left(\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right) - \mathrm{Cov}\left(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)$ is non-negative definite, i.e.

$$\mathbf{a}^\top \left(\mathrm{Cov}\left(\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right) - \mathrm{Cov}\left(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)\right) \mathbf{a} \geq 0$$

for every vector $\mathbf{a}$. Note that

$$\mathbf{a}^\top \left(\mathrm{Cov}\left(\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right) - \mathrm{Cov}\left(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)\right) \mathbf{a} = \mathrm{var}\left(\mathbf{a}^\top \left(\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)\right) - \mathrm{var}\left(\mathbf{a}^\top \left(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)\right).$$

Now we consider the aspect $\psi = \mathbf{a}^\top \boldsymbol{\Psi} = \mathbf{a}^\top (\mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma})$, for which $\hat{\psi} = \mathbf{a}^\top \hat{\boldsymbol{\Psi}} = \mathbf{a}^\top \left(\mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\boldsymbol{\gamma}}\right)$ and $\tilde{\psi} = \mathbf{a}^\top \tilde{\boldsymbol{\Psi}} = \mathbf{a}^\top \mathbf{U}\mathbf{Y}$. Note also that $\mathbf{a}^\top \mathbf{UX} = \mathbf{a}^\top \mathbf{K}$. Then we obtain from the one-dimensional case with $\mathbf{k}^\top = \mathbf{a}^\top \mathbf{K}$, $\mathbf{l}^\top = \mathbf{a}^\top \mathbf{L}$ and $\mathbf{u}^\top = \mathbf{a}^\top \mathbf{U}$ the inequality

$$\mathrm{var}\left(\mathbf{a}^\top \left(\tilde{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)\right) - \mathrm{var}\left(\mathbf{a}^\top \left(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}\right)\right) = \mathrm{var}\left(\tilde{\psi} - \psi\right) - \mathrm{var}\left(\hat{\psi} - \psi\right) \geq 0.$$

$\square$

Note that according to Remark 1 the linear aspect $\mathbf{X}\boldsymbol{\beta}$ of the fixed effects $\boldsymbol{\beta}$ and the random effects $\boldsymbol{\gamma}$ are predictable.

Corollary 3. *The best linear unbiased estimator of $\mathbf{X}\boldsymbol{\beta}$ is $\mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\gamma}}$ is the best linear unbiased predictor of $\boldsymbol{\gamma}$.*

The last result is also presented in Christensen (2002).

Theorem 8. *The mean squared error matrix of the best linear unbiased predictor $\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\boldsymbol{\gamma}}$ of a predictable linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \mathbf{K}\,\mathbf{C}'_{11}\,\mathbf{K}^\top + \mathbf{K}\,\mathbf{C}'_{12}\,\mathbf{L}^\top + \mathbf{L}\,\mathbf{C}'^\top_{12}\,\mathbf{K}^\top + \mathbf{L}\,\mathbf{C}'_{22}\,\mathbf{L}^\top. \qquad (3.51)$$

Proof. Using the property $\mathbf{K} = \mathbf{UX}$ of a predictable linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ we obtain $\hat{\boldsymbol{\Psi}} = \mathbf{UX}\hat{\boldsymbol{\beta}} + \mathbf{L}\hat{\boldsymbol{\gamma}}$ and

$$\begin{aligned}
\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) &= \mathrm{Cov}\left(\mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{L}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right) \\
&= \mathrm{Cov}\left(\mathbf{UX}\hat{\boldsymbol{\beta}} + \mathbf{L}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})\right) \\
&= (\mathbf{U}, \mathbf{L})\,\mathrm{Cov}\begin{pmatrix} \mathbf{X}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{pmatrix} (\mathbf{U}, \mathbf{L})^\top. \qquad (3.52)
\end{aligned}$$

Then it follows directly from the result (3.26) of Lemma 3 that

$$\mathrm{Cov}\begin{pmatrix} \mathbf{X}\hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma} \end{pmatrix} = \begin{pmatrix} \mathbf{X} & 0 \\ 0 & \mathbf{I} \end{pmatrix} \mathbf{C}' \begin{pmatrix} \mathbf{X}^{\top} & 0 \\ 0 & \mathbf{I} \end{pmatrix},$$

which implies

$$\begin{aligned} \mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) &= (\mathbf{U}\mathbf{X}, \mathbf{L})\, \mathbf{C}' \,(\mathbf{U}\mathbf{X}, \mathbf{L})^{\top} \\ &= (\mathbf{K}, \mathbf{L})\, \mathbf{C}' \,(\mathbf{K}, \mathbf{L})^{\top} \\ &= \mathbf{K}\,\mathbf{C}'_{11}\,\mathbf{K}^{\top} + \mathbf{K}\,\mathbf{C}'_{12}\,\mathbf{L}^{\top} + \mathbf{L}\,{\mathbf{C}'_{12}}^{\top}\,\mathbf{K}^{\top} + \mathbf{L}\,\mathbf{C}'_{22}\,\mathbf{L}^{\top}. \end{aligned}$$

$\square$

The result of Theorem 8 is also given in Henderson (1975) without proof.

4 Estimation and Prediction in Hierarchical Random Coefficient Regression Models

This chapter provides best linear unbiased estimators and predictors of linear aspects in hierarchical random coefficient regression models. The problem of estimation of the mean population parameters has been widely considered in the literature (see e.g. Rao (1965), Spjotvoll (1977)). For the prediction of the individual parameters some results are presented in Fedorov and Jones (2005), Fedorov and Hackl (1997), Candel (2009) and Fedorov and Leonov (2013).

4.1 Hierarchical Random Coefficient Regression Models: Model Formulation

In this chapter we consider hierarchical random coefficient regression models, for which on the individual level the j-th observation of the i-th individual is given by the formula

$$Y_{ij} = \mathbf{f}(x_{ij})^\top \boldsymbol{\beta}_i + \varepsilon_{ij}, \quad i = 1, .., n, \quad j = 1, .., m_i, \quad x_{ij} \in \mathcal{X}, \tag{4.1}$$

where n is the number of individuals, m_i is the number of observations at individual i, experimental settings x_{ij} come from the experimental region $\mathcal{X}$. $\mathbf{f} = (f_1, .., f_p)^\top :$ $\mathcal{X} \to \mathbb{R}^p$ is a vector of known regression functions. The image $\mathbf{f}(\mathcal{X}) \subset \mathbb{R}^p$ of the experimental region $\mathcal{X}$ is assumed to be a compact subset of $\mathbb{R}^p$. Observational errors ε_{ij} are assumed to be uncorrelated, have zero mean and a common variance $\sigma^2 > 0$. Individual parameters (individual random effects) $\boldsymbol{\beta}_i = (\beta_{i1}, .., \beta_{ip})^\top$ have unknown (population) mean $\mathrm{E}(\boldsymbol{\beta}_i) = \boldsymbol{\beta} = (\beta_1, .., \beta_p)^\top \in \mathbb{R}^p$ and (population) covariance matrix $\mathrm{Cov}(\boldsymbol{\beta}_i) = \sigma^2 \mathbf{D} \in \mathbb{R}^{p \times p}$ with a given (dispersion) matrix $\mathbf{D}$. Individual parameters for different individuals are uncorrelated with each other and with all observational errors. The dispersion matrix $\mathbf{D}$ may be singular if, for example, some individual parameters $\boldsymbol{\beta}_{i\ell}$ are equal for all individuals (for all $i = 1, ..., n$) and consequently fixed (since $\boldsymbol{\beta}_{i\ell}$ and $\boldsymbol{\beta}_{i'\ell}$ are uncorrelated for $i \neq i'$).

Note that in contrast to the Bayesian model (2.9), considered in section 2.1.2, in the random coefficient regression model (4.1) the expectation $\boldsymbol{\beta}$ of the (individual) random

parameters $\boldsymbol{\beta}_i$ is unknown, which leads to completely different results for optimal designs as it will be shown in the next chapter.

Further a particular case of the model (4.1) will be considered, where the number of observations as well as the experimental settings are the same for all individuals ($m_i = m$, $x_{ij} = x_j$):

$$Y_{ij} = \mathbf{f}(x_j)^\top \boldsymbol{\beta}_i + \varepsilon_{ij}, \quad i = 1, .., n, \quad j = 1, .., m, \quad x_j \in \mathcal{X}. \tag{4.2}$$

In vector notation this model may be represented as

$$\mathbf{Y}_i = \mathbf{F}\boldsymbol{\beta}_i + \boldsymbol{\varepsilon}_i, \tag{4.3}$$

where $\mathbf{Y}_i = (Y_{i1}, ..., Y_{im})^\top$ is the individual vector of observations at the i-th individual, $\mathbf{F} = (\mathbf{f}(x_1), ..., \mathbf{f}(x_m))^\top$ is the individual design matrix, which is the same for all individuals and coincides with the design matrix in the fixed effects model (2.2), $\boldsymbol{\varepsilon}_i = (\varepsilon_{i1}, ..., \varepsilon_{im})^\top$ denotes the corresponding vector of observational errors.

Now the random coefficient regression model (4.2) will be represented as a special case of the linear mixed model (3.1). For this we introduce the centered random effects (individual deviations) $\mathbf{b}_i := \boldsymbol{\beta}_i - \boldsymbol{\beta}$, which separate the individual random effects from the population mean. Then the model (4.2) can be rewritten as

$$Y_{ij} = \mathbf{f}(x_j)^\top \boldsymbol{\beta} + \mathbf{f}(x_j)^\top \mathbf{b}_i + \varepsilon_{ij} , \tag{4.4}$$

which results in the vector notation

$$\mathbf{Y}_i = \mathbf{F}\boldsymbol{\beta} + \mathbf{F}\mathbf{b}_i + \boldsymbol{\varepsilon}_i. \tag{4.5}$$

Further we use the notation $\mathbf{B} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_n^\top)^\top$ and $\mathbf{b} = (\mathbf{b}_1^\top, ..., \mathbf{b}_n^\top)^\top$ for the vectors of all individual parameters and all individual deviations and $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, ..., \boldsymbol{\varepsilon}_n^\top)^\top$ denotes the corresponding error vector. Then the complete observational vector $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_n^\top)^\top$ for all individuals has the form

$$\mathbf{Y} = (\mathbf{I}_n \otimes \mathbf{F})\mathbf{B} + \boldsymbol{\varepsilon} \tag{4.6}$$

according to (4.3) or alternatively (according to (4.5))

$$\mathbf{Y} = (\mathbf{1}_n \otimes \mathbf{F})\boldsymbol{\beta} + (\mathbf{I}_n \otimes \mathbf{F})\mathbf{b} + \boldsymbol{\varepsilon}, \tag{4.7}$$

where $\mathbf{I}_n$ is the $n \times n$ identity matrix, $\mathbf{1}_n = (1, .., 1)^\top$ is the vector of length n with all components equal to 1, " $\otimes$ " denotes the Kronecker product.

Note that the expected value and the covariance matrix of the full observational vector $\mathbf{Y}$ are given by the following formulas:

$$E(\mathbf{Y}) = (\mathbf{1}_n \otimes \mathbf{F})\boldsymbol{\beta},$$

$$\mathrm{Cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_n \otimes (\mathbf{FDF}^\top + \mathbf{I}_m).$$

The individual deviations $\mathbf{b}$ from formula (4.7) have zero mean and their covariance matrix $\mathrm{Cov}\,(\mathbf{b}) = \sigma^2\,\mathbf{I}_n \otimes \mathbf{D}$ is only non-singular if the dispersion matrix $\mathbf{D}$ of the individual deviations has full rank. Therefore, it is convenient to rewrite the model (4.7) in some other form, where the dispersion matrix of random effects in non-singular. Let $q \leq p$ be the rank of the matrix $\mathbf{D}$. Then there exists a $p \times q$ matrix $\mathbf{H}$ with rank q and $\mathbf{D} = \mathbf{HH}^\top$. (In this case the matrix $\mathbf{H}^\top\mathbf{H}$ is non-singular). Now we introduce the random variables

$$\boldsymbol{\gamma}_i := (\mathbf{H}^\top\mathbf{H})^{-1}\mathbf{H}^\top\mathbf{b}_i, \tag{4.8}$$

which satisfy $\mathbf{b}_i = \mathbf{H}\boldsymbol{\gamma}_i$ and have zero expected value and the covariance matrix $\mathrm{Cov}\,(\boldsymbol{\gamma}_i) = \sigma^2\,\mathbf{I}_q$. The latter allows the following representations of the model:

$$\mathbf{Y}_i = \mathbf{F}\boldsymbol{\beta} + \mathbf{FH}\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i \tag{4.9}$$

on the individual level and

$$\mathbf{Y} = (\mathbf{1}_n \otimes \mathbf{F})\boldsymbol{\beta} + (\mathbf{I}_n \otimes (\mathbf{FH}))\boldsymbol{\gamma} + \boldsymbol{\varepsilon} \tag{4.10}$$

on the population level for $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, .., \boldsymbol{\gamma}_n^\top)^\top$.

Now it is easy to see that the model (4.10) satisfies all the conditions of the linear mixed model (3.1) with $E\,(\boldsymbol{\gamma}) = 0$, $\mathbf{G} = \mathrm{Cov}\,(\boldsymbol{\gamma}) = \sigma^2\,\mathbf{I}_{nq}$ and $\mathbf{R} = \mathrm{Cov}\,(\boldsymbol{\varepsilon}) = \sigma^2\,\mathbf{I}_{nm}$. Consequently, for $\mathbf{X} = \mathbf{1}_n \otimes \mathbf{F}$ and $\mathbf{Z} = \mathbf{I}_n \otimes (\mathbf{FH})$ (4.10) is of the form (3.1).

4.2 Predictable Linear Aspects in Hierarchical Random Coefficient Regression Models

In this section linear aspects of the form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{Tb}$ will be investigated. Note that the dispersion matrix $\mathbf{D}$ of random effects may be singular, and then the corresponding

regression model (4.7) will not satisfy the conditions of the linear mixed model (3.1). Therefore, we will also use the alternative form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{L}\boldsymbol{\gamma}$ with $\mathbf{L} = \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H})$ of the linear aspect $\boldsymbol{\Psi}$, which allows to apply the theory developed in the previous chapter.

Lemma 4. *Let* $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{Tb}$ *be a linear aspect in the hierarchical random coefficient regression model. Then it holds*

a) $\boldsymbol{\Psi}$ *is predictable if and only if some* $\tau \times mn$ *matrix* $\mathbf{U}$ *exists with* $\mathbf{U}(\mathbf{1}_n \otimes \mathbf{F}) = \mathbf{K}$;

b) *If* $\mathbf{K} = \mathbf{0}$, $\boldsymbol{\Psi}$ *is predictable for all* $\mathbf{T}$;

c) *If the fixed effects design matrix* $\mathbf{F}$ *has full column rank,* $\boldsymbol{\Psi}$ *is predictable for all* $\mathbf{K}$ *and all* $\mathbf{T}$;

d) *If* $\mathbf{K} = \mathbf{I}_p$, $\boldsymbol{\Psi}$ *is predictable if and only if the fixed effects design matrix* $\mathbf{F}$ *has full column rank.*

Proof.   a) The proof follows directly from Theorem 5 for $\mathbf{X} = \mathbf{1}_n \otimes \mathbf{F}$.

b) The proof follows directly from Remark 1 for $\mathbf{b} = (\mathbf{I}_n \otimes \mathbf{H})\boldsymbol{\gamma}$.

c) For $\mathbf{X} = \mathbf{1}_n \otimes \mathbf{F}$ the aspect $\boldsymbol{\Psi}$ is predictable according to Corollary 1.

d) If $\mathbf{F}$ has full column rank, $\boldsymbol{\Psi}$ is predictable according to part c).

If $\boldsymbol{\Psi}$ is predictable, then there exists some matrix $\mathbf{U}$ with $\mathbf{U}(\mathbf{1}_n \otimes \mathbf{F}) = \mathbf{I}_p$. Consequently, the rank of $\mathbf{U}(\mathbf{1}_n \otimes \mathbf{F})$ equals p, which is only possible if the design matrix $\mathbf{F}$ has full column rank.

$\square$

For further considerations of predictable linear aspects we will use the following properties of sums of matrices, which can be derived using the standard inversion formula (see e.g. Harville (1997), ch. 18, p. 424):

$$\left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1} = \mathbf{I}_m - \mathbf{FH}\left(\mathbf{H}^\top\mathbf{F}^\top\mathbf{FH} + \mathbf{I}_q\right)^{-1}\mathbf{H}^\top\mathbf{F}^\top \tag{4.11}$$

and

$$\left(\mathbf{H}^\top\mathbf{F}^\top\mathbf{FH} + \mathbf{I}_q\right)^{-1} = \mathbf{I}_q - \mathbf{H}^\top\mathbf{F}^\top\left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1}\mathbf{FH}, \tag{4.12}$$

which implies

$$\mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q\right)^{-1} \mathbf{H}^\top = \mathbf{D} - \mathbf{D} \mathbf{F}^\top \left(\mathbf{F} \mathbf{D} \mathbf{F}^\top + \mathbf{I}_m\right)^{-1} \mathbf{F} \mathbf{D}. \tag{4.13}$$

For a non-singular dispersion matrix $\mathbf{D}$, we obtain by the simplified version of the standard inversion formula of sums of matrices (see e.g. Harville (1997), ch. 18, p. 424)

$$\left(\mathbf{F} \mathbf{D} \mathbf{F}^\top + \mathbf{I}_m\right)^{-1} = \mathbf{I}_m - \mathbf{F} \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1} \mathbf{F}^\top. \tag{4.14}$$

Consequently, (4.13) simplifies to

$$\mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q\right)^{-1} \mathbf{H}^\top = \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1}. \tag{4.15}$$

If the information matrix $\mathbf{F}^\top \mathbf{F}$ in the corresponding fixed effects model is non-singular, i.e. the fixed effects design matrix $\mathbf{F}$ has full column rank, we obtain

$$\left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q\right)^{-1} = \mathbf{I}_q - \mathbf{H}^\top \left(\left(\mathbf{F}^\top \mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1} \mathbf{H}, \tag{4.16}$$

which implies

$$\mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q\right)^{-1} \mathbf{H}^\top = \mathbf{D} - \mathbf{D} \left(\left(\mathbf{F}^\top \mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1} \mathbf{D} \tag{4.17}$$

and results in

$$\mathbf{D} - \mathbf{D} \left(\left(\mathbf{F}^\top \mathbf{F}\right)^{-1} + \mathbf{D}\right)^{-1} \mathbf{D} = \left(\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1}\right)^{-1} \tag{4.18}$$

if both matrices $\mathbf{D}$ and $\mathbf{F}^\top \mathbf{F}$ are non-singular.

Theorem 9. *The best linear unbiased predictor of a predictable linear aspect* $\boldsymbol{\Psi} = \mathbf{K} \boldsymbol{\beta} + \mathbf{T} \mathbf{b}$ *is given by* $\hat{\boldsymbol{\Psi}} = \mathbf{K} \hat{\boldsymbol{\beta}} + \mathbf{T} \hat{\mathbf{b}}$, *where*

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{F}^\top \left(\mathbf{F} \mathbf{D} \mathbf{F}^\top + \mathbf{I}_m\right)^{-1} \mathbf{F}\right)^- \mathbf{F}^\top \left(\mathbf{F} \mathbf{D} \mathbf{F}^\top + \mathbf{I}_m\right)^{-1} \bar{\mathbf{Y}} \tag{4.19}$$

for $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{Y}_i$ *and*

$$\hat{\mathbf{b}} = \left(\left(\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top\right) \otimes \left(\mathbf{D} \mathbf{F}^\top \left(\mathbf{F} \mathbf{D} \mathbf{F}^\top + \mathbf{I}_m\right)^{-1}\right)\right) \mathbf{Y}. \tag{4.20}$$

Proof. From Theorem 7 and (3.22) it follows that

$$\hat{\boldsymbol{\beta}} = \left(\frac{1}{n}\mathbf{1}_n^\top \otimes \left(\left(\mathbf{F}^\top \left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1} \mathbf{F}\right)^- \mathbf{F}^\top \left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1}\right)\right) \mathbf{Y}$$

$$= \left(\mathbf{F}^\top \left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1} \mathbf{F}\right)^- \mathbf{F}^\top \left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1} \bar{\mathbf{Y}}.$$

According to Theorem 7 and (3.23)

$$\hat{\boldsymbol{\gamma}} = \left(\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{H}^\top\mathbf{F}^\top\mathbf{FH} + \mathbf{I}_q\right) + \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{I}_q\right)^{-1}$$
$$\cdot \left(\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{H}^\top\mathbf{F}^\top\right)\right) \mathbf{Y}.$$

Using properties of orthogonal idempotent matrices $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ and $\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ it may be easily verified that

$$\left(\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{H}^\top\mathbf{F}^\top\mathbf{FH} + \mathbf{I}_q\right) + \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{I}_q\right)^{-1} =$$
$$\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{H}^\top\mathbf{F}^\top\mathbf{FH} + \mathbf{I}_q\right)^{-1} + \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{I}_q.$$

Then we obtain the following form of the prediction of the random effects $\boldsymbol{\gamma}$:

$$\hat{\boldsymbol{\gamma}} = \left(\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\left(\mathbf{H}^\top\mathbf{F}^\top\mathbf{FH} + \mathbf{I}_q\right)^{-1} \mathbf{H}^\top\mathbf{F}^\top\right)\right) \mathbf{Y}.$$

From $\hat{\mathbf{b}} = (\mathbf{I}_n \otimes \mathbf{H}) \hat{\boldsymbol{\gamma}}$ and the property (4.12) of inverses of sums of matrices it follows that

$$\hat{\mathbf{b}} = \left(\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{DF}^\top - \mathbf{DF}^\top \left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1} \mathbf{FDF}^\top\right)\right) \mathbf{Y}$$

$$= \left(\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{DF}^\top \left(\mathbf{FDF}^\top + \mathbf{I}_m\right)^{-1}\right)\right) \mathbf{Y}.$$

$\square$

Note that also if $\hat{\boldsymbol{\beta}}$ is not unique, the predictor $\hat{\boldsymbol{\Psi}}$ is independent of the choice of the generalized inverse $(\mathbf{F}^\top\mathbf{F})^-$. The latter follows from Theorem 5, which results in the condition $\mathbf{K} = \mathbf{U}(\mathbf{1}_n \otimes \mathbf{F})$ for predictable aspects in the hierarchical random coefficient regression model.

Note also that $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{b}}$ do not depend on the choice of the matrix $\mathbf{H}$.

Theorem 10. *If the fixed effects design matrix* $\mathbf{F}$ *has full column rank, the population parameter* $\boldsymbol{\beta}$ *is estimable and its best linear unbiased estimator is given by*

$$\hat{\boldsymbol{\beta}} = (\mathbf{F}^\top \mathbf{F})^{-1} \mathbf{F}^\top \bar{\mathbf{Y}}. \tag{4.21}$$

Proof. The fixed effects $\boldsymbol{\beta}$ are estimable due to Lemma 4, part c). By the property (4.11) of inverses of sums of matrices we get from (4.19) the equality

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\mathbf{F}^\top \mathbf{F} - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \right)^{-} \\
&\quad \cdot \left(\mathbf{F}^\top - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \mathbf{F}^\top \right) \bar{\mathbf{Y}}.
\end{aligned}
$$

Then for non-singular $\mathbf{F}^\top \mathbf{F}$ we obtain

$$\mathbf{F}^\top \mathbf{F} - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \mathbf{F}^\top \mathbf{F} = \left((\mathbf{F}^\top \mathbf{F})^{-1} + \mathbf{D} \right)^{-1}. \tag{4.22}$$

Consequently, the matrix

$$
\begin{aligned}
&\mathbf{I}_m - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \\
&= \left(\mathbf{F}^\top \mathbf{F} - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \right) (\mathbf{F}^\top \mathbf{F})^{-1}
\end{aligned}
$$

is non-singular. Then it follows that

$$
\begin{aligned}
\hat{\boldsymbol{\beta}} &= \left(\mathbf{F}^\top \mathbf{F} \right)^{-1} \left(\mathbf{I}_m - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \right)^{-1} \\
&\quad \cdot \left(\mathbf{I}_m - \mathbf{F}^\top \mathbf{F} \mathbf{H} \left(\mathbf{H}^\top \mathbf{F}^\top \mathbf{F} \mathbf{H} + \mathbf{I}_q \right)^{-1} \mathbf{H}^\top \right) \mathbf{F}^\top \bar{\mathbf{Y}} \\
&= \left(\mathbf{F}^\top \mathbf{F} \right)^{-1} \mathbf{F}^\top \bar{\mathbf{Y}}.
\end{aligned}
$$

$\square$

Lemma 5. *The mean squared error matrix of the best linear unbiased predictor* $\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{T}\hat{\mathbf{b}}$ *of a predictable linear aspect* $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{T}\mathbf{b}$ *is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = \mathbf{K}\,\tilde{\mathbf{C}}_{11}\,\mathbf{K}^\top + \mathbf{K}\,\tilde{\mathbf{C}}_{12}\,\mathbf{T}^\top + \mathbf{T}\,\tilde{\mathbf{C}}_{12}^\top\,\mathbf{K}^\top + \mathbf{T}\,\tilde{\mathbf{C}}_{22}\,\mathbf{T}^\top \tag{4.23}$$

where

$$\tilde{\mathbf{C}}_{11} = \frac{\sigma^2}{n} \left(\mathbf{F}^\top (\mathbf{F} \mathbf{D} \mathbf{F}^\top + \mathbf{I}_m)^{-1} \mathbf{F} \right)^{-},$$

$$\tilde{\mathbf{C}}_{12} = \tilde{\mathbf{C}}_{21}^\top = -\frac{\sigma^2}{n} \mathbf{1}_n^\top \otimes \left((\mathbf{F}^\top \mathbf{F})^{-} \mathbf{F}^\top \mathbf{F} \mathbf{D} \right)$$

and

$$\tilde{\mathbf{C}}_{22} = \frac{\sigma^2}{n} \left((\mathbf{1}_n \mathbf{1}_n^\top) \otimes \mathbf{D} + (n\mathbf{I}_n - \mathbf{1}_n \mathbf{1}_n^\top) \otimes (\mathbf{D} - \mathbf{D}\mathbf{F}^\top \left(\mathbf{F}\mathbf{D}\mathbf{F}^\top + \mathbf{I}_m\right)^{-1} \mathbf{F}\mathbf{D}) \right).$$

Proof. As it was mentioned at the beginning of this chapter, the linear aspect $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{T}\mathbf{b}$ can be represented in the form $\boldsymbol{\Psi} = \mathbf{K}\boldsymbol{\beta} + \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H})\boldsymbol{\gamma}$. Then it follows from Theorem 8 that the mean squared error matrix of the best linear unbiased predictor $\hat{\boldsymbol{\Psi}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{T}\hat{\mathbf{b}} = \mathbf{K}\hat{\boldsymbol{\beta}} + \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H})\hat{\boldsymbol{\gamma}}$ of $\boldsymbol{\Psi}$ is given by

$$\begin{aligned} \mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) &= \mathbf{K}\,\mathbf{C}'_{11}\,\mathbf{K}^\top + \mathbf{K}\,\mathbf{C}'_{12}\,(\mathbf{I}_n \otimes \mathbf{H}^\top)\mathbf{T}^\top + \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H})\,\mathbf{C}'^\top_{12}\,\mathbf{K}^\top \quad (4.24) \\ &\quad + \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H})\,\mathbf{C}'_{22}\,(\mathbf{I}_n \otimes \mathbf{H}^\top)\mathbf{T}^\top, \end{aligned}$$

where the components $\mathbf{C}'_{22}$, $\mathbf{C}'_{11}$ and $\mathbf{C}'_{12}$ of the matrix $\mathbf{C}'$ defined by (3.20) may be derived directly from formulas (3.15)-(3.17):

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

$$\mathbf{C}'_{11} = \frac{\sigma^2}{n} \left(\mathbf{F}^\top (\mathbf{F}\mathbf{D}\mathbf{F}^\top + \mathbf{I}_m)^{-1} \mathbf{F} \right)^-,$$

$$\mathbf{C}'_{12} = -\frac{\sigma^2}{n} \mathbf{1}_n^\top \otimes ((\mathbf{F}^\top \mathbf{F})^- \mathbf{F}^\top \mathbf{F} \mathbf{H}).$$

The mean squared error matrix (4.24) may be also represented as

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = (\mathbf{K},\ \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H}))\,\mathbf{C}'\,(\mathbf{K},\ \mathbf{T}(\mathbf{I}_n \otimes \mathbf{H}))^\top$$

or equivalently as

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = (\mathbf{K},\ \mathbf{T}) \begin{pmatrix} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_n \otimes \mathbf{H} \end{pmatrix} \mathbf{C}' \begin{pmatrix} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_n \otimes \mathbf{H}^\top \end{pmatrix} (\mathbf{K},\ \mathbf{T})^\top.$$

By formula (4.13) the block matrix

$$\tilde{\mathbf{C}} = \begin{pmatrix} \tilde{\mathbf{C}}_{11} & \tilde{\mathbf{C}}_{12} \\ \tilde{\mathbf{C}}_{21} & \tilde{\mathbf{C}}_{22} \end{pmatrix}$$

may be rewritten as

$$\tilde{\mathbf{C}} = \begin{pmatrix} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_n \otimes \mathbf{H} \end{pmatrix} \mathbf{C}' \begin{pmatrix} \mathbf{I}_p & 0 \\ 0 & \mathbf{I}_n \otimes \mathbf{H}^\top \end{pmatrix}.$$

Then we obtain

$$\mathrm{Cov}(\hat{\boldsymbol{\Psi}} - \boldsymbol{\Psi}) = (\mathbf{K},\, \mathbf{T})\, \tilde{\mathbf{C}}\, (\mathbf{K},\, \mathbf{T})^{\top},$$

which coincides with (4.23). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

For a full column rank matrix $\mathbf{F}$ we obtain the following result using the properties (4.13) and (4.17).

Corollary 4. *If the fixed effects design matrix* $\mathbf{F}$ *has full column rank, the mean squared error matrix* $\tilde{\mathbf{C}}$ *of* $(\hat{\boldsymbol{\beta}}^{\top}, \hat{\mathbf{b}}^{\top})^{\top}$ *simplifies to*

$$\frac{\sigma^2}{n}\begin{pmatrix} (\mathbf{F}^{\top}\mathbf{F})^{-1} + \mathbf{D} & -\mathbf{1}_n^{\top} \otimes \mathbf{D} \\ -\mathbf{1}_n \otimes \mathbf{D} & (\mathbf{1}_n\mathbf{1}_n^{\top}) \otimes \mathbf{D} + (n\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\top}) \otimes \left(\mathbf{D} - \mathbf{D}((\mathbf{F}^{\top}\mathbf{F})^{-1} + \mathbf{D})^{-1}\mathbf{D}\right) \end{pmatrix}.$$

Corollary 5. *If the fixed effects design matrix* $\mathbf{F}$ *has full column rank, the covariance matrix of the best linear unbiased estimator* $\hat{\boldsymbol{\beta}}$ *of the population parameter* $\boldsymbol{\beta}$ *is given by*

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \frac{\sigma^2}{n}\left((\mathbf{F}^{\top}\mathbf{F})^{-1} + \mathbf{D}\right). \qquad\qquad (4.25)$$

Using the property (4.18) we get the special form of the mean squared error matrix $\tilde{\mathbf{C}}$ in the case of a non-singular dispersion matrix of random effects.

Corollary 6. *If the fixed effects design matrix* $\mathbf{F}$ *has full column rank and the dispersion matrix* $\mathbf{D}$ *of random effects is non-singular, the mean squared error matrix* $\tilde{\mathbf{C}}$ *of* $(\hat{\boldsymbol{\beta}}^{\top}, \hat{\mathbf{b}}^{\top})^{\top}$ *simplifies to*

$$\frac{\sigma^2}{n}\begin{pmatrix} (\mathbf{F}^{\top}\mathbf{F})^{-1} + \mathbf{D} & -\mathbf{1}_n^{\top} \otimes \mathbf{D} \\ -\mathbf{1}_n \otimes \mathbf{D} & (\mathbf{1}_n\mathbf{1}_n^{\top}) \otimes \mathbf{D} + (n\mathbf{I}_n - \mathbf{1}_n\mathbf{1}_n^{\top}) \otimes \left(\mathbf{F}^{\top}\mathbf{F} + \mathbf{D}^{-1}\right)^{-1} \end{pmatrix}.$$

4.3 Prediction of Individual Deviations

In this section we introduce the best linear unbiased prediction of the individual deviations $\mathbf{b}$. According to Lemma 4 part b) the individual deviations are predictable for any design matrix $\mathbf{F}$.

Theorem 11. *The best linear unbiased predictor of the individual deviations* $\mathbf{b}_i$ *is given by*

$$\hat{\mathbf{b}}_i = \mathbf{D}\mathbf{F}^{\top}(\mathbf{F}\mathbf{D}\mathbf{F}^{\top} + \mathbf{I}_m)^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}}). \qquad\qquad (4.26)$$

Proof. To make use of the theory developed in the previous section, we represent the individual deviations in the form $\mathbf{b}_i = (\mathbf{e}_i^\top \otimes \mathbf{I}_p)\mathbf{b}$, where $\mathbf{e}_i$ denotes the i-th unit vector of length n. Then it follows from Theorem 9 that

$$
\begin{aligned}
\hat{\mathbf{b}}_i &= (\mathbf{e}_i^\top \otimes \mathbf{I}_p)\,\hat{\mathbf{b}} \\
&= (\mathbf{e}_i^\top \otimes \mathbf{I}_p)\left((\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes \left(\mathbf{D}\mathbf{F}^\top(\mathbf{F}\mathbf{D}\mathbf{F}^\top + \mathbf{I}_m)^{-1}\right)\right)\mathbf{Y} \\
&= \left((\mathbf{e}_i^\top - \frac{1}{n}\mathbf{1}_n^\top) \otimes \left(\mathbf{D}\mathbf{F}^\top(\mathbf{F}\mathbf{D}\mathbf{F}^\top + \mathbf{I}_m)^{-1}\right)\right)\mathbf{Y} \\
&= \mathbf{D}\mathbf{F}^\top(\mathbf{F}\mathbf{D}\mathbf{F}^\top + \mathbf{I}_m)^{-1}(\mathbf{Y}_i - \bar{\mathbf{Y}}).
\end{aligned}
$$

$\square$

For a non-singular dispersion matrix $\mathbf{D}$ we obtain from the property (4.14) of sums of matrices the next result.

Corollary 7. *If the dispersion matrix of random effects is non-singular, the best linear unbiased predictor $\hat{\mathbf{b}}_i$ of the individual deviations $\mathbf{b}_i$ simplifies to*

$$
\hat{\mathbf{b}}_i = \left(\mathbf{F}^\top\mathbf{F} + \mathbf{D}^{-1}\right)^{-1}\mathbf{F}^\top\left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right). \tag{4.27}
$$

In the case of full column rank fixed effects design matrix we will further use the notation $\hat{\boldsymbol{\beta}}_{i;\text{ind}}$ for the individualized estimator based only on observations at the i-th individual:

$$
\hat{\boldsymbol{\beta}}_{i;\text{ind}} = (\mathbf{F}^\top\mathbf{F})^{-1}\mathbf{F}^\top\mathbf{Y}_i. \tag{4.28}
$$

Then due to Theorem 10 the predictor (4.26) may be represented in the following alternative form.

Remark 2. *If the fixed effects design matrix $\mathbf{F}$ has full column rank, the best linear unbiased predictor $\hat{\mathbf{b}}_i$ of the individual deviations $\mathbf{b}_i$ simplifies to*

$$
\hat{\mathbf{b}}_i = \mathbf{D}\left((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D}\right)^{-1}(\hat{\boldsymbol{\beta}}_{i,\text{ind}} - \hat{\boldsymbol{\beta}}). \tag{4.29}
$$

The following theorem gives the mean squared error matrix of the prediction (4.26).

Theorem 12. *The mean squared error matrix of the best linear unbiased prediction* $\hat{\mathbf{b}}$ *of the individual deviations is given by*

$$\text{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \tag{4.30}$$
$$= \sigma^2 \left(\frac{1}{n}\left(\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \mathbf{D} + \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{D} - \mathbf{D}\mathbf{F}^\top\left(\mathbf{F}\mathbf{D}\mathbf{F}^\top + \mathbf{I}_m\right)^{-1}\mathbf{F}\mathbf{D}\right)\right).$$

Proof. The result (4.30) follows from Lemma 5 for the null matrix $\mathbf{K} = \mathbf{0}_{p \times p}$ and the identity matrix $\mathbf{L} = \mathbf{I}_{np}$. $\qquad\square$

Corollary 8. *If the dispersion matrix* $\mathbf{D}$ *of random effects is non-singular, the mean squared error matrix of the best linear unbiased prediction* $\hat{\mathbf{b}}$ *of the individual deviations simplifies to*

$$\text{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right) = \sigma^2 \left(\frac{1}{n}\left(\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \mathbf{D} + \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{F}^\top\mathbf{F} + \mathbf{D}^{-1}\right)^{-1}\right). \tag{4.31}$$

For further considerations we will use the reduced model

$$\mathbf{Y}_i = \mathbf{F}_0\boldsymbol{\beta}_i^0 + \boldsymbol{\varepsilon}_i, \quad i = 1, ..., n, \tag{4.32}$$

where $\boldsymbol{\beta}_i^0 = (\boldsymbol{\beta}_{i1}, ..., \boldsymbol{\beta}_{iq})^\top$, $\mathbf{F}_0 = (\mathbf{f}^0(x_1), ..., \mathbf{f}^0(x_m))^\top$ for $\mathbf{f}^0 = (f_1, .., f_q)^\top$ and $\text{Cov}(\boldsymbol{\beta}_i^0) = \sigma^2\mathbf{D}_0$ for a non-singular $q \times q$ matrix $\mathbf{D}_0$.

Now we consider the special case of the model (4.2) with the the block diagonal form

$$\mathbf{D} = \left(\mathbf{I}_q \; \mathbf{0}_{q \times (p-q)}\right)^\top \mathbf{D}_0 \left(\mathbf{I}_q \; \mathbf{0}_{q \times (p-q)}\right) \tag{4.33}$$

of the dispersion matrix of random effects (i.e. only the first q parameters are random).

Theorem 13. *If the dispersion matrix* $\mathbf{D}$ *of random effects has the block diagonal form (4.33), the best linear unbiased predictor of the individual deviations* $\mathbf{b}_i$ *is given by*

$$\hat{\mathbf{b}}_i = \left(\mathbf{I}_q \; \mathbf{0}_{q \times (p-q)}\right)^\top \hat{\mathbf{b}}_i^0, \tag{4.34}$$

where $\hat{\mathbf{b}}_i^0 = \left(\mathbf{F}_0^\top\mathbf{F}_0 + \mathbf{D}_0^{-1}\right)^{-1}\mathbf{F}_0^\top\left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right)$ *is the best linear unbiased predictor of the individual deviations* $\mathbf{b}_i^0$ *in the reduced model (4.32).*

Proof. For the parameter vector $\boldsymbol{\beta}_i = \left(\boldsymbol{\beta}_i^{0\top}, \boldsymbol{\beta}_i^{1\top}\right)^{\top}$ with $\boldsymbol{\beta}_i^1 = (\boldsymbol{\beta}_{i(q+1)}, ..., \boldsymbol{\beta}_{ip})^{\top}$ and the regression functions $\mathbf{f} = (\mathbf{f}^{0\top}, \mathbf{f}^{1\top})^{\top}$ with $\mathbf{f}^1 = (f_{q+1}, .., f_p)^{\top}$ the model (4.2) may be represented as

$$Y_{ij} = \mathbf{f}^0(x_j)^{\top}\boldsymbol{\beta}_i^0 + \mathbf{f}^1(x_j)^{\top}\boldsymbol{\beta}_i^1 + \varepsilon_{ij}. \tag{4.35}$$

Then for $\mathbf{F}_1 = (\mathbf{f}^1(x_1), ..., \mathbf{f}^1(x_m))^{\top}$ the vector of all observations at individual i has the form

$$\begin{aligned}
\mathbf{Y}_i &= \mathbf{F}_0\boldsymbol{\beta}_i^0 + \mathbf{F}_1\boldsymbol{\beta}_i^1 + \boldsymbol{\varepsilon}_i \tag{4.36}\\
&= (\mathbf{F}_0, \mathbf{F}_1) \begin{pmatrix} \boldsymbol{\beta}_i^0 \\ \boldsymbol{\beta}_i^1 \end{pmatrix} + \boldsymbol{\varepsilon}_i.
\end{aligned}$$

It is easy to see that $\mathrm{Cov}(\boldsymbol{\beta}_i^0) = \mathbf{D}_0$ and $\boldsymbol{\beta}_i^1$ are fixed effects. Hence, for the parameter vector $\boldsymbol{\beta}_i = \left(\boldsymbol{\beta}_i^{0\top}, \boldsymbol{\beta}_i^{1\top}\right)^{\top}$ and the design matrix $\mathbf{F} = (\mathbf{F}_0, \mathbf{F}_1)$ the model (4.36) is of the form (4.3).

For a non-singular matrix $\mathbf{D}_0$ there exists a non-singular matrix $\mathbf{H}_0$ with rank q and $\mathbf{D}_0 = \mathbf{H}_0\mathbf{H}_0^{\top}$. Then the dispersion matrix $\mathbf{D}$ of random effects may be represented as $\mathbf{D} = \mathbf{H}\mathbf{H}^{\top}$ with $\mathbf{H} = (\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)})^{\top}\mathbf{H}_0$.

Using formula (4.26) we obtain

$$\begin{aligned}
\hat{\mathbf{b}}_i &= (\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)})^{\top}\mathbf{H}_0 \left(\mathbf{H}_0^{\top}\mathbf{F}_0^{\top}\mathbf{F}_0\mathbf{H}_0 + \mathbf{I}_q\right)^{-1} \mathbf{H}_0^{\top}\mathbf{F}_0^{\top} \left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right) \\
&= (\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)})^{\top} \left(\mathbf{F}_0^{\top}\mathbf{F}_0 + \mathbf{D}_0^{-1}\right)^{-1} \mathbf{F}_0^{\top} \left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right).
\end{aligned}$$

According to Corollary 7 $\hat{\mathbf{b}}_i^0 = \left(\mathbf{F}_0^{\top}\mathbf{F}_0 + \mathbf{D}_0^{-1}\right)^{-1} \mathbf{F}_0^{\top} \left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right)$ is the best linear unbiased predictor for $\mathbf{b}_i^0$ in the reduced model (4.32), which proves the result. $\square$

The mean squared error matrix of this prediction is given by the next theorem.

Theorem 14. *If the dispersion matrix $\mathbf{D}$ of random effects has the block diagonal form (4.33), the mean squared error matrix of the best linear unbiased predictor $\hat{\mathbf{b}}$ of the individual deviations is given by*

$$\mathrm{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right) = \left(\mathbf{I}_n \otimes (\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)})^{\top}\right) \mathrm{Cov}\left(\hat{\mathbf{b}}_0 - \mathbf{b}_0\right) \left(\mathbf{I}_n \otimes (\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)})\right), \tag{4.37}$$

where $\mathrm{Cov}\left(\hat{\mathbf{b}}_0 - \mathbf{b}_0\right)$ denotes the mean squared error matrix of the best linear unbiased predictor $\hat{\mathbf{b}}_i^0$ of the individual deviations in the reduced model (4.32).

Proof. The proof of this result is similar to the proof of Theorem 13. The form (4.37) follows from formula (4.30) for the mean squared error matrix of the prediction $\hat{\mathbf{b}}$ in the general case, and Corollary 8, which is used to determine the mean squared error matrix $\mathrm{Cov}\left(\hat{\mathbf{b}}_0 - \mathbf{b}_0\right)$ in the reduced model. □

Note that the prediction $\hat{\mathbf{b}}_i$ of the individual deviations as well as its mean squared error matrix $\mathrm{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right)$ are independent of the fixed part $\mathbf{F}_1\boldsymbol{\beta}_i^1$ of the model (4.36). Consequently, the reduced model only has to be investigated if the dispersion matrix of random effects has the block diagonal form (4.33).

Remark 3. *If the dispersion matrix $\mathbf{D}$ of random effects in the model (4.2) is singular, the model can be reparametrized so that the dispersion matrix in the new model has the block diagonal form (4.33).*

The latter statement can be established in the following way: Since the matrix $\mathbf{H}$, which has been introduced in Section 4.1, has full column rank, there exists a $p \times (p-q)$ matrix $\mathbf{H}_1$ so that the matrix $\tilde{\mathbf{H}} = (\mathbf{H}\,\mathbf{H}_1)$ is non-singular (full rank). For the random variables $\tilde{\boldsymbol{\gamma}}_i := (\boldsymbol{\gamma}_i^\top\, \mathbf{0}_{(p-q)}^\top)^\top$, where $\mathbf{0}_{(p-q)}$ denotes the zero vector of length $p - q$, it holds that $\mathbf{b}_i = \tilde{\mathbf{H}}\tilde{\boldsymbol{\gamma}}_i$. Then for $\tilde{\boldsymbol{\gamma}} = (\tilde{\boldsymbol{\gamma}}_1^\top, ..., \tilde{\boldsymbol{\gamma}}_n^\top,)^\top$ the model (4.10) can be rewritten in form

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

where the dispersion matrix of the random effects $\tilde{\boldsymbol{\gamma}}$ has the block diagonal form (4.33) with $\mathbf{D}_0 = \mathbf{I}_q$.

4.4 Prediction of Individual Parameters

In this section we introduce the best linear unbiased predictor of the individual parameters $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{b}_i$. According to part d) of Lemma 4 the latter parameters are only predictable if the design matrix $\mathbf{F}$ is a full column rank matrix. Hence, a special case of the model (4.2) with $\mathrm{rank}(\mathbf{F}) = p$ will be investigated.

Theorem 15. *The best linear unbiased predictor of the individual parameters $\boldsymbol{\beta}_i$ is given by*

$$\hat{\boldsymbol{\beta}}_i = \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} + (\mathbf{F}^\top\mathbf{F})^{-1}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}. \qquad (4.38)$$

Proof. The individual parameters $\boldsymbol{\beta}_i$ can be represented in the form $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \mathbf{b}_i$. Then it follows from Theorems 9 and 10 and Remark 2 that $\boldsymbol{\beta}_i$ is predictable and its best linear unbiased predictor is given by

$$\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i.$$

After applying formula (4.21) we obtain

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_i &= \hat{\boldsymbol{\beta}} + \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}(\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} - \hat{\boldsymbol{\beta}}) \\
&= \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} + (\mathbf{F}^\top\mathbf{F})^{-1}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}.
\end{aligned}
$$

$\square$

Note that the predictor (4.38) of the individual parameters is a weighted average of the individualized estimator $\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}}$ and the best linear unbiased estimator (4.21) of the population parameter and coincides with the empirical Bayes estimator introduced by Bryk and Raudenbush (1992) (see also Candel (2009)) of the individual parameters $\boldsymbol{\beta}_i$ in the special case of a known dispersion matrix of random effects.

Corollary 9. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the best linear unbiased predictor $\hat{\boldsymbol{\beta}}_i$ of the individual parameters $\boldsymbol{\beta}_i$ simplifies to*

$$\hat{\boldsymbol{\beta}}_i = (\mathbf{F}^\top\mathbf{F} + \mathbf{D}^{-1})^{-1}(\mathbf{F}^\top\mathbf{F}\,\hat{\boldsymbol{\beta}}_{i;\mathrm{ind}} + \mathbf{D}^{-1}\hat{\boldsymbol{\beta}}). \tag{4.39}$$

The latter result was derived by Fedorov and Jones (2005) under the additional assumption of normal distribution for multicentre regression models.

The following theorem represents the mean squared error matrix of the prediction $\hat{\mathbf{B}} = (\hat{\boldsymbol{\beta}}_1^\top, ..., \hat{\boldsymbol{\beta}}_n^\top)^\top$ of the individual parameters $\mathbf{B} = (\boldsymbol{\beta}_1^\top, ..., \boldsymbol{\beta}_n^\top)^\top$.

Theorem 16. *The mean squared error matrix of the prediction $\hat{\mathbf{B}} = (\hat{\boldsymbol{\beta}}_1^\top, ..., \hat{\boldsymbol{\beta}}_n^\top)^\top$ of the individual parameters is given by*

$$
\begin{aligned}
&\mathrm{Cov}\left(\hat{\mathbf{B}} - \mathbf{B}\right) \\
&= \sigma^2 \left(\frac{1}{n}\left(\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes (\mathbf{F}^\top\mathbf{F})^{-1} + \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes \left(\mathbf{D} - \mathbf{D}((\mathbf{F}^\top\mathbf{F})^{-1} + \mathbf{D})^{-1}\mathbf{D}\right)\right). \tag{4.40}
\end{aligned}
$$

Proof. The prediction of the individual parameters may be represented as $\hat{\mathbf{B}} = (\mathbf{1}_n \otimes \mathbf{I}_p) \hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}$. Then it follows from Lemma 5 for $\mathbf{K} = \mathbf{1}_n \otimes \mathbf{I}_p$ and $\mathbf{L} = \mathbf{I}_{np}$ that

$$\mathrm{Cov}\left(\hat{\mathbf{B}} - \mathbf{B}\right) = (\mathbf{1}_n \otimes \mathbf{I}_p \; \mathbf{I}_{np}) \, \tilde{\mathbf{C}} \, (\mathbf{1}_n \otimes \mathbf{I}_p \; \mathbf{I}_{np})^\top.$$

Using Corollary 4 and some linear algebra we obtain the result (4.40)   □

Corollary 10. *If the dispersion matrix* $\mathbf{D}$ *of random effects is non-singular, the mean squared error matrix of the prediction* $\hat{\mathbf{B}}$ *simplifies to*

$$
\begin{aligned}
&\mathrm{Cov}\left(\hat{\mathbf{B}} - \mathbf{B}\right) \\
&= \; \sigma^2 \left(\frac{1}{n} \left(\mathbf{1}_n \mathbf{1}_n^\top\right) \otimes (\mathbf{F}^\top \mathbf{F})^{-1} + \left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top\right) \otimes (\mathbf{F}^\top \mathbf{F} + \mathbf{D}^{-1})^{-1}\right).
\end{aligned}
\tag{4.41}
$$

Note that the mean squared error matrix (4.41) can be recognized as a weighted sum of covariance matrix (2.8) in the model without random effects and the mean squared error matrix (2.15) in the Bayesian model.

Example 1. We consider the particular case of random intercepts (random block effects) with an explicit individual constant term $f_1(x) \equiv 1$ for all individuals. Since only the intercept $\boldsymbol{\beta}_{i1}$ is random, the dispersion matrix $\mathbf{D}$ can be written as $\mathbf{D} = d_1 \mathbf{e}_1 \mathbf{e}_1^\top$ and has rank $q = 1$, $q < p$. Then in the reduced model (4.32) the design matrix is given by $\mathbf{F}_0 = \mathbf{1}_m$ and the dispersion matrix is one-dimensional with $\mathbf{D}_0 = d_1$. Consequently, the condition (4.33) is satisfied and according to Theorem 13 the prediction of the individual deviations has the form

$$\hat{\mathbf{b}}_i = \frac{d_1}{1 + m d_1} \mathbf{1}_m^\top \left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right) \mathbf{e}_1. \tag{4.42}$$

The corresponding mean squared error matrix simplifies to

$$\mathrm{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right) = \sigma^2 \left(\frac{d_1}{n}\mathbf{1}_n \mathbf{1}_n^\top + \frac{d_1}{1 + m d_1}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n \mathbf{1}_n^\top)\right) \otimes (\mathbf{e}_1 \mathbf{e}_1^\top), \tag{4.43}$$

which is independent of the design matrix.

The explicit form of the best linear unbiased predictor $\hat{\boldsymbol{\beta}}_i$ of the individual parameters follows from formula $\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} + \hat{\mathbf{b}}_i$:

$$\hat{\boldsymbol{\beta}}_i = \hat{\boldsymbol{\beta}} + \frac{d_1}{1 + m d_1}\mathbf{1}_m^\top \left(\mathbf{Y}_i - \bar{\mathbf{Y}}\right) \mathbf{e}_1. \tag{4.44}$$

Obviously, this result may be derived directly from (4.38). The mean squared error matrix (4.40) of the prediction simplifies to

$$\text{Cov}\left(\hat{\mathbf{B}} - \mathbf{B}\right) = \sigma^2 \left(\frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{F}^\top\mathbf{F})^{-1} + \frac{d_1}{1+md_1}(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{e}_1\mathbf{e}_1^\top)\right) \quad (4.45)$$

for the case of random intercepts. Note that only the predicted intercepts differ for different individuals, while the estimates of the other parameters are the same.

5 Optimal Designs in Hierarchical Random Coefficient Regression Models

In this chapter we consider optimal designs for the prediction of the individual parameters $\boldsymbol{\beta}_i$ and individual deviations $\mathbf{b}_i$ in the model (4.2). For the prediction of the individual parameter some results for optimal designs are provided in Prus and Schwabe (2016) (see also Prus and Schwabe (2011)). The problem of optimal designs for the prediction of the individual deviations has been briefly discussed by Prus and Schwabe (2013). For the estimation of the population mean parameter solutions for the design optimization problem may be found in Fedorov and Hackl (1997), Liski *et al.* (2002) or Entholzner *et al.* (2005).

As it was mentioned in Section 4.1, in model (4.2) the number of observations m as well as the experimental settings $x_1, ..., x_m$ are the same for all individuals. Therefore, the individual (exact and approximate) designs are also the same for all individuals and have the general form (2.16) introduced in the second chapter. $\Xi_{\mathcal{X}}$ and Ξ denote again the sets of all exact and approximate designs of the form (2.16) respectively.

The notations $\mathbf{M}(\xi)$ and $\boldsymbol{\Delta}$ will be used again for the adjusted version of the dispersion matrix of random effects and the standardized individual information matrix in the fixed effects models defined in (2.17) and (2.18) respectively. For simplicity, the matrix $\mathbf{M}(\xi)$ will be further called individual information matrix. Note that for exact designs it holds that $\mathbf{M}(\xi) = \frac{1}{m}\mathbf{F}^\top\mathbf{F}$.

Then for an approximate design ξ the mean squared error matrix of the prediction of the individual parameters can be defined by

$$\mathrm{MSE}_B\,(\xi) = \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top) \otimes \mathbf{M}(\xi)^{-1} + (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}), \quad (5.1)$$

which coincides with the representation (4.40) (divided by the constant factor $\frac{\sigma^2}{m}$) in the special case of an exact design. In this case the individual information matrix $\mathbf{M}(\xi)$ is assumed to be non-singular.

If also the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the mean squared error matrix (5.1) simplifies to

$$\mathrm{MSE}_B\,(\xi) = \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top) \otimes \mathbf{M}(\xi)^{-1} + (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}. \quad (5.2)$$

According to Remark 3, in the case of prediction of the individual deviations we may assume without loss of generality that the dispersion matrix $\mathbf{D}$ is non-singular or has the block diagonal form (4.33). We define the mean squared error matrix for an approximate design for the non-singular case as

$$\mathrm{MSE}_b\left(\xi\right) = \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top) \otimes \mathbf{\Delta} + (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{M}(\xi) + \mathbf{\Delta}^{-1})^{-1}, \tag{5.3}$$

which coincides with (4.31) (if we suppress the constant factor $\frac{\sigma^2}{m}$) in the special case of an exact design.

In the singular (block diagonal) case we analogously define for the adjusted dispersion matrix $\mathbf{\Delta}_0 = m\mathbf{D}_0$ of random effects and the individual information matrix $\mathbf{M}_0(\xi) = \frac{1}{m}\sum_{j=1}^k m_j\,\mathbf{f}_0(x_j)\mathbf{f}_0(x_j)^\top$ (equal to $\frac{1}{m}\mathbf{F}_0^\top\mathbf{F}_0$ for exact designs) in the reduced model (4.32) the mean squared error matrix

$$\mathrm{MSE}_b\left(\xi\right) = \left(\mathbf{I}_n \otimes (\mathbf{I}_q \;\; \mathbf{0}_{q\times(p-q)})^\top\right) \mathrm{MSE}_{b_0}\left(\xi\right) \left(\mathbf{I}_n \otimes (\mathbf{I}_q \;\; \mathbf{0}_{q\times(p-q)})\right), \tag{5.4}$$

where

$$\mathrm{MSE}_{b_0}\left(\xi\right) = \frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top) \otimes \mathbf{\Delta}_0 + (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{M}_0(\xi) + \mathbf{\Delta}_0^{-1})^{-1}. \tag{5.5}$$

Note that in (5.3) and (5.4) the individual information matrix $\mathbf{M}(\xi)$ may be singular.

Using the notation (2.17) and $\mathbf{\Delta}$ as above the covariance matrix (4.25) of the estimator (4.21) of the fixed effects may be generalized to

$$\mathrm{Cov}_\beta(\xi) = \frac{\sigma^2}{nm}\left(\mathbf{M}(\xi)^{-1} + \mathbf{\Delta}\right) \tag{5.6}$$

for an approximate design ξ with a non-singular individual information matrix $\mathbf{M}(\xi)$. Since the factor $\frac{\sigma^2}{nm}$ as well as the adjusted dispersion matrix $\mathbf{\Delta}$ are independent of the designs, it can be concluded that L-optimal designs in the fixed effects model (2.1) are L-optimal for the estimation of the population parameter in the hierarchical random coefficient regression model (4.2). However, the latter statement does not hold for the D-criterion (see Entholzner *et al.* (2005)).

5.1 *D*-Criterion

The determinant criterion in the hierarchical random coefficient regression model (4.2) may be defined analogously to the Bayesian D-criterion (2.9) as a logarithmized determi-

nant of the mean squared error matrix:

$$D_B(\xi) = \ln \det \text{MSE}_B(\xi) \qquad (5.7)$$

and

$$D_b(\xi) = \ln \det \text{MSE}_b(\xi) \qquad (5.8)$$

for the prediction of the individual parameters and individual deviations respectively.

The latter approach makes only sense if the dispersion matrix $\mathbf{D}$ of random effects and consequently the mean squared error matrices $\text{MSE}_B(\xi)$ and $\text{MSE}_b(\xi)$ are non-singular, since otherwise the criterion functions (5.7) and (5.8) would be equal to $-\infty$. For a general dispersion matrix the D-criterion has to be appropriately adjusted. We consider again the case of a non-zero matrix $\mathbf{D}$ with rank q $(1 \le q \le p)$ and use the property that for positive definite symmetric matrices the determinant is equal to the product of the eigenvalues. Then the D-criterion for the individual parameters may be defined as the logarithm of the product of the $(n-1)q+p$ largest eigenvalues of the mean squared error matrix $\text{MSE}_B(\xi)$ and coincides for $q = p$ with the classical D-criterion (5.7).

Definition 13. *For a general dispersion matrix $\mathbf{D}$ with rank q the approximate design $\xi^* \in \Xi$ is called D-optimal for the prediction of the individual parameters if*

$$D_B(\xi^*) = \min_{\xi \in \Xi} D_B(\xi) \qquad (5.9)$$

for

$$D_B(\xi) = \ln \prod_{\iota=1}^{(n-1)q+p} \lambda_\iota(\xi), \qquad (5.10)$$

where $\lambda_1(\xi), ..., \lambda_{(n-1)q+p}(\xi)$ are the $(n-1)q+p$ largest eigenvalues of the mean squared error matrix $\text{MSE}_B(\xi)$ of the prediction of the individual parameters. The criterion function (5.10) is called the (generalized) D-criterion for the prediction of the individual parameters.

The (generalized) D-criterion for the prediction of the individual deviations can be similarly defined as the logarithm of the product of the nq largest eigenvalues of $\text{MSE}_b(\xi)$, which coincides with (5.8) for $q = p$.

Definition 14. *For a general dispersion matrix* $\mathbf{D}$ *with rank* q *the approximate design* $\xi^* \in \Xi$ *is called D-optimal for the prediction of the individual deviations if*

$$\mathrm{D}_b(\xi^*) = \min_{\xi \in \Xi} \mathrm{D}_b(\xi) \tag{5.11}$$

for

$$\mathrm{D}_b(\xi) = \ln \prod_{\iota=1}^{nq} \nu_\iota(\xi), \tag{5.12}$$

where $\nu_1(\xi), ..., \nu_{nq}(\xi)$ *are the* nq *largest eigenvalues of the mean squared error matrix* $\mathrm{MSE}_b(\xi)$ *of the prediction of the individual deviations. The criterion function (5.12) is called the (generalized) D-criterion for the prediction of the individual deviations.*

For further considerations of design criteria we will use the representation

$$\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta} = m\,\mathbf{H}\left(m\,\mathbf{H}^\top\mathbf{M}(\xi)\,\mathbf{H} + \mathbf{I}_q)\right)^{-1}\mathbf{H}^\top, \tag{5.13}$$

which can be verified analogously to (4.17) for the matrix $\mathbf{H}$ defined in Section 4.1 with $\mathbf{D} = \mathbf{H}\mathbf{H}^\top$ and consequently $\boldsymbol{\Delta} = m\,\mathbf{H}\mathbf{H}^\top$. If the dispersion matrix of random effects in non-singular, (5.13) may be rewritten (analogously to (4.18)) as

$$\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta} = (\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1}. \tag{5.14}$$

The following Lemma presents an explicit form of the (generalized) D-criterion for the prediction of the individual parameters.

Lemma 6. *For a general dispersion matrix* $\mathbf{D}$ *with rank* q *the (generalized) D-criterion for the prediction of the individual parameters is equal to*

$$\mathrm{D}_B(\xi) = \ln \det(\mathbf{M}(\xi)^{-1}) + (n-1)\ln \prod_{\ell=1}^{q} \mu_\ell(\xi), \tag{5.15}$$

where $\mu_1(\xi), ..., \mu_q(\xi)$ *are the* q *largest eigenvalues of the matrix* $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$.

Proof. Since the individual information matrix $\mathbf{M}(\xi)$ is positive definite, its inverse $\mathbf{M}(\xi)^{-1}$ has p positive eigenvalues $\eta_1(\xi), ..., \eta_p(\xi)$. Since the adjusted dispersion $\boldsymbol{\Delta}$ has rank q, according to Harville (1997) (ch. 14, p. 256) and formula (5.13) the matrix $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$ is nonnegative definite with the same rank and consequently has q positive eigenvalues $\mu_1(\xi), ..., \mu_q(\xi)$ and $p - q$ eigenvalues equal to zero.

Let $\mathbf{v}_1(\xi), ..., \mathbf{v}_p(\xi)$ be pairwise orthogonal eigenvectors of the matrix $\mathbf{M}(\xi)^{-1}$ to the eigenvalues $\eta_1(\xi), ..., \eta_p(\xi)$ respectively. The only positive eigenvalue of $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top$ is 1 with the corresponding eigenvector $\mathbf{1}_n$. Then $\mathbf{1}_n \otimes \mathbf{v}_1(\xi), ..., \mathbf{1}_n \otimes \mathbf{v}_p(\xi)$ are pairwise orthogonal eigenvectors of $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{M}(\xi)^{-1}$ corresponding to the eigenvalues $\eta_1(\xi), ..., \eta_p(\xi)$ respectively.

It follows from $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{1}_n = \mathbf{0}$ that for $k = 1, ..., p$

$$\mathrm{MSE}_B(\xi)\,(\mathbf{1}_n \otimes \mathbf{v}_k(\xi)) = \left(\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top \otimes \mathbf{M}(\xi)^{-1}\right)(\mathbf{1}_n \otimes \mathbf{v}_k(\xi)) = \eta_k(\xi)\,(\mathbf{1}_n \otimes \mathbf{v}_k(\xi)),$$

which means that $\eta_1(\xi), ..., \eta_p(\xi)$ are eigenvalues of $\mathrm{MSE}_B(\xi)$.

Let $\mathbf{w}_1(\xi), ..., \mathbf{w}_q(\xi)$ be pairwise orthogonal eigenvectors of $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$ corresponding to the eigenvalues $\mu_1(\xi), ..., \mu_q(\xi)$ respectively. The matrix $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)$ has the only positive eigenvalue 1 with the algebraic multiplicity of $n-1$. Then there exist $n-1$ pairwise orthogonal eigenvectors corresponding to this eigenvalue: $\mathbf{h}_1, ..., \mathbf{h}_{n-1}$. Note that these vectors satisfy the condition $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{h}_t = \mathbf{h}_t$. Consequently, for all $t = 1, ..., n-1$ and all $\ell = 1, ..., q$ the vector $\mathbf{h}_t \otimes \mathbf{w}_\ell(\xi)$ is an eigenvector of $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right) \otimes (\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})$ corresponding to $\mu_\ell(\xi)$. It is easy to see that these eigenvectors are orthogonal to each other.

It follows from $\left(\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\right)\mathbf{h}_t = \mathbf{h}_t$ that $\frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top\mathbf{h}_t = \mathbf{0}$ for all $t = 1, ..., n-1$. Consequently, for all $t = 1, ..., n-1$ and all $\ell = 1, ..., q$ the vector $\mathbf{h}_t \otimes \mathbf{w}_\ell(\xi)$ is an eigenvector of $\mathrm{MSE}_B(\xi)$ corresponding to $\mu_\ell(\xi)$ and $\mu_\ell(\xi)$ is an eigenvalue of $\mathrm{MSE}_B(\xi)$ with multiplicity $n-1$.

For all $k = 1, ..., p$, $t = 1, ..., n-1$ and $\ell = 1, ..., q$ the vectors $\mathbf{1}_n \otimes \mathbf{v}_k(\xi)$ and $\mathbf{h}_t \otimes \mathbf{w}_\ell(\xi)$ are orthogonal to each other. Since the the mean squared error matrix $\mathrm{MSE}_B(\xi)$ is nonnegative definite and its rank cannot exceed $(n-1)q + p$, the other $(n-1)(p-q)$ eigenvalues are equal to zero. Hence, the following equality holds:

$$
\begin{aligned}
\prod_{\iota=1}^{(n-1)q+p} \lambda_\iota(\xi) &= \prod_{k=1}^{p} \eta_k(\xi) \prod_{\ell=1}^{q} (\mu_\ell(\xi))^{n-1} \\
&= \det(\mathbf{M}(\xi)^{-1}) \prod_{\ell=1}^{q} (\mu_\ell(\xi))^{n-1}.
\end{aligned}
$$

The logarithm of the right hand side of this equation coincides with the right hand side of (5.15). $\qquad\square$

Note that according to the proof of Lemma 6 the eigenvalues $\mu_1(\xi), ..., \mu_q(\xi)$ of the matrix $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$ are at the same time eigenvalues of the mean squared error matrix $\mathrm{MSE}_B(\xi)$ with multiplicity $n-1$, which means that for every $\mu_\ell(\xi)$ there exist $n-1$ eigenvalues $\lambda_{s_\ell}(\xi)$ with $\mu_\ell(\xi) = \lambda_{s_\ell}(\xi)$, $\ell = 1, ..., q$, $s_\ell = 1, ..., (n-1)q$. Also the eigenvalues, $\eta_1(\xi), ..., \eta_p(\xi)$ of $\mathbf{M}(\xi)^{-1}$ are at the same time eigenvalues of $\mathrm{MSE}_B(\xi)$ with multiplicity 1. The other eigenvalues of the mean squared error matrix are equal to zero.

If the dispersion matrix $\mathbf{D}$ is non-singular, the (generalized) D-criterion simplifies.

Corollary 11. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the (generalized) D-criterion for the prediction of the individual parameters simplifies to*

$$\mathrm{D}_B(\xi) = \ln\det(\mathbf{M}(\xi)^{-1}) + (n-1)\ln\det((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}). \tag{5.16}$$

The last result is a weighted sum of the D-criterion $\ln\det(\mathbf{M}(\xi)^{-1})$ in the fixed effects models and the Bayesian D-criterion $\ln\det((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1})$ represented in Chapter 2 and may be also recognized as a particular case of the compound criteria considered by Cook and Wong (1994) (see also Läuter (1974), Atkinson and Donev (1992), ch. 21, Atkinson *et al.* (2007), ch. 10, 21).

The next lemmas give explicit representations of the (generalized) D-criterion for the prediction of the individual deviations introduced in Definition 14 in both cases of a singular and a non-singular dispersion matrix of random effects. Note that the individual deviations $\mathbf{b}$ are predictable for any individual design matrix $\mathbf{F}$. Consequently, the individual information matrix $\mathbf{M}(\xi)$ may be singular.

Lemma 7. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the (generalized) D-criterion for the prediction of the individual deviations is equal to*

$$\mathrm{D}_b(\xi) = \ln\det(\boldsymbol{\Delta}) + (n-1)\ln\det((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}). \tag{5.17}$$

Proof. From the nonsingularity of the dispersion matrix $\mathbf{D}$ of random effects, which implies $q = p$ follows the nonsingularity of the adjusted dispersion matrix $\boldsymbol{\Delta}$. Then the matrix $\boldsymbol{\Delta}$ is positive definite and consequently has p positive eigenvalues $\tau_1(\xi), ..., \tau_p(\xi)$ with corresponding pairwise orthogonal eigenvectors $\mathbf{u}_1(\xi), ..., \mathbf{u}_p(\xi)$. Analogously to the proof of Lemma 6 it can be shown that $\tau_1(\xi), ..., \tau_p(\xi)$ are at the same time eigenvalues

of the mean squared error matrix $\mathrm{MSE}_b(\xi)$ with pairwise orthogonal eigenvectors $\mathbf{1}_n \otimes \mathbf{u}_1(\xi), ..., \mathbf{1}_n \otimes \mathbf{u}_p(\xi)$.

Since by (5.14) the matrix $(\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}$ is equivalent to the matrix $\boldsymbol{\Delta}-\boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1}+\boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$ for a non-singular adjusted dispersion matrix $\boldsymbol{\Delta}$, we use here the same notations $\mu_1(\xi), ..., \mu_p(\xi)$ and $\mathbf{w}_1(\xi), ..., \mathbf{w}_p(\xi)$ for the p positive eigenvalues of $(\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}$ and the corresponding pairwise orthogonal eigenvectors as in the proof of Lemma 6. According to that proof $\mu_1(\xi), ..., \mu_p(\xi)$ are also eigenvalues of $\mathrm{MSE}_b(\xi)$ with algebraic multiplicity $n-1$ and pairwise orthogonal eigenvectors $\mathbf{h}_t \otimes \mathbf{w}_\ell(\xi)$, $t = 1, ..., n-1$, $\ell = 1, ..., p$.

Since for all $k = 1, ..., p$, $t = 1, ..., n-1$ and $\ell = 1, ..., p$ the vectors $\mathbf{1}_n \otimes \mathbf{u}_k(\xi)$ and $\mathbf{h}_t \otimes \mathbf{w}_\ell(\xi)$ are orthogonal to each other and the dimension of $\mathrm{MSE}_b(\xi)$ is equal to np, there mean squared error matrix has no other eigenvalues. Consequently,

$$\prod_{\iota=1}^{np} \nu_\iota(\xi) \;=\; \prod_{k=1}^{p} \eta_k(\xi) \prod_{\ell=1}^{p} (\mu_\ell(\xi))^{n-1}.$$

Since both matrices $\boldsymbol{\Delta}$ and $(\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}$ are non-singular, their determinants are equal to the products of their eigenvalues, which implies for $q = p$ the result. $\qquad\square$

Note that in the criterion (5.17) the first term is constant and the second term is proportional to the Bayesian D-criterion (2.22).

Corollary 12. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, Bayesian D-optimal designs are D-optimal for the prediction of the individual deviations.*

Now we allow for a singular dispersion matrix of random effects. According to Remark 3 singular dispersion matrices with the special structure (4.33) only have to be considered.

Lemma 8. *If the dispersion matrix $\mathbf{D}$ of random effects is singular and has the block diagonal form (4.33), the (generalized) D-criterion for the prediction of the individual deviations is equal to*

$$D_b(\xi) = \ln \det(\boldsymbol{\Delta}_0) + (n-1) \ln \det((\mathbf{M}_0(\xi) + \boldsymbol{\Delta}_0^{-1})^{-1}), \tag{5.18}$$

where $\mathbf{M}_0(\xi) = \frac{1}{m}\mathbf{F}_0^\top \mathbf{F}_0$ and $\boldsymbol{\Delta}_0 = m\mathbf{D}_0$ are the individual information matrix and the dispersion matrix of random effects in the reduced model (4.32).

Proof. Since the positive eigenvalues of the matrices

$$\text{MSE}_b\left(\xi\right) = \left(\mathbf{I}_n \otimes \left(\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)}\right)^\top\right) \text{MSE}_{b_0}\left(\xi\right) \left(\mathbf{I}_n \otimes \left(\mathbf{I}_q \ \mathbf{0}_{q\times(p-q)}\right)\right)$$

and $\text{MSE}_{b_0}\left(\xi\right)$ are the same and the mean squared error matrix $\text{MSE}_{b_0}\left(\xi\right)$ in the reduced model (4.32) is of the form (5.3), the result follows directly from Definition 14 and Lemma 7. $\qquad\square$

Note that (5.18) coincides with the D-criterion (5.17) for the prediction of the individual deviations in the reduced model (4.32).

Corollary 13. *If the dispersion matrix* $\mathbf{D}$ *of random effects is singular and has the block diagonal form (4.33), D-optimal designs for the prediction of the individual deviations in the reduced model (4.32) are D-optimal for the prediction of the individual deviations in (4.2).*

It follows from Corollaries 12 and 13 that Bayesian D-optimal designs, which minimize the Bayesian D-criterion $\ln\det(\mathbf{M}_0(\xi) + \mathbf{\Delta}_0^{-1})^{-1}$ in the corresponding reduced model, are D-optimal for the prediction of the individual deviations in the random coefficient regression model (4.2).

Since the (generalized) D-criterion for the prediction of the individual deviations does not simplify to some well known criterion, the optimality condition has to be formulated explicitly in this case.

Theorem 17. *The approximate design* $\xi^* \in \Xi$ *is D-optimal for the prediction of the individual parameters if and only if*

$$\mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x) + (n-1)\,\mathbf{f}(x)^\top(\mathbf{\Delta} - \mathbf{\Delta}(\mathbf{M}(\xi^*)^{-1} + \mathbf{\Delta})^{-1}\mathbf{\Delta})\mathbf{f}(x)$$
$$\leq p + (n-1)\operatorname{tr}\left(\mathbf{\Delta}(\mathbf{M}(\xi^*)^{-1} + \mathbf{\Delta})^{-1}\right) \tag{5.19}$$

for all $x \in \mathcal{X}$ *Moreover, for any support point* x_j^* *of* ξ^* *with positive weight* $(m_j^* > 0)$ *equality holds in (5.19).*

Proof. According to Silvey (1980), ch. 3 (see also Fedorov and Hackl (1997), ch. 2) a design ξ^* minimizes a convex criterion function Φ if and only if the directional derivative $F_\Phi(\mathbf{M}(\xi^*), \mathbf{M}(\xi_x))$ of Φ at $\mathbf{M}(\xi^*)$ in the direction of the information matrix $\mathbf{M}(\xi_x) =$

$\mathbf{f}(x)\mathbf{f}(x)^\top$ of the one-point design ξ_x is nonnegative for all $x \in \mathcal{X}$. Moreover, this directional derivative is zero for all support points x_j^* $(m_j^* > 0)$ of the optimal design ξ^*.

According to formula (5.13) the (generalized) D-criterion (5.15) may be represented as the following function of the individual information matrix $\mathbf{M}$:

$$
\begin{aligned}
\Phi(\mathbf{M}) &= \ln\det(\mathbf{M}^{-1}) + (n-1)\ln\det((m\,\mathbf{H}^\top\mathbf{M}\mathbf{H}+\mathbf{I}_q)^{-1}\,m\,\mathbf{H}^\top\mathbf{H}) \\
&= \ln\det(\mathbf{M}^{-1}) + (n-1)\ln\det((m\,\mathbf{H}^\top\mathbf{M}\mathbf{H}+\mathbf{I}_q)^{-1}) + (n-1)\ln\det(m\,\mathbf{H}^\top\mathbf{H}),
\end{aligned}
$$

which is convex as a sum of two convex functions and one constant term. Therefore, the corresponding directional derivative $\mathrm{F}_\Phi(\mathbf{M}(\xi^*), \mathbf{M}(\xi_x))$ has to be nonnegative for all $x \in \mathcal{X}$ and equal zero for all support points of ξ^*.

The first term $\Phi_1(\mathbf{M}) := \ln\det(\mathbf{M}^{-1})$ of the criterion function $\Phi(\mathbf{M})$ coincides with the D-criterion in the fixed effects model (2.1) and is hence convex. In the second term $\Phi_2(\mathbf{M}) := \ln\det(m\,\mathbf{H}^\top\mathbf{M}\mathbf{H}+\mathbf{I}_q)^{-1}$ may be considered as the Bayesian D-criterion in the model

$$
Y_j = \mathbf{f}(x_j)^\top\mathbf{H}\mathbf{c} + \varepsilon_j \tag{5.20}
$$

with $\mathrm{E}(\mathbf{c}) = \mathbf{0}_q$ and $\mathrm{Cov}(\mathbf{c}) = \sigma^2\mathbf{I}_q$ (see e.g. Gladitz and Pilz (1982)) and is therefore convex. Note that (5.20) is a special case of the model (2.9) with regression functions $\mathbf{H}^\top\mathbf{f}$.

Now the directional derivative $\mathrm{F}_\Phi(\mathbf{M}_1, \mathbf{M}_2)$ of Φ at $\mathbf{M}_1$ in the direction of $\mathbf{M}_2$ may be determined as weighted sum of the directional derivatives of its components:

$$
\mathrm{F}_{\Phi_1}(\mathbf{M}_1, \mathbf{M}_2) = p - \mathrm{tr}\,(\mathbf{M}_1^{-1}\mathbf{M}_2)
$$

and

$$
\mathrm{F}_{\Phi_2}(\mathbf{M}_1, \mathbf{M}_2) = \mathrm{tr}\,(m\,\mathbf{H}(m\,\mathbf{H}^\top\mathbf{M}_1\mathbf{H}+\mathbf{I}_q)^{-1}\mathbf{H}^\top(\mathbf{M}_1 - \mathbf{M}_2)).
$$

The derivatives $\mathrm{F}_{\Phi_1}(\mathbf{M}_1, \mathbf{M}_2)$ and $\mathrm{F}_{\Phi_2}(\mathbf{M}_1, \mathbf{M}_2)$ may be represented using the equality (5.13) as

$$
\mathrm{F}_{\Phi_1}(\mathbf{M}(\xi^*), \mathbf{f}(x)\mathbf{f}(x)^\top) = p - \mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x)
$$

and

$$
\begin{aligned}
F_{\Phi_2}(\mathbf{M}(\xi^*), \mathbf{f}(x)\mathbf{f}(x)^\top) &= \operatorname{tr}\left((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\,\mathbf{M}(\xi^*)\right) \\
&\quad - \mathbf{f}(x)^\top(\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\mathbf{f}(x) \\
&= \operatorname{tr}\left(\boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\right) \\
&\quad - \mathbf{f}(x)^\top(\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\mathbf{f}(x)
\end{aligned}
$$

for $\mathbf{M}_1 = \mathbf{M}(\xi^*)$ and $\mathbf{M}_2 = \mathbf{f}(x)\mathbf{f}(x)^\top$, which implies the result (5.19). $\qquad\square$

For a non-singular dispersion matrix $\mathbf{D}$ the optimality condition (5.19) simplifies by formula (5.14) to that of a compound criterion.

Corollary 14. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the approximate design $\xi^* \in \Xi$ is D-optimal for the prediction of the individual parameters if and only if*

$$
\begin{aligned}
&\mathbf{f}(x)^\top \mathbf{M}(\xi^*)^{-1}\mathbf{f}(x) + (n-1)\,\mathbf{f}(x)^\top(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1}\mathbf{f}(x) \\
&\qquad \le\; p + (n-1)\operatorname{tr}\left((\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1}\mathbf{M}(\xi^*)\right) \qquad (5.21)
\end{aligned}
$$

for all $x \in \mathcal{X}$ Moreover, for any support point x_j^* of ξ^* with positive weight $(m_j^* > 0)$ equality holds in (5.21).

Example 1 (cont.) We consider again the particular case of random intercepts. To derive an explicit form of the (generalized) D-criterion (5.15) for the prediction of the individual parameters, we use the property (5.13) of the matrix $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}$. Since $\mathbf{H} = \sqrt{d_1}\mathbf{e}_1$ satisfies the condition $\mathbf{H}\mathbf{H}^\top = \mathbf{D}$, we obtain $\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta} = \frac{md_1}{1+md_1}\mathbf{e}_1\mathbf{e}_1^\top$ with the largest eigenvalue $\lambda_1(\xi) = \frac{md_1}{1+md_1}$. Then the (generalized) D-criterion (5.15) for the prediction of the individual parameters simplifies to

$$
D_B(\xi) = \ln\det(\mathbf{M}(\xi)^{-1}) + (n-1)\ln\frac{md_1}{1+md_1}, \qquad (5.22)
$$

which establishes the following result.

Corollary 15. *D-optimal designs in the fixed effects models are D-optimal for the prediction of the individual parameters in the random intercepts model.*

For the prediction of the individual deviations we obtain the mean squared error matrix $\mathrm{MSE}_b(\xi)$ by formula (5.4) using $\mathbf{\Delta}_0 = md_1$ and $\mathbf{M}_0(\xi) = 1$:

$$\mathrm{MSE}_b(\xi) = \left(\frac{md_1}{n} \mathbf{1}_n \mathbf{1}_n^\top + \frac{md_1}{1 + md_1} (\mathbf{I}_n - \frac{1}{n} \mathbf{1}_n \mathbf{1}_n^\top) \right) \otimes (\mathbf{e}_1 \mathbf{e}_1^\top), \qquad (5.23)$$

which coincides for exact designs with (4.43) divided by the factor $\frac{\sigma^2}{m}$. Since the mean squared error matrix (5.23) is constant, the (generalized) D-criterion (5.12) is independent of the design, which means that all designs of the form (2.16) are equally good for the prediction.

5.2 Linear Criteria

Similarly to the Bayesian L-criterion (2.26), the linear criterion for the prediction of the individual parameters for a single individual i is defined (for exact designs) as the trace of the mean squared error matrix of a vector of linear combinations $\mathcal{L}_i \boldsymbol{\beta}_i$, where $\mathcal{L}_i$ is a specified matrix of dimension $\tau \times p$. Linear combinations of the parameters across the individuals are not the subject of this work and the interest is the same in all individuals ($\mathcal{L}_i = \mathcal{L}$, for $i = 1, ..., n$). Therefore, we may define the L-criterion as the sum of the linear criteria for the individuals:

$$\mathrm{L}_B(\xi) = \sum_{i=1}^n \mathrm{tr}\left(\mathrm{Cov}\,(\mathcal{L}\hat{\boldsymbol{\beta}}_i - \mathcal{L}\boldsymbol{\beta}_i)\right), \qquad (5.24)$$

where $\mathcal{L}$ is some matrix of dimension $\tau \times p$.

Further we suppress for simplicity the constant factor $\frac{\sigma^2}{m}$ and generalize the criterion (5.24) to the case of approximate designs.

Definition 15. *The approximate design $\xi^* \in \Xi$ is called L-optimal with respect to $\mathcal{A} = \mathcal{L}^\top \mathcal{L}$ for the prediction of the individual parameters if*

$$\mathrm{L}_B(\xi^*) = \min_{\xi \in \Xi} \mathrm{L}_B(\xi) \qquad (5.25)$$

for

$$\mathrm{L}_B(\xi) = \mathrm{tr}\left(\mathrm{MSE}_B\,(\xi)\,(\mathbf{I}_n \otimes \mathcal{A})\right). \qquad (5.26)$$

Analogously we define the linear criterion for the prediction of the individual deviations using the corresponding mean squared error matrix:

$$L_b(\xi) = \sum_{i=1}^{n} \operatorname{tr}\left(\operatorname{Cov}\left(\mathcal{L}\hat{\mathbf{b}}_i - \mathcal{L}\mathbf{b}_i\right)\right). \tag{5.27}$$

For approximate designs this criterion can be generalized and represented in the following equivalent form.

Definition 16. *The approximate design $\xi^* \in \Xi$ is called L-optimal with respect to $\mathcal{A}$ for the prediction of the individual deviations if*

$$L_b(\xi^*) = \min_{\xi \in \Xi} L_b(\xi) \tag{5.28}$$

for

$$L_b(\xi) = \operatorname{tr}\left(\operatorname{MSE}_b\left(\xi\right)\left(\mathbf{I}_n \otimes \mathcal{A}\right)\right). \tag{5.29}$$

The next lemma gives a more explicit form of the linear criterion (5.26) for the prediction of the individual parameters.

Lemma 9. *The L-criterion for the prediction of the individual parameters is equal to*

$$L_B(\xi) = \operatorname{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{A}\right) + (n-1)\operatorname{tr}\left((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\mathcal{A}\right). \tag{5.30}$$

Proof. Due to formula (5.1) the *L*-criterion (5.26) may be represented in the form

$$L_B(\xi) = \operatorname{tr}\left(\frac{1}{n}(\mathbf{1}_n\mathbf{1}_n^\top) \otimes (\mathbf{M}(\xi)^{-1}\mathcal{A}) + (\mathbf{I}_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^\top) \otimes ((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\mathcal{A})\right)$$

and results using properties of the Kronecker product in

$$L_B(\xi) = \operatorname{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{A}\right) + (n-1)\operatorname{tr}\left((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\mathcal{A}\right).$$

$\square$

If the dispersion matrix of random effects is non-singular, the criterion (5.30) may be simplified using the equality (5.14).

Corollary 16. *If the dispersion matrix $\mathbf{D}$ of individual effects is non-singular, the L-criterion for the prediction of the individual parameters simplifies to*

$$L_B(\xi) = \operatorname{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{A}\right) + (n-1)\operatorname{tr}\left((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}\mathcal{A}\right). \tag{5.31}$$

In this case the L-criterion for the prediction of the individual parameters is proportional to a weighted average of the linear criterion $\operatorname{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{A}\right)$ in the model without random effects and the corresponding Bayesian linear criterion $\operatorname{tr}\left((\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}\mathcal{A}\right)$. As a consequence, the linear criterion can also be interpreted as a compound criterion.

For the prediction of the individual deviations the two cases of a singular and a non-singular dispersion matrix will be considered separately as it has been done for the (generalized) D-criterion in Section 5.1.

Lemma 10. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the L-criterion for the prediction of the individual deviations is equal to*

$$L_b(\xi) = \operatorname{tr}\left(\boldsymbol{\Delta}\,\mathcal{A}\right) + (n-1)\operatorname{tr}\left((\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}\mathcal{A}\right). \tag{5.32}$$

Proof. The result follows from the form (5.29) of the L-criterion and formula (5.3). □

Note that in the criterion (5.32) the first term $\operatorname{tr}\left(\boldsymbol{\Delta}\,\mathcal{A}\right)$ is constant (independent of the designs) and the second term $(n-1)\operatorname{tr}\left((\mathbf{M}(\xi)+\boldsymbol{\Delta}^{-1})^{-1}\mathcal{A}\right)$ is proportional to the Bayesian L-criterion (2.28).

Corollary 17. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, Bayesian L-optimal designs are L-optimal for the prediction of the individual deviations.*

For $\mathbf{M}_0(\xi)$ and $\boldsymbol{\Delta}_0$ from Lemma 8 an explicit form of the L-criterion for the prediction of the individual deviations in the case of a singular dispersion matrix $\mathbf{D}$ is given by the following lemma.

Lemma 11. *If the dispersion matrix $\mathbf{D}$ of random effects is singular and has the block diagonal form (4.33), the L-criterion for the prediction of the individual deviations is equal to*

$$L_b(\xi) = \operatorname{tr}\left(\boldsymbol{\Delta}_0\,\mathcal{A}_0\right) + (n-1)\operatorname{tr}\left((\mathbf{M}_0(\xi)+\boldsymbol{\Delta}_0^{-1})^{-1}\mathcal{A}_0\right), \tag{5.33}$$

where $\mathcal{A}_0 = \left(\mathbf{I}_q\ \mathbf{0}_{q\times(p-q)}\right)\mathcal{A}\left(\mathbf{I}_q\ \mathbf{0}_{q\times(p-q)}\right)^{\top}$.

Proof. Using (5.4) the L-criterion (5.29) can be rewritten as

$$L_b(\xi) = \operatorname{tr}\left(\operatorname{MSE}_{b_0}(\xi)\left(\mathbf{I}_n\otimes\mathcal{A}_0\right)\right)$$

and results in (5.33) after applying (5.5). □

Note that (5.33) coincides for $\mathcal{A} = \mathcal{A}_0$ with the L-criterion (5.32) in the reduced model (4.32).

Corollary 18. *If the dispersion matrix $\mathbf{D}$ of random effects is singular and has the block diagonal form (4.33), L-optimal designs for the prediction of the individual deviations in the reduced model (4.32) are L-optimal for the prediction of the individual deviations in (4.2).*

Due to Corollaries 17 and 18 Bayesian L-optimal designs, which minimize the Bayesian L-criterion $\operatorname{tr}\left((\mathbf{M}_0(\xi) + \boldsymbol{\Delta}_0^{-1})^{-1}\mathcal{A}_0\right)$, are L-optimal for the prediction of the individual deviations in (4.2).

Remark 4. *The c- and A-criteria for the prediction of the individual parameters and individual deviations are special cases of the linear criteria (5.30), (5.32) and (5.33) with $\mathcal{L} = \mathbf{c}^\top$ and $\mathcal{L} = \mathbf{I}_p$ respectively.*

Another frequently used particular case of the linear criterion is the integrated mean squared error criterion (IMSE-criterion). For the prediction of the individual parameters this criterion will be defined (for exact design) as the sum over all individuals of the integrated mean squared differences between the predicted and the real response, $\mathbf{f}(x)^\top\hat{\boldsymbol{\beta}}_i$ and $\mathbf{f}(x)^\top\boldsymbol{\beta}_i$, respectively:

$$\mathrm{IMSE}_B(\xi) = \sum_{i=1}^{n} \int_{\mathcal{X}} \mathrm{Cov}\left(\mathbf{f}(x)^\top\hat{\boldsymbol{\beta}}_i - \mathbf{f}(x)^\top\boldsymbol{\beta}_i\right)\nu(\mathrm{d}x), \qquad (5.34)$$

where ν denotes a specified weight distribution on the design region $\mathcal{X}$.

Then we suppress the factor $\frac{\sigma^2}{m}$ and generalize the IMSE-criterion (5.34) to the case of approximate designs as it has been done for the linear criterion (5.24).

Definition 17. *The approximate design $\xi^* \in \Xi$ is called IMSE-optimal for the prediction of the individual parameters if*

$$\mathrm{IMSE}_B(\xi^*) = \min_{\xi \in \Xi} \mathrm{IMSE}_B(\xi) \qquad (5.35)$$

for

$$\mathrm{IMSE}_B(\xi) = \operatorname{tr}\left(\mathrm{MSE}_B(\xi)\left(\mathbf{I}_n \otimes \mathcal{V}\right)\right). \qquad (5.36)$$

where $\mathcal{V} = \int_{\mathcal{X}} \mathbf{f}(x)\mathbf{f}(x)^\top\nu(\mathrm{d}x)$.

We define (for exact designs) the IMSE-criterion for the prediction of the individual deviations using the predicted and the real response $\mathbf{f}(x)^\top \hat{\mathbf{b}}_i$ and $\mathbf{f}(x)^\top \mathbf{b}_i$ instead of $\mathbf{f}(x)^\top \hat{\boldsymbol{\beta}}_i$ and $\mathbf{f}(x)^\top \boldsymbol{\beta}_i$:

$$\text{IMSE}_b(\xi) = \sum_{i=1}^n \int_{\mathcal{X}} \text{Cov}\left(\mathbf{f}(x)^\top \hat{\mathbf{b}}_i - \mathbf{f}(x)^\top \mathbf{b}_i\right) \nu(\mathrm{d}x). \tag{5.37}$$

This criterion can also be generalized for approximate designs.

Definition 18. *The approximate design $\xi^* \in \Xi$ is called IMSE-optimal for the prediction of the individual deviations if*

$$\text{IMSE}_b(\xi^*) = \min_{\xi \in \Xi} \text{IMSE}_b(\xi) \tag{5.38}$$

for

$$\text{IMSE}_b(\xi) = \text{tr}\left(\text{MSE}_b(\xi)\left(\mathbf{I}_n \otimes \mathcal{V}\right)\right). \tag{5.39}$$

According to the next lemma the IMSE-criterion for the prediction of the individual parameters is a particular case of the corresponding L-criterion for the special matrix $\mathcal{A} = \mathcal{V}$, which may be interpreted as the individual information matrix for the weight distribution ν (considered as a design).

Lemma 12. *The IMSE-criterion for the prediction of the individual parameters is equal to*

$$\text{IMSE}_B(\xi) = \text{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{V}\right) + (n-1)\,\text{tr}\left((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\,\mathcal{V}\right). \tag{5.40}$$

Proof. The result follows directly from (5.36) after applying formula (5.1). □

Note that the matrix $\mathcal{V}$ is non-negative definite. Then a $\tau \times p$ matrix $\mathcal{L}$ exists with $\mathcal{L}^\top \mathcal{L} = \mathcal{V}$, which means that the IMSE-criterion (5.40) may be represented in the form (5.24) in the case of exact designs.

Corollary 19. *If the dispersion matrix $\mathbf{D}$ of individual effects is non-singular, the IMSE-criterion for the prediction of the individual parameters simplifies to*

$$\text{IMSE}_B(\xi) = \text{tr}\left(\mathbf{M}(\xi)^{-1}\mathcal{V}\right) + (n-1)\,\text{tr}\left((\mathbf{M}(\xi) + \boldsymbol{\Delta}^{-1})^{-1}\mathcal{V}\right). \tag{5.41}$$

The next lemma gives an explicit form of the IMSE-criterion for the prediction of the individual deviations in the case of a non-singular dispersion matrix of random effects.

Lemma 13. *If the dispersion matrix* $\mathbf{D}$ *of random effects is non-singular, the IMSE-criterion for the prediction of the individual deviations is equal to*

$$\mathrm{IMSE}_b(\xi) = \mathrm{tr}\,(\mathbf{\Delta}\,\mathcal{V}) + (n-1)\,\mathrm{tr}\,((\mathbf{M}(\xi) + \mathbf{\Delta}^{-1})^{-1}\,\mathcal{V}). \tag{5.42}$$

Proof. The result follows from (5.39) and (5.3). □

In the next lemma we present an explicit form of the IMSE-criterion for the prediction of the individual deviations in the case of a singular (block diagonal) dispersion matrix $\mathbf{D}$ for $\mathbf{M}_0(\xi)$ and $\mathbf{\Delta}_0$ defined in Lemma 8.

Lemma 14. *If the dispersion matrix* $\mathbf{D}$ *of random effects is singular and has the block diagonal form (4.33), the IMSE-criterion for the prediction of the individual deviations is equal to*

$$\mathrm{IMSE}_b(\xi) = \mathrm{tr}\,(\mathbf{\Delta}_0\,\mathcal{V}_0) + (n-1)\,\mathrm{tr}\,((\mathbf{M}_0(\xi) + \mathbf{\Delta}_0^{-1})^{-1}\,\mathcal{V}_0), \tag{5.43}$$

where $\mathcal{V}_0 = (\mathbf{I}_q\ \mathbf{0}_{q\times(p-q)})\,\mathcal{V}\,(\mathbf{I}_q\ \mathbf{0}_{q\times(p-q)})^\top.$

Proof. Using (5.39) and (5.4) we obtain the following form of the IMSE-criterion:

$$\mathrm{IMSE}_b(\xi) = \mathrm{tr}\,(\mathrm{MSE}_{b_0}(\xi)\,(\mathbf{I}_n \otimes \mathcal{V}_0))\,.$$

Then we apply (5.5) in the last formula and get the result. □

It may be easily shown that $\mathcal{V}_0 = \int_{\mathcal{X}} \mathbf{f}_0(x)\mathbf{f}_0(x)^\top \nu(\mathrm{d}x)$, where $\mathbf{f}_0$ are the regression functions in the reduced model (4.32). Then the IMSE-criterion (5.43) coincides with the criterion (5.42) in the reduced model and is consequently a special case of the L-criterion (5.32) (for $\mathcal{A} = \mathcal{V}_0$). Furthermore, the IMSE-criteria (5.42) and (5.43) may be represented in the form (5.29).

Now we formulate the optimality condition for the linear criteria for the prediction of the individual parameters.

Theorem 18. *The approximate design $\xi^* \in \Xi$ is L-optimal with respect to $\mathcal{A}$ for the prediction of the individual parameters if and only if*

$$
\mathbf{f}(x)^\top \mathbf{M}(\xi^*)^{-1} \mathcal{A}\, \mathbf{M}(\xi^*)^{-1} \mathbf{f}(x) \tag{5.44}
$$
$$
+ (n-1)\, \mathbf{f}(x)^\top \mathbf{M}(\xi^*)^{-1}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1} \boldsymbol{\Delta}\, \mathcal{A}\, \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x)
$$
$$
\leq\ \mathrm{tr}\,(\mathbf{M}(\xi^*)^{-1}\mathcal{A}) + (n-1)\,\mathrm{tr}\,(\boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\mathbf{M}(\xi^*)^{-1}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\,\mathcal{A})
$$

for all $x \in \mathcal{X}$ Moreover, for any support point x_j^* of ξ^* with positive weight $(m_j^* > 0)$ equality holds in (5.44).

Proof. The L-criterion (5.30) may be rewritten in the form

$$
\Phi(\mathbf{M}) = \mathrm{tr}\,(\mathbf{M}^{-1}\mathcal{A}) + (n-1)\,\mathrm{tr}\,(m\mathbf{H}(m\,\mathbf{H}^\top\mathbf{M}\mathbf{H} + \mathbf{I}_q)^{-1}\,\mathbf{H}^\top\mathcal{A})
$$

using formula (5.13). Then this criterion is a convex function of the individual information matrix $\mathbf{M}$ as a weighted sum of the L-criterion $\Phi_1(\mathbf{M}) := \mathrm{tr}\,(\mathbf{M}^{-1}\mathcal{A})$ in the model without random effects and the term $\Phi_2(\mathbf{M}) := \mathrm{tr}\,(m\mathbf{H}(m\,\mathbf{H}^\top\mathbf{M}\mathbf{H} + \mathbf{I}_q)^{-1}\,\mathbf{H}^\top\mathcal{A})$, which may be recognized as the Bayesian L-criterion in the model (5.20).

The directional derivatives of the functions $\Phi_1(\mathbf{M})$ and $\Phi_2(\mathbf{M})$ at $\mathbf{M_1}$ in the direction of $\mathbf{M_2}$ are given by

$$
F_{\Phi_1}(\mathbf{M}_1, \mathbf{M}_2) = \mathrm{tr}\,(\mathbf{M}_1^{-1}\mathcal{A}) - \mathrm{tr}\,(\mathbf{M}_1^{-1}\mathbf{M}_2\,\mathbf{M}_1^{-1}\mathcal{A})
$$

and

$$
F_{\Phi_2}(\mathbf{M}_1, \mathbf{M}_2) = \mathrm{tr}\,(m\,\mathbf{H}(m\,\mathbf{H}^\top\mathbf{M}_1\mathbf{H} + \mathbf{I}_q)^{-1}\mathbf{H}^\top(\mathbf{M}_1 - \mathbf{M}_2)m\,\mathbf{H}(m\,\mathbf{H}^\top\mathbf{M}_1\mathbf{H} + \mathbf{I}_q)^{-1}\mathbf{H}^\top\mathcal{A}).
$$

Further for $\mathbf{M}_1 = \mathbf{M}(\xi^*)$ and $\mathbf{M}_2 = \mathbf{f}(x)\mathbf{f}(x)^\top$ we obtain

$$
F_{\Phi_1}(\mathbf{M}(\xi^*), \mathbf{f}(x)\mathbf{f}(x)^\top) = \mathrm{tr}\,(\mathbf{M}(\xi)^{-1}\mathcal{A}) - \mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}\mathcal{A}\,\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x).
$$

Using (5.13) again we get

$$
F_{\Phi_2}(\mathbf{M}(\xi^*), \mathbf{f}(x)\mathbf{f}(x)^\top)
$$
$$
=\ \mathrm{tr}\,((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\,\mathbf{M}(\xi^*)(\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\,\mathcal{A})
$$
$$
-\ \mathbf{f}(x)^\top((\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\,\mathcal{A}\,(\boldsymbol{\Delta} - \boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta})\,\mathbf{f}(x)
$$
$$
=\ \mathrm{tr}\,(\boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\mathbf{M}(\xi^*)^{-1}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\,\mathcal{A})
$$
$$
-\ \mathbf{f}(x)^\top\mathbf{M}(\xi^*)^{-1}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\boldsymbol{\Delta}\,\mathcal{A}\,\boldsymbol{\Delta}(\mathbf{M}(\xi^*)^{-1} + \boldsymbol{\Delta})^{-1}\mathbf{M}(\xi^*)^{-1}\mathbf{f}(x).
$$

Then the optimality condition (5.44) follows directly from the general equivalence theorem (see e. g. Silvey (1980)). □

According to formula (5.14) the optimality condition (5.44) simplifies to that of a compound criterion, if the dispersion matrix of random effects is non-singular.

Corollary 20. *If the dispersion matrix* $\mathbf{D}$ *of random effects is non-singular, the approximate design* $\xi^* \in \Xi$ *is L-optimal with respect to* $\mathcal{A}$ *for the prediction of the individual parameters if and only if*

$$\mathbf{f}(x)^\top \mathbf{M}(\xi^*)^{-1} \mathcal{A} \, \mathbf{M}(\xi^*)^{-1} \mathbf{f}(x) + (n-1)\, \mathbf{f}(x)^\top (\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1} \mathcal{A} \, (\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1} \mathbf{f}(x)$$
$$\leq \; \operatorname{tr}(\mathbf{M}(\xi^*)^{-1} \mathcal{A}) + (n-1)\operatorname{tr}((\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1}\mathbf{M}(\xi^*)(\mathbf{M}(\xi^*) + \boldsymbol{\Delta}^{-1})^{-1}\mathcal{A}) \quad (5.45)$$

for all $x \in \mathcal{X}$ *Moreover, for any support point* x_j^* *of* ξ^* *with positive weight* $(m_j^* > 0)$ *equality holds in* (5.45).

Remark 5. *The optimality conditions for the IMSE-, c- and A-criteria are special cases of the conditions (5.44)-(5.45) for* $\mathcal{A} = \mathcal{V}$, $\mathcal{A} = \mathbf{c}\mathbf{c}^\top$ *and* $\mathcal{A} = \mathbf{I}_p$ *respectively.*

Example 1 (cont.) In the particular case of random intercepts the L-criterion (5.30) for the prediction of the individual parameters simplifies to

$$\mathrm{L}_B(\xi) = \operatorname{tr}(\mathbf{M}(\xi)^{-1}\mathcal{A}) + \frac{(n-1)\, m d_1}{1 + m d_1}\mathbf{e}_1^\top \mathcal{A}\, \mathbf{e}_1, \quad (5.46)$$

which depends on the dispersion matrix only through an additive constant term.

Corollary 21. *L-optimal designs in the fixed effects models are L-optimal for the prediction of the individual parameters in the random intercepts model.*

According to formula (5.23) the mean squared error matrix of the prediction of the individual deviations is constant in the random intercepts case. Then it follows from the form (5.39) of the L-criterion that the criterion function is independent of the designs and all designs of the form (2.16) are optimal for the prediction.

5.3 Example: Linear Regression with Random Slope

As shown in sections 5.1 and 5.2, the D- or L-optimal designs are independent of the dispersion matrix $\mathbf{D}$ of random effects in the random intercepts model. In the more general situation, where the random coefficients are associated with the effects of explanatory variables, the influence of the dispersion matrix may become crucial. To illustrate this we consider the linear regression model

$$Y_{ij} = \beta_{i1} + \beta_{i2} x_j + \varepsilon_{ij} \tag{5.47}$$

on the experimental region $\mathcal{X} = [0, 1]$. For simplicity we assume the same (constant) intercept $\beta_{i1} \equiv \beta_1$ for all individuals, which implies the dispersion matrix $\mathbf{D} = d_2 \mathbf{e}_2 \mathbf{e}_2^\top$ of the random effects, where $\mathbf{e}_2 = (0, 1)^\top$. The latter assumption accentuates the dependence of the optimal design on the dispersion, when the variance of the slope is large compared to the variance of the baseline (intercept).

Since the model (5.47) is a particular case of (4.2) with regression functions $\mathbf{f}(x) = (1, x)^\top$, we obtain the following general form of the individual information matrix $\mathbf{M}(\xi)$ using (2.17):

$$\mathbf{M}(\xi) = \begin{pmatrix} 1 & \sum_{j=1}^k w_j x_j \\ \sum_{j=1}^k w_j x_j & \sum_{j=1}^k w_j x_j^2 \end{pmatrix}, \tag{5.48}$$

where the proportion w_j of the number m_j of replications at the point x_j is given by $w_j = \frac{m_j}{m}$.

As already mentioned at the beginning of this chapter, the individual information matrix $\mathbf{M}(\xi)$ is assumed to be non-singular in the case of prediction of the individual parameters. Using the equality (5.13) it may be easily proved that the matrix $\mathbf{\Delta} - \mathbf{\Delta}(\mathbf{M}(\xi)^{-1} + \mathbf{\Delta})^{-1}\mathbf{\Delta}$ is non-negative definite. Since $\mathbf{M}(\xi)$ is positive definite, the left hand side of the optimality condition (5.19) is a parabola (in x) with a positive leading term for every dispersion matrix $\mathbf{D}$. Consequently, all D-optimal designs ξ^* for the prediction of the individual parameters may take observations only at the endpoints $x = 0$ and $x = 1$ of the design region and have the general form

$$\xi = \begin{pmatrix} 0 & 1 \\ m - m_1 & m_1 \end{pmatrix}, \tag{5.49}$$

58

where only the optimal number m_1^* of replications at the support point $x = 1$ has to be determined. Then for $w_1 = \frac{m_1}{m}$, $w_1 \in (0,1)$ the individual information matrix has the following explicit form:

$$\mathbf{M}(\xi) = \begin{pmatrix} 1 & w_1 \\ w_1 & w_1 \end{pmatrix}, \tag{5.50}$$

which results according to the (generalized) D-criterion (5.15) for the prediction of the individual parameters in the criterion function

$$\mathrm{D}_B(\xi) = \ln \frac{m^2}{m_1(m - m_1)} + (n - 1) \ln \frac{md_2}{1 + m_1 d_2}. \tag{5.51}$$

For the IMSE-criterion we use the uniform weighting $\nu = \lambda|_{[0,1]}$ on the design region, which results in the positive definite matrix

$$\mathcal{V} = \int_0^1 \mathbf{f}(x)\mathbf{f}(x)^\top \mathrm{d}x = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}. \tag{5.52}$$

Then the left hand side of the optimality condition (5.44) with $\mathcal{A} = \mathcal{V}$ is a parabola with a positive leading term for every dispersion matrix $\mathbf{D}$. Consequently, the IMSE-optimal design will have the general form (5.49) and the individual information matrix $\mathbf{M}(\xi)$ will be of the form (5.50), which implies the following criterion function:

$$\mathrm{IMSE}_B(\xi) = \frac{1}{3} \left(\frac{m^2}{m_1(m - m_1)} + (n - 1)\frac{md_2}{1 + m_1 d_2} \right). \tag{5.53}$$

The optimal number of replications m_1^* depends not only on the slope variance d_2 but also on the number of individuals n and on the number of observations m at each individual. Since we are primarily interested in the dependence of the dispersion matrix, we fix the number of individuals to $n = 100$ and the intra-individual number of observations to $m = 10$. For these given values the behaviour of the optimal number of replications in dependence of the monotonically increasing transformation $\rho = d_2/(1 + d_2)$ of d_2 (intra-class correlation), which is used to cover the whole range of possible values by a finite interval ($\rho \in (0, 1)$), is illustrated graphically by Figure 1.

The optimal number of replications m_1^* increases with d_2 from $m_1^* = 5$ (for $d_2 \to 0$ or $\rho \to 0$) to $m_1^* \approx 9.9$ ($d_2 \to \infty$ or $\rho \to 1$) for the (generalized) D-criterion (solid line) and from $m_1^* = 5$ to $m_1^* \approx 9.1$ for the IMSE-criterion (dashed line).
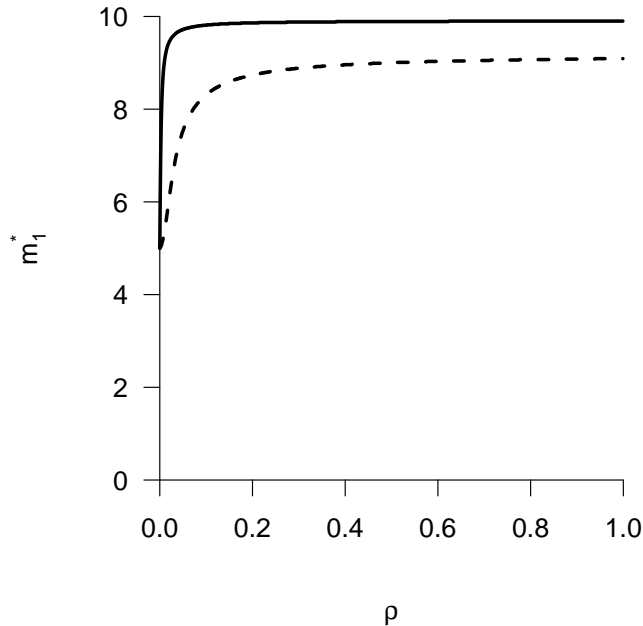
Figure 1: Optimal number of replications m_1^* at $x = 1$: (generalized) D-criterion - solid line; IMSE-criterion - dashed line

Figure 2 illustrates the efficiency the equi-replicated design, which assigns equal numbers of replications at points $x = 0$ and $x = 1$: $m_1 = m/2 = 5$, in dependence of the variance parameter ρ. For the (generalized) D-criterion the efficiency is computed by the formula

$$\text{eff}_D(\xi) = \left(\frac{\exp(D_B(\xi^*))}{\exp(D_B(\xi))} \right)^{\frac{1}{(n-1)q+p}} = \left(\frac{\det(\mathbf{M}(\xi^*))^{-1}(\prod_{\ell=1}^q \mu_\ell(\xi^*))^{n-1}}{\det(\mathbf{M}(\xi))^{-1}(\prod_{\ell=1}^q \mu_\ell(\xi))^{n-1}} \right)^{\frac{1}{(n-1)q+p}}, \quad (5.54)$$

which is homogeneous in terms of the mean squared error matrix according to the general form (5.10) of the criterion. For the IMSE-criterion we apply the standard formula

$$\text{eff}_{\text{IMSE}}(\xi) = \frac{\text{IMSE}_B(\xi^*)}{\text{IMSE}_B(\xi)}, \quad (5.55)$$

which also implies a homogeneous efficiency. The efficiency is equal to 1 for both criteria for $\rho \to 0$ and decreases with d_2 to $\text{eff}_D(\xi) \approx 0.50$ and $\text{eff}_{\text{IMSE}}(\xi) \approx 0.60$ for the (generalized) D- and IMSE-criterion respectively.

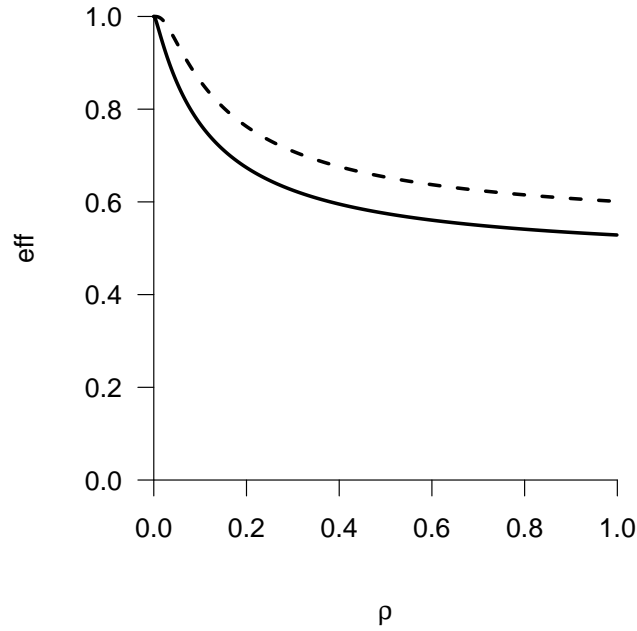To make use of the theory developed in Section 4.3 we rewrite for the prediction of

Figure 2: Efficiency of equi-replicated design: (generalized) D-criterion - solid line; IMSE-criterion - dashed line

the individual deviations the model (5.47) in the form (4.35):

$$Y_{ij} = \beta_{i2}x_j + \beta_{i1} + \varepsilon_{ij}, \quad x_j \in [0, 1]. \tag{5.56}$$

In this model the regresion functions have the form $\mathbf{f}(x_j) = (x_j, 1)^\top$ and the dispersion matrix $\mathbf{D} = d_2\mathbf{e}_1\mathbf{e}_1^\top$ satisfy the condition (4.33). Then according to Corollaries 13 and 18 the IMSE- or D-optimal designs for the prediction of the individual deviations in the reduced model

$$Y_{ij} = \beta_{i2}x_j + \varepsilon_{ij}, \quad x_j \in [0, 1] \tag{5.57}$$

are IMSE- or D-optimal for the prediction of the individual deviations in (5.56). For the reduced model (5.57) we obtain $\mathbf{D}_0 = d_2$, $\mathbf{f}(x_j) = x_j$ and consequently $\mathbf{M}_0(\xi) = \sum_{j=1}^k w_j x_j^2 \geq 0$. For the IMSE-criterion we use again $\nu = \lambda|_{[0,1]}$, which implies $\mathcal{V}_0 = \int_0^1 x^2 \mathrm{d}x = \frac{1}{3}$. Due to Corollaries 12 and 17 Bayesian optimal designs are optimal for the prediction of the individual deviations in the model (5.57). Since the left hand sides of the optimality conditions (2.24) and (2.30) have the form $c\,x^2$ with $c > 0$, the D- and

IMSE-optimal design ξ^* has the form

$$\xi = \begin{pmatrix} 1 \\ m \end{pmatrix}, \tag{5.58}$$

i.e. they take all observations at point $x = 1$. Note that this design illustrates the behavior of the Bayesian optimal designs in the model (5.47) for small values of the intercept variance and leads to a singular individual information matrix, which makes it useless for the prediction of individual parameters.

5.4 Construction of Optimal Designs by Equivariance

We consider a one-to-one transformation $g : \mathcal{X} \to \mathcal{X}_g$ of the experimental region $\mathcal{X}$ ($\mathcal{X}_g = g(\mathcal{X})$), and we assume that the regression functions $\mathbf{f}$ are defined simultaneously on both experimental regions ($\mathcal{X}$ and $\mathcal{X}_g$) and are linearly equivariant with respect to the transformation g. For the latter the existence of a non-singular $p \times p$ matrix $\mathbf{Q}_g$ such that

$$\mathbf{f}(g(x)) = \mathbf{Q}_g \mathbf{f}(x) \tag{5.59}$$

for all $x \in \mathcal{X}$ is required (see e.g. Schwabe (1996), ch. 3). Then the corresponding transformation of an approximate design ξ may be defined as

$$g : \xi = \begin{pmatrix} x_1, ..., x_k \\ m_1 , ..., m_k \end{pmatrix} \to \xi^g = \begin{pmatrix} g(x_1), ..., g(x_k) \\ m_1 , ..., m_k \end{pmatrix}, \tag{5.60}$$

where the frequencies m_j are the same for both designs ξ and ξ^g and only the design points x_j are transformed.

Further we add explicitly the dispersion matrix $\mathbf{D}$ to the definitions of the mean squared error matrices of the prediction of the individual parameters and individual deviations as well as to the definitions of the design criteria in order to indicate the influence of the dispersion matrix.

Lemma 15. *The following relation between the mean squared error matrices for the prediction of the individual parameters with respect to the initial and the transformed approximate designs ξ and ξ^g respectively holds:*

$$\mathrm{MSE}_B(\xi^g, \mathbf{D}_g) = (\mathbf{I}_n \otimes \mathbf{Q}_g^{-\top}) \, \mathrm{MSE}_B(\xi, \mathbf{D}) \, (\mathbf{I}_n \otimes \mathbf{Q}_g^{-1}), \tag{5.61}$$

where $\mathbf{D}_g = \mathbf{Q}_g^{-\top} \mathbf{D} \mathbf{Q}_g^{-1}$ *and* $\mathbf{Q}_g^{-\top} = (\mathbf{Q}_g^{\top})^{-1} = (\mathbf{Q}_g^{-1})^{\top}$.

Proof. For this proof we define the transformed regression functions as

$$\mathbf{f}^g(x) := \mathbf{f}(g(x)) = \mathbf{Q}_g \, \mathbf{f}(x).$$

Then we obtain the following results for the individual information matrix of the transformed design ξ^g:

$$\begin{aligned}
\mathbf{M}(\xi^g) &= \frac{1}{m} \sum_{j=1}^{k} m_j \, \mathbf{f}^g(x_j) \, \mathbf{f}^g(x_j)^{\top} \\
&= \frac{1}{m} \sum_{j=1}^{k} m_j \, \mathbf{Q}_g \, \mathbf{f}(x_j) \, \mathbf{f}(x_j)^{\top} \mathbf{Q}_g^{\top} \\
&= \mathbf{Q}_g \, \mathbf{M}(\xi) \, \mathbf{Q}_g^{\top}
\end{aligned}$$

and consequently

$$\mathbf{M}(\xi^g)^{-1} = \mathbf{Q}_g^{-\top} \, \mathbf{M}(\xi)^{-1} \, \mathbf{Q}_g^{-1}.$$

For $\boldsymbol{\Delta}_g = m \, \mathbf{D}_g = \mathbf{Q}_g^{-\top} \boldsymbol{\Delta} \, \mathbf{Q}_g^{-1}$ it may be seen that

$$\left(\mathbf{M}(\xi^g)^{-1} + \boldsymbol{\Delta}_g\right)^{-1} = \mathbf{Q}_g \left(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta}\right)^{-1} \mathbf{Q}_g^{\top},$$

which results in

$$\boldsymbol{\Delta}_g - \boldsymbol{\Delta}_g \left(\mathbf{M}(\xi^g)^{-1} + \boldsymbol{\Delta}_g\right)^{-1} \boldsymbol{\Delta}_g = \mathbf{Q}_g^{-\top} \left(\boldsymbol{\Delta} - \boldsymbol{\Delta} \left(\mathbf{M}(\xi)^{-1} + \boldsymbol{\Delta}\right)^{-1} \boldsymbol{\Delta}\right) \mathbf{Q}_g^{-1}.$$

Then using formula (5.1) for the mean squared error matrix of the prediction of the individual parameters we get (5.61). □

The next result for the mean squared error matrix of the prediction of the individual deviations is a by-product of the proof of Lemma 15.

Lemma 16. *The following relation between the mean squared error matrices for the prediction of the individual deviations with respect to the initial and the transformed approximate designs ξ and ξ^g respectively holds:*

$$\mathrm{MSE}_b(\xi^g, \mathbf{D}_g) = (\mathbf{I}_n \otimes \mathbf{Q}_g^{-\top}) \, \mathrm{MSE}_b(\xi, \mathbf{D}) \, (\mathbf{I}_n \otimes \mathbf{Q}_g^{-1}). \tag{5.62}$$

Lemmas 15 and 16 imply the following result for the D-optimal designs.

Theorem 19. *If the approximate design ξ^* is D-optimal for the prediction of the individual parameters for the dispersion matrix $\mathbf{D}$ on the experimental region $\mathcal{X}$, then the approximate design $(\xi^*)^g$ is D-optimal for the prediction of the individual parameters for the dispersion matrix $\mathbf{D}_g$ on the experimental region $\mathcal{X}_g = g(\mathcal{X})$.*

Proof. For this proof the criterion function (5.15) of the (generalized) D-criterion has to be represented in an appropriate form. To do this we use again formula (5.13). Since the positive eigenvalues of the matrices $\mathbf{H}(\mathbf{H}^\top\mathbf{M}(\xi)\mathbf{H}+\mathbf{I}_q)^{-1}\mathbf{H}^\top$ and $(\mathbf{H}^\top\mathbf{M}(\xi)\mathbf{H}+\mathbf{I}_q)^{-1}\mathbf{H}^\top\mathbf{H}$ are the same and the matrix $\mathbf{H}^\top\mathbf{H}$ has full rank, the (generalized) D-criterion can be written as

$$
\begin{aligned}
\mathrm{D}_B(\xi,\mathbf{D}) &= \ln\det(\mathbf{M}(\xi)^{-1}) + (n-1)\ln\det((\mathbf{H}^\top\mathbf{M}(\xi)\mathbf{H}+\mathbf{I}_q)^{-1}\mathbf{H}^\top\mathbf{H}) \\
&= -\ln\det(\mathbf{M}(\xi)) + (n-1)(\ln\det(\mathbf{H}^\top\mathbf{H}) - \ln\det(\mathbf{H}^\top\mathbf{M}(\xi)\mathbf{H}+\mathbf{I}_q)).
\end{aligned}
$$

Now let $\mathbf{H}_g = \mathbf{Q}_g^{-\top}\mathbf{H}$. Then the matrix $\mathbf{H}_g$ has full column rank q and satisfies the condition $\mathbf{H}_g\mathbf{H}_g^\top = \mathbf{D}_g$. Consequently, the (generalized) D-criterion (5.15) for the transformed design ξ^g on the experimental region $\mathcal{X}_g$ may be represented as

$$
\mathrm{D}_B(\xi^g,\mathbf{D}_g) = -\ln\det(\mathbf{M}(\xi^g)) + (n-1)(\ln\det(\mathbf{H}_g^\top\mathbf{H}_g) - \ln\det(\mathbf{H}_g^\top\mathbf{M}(\xi^g)\mathbf{H}_g+\mathbf{I}_q))
$$

and can be further transformed to

$$
\begin{aligned}
\mathrm{D}_B(\xi^g,\mathbf{D}_g) &= -\ln\det(\mathbf{Q}_g\mathbf{M}(\xi)\mathbf{Q}_g^\top) + (n-1)(\ln\det(\mathbf{H}_g^\top\mathbf{H}_g) - \ln\det(\mathbf{H}^\top\mathbf{M}(\xi)\mathbf{H}+\mathbf{I}_q)) \\
&= -\ln\det(\mathbf{M}(\xi)) + (n-1)(\ln\det(\mathbf{H}^\top\mathbf{H}) - \ln\det(\mathbf{H}^\top\mathbf{M}(\xi)\mathbf{H}+\mathbf{I}_q)) \\
&\quad + (n-1)(\ln\det(\mathbf{H}_g^\top\mathbf{H}_g) - \ln\det(\mathbf{H}^\top\mathbf{H})) - \ln\det(\mathbf{Q}_g\mathbf{Q}_g^\top) \\
&= \mathrm{D}_B(\xi,\mathbf{D}) + (n-1)(\ln\det(\mathbf{H}_g^\top\mathbf{H}_g) - \ln\det(\mathbf{H}^\top\mathbf{H})) - 2\ln|\det(\mathbf{Q}_g)|,
\end{aligned}
$$

where on the right hand side only the first term depends on the design. $\qquad\square$

A similar result may be formulated for the prediction of the individual deviations. Due to Corollaries 13 and 18 and Remark 3 it is sufficient to consider only the case of a non-singular dispersion matrix of random effects.

Theorem 20. *If the approximate design ξ^* is D-optimal for the prediction of the individual deviations for the dispersion matrix $\mathbf{D}$ on the experimental region $\mathcal{X}$, then the approximate design $(\xi^*)^g$ is D-optimal for the prediction of the individual deviations for the dispersion matrix $\mathbf{D}_g$ on the experimental region $\mathcal{X}_g = g(\mathcal{X})$.*

Proof. From (5.8) and (5.62) it follows directly that

$$\mathrm{D}_b(\xi^g, \mathbf{D}_g) = \mathrm{D}_b(\xi, \mathbf{D}) - 2\ln\det|\mathbf{Q}_g|,$$

where the second term on the right hand side is independent of the design. $\square$

Similar results may be also formulated for the L-criterion. However, in this case also the matrix $\mathcal{L}$ in (5.26) and (5.29) has to be appropriately transformed.

Theorem 21. *If the approximate design ξ^* is L-optimal for the prediction of the individual parameters for the dispersion matrix $\mathbf{D}$ on the experimental region $\mathcal{X}$ with respect to the linear transformation $\mathcal{L}$, then the design $(\xi^*)^g$ is L-optimal for the prediction of the individual parameters for the dispersion matrix $\mathbf{D}_g$ on the experimental region $\mathcal{X}_g = g(\mathcal{X})$ with respect to the linear transformation $\mathcal{L}_g = \mathcal{L}\,\mathbf{Q}_g^\top$.*

Proof. Using properties of the trace of a matrix we get from the L-criterion (5.26) and formula (5.61) the following relation

$$
\begin{aligned}
\mathrm{L}_B(\xi^g, \mathbf{D}_g, \mathcal{L}_g) &= \mathrm{tr}\left(\left(\mathbf{I}_n \otimes \mathbf{Q}_g^{-\top}\right)\mathrm{MSE}_B\left(\xi, \mathbf{D}\right)\left(\mathbf{I}_n \otimes \mathbf{Q}_g^{-1}\right)\left(\mathbf{I}_n \otimes \left(\mathcal{L}_g^\top \mathcal{L}_g\right)\right)\right) \\
&= \mathrm{tr}\left(\mathrm{MSE}_B\left(\xi, \mathbf{D}\right)\left(\mathbf{I}_n \otimes \mathbf{Q}_g^{-1}\right)\left(\mathbf{I}_n \otimes \left(\mathbf{Q}_g\,\mathcal{L}^\top \mathcal{L}\,\mathbf{Q}_g^\top\right)\right)\left(\mathbf{I}_n \otimes \mathbf{Q}_g^{-\top}\right)\right) \\
&= \mathrm{tr}\left(\mathrm{MSE}_B\left(\xi, \mathbf{D}\right)\left(\mathbf{I}_n \otimes (\mathcal{L}^\top \mathcal{L})\right)\right) \\
&= \mathrm{L}_B(\xi, \mathbf{D}, \mathcal{L}),
\end{aligned}
$$

which proves the result. $\square$

Theorem 22. *If the approximate design ξ^* is L-optimal for the prediction of the individual deviations for the dispersion matrix $\mathbf{D}$ on the experimental region $\mathcal{X}$ with respect to the linear transformation $\mathcal{L}$, then the design $(\xi^*)^g$ is L-optimal for the prediction of the individual deviations for the dispersion matrix $\mathbf{D}_g$ on the experimental region $\mathcal{X}_g = g(\mathcal{X})$ with respect to the linear transformation $\mathcal{L}_g = \mathcal{L}\,\mathbf{Q}_g^\top$.*

Proof. Analogously to the proof of Theorem 21 it can be shown using the L-criterion (5.29) for the prediction of the individual deviations and formula (5.62) that

$$\mathrm{L}_b(\xi^g, \mathbf{D}_g, \mathcal{L}_g) = \mathrm{L}_b(\xi, \mathbf{D}, \mathcal{L}),$$

which implies the result. □

The IMSE-criterion for the prediction of the individual parameters or individual deviations is a special case of the L-criterion with $\mathcal{A} = \mathcal{V}$ (see Lemma 12). As already mentioned in Section 5.2, there always exists some matrix $\mathcal{L}$ with $\mathcal{L}^\top \mathcal{L} = \mathcal{V}$. If the weighting measure ν will be transformed to its image ν^g, we obtain $\mathcal{V}_g = \int_{\mathcal{X}_g} \mathbf{f}(x)\,\mathbf{f}(x)^\top \nu^g(\mathrm{d}x) = \int_{\mathcal{X}} \mathbf{Q}_g\,\mathbf{f}(x)\,\mathbf{f}(x)^\top \mathbf{Q}_g^\top \nu(\mathrm{d}x) = \mathbf{Q}_g\,\mathcal{V}\,\mathbf{Q}_g^\top$. Then for every matrix $\mathcal{L}$ with $\mathcal{L}^\top \mathcal{L} = \mathcal{V}$ the matrix $\mathcal{L}_g = \mathcal{L}\,\mathbf{Q}_g^\top$ satisfies the condition $\mathcal{L}_g^\top \mathcal{L}_g = \mathcal{V}_g$, which establishes the next result.

Corollary 22. *If the approximate design ξ^* is IMSE-optimal for the prediction of the individual parameters (or individual deviations) for the dispersion matrix $\mathbf{D}$ on the experimental region $\mathcal{X}$ with respect to the measure ν, then the design $(\xi^*)^g$ is IMSE-optimal for the prediction of the individual parameters (or individual deviations) for the dispersion matrix $\mathbf{D}_g$ on the experimental region $\mathcal{X}_g = g(\mathcal{X})$ with respect to the measure ν^g.*

Example 2. We consider again the linear regression model (5.47) on the experimental region $\mathcal{X} = [0, 1]$. For the IMSE-criterion the uniform weighting $\nu = \lambda|_{[0,1]}$ will be used. According to Section 5.3 the D- and IMSE-optimal designs for the prediction of the individual parameters are for every dispersion matrix $\mathbf{D}$ of the form (5.49):

$$\xi^* = \begin{pmatrix} 0 & 1 \\ m_0^* & m_1^* \end{pmatrix}, \tag{5.63}$$

where $m_0^* = m - m_1^*$ and m_1^* are the optimal numbers of replications at the endpoints 0 and 1 of the experimental region. Since the left hand sides of the optimality conditions (2.24) and (2.30) are parabolas with positive leading terms for every non-singular dispersion matrix $\mathbf{D}$, Bayesian optimal designs, which are optimal for the prediction of the individual deviations, also have the form (5.49).

Further we assume that the intercepts β_{i1} and slopes β_{i2} are random and independent of each other, which implies a diagonal form of the dispersion matrix $\mathbf{D} = \mathrm{diag}\,(d_1, d_2)$,

where d_1 and d_2 are the variances of the intercepts and slopes respectively. Then optimal designs for the predictions of the individual parameters as well as individual deviations can be obtained numerically for any given values of d_1, d_2, m, and n for both design criteria.

Now we consider the linear transformation $g : [0,1] \rightarrow [0,a]$ with $g(x) = a\,x$ for some $a > 0$. According to (5.59) the dispersion matrix $\mathbf{D}$ is transformed via $\mathbf{Q}_g = \mathrm{diag}\,(1,a)$ to $\mathbf{D}_g = \mathrm{diag}\,(d_1, d_2/a^2)$. For the weighting measure ν we obtain the image $\nu^g = \frac{1}{a}\lambda|_{[0,a]}$, where $\lambda|_{[0,a]}$ is Lebesgue measure on $[0,a]$. Then according to Theorems 19 and 20 and Corollary 22 D- and IMSE-optimal designs for the prediction of the individual parameters as well as individual deviations in the linear regression model (5.47) on the experimental region $[0,a]$ are of the form

$$(\xi^*)^g = \begin{pmatrix} 0 & a \\ m_0^* & m_1^* \end{pmatrix},$$

where the optimal numbers of replications m_0^* and m_1^* are the same as in (5.63).

Next we consider a (finite) group G of transformations $g : \mathcal{X} \rightarrow \mathcal{X}$ of the experimental region onto itself, which satisfy the equivariance condition (5.59) of the regression functions as well as the invariance condition $\mathbf{D}_g = \mathbf{D}$ of the dispersion matrix for all $g \in G$.

Theorem 23. *Let G be a finite group of transformations on $\mathcal{X}$, for which the regression functions $\mathbf{f}$ are linearly equivariant and the dispersion matrix $\mathbf{D}$ is invariant.*
If ξ^* is D-optimal on $\mathcal{X}$ for the prediction of the individual parameters (or individual deviations) for the dispersion $\mathbf{D}$, then the design $\bar{\xi}^* = \frac{1}{\#G}\sum_{g \in G}(\xi^*)^g$ is also D-optimal on $\mathcal{X}$ for the prediction of the individual parameters (or individual deviations) for the dispersion matrix $\mathbf{D}$.

Proof. Let the design ξ^* be D-optimal for the prediction of the individual parameters on the experimental region $\mathcal{X}$ for the dispersion $\mathbf{D}$. Then it follows from Theorem 19 and the invariance of the dispersion matrix that the transformed designs $(\xi^*)^g$ are also D-optimal for the prediction of the individual parameters on $\mathcal{X}$ for $\mathbf{D}$ for all $g \in G$. Consequently,

$$\mathrm{D}_B(\xi^*, \mathbf{D}) = \mathrm{D}_B((\xi^*)^g, \mathbf{D}), \ \forall\, g \in G. \tag{5.64}$$

From the convexity of the criterion function we obtain

$$D_B(\bar{\xi}^*, \mathbf{D}) \leq D_B(\xi^*, \mathbf{D}), \tag{5.65}$$

which implies the D-optimality of the design $\bar{\xi}^*$ for the prediction of the individual parameters on $\mathcal{X}$ for the dispersion $\mathbf{D}$.

For the prediction of the individual deviations we use Theorem 20 instead of Theorem 19 and obtain the optimality of the design $\bar{\xi}^*$ from

$$D_b(\xi^*, \mathbf{D}) = D_b((\xi^*)^g, \mathbf{D}), \ \forall\, g \in G \tag{5.66}$$

and

$$D_b(\bar{\xi}^*, \mathbf{D}) \leq D_b(\xi^*, \mathbf{D}). \tag{5.67}$$

$\square$

For the linear criteria we additionally need the invariance of the transformation matrix $\mathcal{A} = \mathcal{L}^\top \mathcal{L}$: $\mathcal{A}_g = \mathcal{L}_g^\top \mathcal{L}_g = \mathcal{A}$ for all $g \in G$. As the L-criteria (5.30), (5.32) and (5.33) are explicitly defined by means of the transformation matrix $\mathcal{A}$ the notations $L_B(\xi, \mathbf{D}, \mathcal{A})$ and $L_b(\xi, \mathbf{D}, \mathcal{A})$ may be used for the linear criteria for the predictions of the individual parameters and individual deviations respectively.

Theorem 24. *Let G be a finite group of transformations on $\mathcal{X}$, for which the regression functions $\mathbf{f}$ are linearly equivariant and the dispersion matrix $\mathbf{D}$ as well as the transformation matrix $\mathcal{A}$ are invariant.*
If ξ^* is L-optimal on $\mathcal{X}$ for the prediction of the individual parameters (or individual deviations) for the dispersion $\mathbf{D}$ with respect to the transformation matrix $\mathcal{A}$, then the design $\bar{\xi}^* = \frac{1}{\#G}\sum_{g \in G}(\xi^*)^g$ is also L-optimal on $\mathcal{X}$ for the prediction of the individual parameters (or individual deviations) for the dispersion matrix $\mathbf{D}$ with respect to $\mathcal{A}$.

Proof. Let the design ξ^* be L-optimal for the prediction of the individual parameters on the experimental region $\mathcal{X}$ for the dispersion $\mathbf{D}$ and the transformation matrix $\mathcal{A}$. By Theorem 21 the transformed designs $(\xi^*)^g$ are also L-optimal for the prediction of the individual parameters on $\mathcal{X}$ for $\mathbf{D}$ and $\mathcal{A}$ for all $g \in G$. Then

$$L_B(\xi^*, \mathbf{D}, \mathcal{A}) = L_B((\xi^*)^g, \mathbf{D}, \mathcal{A}), \ \forall\, g \in G \tag{5.68}$$

and since the L-criterion is convex,

$$\mathrm{L}_B(\bar{\xi}^*, \mathbf{D}, \mathcal{A}) \le \mathrm{L}_B(\xi^*, \mathbf{D}, \mathcal{A}), \tag{5.69}$$

which implies the L-optimality of the design $\bar{\xi}^*$.

For the prediction of the individual deviation we use Theorem 22 instead of Theorem 21. $\qquad \square$

Remark 6. *The symmetrized design $\bar{\xi}^*$ is invariant with respect to G, i. e. $\bar{\xi}^* = (\bar{\xi}^*)^g$ for all $g \in G$.*

The last statement follows from group properties and allows to reduce the search for an optimal design to the class of invariant designs.

For the IMSE-criterion we need the invariance of the measure ν: $\nu^g = \nu$, which implies $\mathcal{V}_g = \mathcal{V}$. Since in both cases of prediction of the individual parameters and individual deviations the IMSE-criterion is a special case of the linear criterion (for $\mathcal{A} = \mathcal{V}$), we obtain the next result.

Corollary 23. *Let G be a finite group of transformations on $\mathcal{X}$, for which the regression functions $\mathbf{f}$ are linearly equivariant and the dispersion matrix $\mathbf{D}$ as well as the measure ν are invariant.*
If ξ^* is IMSE-optimal on $\mathcal{X}$ for the prediction of the individual parameters (or individual deviations) for the dispersion $\mathbf{D}$ with respect to the measure ν, then the design $\bar{\xi}^* = \frac{1}{\#G} \sum_{g \in G} (\xi^*)^g$ is also IMSE-optimal on $\mathcal{X}$ for the prediction of the individual parameters (or individual deviations) for the dispersion matrix $\mathbf{D}$ with respect to the measure ν.

Example 3. Here we consider the linear regression model (5.47) on the symmetric experimental region $[-a, a]$ for some $a > 0$. We assume the dispersion matrix $\mathbf{D}$ of random effects to be diagonal and use the uniform weighting $\nu = \frac{1}{2a}\lambda|_{[-a,a]}$ for the IMSE-criterion. Since the left hand sides of the optimality conditions (5.19) and (5.44) as well as (2.24) and (2.30) are parabolas (in x) with positive leading terms, D- and IMSE-optimal designs take all observation at the endpoints of the experimental region and consequently have the general form

$$\xi^* = \begin{pmatrix} -a & a \\ m_0^* & m_1^* \end{pmatrix}, \tag{5.70}$$

where $m_0^* = m - m_1^*$ and m_1^* are the optimal numbers of observations at points $-a$ and a in both cases of prediction of the individual parameters and individual deviations.

Further we consider the group $G = \{id, g\}$ with $g(x) = -x$ and $id(x) = x$. Then we obtain $\mathbf{Q}_g = \operatorname{diag}(1, -1)$, and $\mathbf{Q}_{id}$ is equal to the identity matrix. Consequently, every diagonal dispersion matrix is invariant with respect to G i.e. $\mathbf{D}_g = \mathbf{D}$. The uniform weighting ν is transformed to itself by both functions g and id from the group G.

The symmetrized design $\bar{\xi}^*$ is then of the form (5.70) and results in the invariant design

$$\bar{\xi}^* = \begin{pmatrix} -a & a \\ m/2 & m/2 \end{pmatrix}.$$

From Theorem 23 and Corollary 23 follows the D- and IMSE-optimality of this design for both predictions.

Example 4. In this example the quadratic regression model

$$Y_{ij} = \beta_{i1} + \beta_{i2} x_j + \beta_{i3} x_j^2 + \varepsilon_{ij} \tag{5.71}$$

on the standard symmetric design region $\mathcal{X} = [-1, 1]$ with a chessboard structured dispersion matrix

$$\mathbf{D} = \begin{pmatrix} d_{11} & 0 & d_{13} \\ 0 & d_{22} & 0 \\ d_{13} & 0 & d_{33} \end{pmatrix}$$

will be considered. For the IMSE-criterion we apply the uniform weighting $\nu = \frac{1}{2}\lambda|_{[-1,1]}$, which results in the nonsingular matrix

$$\mathcal{V} = \begin{pmatrix} 2 & 0 & \frac{2}{3} \\ 0 & \frac{2}{3} & 0 \\ \frac{2}{3} & 0 & \frac{2}{5} \end{pmatrix}.$$

Then the left hand sides of the optimality conditions (5.19) and (2.24) for the D-criteria as well as the conditions (5.44) and (2.30) with $\mathcal{A} = \mathcal{V}$ for the IMSE-criteria are (in x) polynomial functions of degree four with positive leading terms. Consequently, the corresponding optimal designs are supported by not more than three design points including

70

the two endpoints of the experimental region:

$$\xi^* = \begin{pmatrix} -1 & b & 1 \\ m_1 & m - m_1 - m_2 & m_2 \end{pmatrix}, \tag{5.72}$$

where $b \in [-1, 1]$ may be different for different design criteria.

Now we consider the sign change group G introduced in Example 3. Since $\mathbf{Q}_g = \mathrm{diag}\,(1, -1, 1)$ and $\mathbf{Q}_{id} = \mathbf{I}_3$, the dispersion matrix $\mathbf{D}$ is invariant for both transformations in G. The invariance of the weighting measure ν may also be confirmed. Then by Theorem 23 and Corollary 23 designs of the general form

$$\bar{\xi}^* = \begin{pmatrix} -1 & 0 & 1 \\ m_1^* & m - 2m_1^* & m_1^* \end{pmatrix} \tag{5.73}$$

are optimal for the predictions of the individual parameters and individual deviations for the D- and IMSE-criteria. The optimal number of observations m_1^* at points $x = 1$ and $x = -1$ depends on the design criterion as well as on the entries of the dispersion matrix and on the number of individuals.

Using Theorems 19 and 20 and Corollary 22 it may be shown that the IMSE- (with uniform ν) and D-optimal designs for the predictions of the individual parameters and individual deviations in the model (5.71) on any symmetric interval $[-a, a]$ will be of the form

$$\xi^* = \begin{pmatrix} -a & 0 & a \\ m_1^* & m - 2m_1^* & m_1^* \end{pmatrix},$$

where the optimal frequency m_1^* is the same as in (5.73).

6 Multi-Group Models

6.1 Estimation and Prediction in Multi-Group Models

In this part the more general model with s different groups of individuals will be considered. In this case the j-th observation at individual i in group l is given by

$$Y_{lij} = \mathbf{f}(x_{lj})^\top \boldsymbol{\beta}_{li} + \varepsilon_{lij}, \quad l = 1, .., s, \quad i = 1, .., n_l, \quad j = 1, .., m, \quad x_{lj} \in \mathcal{X}, \tag{6.1}$$

where n_l is the number of individuals in group l ($n_1, ..., n_s$ satisfy $\sum_{l=1}^s n_l = n$) and m is the number of observations at an individual, which are assumed to be the same for all individuals in all groups. Also the regression functions $\mathbf{f} = (f_1, .., f_p)^\top$ are the same. The design support points x_{lj} may differ between the groups and come from the experimental region $\mathcal{X}$. The image $\mathbf{f}(\mathcal{X}) \subset \mathbb{R}^p$ of $\mathcal{X}$ is assumed to be a compact set. ε_{lij} denote the observational errors with common variance $\sigma^2 > 0$. The individual parameters $\boldsymbol{\beta}_{li} = (\beta_{li1}, .., \beta_{lip})^\top$ have unknown group mean $\boldsymbol{\beta}^l \in \mathbb{R}^p$ and covariance matrix $\sigma^2 \mathbf{D} \in \mathbb{R}^{p \times p}$ with a given dispersion matrix $\mathbf{D}$. All individual parameters and all observational errors are uncorrelated.

For any group the model (6.1) coincides with the random coefficient regression model (4.2) introduced in Section 4.1. Hence, in group l the vector $\mathbf{Y}_l = (\mathbf{Y}_{l1}^\top, ..., \mathbf{Y}_{ln_l}^\top)^\top$ of observations at all individuals for $\mathbf{Y}_{li} = (\mathbf{Y}_{li1}^\top, ..., \mathbf{Y}_{lim}^\top)^\top$ may be represented in the form (4.10):

$$\mathbf{Y}_l = (\mathbf{1}_{n_l} \otimes \mathbf{F}_l) \boldsymbol{\beta}^l + (\mathbf{I}_{n_l} \otimes (\mathbf{F}_l \mathbf{H})) \boldsymbol{\gamma}_l + \boldsymbol{\varepsilon}_l, \tag{6.2}$$

where $\mathbf{F}_l = (\mathbf{f}(x_{l1}), ..., \mathbf{f}(x_{lm}))^\top$ denotes the individual design matrix for the individuals from group l and $\boldsymbol{\gamma}_l = \mathbf{I}_{n_l} \otimes ((\mathbf{H}^\top \mathbf{H})^{-1} \mathbf{H}^\top) \mathbf{b}_l$ are the random effects in group l (similarly to the individual design matrix $\mathbf{F}$ and the random effects $\boldsymbol{\gamma}$ in (4.10) respectively). The vector of individual deviations in group l is given by $\mathbf{b}_l = (\mathbf{b}_{l1}^\top, ..., \mathbf{b}_{ln_l}^\top)^\top$ for $\mathbf{b}_{li} = \boldsymbol{\beta}_{li} - \boldsymbol{\beta}^l$. $\boldsymbol{\varepsilon}_l$ is the vector of the observational errors in the group.

Then the complete observational vector $\mathbf{Y} = (\mathbf{Y}_1^\top, ..., \mathbf{Y}_s^\top)^\top$ for all groups has the form

$$\mathbf{Y} = \operatorname{diag}(\mathbf{1}_{n_l} \otimes \mathbf{F}_l)_{l=1,...,s} \boldsymbol{\beta} + \operatorname{diag}(\mathbf{I}_{n_l} \otimes (\mathbf{F}_l \mathbf{H}))_{l=1,...,s} \boldsymbol{\gamma} + \boldsymbol{\varepsilon}, \tag{6.3}$$

where $\boldsymbol{\beta} = (\boldsymbol{\beta}^{1\top}, ..., \boldsymbol{\beta}^{s\top})^\top$, $\boldsymbol{\gamma} = (\boldsymbol{\gamma}_1^\top, ..., \boldsymbol{\gamma}_s^\top)^\top$, $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}_1^\top, ..., \boldsymbol{\varepsilon}_s^\top)^\top$ and $\operatorname{diag}(\mathbf{A}_l)_{l=1,...,s}$ denotes the block diagonal matrix with blocks $\mathbf{A}_1, ..., \mathbf{A}_s$. Note that $\boldsymbol{\gamma}$ satisfies the condition $\mathbf{b} = (\mathbf{I}_n \otimes \mathbf{H}) \boldsymbol{\gamma}$ for $\mathbf{b} = (\mathbf{b}_1^\top, ..., \mathbf{b}_s^\top)^\top$.

Note that for group l the model (6.1) may also be represented as

$$\mathbf{Y}_l = (\mathbf{1}_{n_l} \otimes \mathbf{F}_l)\boldsymbol{\beta}^l + (\mathbf{I}_{n_l} \otimes \mathbf{F}_l)\mathbf{b}_l + \boldsymbol{\varepsilon}_l, \tag{6.4}$$

which is similar to (4.7) in part 4.1, and consequently in the form

$$\mathbf{Y} = \text{diag}\,(\mathbf{1}_{n_l} \otimes \mathbf{F}_l)_{l=1,\ldots,s}\,\boldsymbol{\beta} + \text{diag}\,(\mathbf{I}_{n_l} \otimes \mathbf{F}_l)_{l=1,\ldots,s}\,\mathbf{b} + \boldsymbol{\varepsilon}, \tag{6.5}$$

for the whole population.

It is easy to see that the particular case of the model (6.1) with identical groups (when it is known that the group means coincide, $\boldsymbol{\beta}^l \equiv \boldsymbol{\beta}^0$) and the same group designs $x_{lj} \equiv x_j$ coincides with the model (4.2).

Note that the form (6.3) of the multi-group model satisfies the conditions of the linear mixed model (3.1) for $\mathbf{X} = \text{diag}\,(\mathbf{1}_{n_l} \otimes \mathbf{F}_l)_{l=1,\ldots,s}$, $\mathbf{Z} = \text{diag}\,(\mathbf{I}_{n_l} \otimes \mathbf{F}_l\mathbf{H})_{l=1,\ldots,s}$, $\mathbf{G} = \text{Cov}(\boldsymbol{\gamma}) = \sigma^2\mathbf{I}_{nq}$ and $\mathbf{R} = \text{Cov}(\boldsymbol{\varepsilon}) = \sigma^2\mathbf{I}_{np}$, where $n = \sum_{l=1}^{s} n_l$. The next theorem represents the best linear unbiased predictor of the individual deviations.

Theorem 25. *The best linear unbiased predictor of the individual deviations* $\mathbf{b}_{li}$ *in group* l *is given by*

$$\hat{\mathbf{b}}_{li} = \mathbf{D}\mathbf{F}_l^\top(\mathbf{F}_l\mathbf{D}\mathbf{F}_l^\top + \mathbf{I}_m)^{-1}(\mathbf{Y}_{li} - \bar{\mathbf{Y}}_l), \tag{6.6}$$

where $\bar{\mathbf{Y}}_l = \frac{1}{n_l}\sum_{i=1}^{n_l}\mathbf{Y}_{li}$ *denotes the mean observational vector in group* l .

Proof. Using the form (6.3) of the multi-group model and Corollary 3 it can be shown that

$$\hat{\boldsymbol{\gamma}} = \text{diag}\,\left((\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_n\mathbf{1}_{n_l}^\top) \otimes ((\mathbf{H}^\top\mathbf{F}_l^\top\mathbf{F}_l\mathbf{H} + \mathbf{I}_q)^{-1}\mathbf{H}^\top\mathbf{F}_l^\top)\right)_{l=1,\ldots,s}\mathbf{Y}$$

then using $\hat{\mathbf{b}} = (\mathbf{I}_n \otimes \mathbf{H})\hat{\boldsymbol{\gamma}}$ we get for $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_1^\top, ..., \hat{\mathbf{b}}_s^\top)^\top$ and $\hat{\mathbf{b}}_l = (\hat{\mathbf{b}}_{l1}^\top, ..., \hat{\mathbf{b}}_{ln_l}^\top)^\top$ the equations

$$\hat{\mathbf{b}} = \text{diag}\,\left((\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes \left(\mathbf{D}\mathbf{F}_l^\top(\mathbf{F}_l\mathbf{D}\mathbf{F}_l^\top + \mathbf{I}_m)^{-1}\right)\right)_{l=1,\ldots,s}\mathbf{Y}$$

and

$$\hat{\mathbf{b}}_l = \left((\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes \left(\mathbf{D}\mathbf{F}_l^\top(\mathbf{F}_l\mathbf{D}\mathbf{F}_l^\top + \mathbf{I}_m)^{-1}\right)\right)\mathbf{Y}_l$$

and consequently the result (6.6). $\qquad\square$

The next theorem gives the mean squared error matrix of the best linear unbiased prediction $\hat{\mathbf{b}} = (\hat{\mathbf{b}}_1^\top, ..., \hat{\mathbf{b}}_s^\top)^\top$ for $\hat{\mathbf{b}}_l = (\hat{\mathbf{b}}_{l1}^\top, ..., \hat{\mathbf{b}}_{ln_l}^\top)^\top$.

Theorem 26. *The mean squared error matrix of the prediction $\hat{\mathbf{b}}$ of the individual deviations is given by*

$$
\mathrm{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right) = \sigma^2 \operatorname{diag}\left(\left(\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top\right) \otimes \left(\mathbf{D} - \mathbf{D}\mathbf{F}_l^\top\left(\mathbf{F}_l\mathbf{D}\mathbf{F}_l^\top + \mathbf{I}_m\right)^{-1}\mathbf{F}_l\mathbf{D}\right)\right.
$$
$$
\left. + \frac{1}{n_l}\left(\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top\right) \otimes \mathbf{D}\right)_{l=1,...,s} \tag{6.7}
$$

Proof. The result follows from Theorem 8 for $\mathbf{K} = \mathbf{0}$ and $\mathbf{L} = (\mathbf{I}_n \otimes \mathbf{H})$ since $\hat{\mathbf{b}} = (\mathbf{I}_n \otimes \mathbf{H})\hat{\boldsymbol{\gamma}}$. $\square$

Note that if the dispersion matrix $\mathbf{D}$ of random effects has the block diagonal form (4.33), the prediction (6.6) may be represented as a transformation of the prediction of the individual deviations in the corresponding reduced model (analogously to (4.34)). This case will not be considered explicitly. Instead, we assume the dispersion matrix to be non-singular.

Corollary 24. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the best linear unbiased predictor $\hat{\mathbf{b}}_{li}$ of the individual deviations simplifies to*

$$
\hat{\mathbf{b}}_{li} = (\mathbf{F}_l^\top\mathbf{F}_l + \mathbf{D}^{-1})^{-1}\mathbf{F}_l^\top(\mathbf{Y}_{li} - \bar{\mathbf{Y}}_l). \tag{6.8}
$$

Corollary 25. *If the dispersion matrix $\mathbf{D}$ of random effects is non-singular, the mean squared error matrix of the prediction $\hat{\mathbf{b}}$ of the individual deviations simplifies to*

$$
\mathrm{Cov}\left(\hat{\mathbf{b}} - \mathbf{b}\right) \tag{6.9}
$$
$$
= \sigma^2 \operatorname{diag}\left(\left(\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top\right) \otimes \left(\mathbf{F}_l^\top\mathbf{F}_l + \mathbf{D}^{-1}\right)^{-1} + \frac{1}{n_l}\left(\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top\right) \otimes \mathbf{D}\right)_{l=1,...,s}.
$$

Further we consider the individual parameters $\boldsymbol{\beta}_{li}$, which are all predictable (according to Corollary 1) only if all design matrices $\mathbf{F}_l$ are of full column rank. The next theorem represents the best linear unbiased predictor of the individual parameters under the full column rank assumption for the best linear unbiased estimator $\hat{\boldsymbol{\beta}}^l = (\mathbf{F}_l^\top\mathbf{F}_l)^{-1}\mathbf{F}_l^\top\bar{\mathbf{Y}}_l$ of the group mean $\boldsymbol{\beta}^l$ and the individualized estimator $\hat{\boldsymbol{\beta}}_{li;\mathrm{ind}} = (\mathbf{F}_l^\top\mathbf{F}_l)^{-1}\mathbf{F}_l^\top\mathbf{Y}_{li}$ depending only on the observations at the i-th individual in the l-th group respectively.

Theorem 27. *The best linear unbiased predictor of the individual parameters $\boldsymbol{\beta}_{li}$ in group l is given by*

$$\hat{\boldsymbol{\beta}}_{li} = \mathbf{D}((\mathbf{F}_l^\top \mathbf{F}_l)^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}_{li;\text{ind}} + (\mathbf{F}_l^\top \mathbf{F}_l)^{-1}((\mathbf{F}_l^\top \mathbf{F}_l)^{-1} + \mathbf{D})^{-1}\hat{\boldsymbol{\beta}}^l. \qquad (6.10)$$

Proof. The total vector of all individual parameters $\mathbf{B}$ may be represented in the form $\mathbf{B} = \text{diag}(\mathbf{1}_{n_l} \otimes \mathbf{I}_p)_{l=1,\ldots,s}\,\boldsymbol{\beta} + (\mathbf{I}_n \otimes \mathbf{H})\boldsymbol{\gamma}$. Then the result follows from Theorem 6 for $\hat{\mathbf{B}} = (\hat{\mathbf{B}}_1^\top, \ldots, \hat{\mathbf{B}}_s^\top)^\top$ and $\hat{\mathbf{B}}_l = (\hat{\boldsymbol{\beta}}_{l1}^\top, \ldots, \hat{\boldsymbol{\beta}}_{ln_l}^\top)^\top$. $\qquad\square$

The mean squared error matrix of the prediction $\hat{\mathbf{B}} = (\hat{\mathbf{B}}_1^\top, \ldots, \hat{\mathbf{B}}_s^\top)^\top$ of the full vector of all individual parameters $\mathbf{B} = (\mathbf{B}_1^\top, \ldots, \mathbf{B}_s^\top)^\top$ is presented for $\hat{\mathbf{B}}_l = (\hat{\boldsymbol{\beta}}_{l1}^\top, \ldots, \hat{\boldsymbol{\beta}}_{ln_l}^\top)^\top$ in the next theorem.

Theorem 28. *The mean squared error matrix of the prediction $\hat{\mathbf{B}}$ of the individual parameters vector is given by*

$$
\begin{aligned}
\text{Cov}\left(\hat{\mathbf{B}} - \mathbf{B}\right) \;=\;& \sigma^2 \,\text{diag}\left(\left(\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top\right) \otimes \left(\mathbf{D} - \mathbf{D}((\mathbf{F}_l^\top \mathbf{F}_l)^{-1} + \mathbf{D})^{-1}\mathbf{D}\right) \right. \quad (6.11) \\
& \left. + \frac{1}{n_l}(\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes (\mathbf{F}_l^\top \mathbf{F}_l)^{-1} \right)_{l=1,\ldots,s} .
\end{aligned}
$$

Proof. The result follows directly from Lemma 1 for $\mathbf{K} = \text{diag}(\mathbf{1}_{n_l} \otimes \mathbf{I}_p)_{l=1,\ldots,s}$ and $\mathbf{L} = \mathbf{I}_n \otimes \mathbf{H}$, since $\mathbf{B} = \text{diag}(\mathbf{1}_{n_l} \otimes \mathbf{I}_p)_{l=1,\ldots,s}\,\boldsymbol{\beta} + (\mathbf{I}_n \otimes \mathbf{H})\boldsymbol{\gamma}$. $\qquad\square$

6.2 Optimal Designs

In this section we consider optimal designs for the prediction of the individual parameters $\boldsymbol{\beta}_{li}$ and individual deviations $\mathbf{b}_{li}$ for $i = 1, \ldots, n_l$ and $l = 1, \ldots, s$ in the multi-group model (6.1). We define (exact) design ς in the multi-group model as

$$\varsigma = \begin{pmatrix} \xi_1 & \cdots & \xi_s \\ n_1 & \cdots & n_s \end{pmatrix} \qquad (6.12)$$

with group sizes $n_1, \ldots, n_s$ and the group-designs

$$\xi_l = \begin{pmatrix} x_{l1} & \cdots & x_{lk_l} \\ m_{l1} & \cdots & m_{lk_l} \end{pmatrix}, \qquad (6.13)$$

where $x_{l1}, ..., x_{lk_l}$ are distinct experimental settings for group l and $m_{l1}, ..., m_{lk_l}$ are the corresponding numbers of replications with $\sum_{j=1}^{k_l} m_{lj} = m$ for all $l = 1, ..., s$. Note that the group-design ξ_l is the same for all individuals in the group and consequently has the form (2.16) for all $l = 1, ..., s$.

For approximate designs, which are also of the form (6.13), the values m_{lj} are not necessarily integer. We denote by $\Xi_{\mathcal{X}}$ and Ξ the sets of all exact and approximate designs of the form (6.12) respectively.

Analogously to the one-group case (4.2) we define the mean squared error matrix of the prediction of the individual parameters for an approximate design ς by

$$
\begin{aligned}
\mathbf{MSE}_B(\varsigma) \;=\; \operatorname{diag}\Bigg(& (\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes (\mathbf{\Delta} - \mathbf{\Delta}(\mathbf{M}_l(\xi_l)^{-1} + \mathbf{\Delta})^{-1}\mathbf{\Delta}) \quad (6.14)\\
& + \frac{1}{n_l}(\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes \mathbf{M}_l(\xi_l)^{-1} \Bigg)_{l=1,...,s},
\end{aligned}
$$

where $\mathbf{\Delta} = m\,\mathbf{D}$ is the adjusted dispersion matrix of random effects and

$$
\mathbf{M}_l(\xi_l) = \frac{1}{m}\sum_{j=1}^{k_l} m_{lj}\,\mathbf{f}(x_{lj})\mathbf{f}(x_{lj})^\top \tag{6.15}
$$

denotes the standardized individual information matrix in the group l. Then it is easy to see that

$$
\mathbf{MSE}_B(\varsigma) = \operatorname{diag}\left(\mathrm{MSE}_{B_l}(\xi_l)\right)_{l=1,...,s}, \tag{6.16}
$$

where

$$
\mathrm{MSE}_{B_l}(\xi_l) = \frac{1}{n_l}(\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top)\otimes\mathbf{M}_l(\xi_l)^{-1}+(\mathbf{I}_{n_l}-\frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top)\otimes(\mathbf{\Delta}-\mathbf{\Delta}(\mathbf{M}_l(\xi_l)^{-1}+\mathbf{\Delta})^{-1}\mathbf{\Delta}) \tag{6.17}
$$

is the mean squared error matrix within group l. (6.17) coincides with (5.1) for $\mathbf{M}_l(\xi_l) = \mathbf{M}(\xi)$ and $n_l = n$. Note that for exact designs the mean squared error matrix (6.14) is in accordance with (6.11) if we suppress the constant factor $\frac{\sigma^2}{m}$.

As it was already mentioned in Section 6.1, in the case of prediction of the individual deviations we assume the dispersion matrix $\mathbf{D}$ to be non-singular. Then we define the mean squared error matrix of the prediction of the individual deviations as

$$
\mathbf{MSE}_b(\varsigma) = \operatorname{diag}\left(\frac{1}{n_l}(\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes \mathbf{\Delta} + (\mathbf{I}_{n_l} - \frac{1}{n_l}\mathbf{1}_{n_l}\mathbf{1}_{n_l}^\top) \otimes (\mathbf{M}_l(\xi_l) + \mathbf{\Delta}^{-1})^{-1}\right)_{l=1,...,s}. \tag{6.18}
$$

For this mean squared error matrix it holds that

$$\mathbf{MSE}_b \left(\varsigma \right) = \operatorname{diag} \left(\operatorname{MSE}_{b_l} \left(\xi_l \right) \right)_{l=1,\dots,s}, \tag{6.19}$$

where

$$\operatorname{MSE}_{b_l} \left(\xi_l \right) = \frac{1}{n_l} (\mathbf{1}_{n_l} \mathbf{1}_{n_l}^\top) \otimes \boldsymbol{\Delta} + (\mathbf{I}_{n_l} - \frac{1}{n_l} \mathbf{1}_{n_l} \mathbf{1}_{n_l}^\top) \otimes (\mathbf{M}_l(\xi_l) + \boldsymbol{\Delta}^{-1})^{-1} \tag{6.20}$$

is in accordance with (5.3) for $\mathbf{M}_l(\xi_l) = \mathbf{M}(\xi)$ and $n_l = n$. In the case of exact designs the mean squared error matrix (6.18) coincides with (6.9) (divided by $\frac{\sigma^2}{m}$).

Now design criteria for predictions in the multi-group model can be defined. For this we consider the case of a general dispersion matrix $\mathbf{D}$ of random effects with rank $q \leq p$.

Definition 19. *For a general dispersion matrix $\mathbf{D}$ with rank q the approximate design $\varsigma^* \in \boldsymbol{\Xi}$ is called D-optimal for the prediction of the individual parameters in the multi-group model if*

$$\mathcal{D}_B(\varsigma^*) = \min_{\varsigma \in \boldsymbol{\Xi}} \mathcal{D}_B(\varsigma) \tag{6.21}$$

for

$$\mathcal{D}_B(\varsigma) = \ln \prod_{\tau=1}^{(n-s)q+sp} \eta_\tau(\varsigma), \tag{6.22}$$

where $\eta_1(\varsigma), \dots, \eta_{(n-s)q+sp}(\varsigma)$ are the $(n-s)q+sp$ largest eigenvalues of the mean squared error matrix $\mathbf{MSE}_B(\varsigma)$ of the prediction of the individual parameters. The criterion function (6.22) is called the (generalized) D-criterion for the prediction of the individual parameters in the multi-group model.

According to the next lemma the (generalized) D-criterion for the prediction of the individual parameters in the multi-group model can be represented as the sum of the corresponding D-criteria in the one-group model (4.2).

Lemma 17. *The (generalized) D-criterion $\mathcal{D}_B(\varsigma)$ for the prediction of the individual parameters in the multi-group model is equal to*

$$\mathcal{D}_B(\varsigma) = \sum_{l=1}^s \mathrm{D}_{B_l}(\xi_l), \tag{6.23}$$

for

$$D_{B_l}(\xi_l) = \ln \prod_{\iota=1}^{(n_l-1)q+p} \lambda_\iota(\xi_l),$$

where $\lambda_1(\xi_l), ..., \lambda_{(n_l-1)q+p}(\xi_l)$ are the $(n_l-1)q + p$ largest eigenvalues of $\mathrm{MSE}_{B_l}(\xi_l)$.

Proof. From properties of block diagonal matrices it follows that the eigenvalues of the mean squared error $\mathrm{MSE}_{B_l}(\xi_l)$ are at the same time the eigenvalues of $\mathbf{MSE}_B(\varsigma)$ for all $l = 1, ..., s$. According to the proof of Lemma 6 the matrix $\mathrm{MSE}_{B_l}(\xi_l)$ has $(n_l-1)q + p$ positive eigenvalues (for all $l = 1, ..., s$). Then it follows from properties of the logarithm that

$$\ln \prod_{\tau=1}^{(n-s)q+sp} \eta_\tau(\varsigma) = \sum_{l=1}^{s} \ln \prod_{\iota=1}^{(n_l-1)q+p} \lambda_\iota(\xi_l),$$

where $\lambda_1(\xi_l), ..., \lambda_{(n_l-1)q+p}(\xi_l)$ are the $(n_l-1)q + p$ positive (largest) eigenvalues of $\mathrm{MSE}_{B_l}(\xi_l)$. Using the form (5.10) of the (generalized) D-criterion in the one-group model (4.2) we obtain the result. $\square$

Definition 20. *For a general dispersion matrix $\mathbf{D}$ with rank q the approximate design $\varsigma^* \in \boldsymbol{\Xi}$ is called D-optimal for the prediction of the individual deviations in the multi-group model if*

$$\mathcal{D}_b(\varsigma^*) = \min_{\varsigma \in \boldsymbol{\Xi}} \mathcal{D}_b(\varsigma) \tag{6.24}$$

for

$$\mathcal{D}_b(\varsigma) = \ln \prod_{\tau=1}^{nq} \upsilon_\tau(\varsigma), \tag{6.25}$$

where $\upsilon_1(\varsigma), ..., \upsilon_{nq}(\varsigma)$ are the nq largest eigenvalues of the mean squared error matrix $\mathbf{MSE}_b(\varsigma)$ of the prediction of the individual deviations. The criterion function (6.25) is called the (generalized) D-criterion for the prediction of the individual deviations in the multi-group model.

The relation between the (generalized) D-criteria for the prediction of the individual deviations in the multi-group model and the model (4.2) is illustrated by the next lemma.

Lemma 18. *The (generalized) D-criterion $\mathcal{D}_b(\varsigma)$ for the prediction of the individual deviations in the multi-group model is equal to*

$$\mathcal{D}_b(\varsigma) = \sum_{l=1}^{s} D_{b_l}(\xi_l), \tag{6.26}$$

for

$$D_{b_l}(\xi_l) = \ln \prod_{\iota=1}^{n_l q} \nu_\iota(\xi_l),$$

where $\nu_1(\xi_l), ..., \nu_{n_l q}(\xi_l)$ *are the* $n_l q$ *largest eigenvalues of* $\mathrm{MSE}_{b_l}(\xi_l)$.

Proof. The result (6.26) follows from properties of block diagonal matrices and the form (5.12) fo the (generalized) D-criterion (analogously to the proof of Lemma 17). $\qquad\square$

Note that the (generalized) D-criteria $D_{B_1}(\xi_1), ..., D_{B_s}(\xi_s)$ for the prediction of the individual parameters as well as the criteria $D_{b_1}(\xi_1), ..., D_{b_s}(\xi_s)$ for the prediction of the individual deviations are identical among the groups and independent of each other. In the particular situation, where the group sizes n_l are fixed, the group designs ξ_l can be determined separately for each group using the theory developed in the previous chapter. Then the resulting optimal designs are given by

$$\varsigma = \begin{pmatrix} \xi(n_1) & ... & \xi(n_s) \\ n_1 & ... & n_s \end{pmatrix}, \tag{6.27}$$

where $\xi(n_l)$ are the optimal group designs, which have the same general form for all groups.

Now the L-criterion for the prediction of the individual parameters in the multi-group model will be introduced. For exact designs and a specified $\tau \times p$ matrix $\mathcal{L}$ the criterion may be defined as the sum of the linear criteria in the groups:

$$\mathbf{L}_B(\varsigma) = \sum_{l=1}^{s} \sum_{i=1}^{n_l} \mathrm{tr}\left(\mathrm{Cov}\left(\mathcal{L}\hat{\boldsymbol{\beta}}_{li} - \mathcal{L}\boldsymbol{\beta}_{li}\right)\right), \tag{6.28}$$

For approximate designs we generalize this criterion and represent it in the following equivalent form.

Definition 21. *The approximate design* $\varsigma^* \in \boldsymbol{\Xi}$ *is called L-optimal with respect to* $\mathcal{A} = \mathcal{L}^\top \mathcal{L}$ *for the prediction of the individual parameters in the multi-group model if*

$$\mathbf{L}_B(\varsigma^*) = \min_{\varsigma \in \boldsymbol{\Xi}} \mathbf{L}_B(\varsigma) \tag{6.29}$$

for

$$\mathbf{L}_B(\xi) = \mathrm{tr}\left(\mathbf{MSE}_B(\varsigma)\left(\mathbf{I}_n \otimes \mathcal{A}\right)\right). \tag{6.30}$$

The next lemma gives an explicit form of the criterion (6.30).

Lemma 19. *The L-criterion* $\mathbf{L}_B(\varsigma)$ *for the prediction of the individual parameters in the multi-group model is equal to*

$$\mathbf{L}_B(\varsigma) = \sum_{l=1}^{s} \mathrm{L}_{B_l}(\xi_l). \tag{6.31}$$

Proof. The result follows directly from the definitions (6.30) and (5.26) of the L-criteria in the multi-group model and the model (4.2) and formula (6.14). $\qquad\square$

The linear criterion for the prediction of the individual deviations can be defined for exact designs similarly to (6.28):

For approximate designs we generalize this criterion on the following way.

Definition 22. *The approximate design* $\varsigma^* \in \boldsymbol{\Xi}$ *is called L-optimal with respect to* $\mathcal{A}$ *for the prediction of the individual deviations in the multi-group model if*

$$\mathbf{L}_b(\varsigma^*) = \min_{\varsigma \in \boldsymbol{\Xi}} \mathbf{L}_b(\varsigma) \tag{6.32}$$

for

$$\mathbf{L}_b(\varsigma) = \mathrm{tr}\, \left(\mathbf{MSE}_b\,(\varsigma)\,(\mathbf{I}_n \otimes \mathcal{A}) \right). \tag{6.33}$$

The criterion function (6.33) is called the L-criterion for the prediction of the individual deviations in the multi-group model.

The next lemma represents a more explicit form of the L-criterion (6.33).

Lemma 20. *The L-criterion* $\mathcal{L}_b(\varsigma)$ *for the prediction of the individual deviations in the multi-group model is equal to*

$$\mathbf{L}_b(\varsigma) = \sum_{l=1}^{s} \mathrm{L}_{b_l}(\xi_l). \tag{6.34}$$

Proof. The result (6.34) follows from the forms (6.33) and (5.29) of the linear criteria in the multi-group and one-group models and formula (6.18) for the mean squared error matrix of the prediction of the individual deviations. $\qquad\square$

Since the linear criteria $\mathrm{L}_{B_1}(\xi_1), ..., \mathrm{L}_{B_s}(\xi_s)$ and $\mathrm{L}_{b_1}(\xi_1), ..., \mathrm{L}_{b_s}(\xi_s)$ for the predictions of the individual parameters and individual deviations respectively are identical among the groups and independent of each other, the group designs ξ_l can be determined separately for each group using the theory developed in Chapter 5 and the resulting optimal designs are of the form (6.27) if the group sizes n_l are fixed.

7 Discussion and Outlook

In this chapter the results obtained will be summarized and discussed. Then a short outlook on further open some problems will be given.

7.1 Summary and Discussion

The purpose of this thesis was to develop an analytical approach for determining designs, which are optimal for prediction of linear aspects, in particular of individual parameters and individual deviations, in hierarchical random coefficient regression models. In this context a generalized D-criterion, which is an extension of the determinant criterion in the situation of a singular covariance matrix of random effects, and linear criteria, in particular the integrated mean squared error criterion, have been considered. For the prediction of the individual deviations it has been established that Bayesian optimal designs retain their optimality if the dispersion matrix of random effects is non-singular. In the case of a singular dispersion matrix, where only parts of the parameter vector are random, the design criteria simplify to those in the reduced model, which takes only the random parameters in account. Consequently the Bayesian optimal designs in the reduced model are optimal for the prediction. For the prediction of the individual parameters both criteria result in a compound criterion: a weighted sum of the corresponding criterion in the fixed effects model and the Bayesian criterion. For these criteria optimality conditions, which result from the general equivalence theorem, have been formulated. In the particular case of random intercepts (random block effects) optimal designs in the models without random effects remain optimal for the prediction of the individual parameters. For the prediction of the individual deviations all designs are equally good in the random block effects models. The analytical results are illustrated by the example of straight line regression. Additionally some examples are given for construction of optimal designs by invariance and equivariance. Finally the more general case of hierarchical random coefficient regression, multi-group models, has been considered. If the group sizes are fixed, the statistical analysis in these models can be performed in each group separately.

7.2 Further Considerations

In the hierarchical random coefficient regression discussed in this thesis the dispersion matrix of random effects is assumed to be known and the obtained optimal designs for the prediction are only locally optimal for a specific dispersion. The assumption of a prior knowledge of the covariance structure may be inappropriate for many populations in real applications, where the dispersion matrix is generally not given. Robustness and sensitivity with respect to the variance parameters should be investigated in the future.

Besides the generalized determinant and linear criteria considered in this work other optimality criteria like eigenvalue and minimax criteria can be a topic of future research. Moreover some other criteria, which lead to more robust designs with respect to the dispersion, can be developed. Furthermore some particular linear criteria like c-criterion can be investigated in more detail as it has been done by Elfving (1952) for the classical regression. Optimal designs for inter- and extrapolation, discussed by Kiefer and Wolfowitz (1964a) and Kiefer and Wolfowitz (1964b) (see also Kiefer and Wolfowitz (1965)) for the fixed effects model, may be another object of future investigations. Note that the inter- and extrapolation can be considered as a special c-criterion. However, in this case the search for optimal designs is focused on the class of "non-singular" designs, for which the fixed effects design matrix has full column rank. Another possibility is to represent the inter- and extrapolation in the form of a linear aspect, which is predictable for any fixed effect design matrix, and consider the corresponding mean squared error as a design criterion.

In this thesis mainly the hierarchical models with identical designs for all observational units have been considered. The only exception is the multi-group case briefly discussed in Chapter 6. For the multi-group models the design criteria have been introduced; however, no general optimality conditions could be formulated. Only for the particular case of fixed group sizes optimal designs have been obtained. These designs follow directly from the optimal designs inside the groups, i.e. the statistical analysis can be performed in each group separately. The general multi-group models, where the group sizes are not fixed and have to be optimized, will be considered in the future in more detail. Furthermore, other special cases of the hierarchical random coefficient regression, where different designs for different individuals are allowed, are relevant for many statistical applications. For

instance, one-point designs, where all observations at an observational unit are made at the same design point, though the design points are different for different units, are popular in engineering (see e.g. Weaver and Meeker (2014)). Moreover, designs with small numbers of observation per individual, considered by Graßhoff *et al.* (2012) for the estimation of the population parameters, are relevant for pharmacological studies.

The design optimality conditions presented in this work have been formulated for approximate designs, which are generally not applicable in populations with small sample sizes within individuals. The problem of generating exact designs in the fixed effects models has been well considered in the literature (see e.g. Harman *et al.* (2015)). Construction of exacts designs for predictions in random coefficient regression models seems to be much more complicated, which is a challenge for future investigations.

References

Atkinson, A. C. and Donev, A. N. (1992). *Optimum Experimental Designs.* Oxford University Press, Oxford.

Atkinson, A. C., Donev, A. N., and Tobias, R. D. (2007). *Optimum Experimental Designs, with SAS.* Oxford University Press, Oxford.

Bryk, A. S. and Raudenbush, S. W. (1992). *Hierarchical Linear Models: Applications and Data Analysis Methods.* Sage, Newbury Park.

Candel, M. J. (2009). Optimal designs for empirical Bayes estimators of individual linear and quadratic growth curves in linear mixed models. *Statistical Methods in Medical Research*, **18**, 397–419.

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

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

Elfving, G. (1952). Optimum allocation in linear regression theory. *The Annals of Mathematical Statistics*, **23**, 255–262.

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

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

Fedorov, V. and Jones, B. (2005). The design of multicentre trials. *Statistical Methods in Medical Research*, **14**, 205–248.

Fedorov, V. and Leonov, S. (2013). *Optimal Design for Nonlinear Response Models.* CRC Press, Boca Raton.

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

Graßhoff, U., Doebler, A., Holling, H., and Schwabe, R. (2012). Optimal design for linear regression models in the presence of heteroscedasticity caused by random coefficients. *Journal of Statistical Planning and Inference*, **142**, 1108–1113.

Harman, R., Bachratá, A., and Filová, L. (2015). Construction of efficient experimental designs under multiple resource constraints. *Applied Stochastic Models in Business and Industry*, DOI: 10.1002/asmb.2117.

Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

Haslett, S. J., Isotalo, J., Liu, Y., and Puntanen, S. (2014). Equalities between OLSE, BLUE and BLUP in the linear model. *Statistical Papers*, **55**, 543–561.

Henderson, C. R. (1963). *Selection index and expected genetic advance. In: Statistical Genetic and Plant Breeding*. National Academy of Science - National Research Council, Washington.

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

Henderson, C. R. (1984). *Applications of Linear Models in Animal Breeding*. University of Guelph, Guelph, Canada.

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.

Isotalo, J., Puntanen, S., and Styan, G. P. H. (2011). *Matrix Tricks for Linear Statistical Models*. Springer, Berlin.

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

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

Kiefer, J. and Wolfowitz, J. (1964a). Optimum extrapolation and interpolation designs i. *Annals of the Institute of Statistical Mathematics*, **16**, 79–108.

Kiefer, J. and Wolfowitz, J. (1964b). Optimum extrapolation and interpolation designs ii. *Annals of the Institute of Statistical Mathematics*, **16**, 295–303.

Kiefer, J. and Wolfowitz, J. (1965). On a theorem of Hoel and Levine on extrapolation designs. *The Annals of Mathematical Statistics*, **36**, 1627–1655.

Läuter, E. (1974). Experimental design in a class of models. *Mathematische Operations-forschung und Statistik*, **5**, 379–398.

Liski, E. P., Mandal, N. K., Shah, K. R., and Sinha, B. K. (2002). *Topics in Optimal Design*. Springer, New York.

Molenberghs, G. and Verbeke, G. (2001). *Linear Mixed Models for Longitudinal Data*. Springer, New York.

Pazman, A. (1986). *Foundations of optimum Experimental Design*. D. Reidel Publishing Company, New York.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed Effects Models in S and S-Plus*. Springer, New York.

Prus, M. and Schwabe, R. (2011). Optimal designs for individual prediction in random coefficient regression models. In V. Melas, G. Nachtmann, and D. Rasch, editors, *Optimal Design of Experiments - Theory and Application: Proceedings of the International Conference in Honor of the late Jagdish Srivastava*, pages 122–129. Center of Experimental Design, University of Natural Resources and Life Sciences, Vienna.

Prus, M. and Schwabe, R. (2013). Optimal designs for the prediction of individual effects in random coefficient regression. In D. Uciński, A. C. Atkinson, and M. Patan, editors, *mODa 10 - Advances in Model-Oriented Design and Analysis*, pages 211–218. Springer, Cham.

Prus, M. and Schwabe, R. (2016). Optimal designs for the prediction of individual parameters in hierarchical models. *Journal of the Royal Statistical Society: Series B*, DOI: 10.1111/rssb.12105.

Pukelsheim, F. (1993). *Optimal Design of Experiments*. Wiley, New York.

Rao, C. R. (1965). The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika*, **52**, 447–458.

Rao, J. N. K. (2003). *Small Area Estimation*. Wiley, New York.

Schmelter, T. (2007). *Experimental Design For Mixed Models With Application to Population Pharmacokinetic Studies*. Ph.D. thesis, Otto-von-Guericke University Magdeburg.

Schwabe, R. (1996). *Optimum Designs for Multi-Factor Models*. Springer, New York.

Silvey, S. D. (1980). *Optimal Design*. Chapman & Hall, London.

Spjotvoll, E. (1977). Random coefficients regression models. a review. *Mathematische Operationsforschung und Statistik, Series Statistics*, **8**, 69–93.

Weaver, B. P. and Meeker, W. Q. (2014). Methods for planning repeated measures accelerated degradation tests. *Applied Stochastic Models in Business and Industry*, **30**, 658–671.