COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

TREND RESISTANT EXPERIMENTAL DESIGNS

MASTER'S THESIS

2014 Bc. Samuel ROSA

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

TREND RESISTANT EXPERIMENTAL DESIGNS

MASTER'S THESIS

Study programme:	Economic and Financial Mathematics
Branch of study:	1114 Applied mathematics
Department:	Department of Applied Mathematics and Statistics
Supervisor:	doc. Mgr. Radoslav Harman, PhD.

Bratislava 2014 **Bc. Samuel ROSA**

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

NÁVRHY EXPERIMENTOV ODOLNÉ VOČI TRENDU

DIPLOMOVÁ PRÁCA

Študijný program:	Ekonomická a finančná matematika
Študijný odbor:	1114 Aplikovaná matematika
Školiace pracovisko:	Katedra aplikovanej matematiky a štatistiky
Vedúci práce:	doc. Mgr. Radoslav Harman, PhD.

Bratislava 2014 **Bc. Samuel ROSA**

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Samuel Rosa
Study programme:	Economic and Financial Mathematics (Single degree study, master II. deg., full time form)
Field of Study:	9.1.9. Applied Mathematics
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Trend resistant experimental designs

Aim: The aim of the thesis is to find a method of constructing efficient experimental designs for the estimation of treatment contrasts in the presence of a nuisance trend.

Supervisor:	doc. Mgr. Radoslav Harman, PhD.
Department:	FMFI.KAMŠ - Department of Applied Mathematics and Statistics
Head of department:	prof. RNDr. Daniel Ševčovič, CSc.

Assigned: 25.01.2013

Approved: 04.02.2013 prof. RNDr. Daniel Ševčovič, CSc.
Guarantor of Study Programme

.....
Student

.....
Supervisor

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Samuel Rosa
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický
Sekundárny jazyk: slovenský

Názov: Návrhy experimentov odolné voči trendu / *Trend resistant experimental designs*

Cieľ: Cieľom diplomovej práce je nájsť metódu konštrukcie efektívnych návrhov experimentov pre odhad kontrastov ošetrení v prípade, že postupnosť pozorovaní vykazuje systematický, no neznámy trend.

Vedúci: doc. Mgr. Radoslav Harman, PhD.
Katedra: FMFI.KAMŠ - Katedra aplikovanej matematiky a štatistiky
Vedúci katedry: prof. RNDr. Daniel Ševčovič, CSc.

Dátum zadania: 25.01.2013

Dátum schválenia: 04.02.2013 prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

.....
študent

.....
vedúci práce

Acknowledgements I would like to express my gratitude to my supervisor doc. Mgr. Radoslav Harman, PhD. for his help and guidance.

Abstract

Bc. Samuel Rosa: Trend resistant experimental designs [Diploma thesis], Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics; Supervisor: doc. Mgr. Radoslav Harman, PhD., Bratislava, 2014, 75 pg.

Suppose that we intend to perform a sequence of independent trials, each with one of v treatments. Let the first treatment be a control and let the effects of the treatments be denoted by $\tau_1, ..., \tau_v$. The mean value of the response of each trial is assumed to be equal to the sum of the effect of the treatment selected for the trial, and the effect of a nuisance time trend. In the thesis, we give a class of optimal approximate designs for the estimation of the set of contrasts $\tau_2 - \tau_1$, ..., $\tau_v - \tau_1$, with respect to any of the Kiefer's ϕ_p-optimality criteria, $p \in [-\infty, 0]$. These criteria include the widely used criteria of D-, A- and E-optimality. We demonstrate that the results can be used to generate efficient exact designs. Furthermore, we show that our results hold in a model with general nuisance effect.

Keywords: Design of experiments, Trend resistant design, Optimal designs

Abstrakt

Bc. Samuel Rosa: Návrhy experimentov odolné voči trendu [Diplomová práca], Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky, Katedra aplikovanej matematiky a štatistiky; vedúci práce: doc. Mgr. Radoslav Harman, PhD., Bratislava, 2014, 75 str.

Uvažujme experiment, v ktorom chceme vykonať sériu nezávislých pokusov, každý s jedným z v ošetrení. Predpokladajme, že prvé ošetrenie je kontrolné a označme vplyvy ošetrení ako $\tau_1, ..., \tau_v$. Predpokladajme tiež, že stredná hodnota pozorovania v každom pokuse je rovná súčtu vplyvu zvoleného ošetrenia a rušivého vplyvu časového trendu. V diplomovej práci sme odvodili triedu optimálnych približných návrhov na štatistický odhad množiny kontrastov $\tau_2 - \tau_1$, ..., $\tau_v - \tau_1$, vzhľadom na ľubovoľné z Kieferových kritérií ϕ_p-optimality pre $p \in [-\infty, 0]$. Tieto kritéria zahŕňajú bežne používané kritéria D-, A- a E-optimality. Ukážeme, že tieto výsledky môžu byť použité na tvorbu exaktných dizajnov. Navyše ukážeme, že naše výsledky platia aj pre model so všeobecným rušivým vplyvom.

Kľúčové slová: Navrhovanie experimentov, Návrhy experimentov odolné voči trendu, Optimálne návrhy

Contents

Notation

$\mathcal{S}(A)$	the column space of matrix A, i.e. the linear space generated by the columns of A
$\mathfrak{S}_+^k$	the set of all $k \times k$ symmetric nonnegative definite matrices
$\succeq$	the Loewner ordering; $A \succeq B \Leftrightarrow A - B \in \mathfrak{S}_+^k$
1_n	the $n \times 1$ vector of ones
0_n	the $n \times 1$ vector of zeros
e_t	the t-th standard unit vector, i.e. a vector with the t-th element equal to one, and all others equal to zero
I_n	the $n \times n$ identity matrix
J_n	the $n \times n$ matrix of ones
$J_{m\times n}$	the $m \times n$ matrix of ones
$0_{m\times n}$	the $m \times n$ matrix of zeros

Introduction

The statisticians are often consulted when the researchers wish to analyse results of an experiment. One might think that this is the only role of statistics in performing an experiment: processing the data after the experiment took place. But that is not the case. The statistics may be used, and often should be used even before an experiment is performed, in designing of the experiment. It is possible to significantly improve the amount of information we get from an experiment just by properly choosing how it will be carried out. This process is studied by a discipline of statistics called *design of experiments* or *experimental design*.

Even in a simple task of weighing multiple objects, we may significantly reduce the uncertainty about the results. Suppose a chemist wishes to weigh eight light objects on a two pan scale. The traditional method of estimating the weights of these objects in eight weighings would be to weigh each object separately. However, by placing each of the objects in one of the two pans in every weighing in a particular way, one may reduce the standard error of the results to a quarter of the value given by the traditional method. This observation was provided in 1944 by Hotelling ([13]). The oldest papers considering experimental design date back to the beginning of the 20th century, e.g. [23], a paper by the famous 'Student'; as one might learn from [3]. Since then, the design of experiments has experienced a steady progress with fundamental contributions by R. Fisher (see [8]).

A design of an experiment determines how many trials are to be performed under which experimental conditions. However, to work with designs more easily, the statisticians came up with the notion of *approximate design*, which is a relaxation of the original concept. The original concept came to be known as an *exact design*. There is a large amount of literature on the subject of design of experiments, including many textbooks, e.g. [19], [20] and [2]. The reader may find a summary of the history of the experimental design in the work [3], although the paper focuses on the papers published in Biometrika.

Any experiment that consists of multiple trials performed in a time sequence may be subject to a nuisance time trend. In this paper we will focus on such experiments. The effect of a time trend might come from ageing of the material (for example in agri-

cultural experiments), heating or wearing down of the experimental devices, changes in the temperature of the environment or even fatigue of the researchers, or many other possible influences. Thus, it is important to be able to provide designs that perform well under the presence of a time trend - designs that are *trend-resistant*. The trend-resistant designs have been a subject of the design literature for a long time, e.g. in the publications [6], [5], [4] or [14].

The aim of this thesis was to study a model which describes the experiments under the presence of a nuisance time trend. Then we aimed to provide optimal designs for estimating effects of the treatments with comparison to a control treatment, using the modern approximate theory. Finally, we wished to apply our results to construct efficient exact designs for the studied model. That is, we aimed at providing efficient trend-resistant designs for estimating the treatment contrasts.

The motivation for the model that we will study came from the manuscript [10]. Unlike our work, this manuscript did not aim to provide optimal approximate designs. Instead, it provided an algorithm (branch-and-bound) for computing optimal exact designs. The algorithm was then demonstrated on a model very similar to ours.

In the first chapter, we will introduce the reader to design of experiments, following the monograph [20]. Using a linear regression model, we will formalize the difference between an exact and an approximate design and analyse the latter further. Some basic definitions and properties of the designs will be provided. Then we will examine how the quality of the designs is measured: we will introduce the optimality criteria, namely the Kiefer's ϕ_p-optimality criteria (see [16]).

We will provide the main results of this work in the second chapter. In this chapter, we will introduce the model that will be studied. Then we will examine how the approximate designs behave in the model. We will define some classes of designs and we will analyse them in detail. Finally, using the defined designs and their examined properties, we will provide a class of designs optimal with respect to the Kiefer's optimality criteria.

In the third chapter, we will propose a method for generating exact designs employ-

ing the results given in Chapter 2. We will show that these generated designs tend to be efficient. Following our observations, we will demonstrate that a highly efficient (nearly optimal) design may be quickly obtained using the proposed method. Moreover, we will show that our results can be applied to measure the efficiency of a given exact design with respect to any of the ϕ_p-optimality criteria.

In the fourth chapter, we will note that our results can be easily generalized. We will show that the same results hold for a model with a general nuisance effect, e.g. the effect of a space trend.

To our knowledge, there are no general results in the literature regarding optimal approximate designs for the model we studied. Therefore we opine that the main contribution of our work is in the results we provided and proved in the second chapter, and generalized in the fourth chapter. That is, we consider the key results of our work to be the obtained class of ϕ_p-optimal approximate designs for the model in question and for the model with a general nuisance trend.

1 Design of Experiments

1.1 Basic Definitions

When we are about to perform an experiment with a given number of trials and a given set of possible experimental conditions, we use design of experiments to choose the 'best' experimental conditions for the trials. The 'best' experimental conditions mean the ones that together produce the most information on the parameters of interest. In this chapter we will formalize this vague definition.

In this work, we will examine designs of experiments for linear regression models. Consider an experiment of N trials with real-valued observations $Y_1, \dots, Y_N$, which depend on experimental conditions $x_1, \dots, x_N$. We may model this experiment using a linear regression of the form

$$Y_i = f^T(x_i)\beta + \varepsilon_i, \quad i = 1, \dots, N. \tag{1.1}$$

Let $\mathfrak{X}$ be the set of all permissible experimental conditions for the experiment, i.e. the experimental conditions $x_1, \dots, x_N$ need to be the elements of $\mathfrak{X}$. The vector $\beta \in \mathbb{R}^m$ represents unknown parameters of the model and the mapping $f : \mathfrak{X} \to \mathbb{R}^m$ assigns regressors to the experimental conditions; $\varepsilon_1, \dots, \varepsilon_N$ are independent and identically distributed random errors with $E(\varepsilon_i) = 0$ and $\mathrm{Var}(\varepsilon_i) = \sigma^2 < \infty$ for every $i \in \{1, \dots, N\}$.

We can express the model (1.1) in the vector form

$$Y = F\beta + \varepsilon, \tag{1.2}$$

where $Y = \big(Y_1, \dots, Y_N\big)^T$, $F = \big(f(x_1), \dots, f(x_N)\big)^T$ and $\varepsilon = \big(\varepsilon_1, \dots, \varepsilon_N\big)^T$.

We assume that the objective of the experiment is to estimate some system $A^T\beta$ of unknown parameters β, where A is an $m \times s$ matrix, $s \leq m$. That is, we aim to estimate s given linear combinations of the unknown parameters of the model. Often we are interested in a specific type of linear combinations, called contrasts.

Definition 1.1. *A* contrast *is a linear combination $a^T\beta$, such that $1^T a = 0$, i.e. the sum of the coefficients of the combination is zero.*

Before we choose the design of an experiment, we do not know the exact experimental conditions x_i for all trials; we know only that they are to be chosen from the set $\mathfrak{X}$.

Then the design of experiment determines which experimental conditions x from the set $\mathfrak{X}$ we will choose. Because of this nature of the design of experiments, we call the experimental conditions the design points.

Note that the numbering of trials is arbitrary, it does not contain any information - all the information about a particular trial i is included in the experimental condition x_i, e.g. if we want to consider time dependence, we need to include it in the experimental conditions. That is, only the number of trials under a specific experimental condition is of interest, not the actual indices of the trials under the experimental condition. This brings up a natural definition of the experimental design as a function which specifies how many trials we are to perform under particular experimental conditions.

Definition 1.2. *An* exact design of experiment *is a function* $\xi : \mathfrak{X} \to \{0, 1, 2 \ldots\}$. *The value* $\xi(x)$ *denotes the number of trials to be performed under experimental conditions* x.

As the attribute 'exact' suggests, it is not a unique definition of experimental designs. Kiefer introduced another view on designs of experiments (see [15]). He considered approximate designs, which represent real-valued weights for the experimental conditions.

Definition 1.3. *An* approximate design of experiment *is a function* $\xi : \mathfrak{X} \to [0, \infty)$ *that satisfies* $\sum_{x \in \mathfrak{X}} \xi(x) = 1$. *The value* $\xi(x)$ *represents the relative proportion of trials to be performed under the experimental conditions* x. *The value* $\xi(x)$ *is called the weight of the design point* x.

The notion of the approximate designs is a generalization of the notion of the exact designs. That is, any exact design may be represented by an approximate one by choosing the suitable weights. Moreover, the approximate designs allow for any weights of the design points, not only non-negative integers.

Similarly to the definition of $\mathfrak{X}$ as the set of all permissible experimental conditions for the experiment, let Ξ be the set of all approximate designs ξ permissible for the experiment.

In this work we will mostly use the approximate designs and therefore, by a design of experiment we will mean the approximate design. Only in the parts where both

approximate and exact designs are used, we will use the full terminology, in order to differentiate between these two types of designs.

To provide some characterization of approximate designs, we will follow the book [20].

1.2 Information Matrix

Now that we have established what design of experiments is, we need to determine how to choose the 'best' design for a given experiment. In order to do so, we will introduce the notion of *information matrix* as a measure of the amount of information a design provides. But let us first define a moment matrix.

Definition 1.4. Moment matrix *of a design ξ is a matrix*

$$M(\xi) = \sum_{x \in \mathfrak{X}} \xi(x) f(x) f^T(x).$$

Remark 1.5. *We may observe the following properties of moment matrices:*
(i) The mapping $\xi \mapsto M(\xi)$ is linear, i.e. $M(a\xi_1 + b\xi_2) = aM(\xi_1) + bM(\xi_2)$ for designs ξ_1, ξ_2 and constants a, b.
(ii) Given any design ξ, the moment matrix $M(\xi)$ is a nonnegative definite symmetric matrix.

Definition 1.6 (by [9])**.** Information matrix *for a nonnegative definite matrix M and for estimating the system $A^T\beta$, where A is an $m \times s$ matrix of full column rank s, is the matrix $N_A(M)$ that satisfies*

$$N_A(M) = \min_{L \in \mathbb{R}^{s \times m} : LA = I_s} LML^T, \tag{1.3}$$

where the minimum is taken with respect to the Loewner ordering.

The definition of Loewner ordering is provided in Appendix, Definition A.6.

The justification of this definition of the information matrix can be found in [20]. For proper statistical inference, we usually need the system of parameters of interest to be estimable. We will show in Lemma 1.12 that in such situations we get a closed form expression of the information matrix. This closed form is closely related to the Fisher information matrix and it is proportional to the inverse of the dispersion matrix

for the least-square estimators of the system of interest (see [18] or [20]). It follows that it indeed is meaningful to say that the information matrix measures the amount of information the experiment provides on the parameters of interest.

We are usually interested in $N_A(M(\xi))$, i.e. the information matrix for a moment matrix of a design ξ. For brevity we will use a shorter notation $N_A(\xi)$ instead of $N_A(M(\xi))$. To simplify the notations we will even often omit the argument (ξ) where it is not necessary to preserve.

We note that the definition of the information matrix is correct, although the information matrix is defined by minimizing some matrix expression with respect to a partial ordering (the Loewner ordering is not total). That is, there exists a unique solution $\tilde{L}$ of the minimization problem (1.3), as shown in Proposition 1.7.

Throughout this work, we will use the generalized inverse matrices. Given any matrix X, the symbol X^- denotes a generalized inverse of X. For the definition of the generalized inverses and their basic properties, see the Appendix.

Proposition 1.7 (from [20])**.** *For any nonnegative definite matrix M and any full column rank matrix A there exists a unique solution of (1.3), i.e. there exist a unique $N_A(M)$. Moreover, the information matrix may be represented as*

$$N_A(M) = LML^T - LMP^T(PMP^T)^-PML^T,$$

where L is a left inverse of A and $P = I - AL$. The representation does not depend on the choice of the left and the generalized inverse.

Proof. See [20]. □

Proposition 1.8. *Given any nonnegative definite matrix M, and any matrix A of full column rank, the information matrix $N_A(M)$ is a nonnegative definite symmetric matrix.*

Proof. The proof is quite straightforward. Since the information matrix is defined by minimizing LML^T it can be written as $\tilde{L}M\tilde{L^T}$ for some matrix $\tilde{L}$. Then it is symmetric, because M is symmetric.

Let x be a vector, then $x^TN_A(M)x = x^T\tilde{L}M\tilde{L^T}x = y^TMy$, where we denoted $y := \tilde{L^T}x$. Then $x^TN_A(M)x \geq 0$, because it is equal to y^TMy for some y and M is nonnegative definite. □

We note that in many publications (e.g. [18]) the information matrix is defined as the moment matrix, i.e. $N(\xi) = \sum_{x \in \mathfrak{X}} \xi(x) f(x) f^T(x)$. That is because these publications examine the experiments, where we aim to estimate the vector β of all parameters, i.e. $A = I_m$. The following Lemma will show that in such experiments the moment matrix and the information matrix are indeed the same.

Proposition 1.9. *Let $A = I_m$. Then $N_A(M) = M$ for any nonnegative definite M.*

Proof. This proposition follows directly from Definition 1.6

$$N_A(M) = \min_{L \in \mathbb{R}^{s \times m} : LA = I_s} LML^T.$$

Once we set $A = I_m$ we get $N_A(M) = \min_{L \in \mathbb{R}^{m \times m} : L = I_m} LML^T = I_m M I_m = M.$ □

The information mapping $M \mapsto N_A(M)$ enjoys many properties (see [20]), we will make use of its concavity.

Lemma 1.10. *Let A be an $m \times s$ matrix with rank s. Then the mapping $M \mapsto N_A(M)$ is matrix concave, i.e. $N_A\big((1-\alpha)M_1 + \alpha M_2\big) \succeq (1-\alpha) N_A(M_1) + \alpha N_A(M_2)$ for $\alpha \in (0,1)$, where $\succeq$ denotes the Loewner ordering.*

Proof. See [20]. □

In this work, we will be interested in designs under which the vector $A^T\beta$ is estimable.

Definition 1.11 (by [20])**.** *The system $A^T\beta$ is said to be* estimable *under the design ξ if there exists an unbiased linear estimator of $A^T\beta$ under ξ, i.e. if there exists an $s \times N$ matrix L such that $\mathrm{E}_\xi(LY) = A^T\beta$.*

The book [20] provides us with an estimability condition as well as a closed form expression of the information matrix in such case. The estimability condition makes use of the column space of a matrix. Given any matrix B, its column space $\mathcal{S}(B)$ is the set of all linear combinations of the columns of B.

Lemma 1.12. *(i) The system $A^T\beta$ is estimable under a design ξ if and only if it satisfies the estimability condition $\mathcal{S}(A) \subseteq \mathcal{S}(M(\xi))$.*
(ii) The information matrix $N_A(\xi)$ is non-singular if and only if $A^T\beta$ is estimable. In that case we can provide an explicit form of $N_A(\xi)$:

$$N_A(\xi) = (A^T M(\xi)^- A)^{-1},$$

where $M(\xi)^-$ is a generalized inverse of $M(\xi)$.

Proof. For entire proof, see [20]. Here we will only clarify one aspect of the formula $N_A(\xi) = (A^T M(\xi)^- A)^{-1}$. We know that in general the generalized inverse $M(\xi)^-$ is not uniquely determined, therefore it seems that neither the information matrix is uniquely determined. That is not the case. Once $A^T\beta$ is estimable, from (i) we get $\mathcal{S}(A) \subseteq \mathcal{S}(M(\xi))$, and since $M(\xi)$ is symmetric we get $\mathcal{S}(A) \subseteq \mathcal{S}(M^T(\xi))$. Therefore we may use Lemma A.8 and we obtain that $N_A(\xi) = (A^T M(\xi)^- A)^{-1}$ is invariant to the choice of the generalized inverse $M(\xi)^-$. □

We note that in the case that the system $A^T\beta$ is estimable, the information matrix $N_A(\xi)$ is positive definite (because it is nonnegative definite and nonsingular).

We may define the set of all matrices (or designs) that satisfy the estimability condition.

Definition 1.13 (by [20])**.** *Let A be an $m \times s$ matrix. The* feasibility cone $\mathfrak{C}(A)$ *for the system $A^T\beta$ is the set of all nonnegative definite matrices M that satisfy $\mathcal{S}(A) \subseteq \mathcal{S}(M)$, i.e.*

$$\mathfrak{C}(A) = \{M \in \mathfrak{S}_+^m | \mathcal{S}(A) \subseteq \mathcal{S}(M)\}.$$

We say that a matrix M is feasible *for $A^T\beta$ if $M \in \mathfrak{C}(A)$ and we say that a design ξ is* feasible *for $A^T\beta$ if its moment matrix $M(\xi)$ is feasible.*

The book [20] provides some basic properties of the feasibility cone.

Proposition 1.14 (by [20])**.** *The feasibility cone is a convex cone, i.e. it satisfies*
(i) for any $\alpha > 0$ if $M \in \mathfrak{C}(A)$, then $\alpha M \in \mathfrak{C}(A)$;
(ii) for any $\alpha \in (0,1)$ if $M_1, M_2 \in \mathfrak{C}(A)$, then $\alpha M_1 + (1-\alpha)M_2 \in \mathfrak{C}(A)$.

Following [18] we define the ellipsoid of concentration.

Definition 1.15 (by [18])**.** *Let A be an $m \times s$ matrix. The* ellipsoid of concentration *for the parameter system $A^T\beta$ under the design ξ is the set $\{x \in \mathbb{R}^s | x^T N_A(\xi) x \leq 1\}$.*

We note that the ellipsoid of concentration is an actual ellipsoid only when the system $A^T\beta$ is estimable. Otherwise it becomes a cylinder.

The ellipsoid of concentration directly relates to the confidence region (or confidence ellipsoid) for an estimable system $A^T\beta$. Under the normality assumption $\varepsilon \sim \mathrm{N}(0, \sigma^2 I_N)$, the least-square estimator $A^T\hat{\beta}$ has the normal distribution with variance $\sigma^2/N \cdot N_A^{-1}(\xi)$ (see [18] or [20]). Thus the confidence region for $A^T\beta$ attains the form $\{x \in \mathbb{R}^s | (x - A^T\hat{\beta})^T N_A(\xi)(x - A^T\hat{\beta}) \leq K\}$, where $K \in \mathbb{R}$ is some constant chosen such that the confidence ellipsoid has the required confidence level. Then by normalizing and shifting the confidence region to the origin, we get the ellipsoid of concentration. That means that the ellipsoid of concentration represents how accurate the experiment is under a given design. The smaller the ellipsoid is, the more confident we may be about the results.

1.3 Optimality Criteria

Although we know that the information matrix represents the amount of information we get from an experiment, we still do not know how to compare different information matrices. To do so, statisticians introduced some real-valued optimality criteria, with D-, A- and E- optimality being ones of the most widely used.

Given an $s \times s$ information matrix $N_A(\xi)$ we denote its eigenvalues $\lambda_1, \ldots, \lambda_s$, where every eigenvalue is repeated in the sequence according to its multiplicity.

We note that the proper notation would be $\lambda_1(N_A(\xi))$ etc., but that would result in confusing expressions, therefore we will often omit the argument $(N_A(\xi))$. We will use this notation for eigenvalues of information matrices throughout the entire work.

Definition 1.16 (by [20]). *Let ϕ be a function $\phi : \mathfrak{S}_+^s \to \mathbb{R}$. Then a design ξ^* is said to be ϕ-optimal if it maximizes the function $\phi\big(N_A(\xi)\big)$ of a $s \times s$ information matrix, among all designs $\xi \in \Xi$.*

Moreover a design ξ^* is said to be

D-optimal, *if it maximizes the determinant* $\Big(\det\big(N_A(\xi)\big)\Big)^{1/s}$

A-optimal, *if it maximizes the harmonic mean of the eigenvalues of the information matrix* $\frac{s}{1/\lambda_1 + \ldots + 1/\lambda_s}$

E-optimal, *if it maximizes the smallest eigenvalue* $\lambda_{\min}(N_A(\xi))$

among all designs $\xi \in \Xi$.

We note that the expression $\Big(\det\big(N_A(\xi)\big)\Big)^{1/s}$ may be written as $\Big(\prod_i \lambda_i\Big)^{1/s}$ and hence the D-optimal design maximizes the geometric mean of the eigenvalues of the information matrix. Moreover A-optimality criterion can be expressed as $s/\text{trace}(N_A^{-1}(\xi))$.

The D-, A- and E-optimality criteria each have some statistical meaning under the normality assumption, provided in [20] and [18]. The D-optimality criterion is inversely proportional to the volume of the ellipsoid of concentration; the A-optimality criterion represents the average variance of the estimators for the contrasts of interest; and the E-optimality criterion represents the worst possible variance among all linear combinations of the system of interest $x^T A^T \hat{\beta}$ with $||x|| \leq 1$.

Let $A^T\beta$ be estimable and let the normality condition be satisfied, i.e. $\varepsilon \sim \mathrm{N}(0, \sigma^2 I_N)$. Then, as we stated earlier, the least-square estimator $A^T\hat{\beta}$ has variance $\sigma^2/N \cdot N_A^{-1}(\xi)$, which is equal to $\sigma^2/N \cdot A^T M^-(\xi)A$. Therefore we may observe the following:

1. The concentration ellipsoid for $A^T\beta$ has volume proportional to $\det^{-1/2}(N_A(\xi))$ (see [18], [20]). Since the value of the D-optimality criterion is proportional to $\det(N_A(\xi))$, a high value of the criterion implies a small volume of the ellipsoid of concentration for $A^T\beta$. Thus a high value of the D-optimality criterion implies a high confidence in the results.

2. Let us denote the columns of the matrix A as $A = \big(a_1, \ldots, a_s\big)$. Then the inverse of the A-optimality criterion may be expressed as

$$\frac{1}{\phi_{-1}(N_A(\xi))} = \frac{1}{s}\text{trace}(N_A^{-1}(\xi)) = \frac{1}{s}\text{trace}A^T M^-(\xi)A = \frac{1}{s}\sum_{i=1}^{s} a_i^T M^-(\xi) a_i. \quad (1.4)$$

   Since the least-squares estimator $A^T\hat{\beta}$ has variance $\sigma^2/N \cdot A^T M^-(\xi)A$, it follows that (1.4) is the average of the standardized variances of the least-square estimators for the linear combinations $a_1^T\beta$, ..., $a_s^T\beta$. That means, high A-optimality criterion implies low average variance of the estimators for the contrasts of interest.

3. The inverse of the smallest eigenvalue of the information matrix may be expressed as

$$\frac{1}{\lambda_{\min}(N_A(\xi))} = \lambda_{\max}(N_A^{-1}(\xi)) = \lambda_{\max}(A^T M^-(\xi)A) = \max_{x\in\mathbb{R}^s, ||x||=1} x^T A^T M^-(\xi) A x.$$

Apparently, $x^T A^T M^-(\xi) A x$ is the normalized variance of the linear combination $x^T A^T \hat{\beta}$. Therefore high E-optimality criterion guarantees low maximum variance among all possible linear combination of the estimators for the contrasts of interest $x^T A^T \beta$ such that $||x|| = 1$.

For the results to be sensible, we usually demand for the function ϕ to satisfy some properties. We call such functions the information functions.

Definition 1.17 (by [20]). *A function $\phi : \mathfrak{S}_+^s \to \mathbb{R}$ is said to be*

isotonic *with respect to the Loewner ordering if it satisfies $A \succeq B \succeq 0 \Rightarrow \phi(A) \geq \phi(B)$,*

concave *when it satisfies $\phi\big((1-\alpha)A + \alpha B\big) \geq (1-\alpha)\phi(A) + \alpha\phi(B)$ for all $\alpha \in (0,1)$, and $A, B \in \mathfrak{S}_+^s$*

positively homogenous *when it satisfies $\phi(\alpha A) = \alpha\phi(A)$ for all $\alpha > 0$, $A \in \mathfrak{S}_+^s$,*

superadditive *when it satisfies $\phi(A+B) \geq \phi(A) + \phi(B)$ for all $A, B \in \mathfrak{S}_+^s$,*

nonnegative *when it satisfies $\phi(A) \geq 0$ for all $A \in \mathfrak{S}_+^s$,*

nonconstant *when there exist $A, B \in \mathfrak{S}_+^s$ such that $\phi(A) \neq \phi(B)$,*

upper semicontinuous *when the upper level sets $\{A \in \mathfrak{S}_+^s | \phi(A) \geq \alpha\}$ are closed for all $\alpha \in \mathbb{R}$*

A function $\phi : \mathfrak{S}_+^s \to \mathbb{R}$ is called an information function *if it is positively homogenous, superadditive, nonnegative, nonconstant and upper semicontinuous.*

The book [20] provides many additional properties of the information functions, we will make use of one of them, stated in the following lemma.

Lemma 1.18. *Let ϕ be a positively homogenous function from $\mathfrak{S}_+^s$ to $\mathbb{R}$. Then ϕ is superadditive if and only if ϕ is concave.*

Proof. In [20]. □

A generalization of the D-, A- and E- optimality criteria are the Kiefer's ϕ_p-optimality criteria.

Definition 1.19 (by [20]). *Let $p \in [-\infty, 0]$. Then a design ξ^* is said to be ϕ_p-optimal if it maximizes the value of Kiefer's optimality criterion $\phi_p(N_A(\xi))$. Let N be a nonnegative definite $s \times s$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_s$, where $\lambda_{\min}$ is the smallest eigenvalue of N. Then the ϕ_p criterion is defined as*

$$\phi_p(N) = \begin{cases} \left(\frac{1}{s}\sum_{j=1}^{s} \lambda_j^p\right)^{1/p}, & p \in (-\infty, 0) \\ \left(\prod_{j=1}^{s} \lambda_j\right)^{1/s}, & p = 0 \\ \lambda_{\min}, & p = -\infty \end{cases}$$

Remark 1.20. *With $p = 0$, -1 and $-\infty$ in ϕ_p-criterion we get the D-, A- and E-optimality criterion, respectively.*

We remark that there exists a definition of Kiefer's optimality criteria for $p \in (0, 1]$ but these criteria are seldom used and therefore we will not investigate them in this work.

The ϕ_p criteria not only satisfy the properties of information functions, but for $p > -\infty$ they are also strictly concave on the set of positive definite matrices.

Lemma 1.21. *The ϕ_p criterial functions are information functions in the sense of Definition 1.17. Furthermore, for $p \in (-\infty, 0]$ the functions ϕ_p are strictly concave on the set of positive definite matrices.*

Proof. In [20]. □

We note that the technical definitions of the ϕ-optimality criteria are different throughout the experimental design literature. However they always provide the same ordering of designs with respect to the particular criteria. The main difference is usually that they define ϕ-optimal design as one that *minimizes* the value of $\phi(N_A)$, where the functions ϕ are 'inverse' to the ones we use in the sense that they provide inverse ordering of matrices.

For example in the publication [18], a D-optimal design is a design that minimizes the function $-\ln\det(N_A(\xi))$. It is obvious that given two positive definite $s \times s$ matrices N_1, N_2, $(\det(N_1))^{1/s} > (\det(N_2))^{1/s}$ if and only if $-\ln\det(N_1) < -\ln\det(N_2)$. Thus this definition is equivalent to the one we provided, with respect to the ordering of the information matrices.

2 Trend-resistant optimal designs

2.1 Introduction

In this chapter we will study the experiments that are performed in a time sequence, where the effect of a time trend is a nuisance. We will provide optimal designs for estimating a particular system of parameters in such experiments. Such designs are called *trend-resistant designs*, because they aim to provide experiments that are resistant to the nuisance trend. The interest in such designs dates back to the mid-20th century, e.g. in paper [6].

The paper [6] studies experiments for ordering the treatments under the presence of a time trend modelled by low-order polynomials. These experiments are represented by an experiment of determining the best method of processing wool. We have a certain number of bulks of wool and in each week one bulk is processed with a chosen treatment applied. Since the experiment takes multiple weeks to perform, the wool ages and thus it affects the results of the particular trials. The effect of the degradation of the wool is assumed to be a nuisance effect.

The authors provide designs that are *orthogonal*, or nearly orthogonal, to the time trend (represented by the low-order polynomials). The orthogonality means that the scalar product of the designs and a polynomial is zero for all the treatments. It can be represented by $\sum_u \xi(t,u)p_k(u) = 0$, where $\xi(t,u)$ is the design value for the t-th treatment and the u-th time, p_k is a k-th level polynomial and the sum is through all times u. The designs orthogonal to the time trend may be called *trend-free designs*, because they eliminate the effect of the trend. We will further examine the orthogonal designs in Subsection 2.3.

The approach of providing designs (nearly) orthogonal to the time trend, represented by low-order polynomials, was later studied in many publications, mostly for factorial experiments, e.g. [5], [4], [14]. We note that in the *factorial experiments* the experimental conditions are determined by multiple variables that attain only a limited number of values. These variables are called factors and their values are called levels. For example, the factors might be agricultural varieties or brands of fertilizers. More

information on factorial experiments can be found in many publications, e.g. [12], [7], [21] or [17].

The paper [5] reviews two methods for constructing trend-resistant factorial designs. The paper [4] provides a method for constructing trend-resistant designs for factorial experiments as a generalization of some earlier results. In the paper [14], some factorial experiments with two or three levels under the presence of a linear or quadratic trend are examined.

But nuisance time trend was not studied only in factorial experiments. The paper [22] gives a nearly orthogonal design in the presence of a linear trend for experiment of particle size estimation. Here, they follow the approach of [6] for ordering of treatments.

A different method of obtaining trend-resistant designs may be found in the paper [1]. The paper provides an algorithm for constructing trend-resistant D-optimal designs. This approach is especially useful in situations where the methods of constructing designs orthogonal to time trend cannot be used. For example, the theoretical results on orthogonal designs require the number of design points to be a multiple of the number of treatments, the time points to be evenly spaced and the time trend needs to be represented by a polynomial. These conditions may not hold. As an example, the authors of the paper [1] argue, that the ageing of wool in the example given by [6] may be an exponential decay rather than a low-order polynomial decay.

The reader may find a survey of the literature on the trend-resistant experimental designs in the papers [3], [5] or [1].

In this work we will consider an experiment similar to the one in [6], i.e. we will consider an experiment consisting of a time sequence of treatments and we will determine which treatments to choose in which times.

Unlike [6] and other papers mentioned earlier, we will not examine combinatorial methods for constructing exact trend-resistant methods. Neither will we provide an algorithm for constructing such designs, like [1] or [10]. Instead, we will use the theory of approximate designs and provide ϕ_p-optimal approximate designs for the model specified in the following subsection.

2.2 Model

Let us consider an industrial experiment of improving the quality of the aluminium foam. We note that the aluminium foam is a material made of aluminium which has the structure of a foam, i.e. high fraction of the material's volume consists of pores. The experimenters wish to develop a component made of aluminium foam, which is more resistant to pressure than the original component. They developed two new components - one with added magnesium and one with added silicon.

The researchers aim to determine whether these new components are more resistant to pressure than the original one. They will test 150 components by using a device, which applies pressure to a component and then they will measure the deformation of the component. However, the device allows to test only one component at a time. Moreover, the experimenters suspect that the device may heat up or even become deformed in time, thus introducing a nuisance time trend to the experiment.

To sum up, the researchers need to choose which components to test in which times to provide the best result (estimate whether the new components have better quality than the original one) under the presence of a nuisance time trend.

We may describe the aforementioned experiment (and many other experiments of similar form) by a model, which is a special case of the model (1.1). We consider an experiment of N trials with v treatments ($v \geq 2$) in n time moments under the presence of a time trend. For each trial $i \in \{1, \ldots, N\}$ we will select in which time $u(i) \in \{1, \ldots, n\}$ will this trial be performed, and we select which treatment $t(i) \in \{1, \ldots, v\}$ to apply. The effect of the treatment $t(i)$ is denoted $\tau_{t(i)} \in \mathbb{R}$. We are interested only in the treatment effects, the effect of the time trend is considered to be a nuisance. The model may be expressed as

$$Y_i = \tau_{t(i)} + \theta_1 h_1(u(i)) + \ldots + \theta_d h_d(u(i)) + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.1}$$

where $\theta_1, \ldots, \theta_d$ are the parameters of the trend and $h_1, \ldots, h_d$ are given regressors of the time trend $h_k : \mathbb{R} \to \mathbb{R}$, and $\varepsilon_1, \ldots, \varepsilon_N$ are independent and identically distributed random errors with zero mean and the same variance $\sigma^2 \in (0, \infty)$.

The set of all permissible experimental conditions is $\mathfrak{X} = \{1, \ldots, v\} \times \{1, \ldots, n\}$ and an (approximate) design ξ is a function $\xi : \{1, \ldots, v\} \times \{1, \ldots, n\} \to [0, 1]$, where $\xi(t, u), t \in \{1, \ldots, v\}, u \in \{1, \ldots, n\}$, is the relative proportion (weight) of the trials

to be performed in time u under the treatment t. Similarly an exact design ξ is any function $\xi : \mathfrak{X} \to \{0, 1, 2, \ldots\}$.

The model (2.1) can be written as a special case of the general regression model (1.1), where $f(t, u) = \left(e_t^T, h_1(u), \ldots, h_d(u)\right)^T$ and $\beta = \left(\tau_1, \ldots, \tau_v, \theta_1, \ldots, \theta_d\right)^T$. We remind the reader that the general model has the form

$$Y_i = f^T(x_i)\beta + \varepsilon_i, \quad i = 1, \ldots, N.$$

We suppose that only the contrasts $\tau_2 - \tau_1, \ldots, \tau_v - \tau_1$ are of interest, i.e. we choose the first treatment to be the control treatment and we aim to estimate the effects of the other treatments compared to the effect of the control. We can express this system of contrasts as $Q^T\tau$, where Q is a $v \times (v-1)$ matrix

$$Q = (-1_{v-1}, I_{v-1})^T$$

and $\tau = (\tau_1, \ldots, \tau_v)^T$. In the terms of Section 1, we aim to estimate the system $A^T\beta$, where $A = \left(Q^T, 0_{v-1\times d}\right)^T$ is a $(v+d) \times (v-1)$ matrix.

From now on, we will examine the model (2.1).

When considering the experiment of improving the aluminium foam, we have $N = 150$ trials in $n = 150$ time moments and consider $v = 3$ treatments: the original component and the two newly developed components. The original component represents the control treatment and therefore it is labelled as the first treatment. The set of the permissible experimental conditions is $\mathfrak{X} = \{1, 2, 3\} \times \{1, \ldots, 150\}$. The effects of the treatments τ_1, τ_2 and τ_3 are the measures of deformations of the three types of components. We are interested in estimating the improvement in the resistance to pressure for the new components, i.e. we wish to estimate $\tau_2 - \tau_1$ and $\tau_3 - \tau_1$.

We may express the model (2.1) as

$$Y_i = \tau_{t(i)} + f_\theta^T(u(i))\theta + \varepsilon_i, \quad i = 1, \ldots, N, \tag{2.2}$$

where $f_\theta(u) := \left(h_1(u), \ldots, h_d(u)\right)^T$ and $\theta = \left(\theta_1, \ldots, \theta_d\right)^T$. Then we can calculate the moment matrix for this model.

Lemma 2.1. *Let ξ be a design for the model (2.2), then its moment matrix is*

$$M(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{12}^T(\xi) & M_{22}(\xi) \end{bmatrix},$$

where

$$M_{11}(\xi) = \operatorname{diag}\left(\sum_{u=1}^{n}\xi(1,u),\ldots,\sum_{u=1}^{n}\xi(v,u)\right),$$

$$M_{12}(\xi) = \left(\sum_{u=1}^{n}\xi(1,u)f_\theta(u),\ldots,\sum_{u=1}^{n}\xi(v,u)f_\theta(u)\right)^T,$$

$$M_{22}(\xi) = \sum_{u=1}^{n}\left(\sum_{t=1}^{v}\xi(t,u)\right)f_\theta(u)f_\theta^T(u).$$

Proof. The moment matrix is defined as $M(\xi) = \sum_{x\in\mathfrak{X}}\xi(x)f(x)f^T(x)$, which is in our case $M(\xi) = \sum_{t,u}\xi(t,u)\big(e_t^T, f_\theta^T(u)\big)^T\big(e_t^T, f_\theta^T(u)\big)$. We can express the moment matrix in block form

$$M(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{12}^T(\xi) & M_{22}(\xi)\end{bmatrix},$$

where $M_{11}(\xi)$ is a $v\times v$ matrix and $M_{22}(\xi)$ is a $d\times d$ matrix. Now all we need to do, is to calculate the blocks:

$$M_{11}(\xi) = \sum_{t,u}\xi(t,u)e_te_t^T = \operatorname{diag}\left(\sum_{u=1}^{n}\xi(1,u),\ldots,\sum_{u=1}^{n}\xi(v,u)\right),$$

$$M_{12}(\xi) = \sum_{t,u}\xi(t,u)e_tf_\theta^T(u) = \begin{bmatrix}\sum_{u=1}^{n}\xi(1,u)f_\theta^T(u)\\ \vdots \\ \sum_{u=1}^{n}\xi(v,u)f_\theta^T(u)\end{bmatrix}$$

and

$$M_{22}(\xi) = \sum_{t,u}\xi(t,u)f_\theta(u)f_\theta^T(u) = \sum_{u=1}^{n}\left(\sum_{t=1}^{v}\xi(t,u)\right)f_\theta(u)f_\theta^T(u).$$

□

2.3 Properties of the Experimental Designs in the Model

We will provide some results that will allow us to better comprehend and work with the model (2.1). First, we will characterize the estimability of $A^T\beta$, using the notion of the *Schur complement* $M_\tau = M_{11} - M_{12}M_{22}^-M_{21}$ for the moment matrix M. For more information on the Schur complement see the Appendix.

Lemma 2.2 (from [10])**.** *Let $A = \big(Q^T, 0_{s\times d}\big)^T$. Then $A^T\beta$ is estimable in the model (2.1) if and only if $\mathcal{S}(M_\tau)\subseteq\mathcal{S}(Q)$.*

Proof. We need to prove that the condition $\mathcal{S}(A) \subseteq \mathcal{S}(M)$ from Lemma 1.12 is equivalent to $\mathcal{S}(M_\tau) \subseteq \mathcal{S}(Q)$.

First we will prove that $(I - M_{22}M_{22}^-)M_{21} = 0$. We denote $X := I - M_{22}M_{22}^-$, then $XM_{22} = M_{22} - M_{22}M_{22}^-M_{22} = M_{22} - M_{22} = 0$. From Lemma A.4 we know that $\mathcal{S}(M_{21}) \subseteq \mathcal{S}(M_{22})$ and thus there exists a matrix Y, such that $M_{21} = M_{22}Y$. Therefore $(I - M_{22}M_{22}^-)M_{21} = (I - M_{22}M_{22}^-)M_{22}Y = 0Y = 0$.

Let us consider the generalized inverse G of M from Lemma A.13.

$$G = \begin{bmatrix} M_\tau^- & -M_\tau^- M_{12}M_{22}^- \\ -M_{22}^- M_{21}M_\tau^- & M_{22}^- + M_{22}^- M_{21}M_\tau^- M_{12}M_{22}^- \end{bmatrix}.$$

The condition $\mathcal{S}(A) \subseteq \mathcal{S}(M)$ can be written as: there exists a matrix X such that $A = MX$. From Lemma A.9 we know that the matrix X exists if and only if A satisfies $A = MGA$. Given $A = \left(Q^T, 0\right)^T$ we may express $A = MGA$ as

$$\begin{bmatrix} Q \\ 0 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} M_\tau^- & -M_\tau^- M_{12}M_{22}^- \\ -M_{22}^- M_{21}M_\tau^- & M_{22}^- + M_{22}^- M_{21}M_\tau^- M_{12}M_{22}^- \end{bmatrix} \begin{bmatrix} Q \\ 0 \end{bmatrix}$$

which is equivalent to

$$\begin{bmatrix} Q \\ 0 \end{bmatrix} = \begin{bmatrix} M_{11} & M_{12} \\ M_{21} & M_{22} \end{bmatrix} \begin{bmatrix} M_\tau^- Q \\ -M_{22}^- M_{21}M_\tau^- Q \end{bmatrix}.$$

Thus we have two matrix equations. The first one is

$$Q = (M_{11} - M_{12}M_{22}^- M_{21})M_\tau^- Q,$$

i.e. $Q = M_\tau M_\tau^- Q$ and that is equivalent to: there exist a matrix Y such that $Q = M_\tau Y$. The second equation is

$$0 = M_{21}M_\tau^- Q - M_{22}M_{22}^- M_{21}M_\tau^- Q. \tag{2.3}$$

The right-hand side of the equation (2.3) is equal to $(I - M_{22}M_{22}^-)M_{21}M_\tau^- Q$, which is equal to 0 as we proved in the first part of this proof. Hence the second equation always holds.

The condition $\mathcal{S}(A) \subseteq \mathcal{S}(M)$ is equivalent to $Q = M_\tau Y$ for some Y, therefore it is equivalent to $\mathcal{S}(M_\tau) \subseteq \mathcal{S}(Q)$, which was to be demonstrated.

□

Once we know which designs are feasible, we may provide their information matrices.

Lemma 2.3 (from [10]). *Let $A = \big(Q^T, 0_{v-1\times d}\big)^T$ and let $A^T\beta$ be estimable under a given design ξ. Then the information matrix $N_A(\xi)$ can be expressed as $N_A(\xi) = (Q^T M_\tau^-(\xi)Q)^{-1}$, where $M_\tau(\xi) = M_{11}(\xi) - M_{12}(\xi)M_{22}^-(\xi)M_{21}(\xi)$ is the Schur complement for $M(\xi)$.*

Proof. From Lemma 1.12 we know that when $A^T\beta$ is estimable, the information matrix can be written as $N_A(\xi) = (A^T M(\xi)^- A)^{-1}$. That is in our case

$$N_A(\xi) = \Big(\big(Q^T, 0_{v-1\times d}\big)M(\xi)^-\big(Q^T, 0_{v-1\times d}\big)^T\Big)^{-1} = (Q^T(M^-)_{11}Q)^{-1},$$

where $(M^-)_{11}$ is the top left block of M^- expressed in block form. Now we need to prove, that there exists a generalized inverse M^-, such that the Schur complement M_τ^- is its top left block. But we know that from Lemma A.13. □

Using Lemma 2.3 we may better understand the emphasis on the orthogonality of the designs in many publications. In model (2.1), a design ξ is *orthogonal* to the time trend if it satisfies $\sum_{u=1}^n \xi(t,u)h_k(u) = 0$ for any $t \in \{1,\dots,v\}$ and $k \in \{1,\dots,d\}$. In other words, the design needs to satisfy $\sum_{u=1}^n \xi(t,u)f_\theta(u) = 0_d$ for any t. Note that the results on orthogonality require for the time regressors h_k to be polynomials.

Following the reasoning of [20], we will examine the orthogonal designs in the most basic situation, where we aim to estimate all treatment effects on their own, i.e. $A = \big(Q, 0\big)^T$ and $Q = I_v$. Let ξ be a feasible design. Then it must satisfy $\mathcal{S}(Q) \subseteq \mathcal{S}(M_\tau(\xi))$. Since $Q = I_v$ is a nonsingular matrix, the Schur complement M_τ needs to be nonsingular too. Therefore the information matrix N_A is $N_A = (I_v M_\tau^- I_v)^{-1} = M_\tau$, because $M_\tau^- = M_\tau^{-1}$ for a nonsingular M_τ. Now the form of the Schur complement gives us an interesting statistical interpretation. The Schur complement $M_\tau = M_{11} - M_{12}M_{22}^-M_{21}$ consists of two terms: M_{11} and $M_{12}M_{22}^-M_{21}$. The term M_{11} expresses the information matrix for estimating the treatment effects in the absence of the nuisance time trend (see Proposition 1.9). The second term, $M_{12}M_{22}^-M_{21}$, represents the loss of information due to the presence of the nuisance time trend.

Recall that the block M_{12} has the following form

$$M_{12} = \begin{bmatrix} \sum_{u=1}^{n} \xi(1,u) f_\theta^T(u) \\ \vdots \\ \sum_{u=1}^{n} \xi(v,u) f_\theta^T(u) \end{bmatrix}.$$

Therefore, once we obtain an orthogonal design, we get $M_{12} = 0$. This has some useful consequences.

Firstly, such design eliminates the loss of information caused by the nuisance time trend. Thus we get more information on the parameters of interest. Consequently, it can be expected that the designs orthogonal to the time trend will perform well with respect to optimality criteria.

Secondly, calculating the information matrix for such designs is much easier. Instead of having to calculate a generalized inverse of M_{22} and then working with a general matrix M_τ, using an orthogonal design, we get $M_\tau = M_{11} - 0 = M_{11}$. The second advantage was especially useful historically, when the computer capabilities did not allow us to work with matrices as easily as now.

The drawback of using the orthogonal designs is clear from the above analysis: they eliminate the *loss* of information by removing $M_{12}M_{22}^{-}M_{21}$, but they might not *provide* enough information represented by the term M_{11}. Naturally, the researchers studied the overall amount of information given by the orthogonal designs and they showed that these designs enjoy useful statistical properties. Namely, in the paper [6] the authors showed that in the experiment they considered, the orthogonal designs are A-optimal.

Later, we will use the regular reparametrization of the nuisance factors, to make the computations easier.

Lemma 2.4 (from [10]). *Let $A = \left(Q^T, 0\right)^T$, where Q is a $v \times s$ matrix of full column rank s. Then a regular reparametrization of nuisance parameters by changing f_θ to $\tilde{f}_\theta = Rf_\theta$ in model (2.1) does not change the information matrix $N_A(\xi)$ for any design ξ.*

Proof. After the reparametrization, the blocks of the moment matrix given by Lemma

2.1, for a fixed design ξ change to $\tilde{M}_{11} = M_{11}$,

$$\begin{aligned}\tilde{M}_{12} &= \left(\sum_{i=1}^{n} \xi(1,i) R f_\theta(i), \ldots, \sum_{u=1}^{n} \xi(v,i) R f_\theta(i)\right)^T \\ &= \left(\sum_{i=1}^{n} \xi(1,i) f_\theta(i), \ldots, \sum_{u=1}^{n} \xi(v,i) f_\theta(i)\right)^T R^T \\ &= \left(R\left(\sum_{i=1}^{n} \xi(1,i) f_\theta(i), \ldots, \sum_{u=1}^{n} \xi(v,i) f_\theta(i)\right)\right)^T = M_{12} R^T\end{aligned}$$

and

$$\tilde{M}_{22} = \sum_{i=1}^{n} \left(\sum_{t=1}^{v} \xi(t,i)\right) R f_\theta(i) f_\theta^T(i) R^T = R M_{22} R^T.$$

From Proposition 1.7 we know that the information matrix may be expressed as

$$N_A(M) = LML^T - LMP^T(PMP^T)^- PML^T$$

for $P = I - AL$ where L is any left inverse of A. Given $A = \big(Q^T, 0\big)^T$, we set $L = \big(K, 0\big)$, where K is some left inverse of Q. Since Q has full column rank, there indeed exists a left inverse K. We show that L is a left inverse matrix of A by checking $LA = I$:

$$LA = \big(K, 0\big)\big(Q^T, 0\big)^T = KQ = I.$$

For our choice of L we get

$$P = I - AL = \begin{bmatrix} I - QK & 0 \\ 0 & I \end{bmatrix}.$$

Then

$$LML^T = \begin{bmatrix} K & 0 \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \begin{bmatrix} K^T \\ 0 \end{bmatrix} = K M_{11} K^T,$$

$$\begin{aligned} LMP^T &= \begin{bmatrix} K & 0 \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \begin{bmatrix} I - K^T Q^T & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} K M_{11} & K M_{12} \end{bmatrix} \begin{bmatrix} I - K^T Q^T & 0 \\ 0 & I \end{bmatrix} \\ &= \begin{bmatrix} K M_{11}(I - K^T Q^T), & K M_{12} \end{bmatrix}\end{aligned}$$

and

$$PMP^T = \begin{bmatrix} I-QK & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} M_{11} & M_{12} \\ M_{12}^T & M_{22} \end{bmatrix} \begin{bmatrix} I-K^TQ^T & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} (I-QK)M_{11}(I-K^TQ^T) & (I-QK)M_{12} \\ M_{12}^T(I-K^TQ^T) & M_{22} \end{bmatrix}.$$

Similarly, $L\tilde{M}L^T = K\tilde{M}_{11}K^T = KM_{11}K^T = LML^T$,

$$\begin{aligned} L\tilde{M}P^T &= \left[K\tilde{M}_{11}(I-K^TQ^T), \quad K\tilde{M}_{12}\right] \\ &= \left[KM_{11}(I-K^TQ^T), \quad KM_{12}R^T\right] \\ &= \left[KM_{11}(I-K^TQ^T), \quad KM_{12}\right] \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix} \\ &= LMP^T \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix} \end{aligned}$$

and

$$\begin{aligned} P\tilde{M}P^T &= \begin{bmatrix} (I-QK)\tilde{M}_{11}(I-K^TQ^T) & (I-QK)\tilde{M}_{12} \\ \tilde{M}_{12}^T(I-K^TQ^T) & \tilde{M}_{22} \end{bmatrix} \\ &= \begin{bmatrix} (I-QK)M_{11}(I-K^TQ^T) & (I-QK)M_{12}R^T \\ RM_{12}^T(I-K^TQ^T) & RM_{22}R^T \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} \begin{bmatrix} (I-QK)M_{11}(I-K^TQ^T) & (I-QK)M_{12} \\ M_{12}^T(I-K^TQ^T) & M_{22} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix} \\ &= \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} PMP^T \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix}. \end{aligned}$$

Using Lemma A.11 we get

$$(P\tilde{M}P^T)^- = \begin{bmatrix} I & 0 \\ 0 & (R^T)^{-1} \end{bmatrix} \left(PMP^T\right)^- \begin{bmatrix} I & 0 \\ 0 & R^{-1} \end{bmatrix}.$$

Hence

$$N_A(\tilde{M}) = LML^T - LMP^T \begin{bmatrix} I & 0 \\ 0 & R^T \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & (R^T)^{-1} \end{bmatrix} \left(PMP^T\right)^- \begin{bmatrix} I & 0 \\ 0 & R^{-1} \end{bmatrix} \begin{bmatrix} I & 0 \\ 0 & R \end{bmatrix} PML^T$$
$$= LML^T - LMP^T(PMP^T)^- PML^T = N_A(M),$$

which we needed to prove.

□

2.4 Elementary Designs and γ-designs

In this subsection we will introduce two notions, elementary designs and γ-designs, which we will use extensively later.

Definition 2.5. *Let $\gamma \in (0,1)$. Then* γ-design *is any design ξ_γ that satisfies*

$$\xi_\gamma(1,u) = \gamma \sum_{t=1}^{v} \xi_\gamma(t,u) \text{ for any } u \in \{1,\ldots,n\} \tag{2.4}$$

$$\xi_\gamma(2,u) = \ldots = \xi_\gamma(v,u) \text{ for any } u \in \{1,\ldots,n\} \tag{2.5}$$

Definition 2.6. Elementary design *is any design that for some $j \in \{1,\ldots,n\}$ satisfies*

$$\xi(t,j) > 0 \text{ for all } t \in \{1,\ldots,v\}, \tag{2.6}$$

$$\xi(t,u) = 0 \text{ for all } u \neq j \text{ and for all } t \in \{1,\ldots,v\}. \tag{2.7}$$

Hence γ-designs are designs, which in each time u assign to the first (control) treatment proportion γ of the total weight assigned to the time u. The rest of the weight is distributed uniformly among the rest of the treatments. Therefore γ represents the total proportion of the trials assigned to the control treatment. Elementary designs are such designs that assign all weight to only one time and furthermore they assign non-zero weight to each treatment.

We note that an *elementary γ-design* is a design, which is both an elementary design and a γ-design. For a fixed γ we may number the elementary γ-designs as follows: let $j \in \{1,\ldots,n\}$, then the j-th elementary γ-design is the one that assigns all its weight to time j. Since γ is fixed, this notation uniquely determines the elementary γ-design.

Proposition 2.7. *Let $\gamma \in (0,1)$. Then the j-th elementary γ-design ξ attains the values*

$$\begin{aligned} &\xi(1,j) = \gamma, \quad \xi(2,j) = \ldots = \xi(v,j) = \frac{1-\gamma}{v-1} \text{ and} \\ &\xi(t,u) = 0 \text{ for } u \neq j \text{ and } t \in \{1,\ldots,v\}. \end{aligned} \tag{2.8}$$

Proof. The last part is a natural consequence of ξ being an elementary design.

The j-th elementary γ-design must satisfy $\xi(2,j) = \ldots = \xi(v,j) =: w_j$ and $\xi(1,j) = \gamma \sum_t \xi(t,j)$. Since it is a design with non-zero values only in time j, it satisfies $\sum_t \xi(t,j) = 1$ and therefore

$$\xi(1,j) = \gamma \quad \text{and} \quad w_j = \frac{1-\gamma}{v-1}.$$

□

Any γ-design can be constructed as a convex combination of elementary γ-designs. That is why we call them elementary.

Proposition 2.8. *Let $\gamma \in (0,1)$ and let ξ_γ be a γ-design. Then there exists a set of elementary γ-designs such that ξ_γ can be expressed as their convex combination.*

Proof. The total weight of ξ_γ in time u is the weight of u-th elementary γ-design in the convex combination. □

In the following Lemma we provide the form of the Schur complement $M_\tau(\xi)$ for any elementary design ξ.

Lemma 2.9. *Let $f_\theta(j) \neq 0_d$ and let ξ be an elementary design that assigns all its weight to a time j, then the Schur complement for its moment matrix is*

$$M_\tau(\xi) = \operatorname{diag}\Big(\xi(1,j), \ldots, \xi(v,j)\Big) - \Big(\xi(1,j),\ldots,\xi(v,j)\Big)^T \Big(\xi(1,j),\ldots,\xi(v,j)\Big).$$

Proof. First, we will calculate the moment matrix of the elementary design using their form given in Lemma 2.1.

$$M_{11} = \operatorname{diag}\left(\xi(1,j),\ldots,\xi(v,j)\right),$$

$$M_{12} = \begin{bmatrix} \xi(1,j) f_\theta^T(j) \\ \vdots \\ \xi(v,j) f_\theta^T(j) \end{bmatrix}$$

and $M_{22} = \sum_t \xi(t,j) f_\theta(j) f_\theta^T(j) = f_\theta(j) f_\theta^T(j)$, because $\sum_t \xi(t,j) = 1$. We note that $\operatorname{rank}(M_{22}) = 1$ and since $\xi(t,j) > 0$ for all t, the matrix M_{11} is nonsingular.

There exists $k \in \{1, \ldots, d\}$ such that $f_\theta(j)_k \neq 0$ (we use $f_\theta(j)_k$ to denote the k-th element of $f_\theta(j)$). We may use Lemma A.10 to get a generalized inverse of a matrix with rank 1 with a nonzero element on position $(1,1)$; and Lemma A.11 to permute the rows and columns of the matrix, to obtain a generalized inverse of a matrix with rank 1 with a nonzero element on the position (k,k). Therefore the matrix $\operatorname{diag}\left(0_{k-1}^T, f_\theta(j)_k^{-1}, 0_{d-k}^T\right)$ is a generalized inverse of M_{22}. Here the notation is correct only if $1 < k < d$, but in the cases of $k = 1$ and $k = d$, the situation is analogous.

Now we may compute

$$
\begin{aligned}
M_{12} M_{22}^- M_{12}^T &= \begin{bmatrix} \xi(1,j) f_\theta^T(j) \\ \vdots \\ \xi(v,j) f_\theta^T(j) \end{bmatrix} \operatorname{diag}\left(0_{k-1}^T, f_\theta(j)_k^{-1}, 0_{d-k}^T\right) \begin{bmatrix} \xi(1,j) f_\theta(j) & \ldots & \xi(v,j) f_\theta(j) \end{bmatrix} \\
&= \begin{bmatrix} 0 & \ldots & 0 & \xi(1,j)/f_\theta(j)_k & 0 & \ldots & 0 \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ 0 & \ldots & 0 & \xi(v,j)/f_\theta(j)_k & 0 & \ldots & 0 \end{bmatrix} \begin{bmatrix} \xi(1,j) f_\theta(j) & \ldots & \xi(v,j) f_\theta(j) \end{bmatrix} \\
&= \begin{bmatrix} \xi(1,j)^2 & \xi(1,j)\xi(2,j) & \ldots & \xi(1,j)\xi(v,j) \\ \vdots & \ddots & & \vdots \\ \xi(v,j)\xi(1,j) & \xi(v,j)\xi(2,j) & \ldots & \xi(v,j)^2 \end{bmatrix}.
\end{aligned}
$$

From this we directly get the Schur complement

$$
M_\tau = M_{11} - M_{12} M_{22}^- M_{12}^T = \begin{bmatrix} \xi(1,j) - \xi(1,j)^2 & -\xi(1,j)\xi(2,j) & \ldots & -\xi(1,j)\xi(v,j) \\ \vdots & \ddots & & \vdots \\ -\xi(v,j)\xi(1,j) & -\xi(v,j)\xi(2,j) & \ldots & \xi(v,j) - \xi(v,j)^2 \end{bmatrix}.
$$

□

Note that the Schur complements M_τ for elementary designs do not depend on the time regressors $h_1, \ldots, h_d$ and thus (if elementary designs are feasible for A) the information matrices for elementary designs do not depend on time regressors.

We will prove that all the elementary designs are feasible for our A.

Theorem 2.10. *Let $A = \left(Q^T, 0_{v-1 \times d}\right)^T$, where $Q = (-1_{v-1}, I_{v-1})^T$ and let ξ be an elementary design. Then the system $A^T\beta$ is estimable under ξ.*

Proof. We consider two alternatives:

a) $f_\theta(j) = 0_d$. Then

$$M = \begin{bmatrix} M_{11} & 0 \\ 0 & 0 \end{bmatrix}.$$

Since ξ is an elementary design, M_{11} is nonsingular. Then there exists a matrix X_{11} such that $M_{11}X_{11} = Q$. Let X be $X = \left(X_{11}^T, 0\right)^T$. Then $MX = A$, which means that $\mathcal{S}(A) \subseteq \mathcal{S}(M)$.

b) $f_\theta(j) \neq 0_d$. We will prove that $\mathcal{S}(M_\tau) \subseteq \mathcal{S}(Q)$. From Lemma 2.9 we know the form of the Schur complement M_τ

$$M_\tau = \begin{bmatrix} \xi(1,j) - \xi(1,j)^2 & -\xi(1,j)\xi(2,j) & \dots & -\xi(1,j)\xi(v,j) \\ \vdots & \ddots & & \vdots \\ -\xi(v,j)\xi(1,j) & -\xi(v,j)\xi(2,j) & \dots & \xi(v,j) - \xi(v,j)^2 \end{bmatrix}.$$

The condition $\mathcal{S}(M_\tau) \subseteq \mathcal{S}(Q)$ is equivalent to: there exists a matrix X such that $Q = M_\tau X$. Using Lemma A.9 we get that we need to prove $M_\tau G Q = Q$, where G is a generalized inverse of M_τ. In order to do so, we first need a generalized inverse G. We propose that $G := \operatorname{diag}\left(1/\xi(1,j), \dots, 1/\xi(v,j)\right)$ is a generalized inverse matrix of M_τ. It is easy to verify that by computing $M_\tau G M_\tau$.

Let us denote $w := \left(\xi(1,j), \dots, \xi(v,j)\right)^T$. Then $M_\tau = \operatorname{diag}(w) - ww^T$ and $G = \left(\operatorname{diag}(w)\right)^{-1}$, which we will denote as $G = \operatorname{diag}^{-1}(w)$. Thus

$$\begin{aligned} M_\tau G M_\tau &= \left(\operatorname{diag}(w) - ww^T\right) \operatorname{diag}^{-1}(w) \left(\operatorname{diag}(w) - ww^T\right) \\ &= \operatorname{diag}(w) \operatorname{diag}^{-1}(w) \operatorname{diag}(w) - \operatorname{diag}(w) \operatorname{diag}^{-1}(w) ww^T - \\ &\quad - ww^T \operatorname{diag}^{-1}(w) \operatorname{diag}(w) + ww^T \operatorname{diag}^{-1}(w) ww^T \\ &= \operatorname{diag}(w) - 2ww^T + ww^T \operatorname{diag}^{-1}(w) ww^T. \end{aligned}$$

Furthermore, $ww^T \operatorname{diag}^{-1}(w) ww^T = w1_v^T ww^T = ww^T$, because $1_v^T w = \sum_{t=1}^{v} \xi(t,j) = 1$ for an elementary design. Therefore $M_\tau G M_\tau = \operatorname{diag}(w) - 2ww^T + ww^T = M_\tau$.

Now we may calculate

$$\begin{aligned} M_\tau GQ &= \left(\operatorname{diag}(w) - ww^T\right) \operatorname{diag}^{-1}(w) \begin{bmatrix} -1_{v-1}^T \\ I_{v-1} \end{bmatrix} \\ &= \left(I_v - w1_v^T\right) \begin{bmatrix} -1_{v-1}^T \\ I_{v-1} \end{bmatrix} \\ &= \begin{bmatrix} -1_{v-1}^T \\ I_{v-1} \end{bmatrix} - w0_{v-1}^T = Q. \end{aligned}$$

□

From Theorem 2.10 it follows that all γ-designs are feasible.

Corollary 2.11. *Let $\gamma \in (0,1)$ and let $A = \left(Q^T, 0_{v-1\times d}\right)^T$, where $Q = (-1_{v-1}, I_{v-1})^T$. Then any γ-design is feasible for the system $A^T\beta$.*

Proof. Let ξ_γ be a γ-design. From Proposition 2.8 we know that any γ-design can be constructed as a convex combination of elementary γ-designs. Since the relationship between ξ and $M(\xi)$ is linear, the moment matrix of any γ-designs can be expressed as a convex combinations of some moment matrices of elementary γ-designs.

Formally, we may express this as: there exist $\alpha_1, \ldots, \alpha_n \in [0,1]$ such that

$$\xi_\gamma = \sum_i \alpha_i \xi_i^{(e)}, \quad \sum_i \alpha_i = 1$$

where $\xi_i^{(e)}$ is the i-th elementary γ-design. Since $M(\xi)$ is linear in ξ, it follows that

$$M(\xi_\gamma) = \sum_i \alpha_i M(\xi_i^{(e)}).$$

Moreover, from Theorem 2.10 we get that all elementary γ-designs are feasible for $A^T\beta$. We know from Proposition 1.14 that the set of all feasible matrices is convex. Therefore, $M(\xi_\gamma)$ lies in the feasibility cone, i.e. ξ_γ is feasible. □

Since we know that the elementary γ-designs are feasible, we may compute their information matrices $N_A(\xi) = (Q^T M_\tau(\xi)^- Q)^{-1}$.

Lemma 2.12. *Let $\gamma \in (0,1)$ and let ξ be an elementary γ-design. Then its information matrix is $N_A(\xi) = \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}$ with eigenvalues*

$$\lambda_1 = \frac{1-\gamma}{v-1} \quad \textit{with multiplicity } v-1 \tag{2.9}$$

and

$$\lambda_2 = \frac{1-\gamma}{v-1} - \frac{(1-\gamma)^2}{v-1} = \frac{\gamma(1-\gamma)}{v-1} \quad \textit{with multiplicity } 1. \tag{2.10}$$

Proof. From Proposition 2.7 we know that

$$\xi(1,j) = \gamma \quad \text{and} \quad \xi(2,j) = \ldots = \xi(v,j) = \frac{1-\gamma}{v-1}.$$

We will consider two cases:

a) $f_\theta(j) \neq 0$. Then from Lemma 2.9 it follows that

$$M_\tau = \operatorname{diag}\Big(\xi(1,j), \ldots, \xi(v,j)\Big) - \Big(\xi(1,j), \ldots, \xi(v,j)\Big)^T \Big(\xi(1,j), \ldots, \xi(v,j)\Big).$$

We know the values of $\xi(t,j)$ and therefore

$$M_\tau = \begin{bmatrix} \gamma - \gamma^2 & -\frac{\gamma(1-\gamma)}{v-1} 1_{v-1}^T \\ -\frac{\gamma(1-\gamma)}{v-1} 1_{v-1} & \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1} \end{bmatrix}.$$

In order to calculate the information matrix N_A we need to find a generalized inverse M_τ^-. Let us denote the bottom right block of M_τ as $D := \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}$. The generalized inverse can be calculated directly by using Lemma A.10(ii), provided that M_τ has rank $v-1$ and D has full rank. That is quite straightforward to prove. When we take the sum of all elements of any column of M_τ, we get

$$\gamma - \gamma^2 - (v-1)\frac{\gamma(1-\gamma)}{v-1} = 0$$

or

$$-\frac{\gamma(1-\gamma)}{v-1} + \frac{1-\gamma}{v-1} - (v-1)\left(\frac{1-\gamma}{v-1}\right)^2 = \frac{(1-\gamma)^2}{v-1} - \frac{(1-\gamma)^2}{v-1} = 0,$$

which is zero in either case. Thus the sum of all rows of M_τ is a row full of zeros, which means that the rows are linearly dependent, i.e. M_τ is singular. Furthermore using Lemma A.5, D has eigenvalues $\lambda_1 = \frac{1-\gamma}{v-1}$ with multiplicity $v-1$ and $\lambda_2 = \frac{1-\gamma}{v-1} - \frac{(1-\gamma)^2}{v-1} = \frac{\gamma(1-\gamma)}{v-1}$ with multiplicity 1. Neither of these eigenvalues is zero, i.e. D has rank $v-1$, and therefore M_τ has rank at least $v-1$. Since we proved that M_τ is nonsingular, it has rank $v-1$. That means, we have satisfied the assumptions of Lemma A.10 and the matrix

$$\begin{bmatrix} 0 & 0_{v-1}^T \\ 0_{v-1} & D^{-1} \end{bmatrix}$$

is a generalized inverse of M_τ. Now we can calculate $Q^T M_\tau^- Q$:

$$\begin{aligned} Q^T M_\tau^- Q &= \Big(-1_{v-1}, I_{v-1}\Big) \begin{bmatrix} 0 & 0_{v-1}^T \\ 0_{v-1} & D^{-1} \end{bmatrix} \Big(-1_{v-1}, I_{v-1}\Big)^T \\ &= \begin{bmatrix} 0_{v-1} & D^{-1} \end{bmatrix} \begin{bmatrix} -1_{v-1}^T \\ I_{v-1} \end{bmatrix} \\ &= D^{-1}. \end{aligned}$$

Then

$$N_A = (Q^T M_\tau^- Q)^{-1} = D = \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}.$$

b) $f_\theta(j) = 0$. Then the blocks of the moment matrix of the j-th elementary γ-design are: $M_{12} = 0$, $M_{22} = 0$ and $M_{11} = \operatorname{diag}\Big(\gamma, \frac{1-\gamma}{v-1}, \ldots, \frac{1-\gamma}{v-1}\Big)$. Thus $M_\tau = M_{11}$ and $M_\tau^- = (M_{11})^{-1} = \operatorname{diag}\Big(\gamma^{-1}, \frac{v-1}{1-\gamma}, \ldots, \frac{v-1}{1-\gamma}\Big)$. Then

$$\begin{aligned} Q^T M_\tau^- Q &= \Big(-1_{v-1}, I_{v-1}\Big) \operatorname{diag}\Big(\gamma^{-1}, \frac{v-1}{1-\gamma}, \ldots, \frac{v-1}{1-\gamma}\Big) \Big(-1_{v-1}, I_{v-1}\Big)^T \\ &= \Big(-1_{v-1}, I_{v-1}\Big) \Big(-\frac{1}{\gamma} 1_{v-1}, \frac{v-1}{1-\gamma} I_{v-1}\Big)^T \\ &= \frac{v-1}{1-\gamma} I_{v-1} + \frac{1}{\gamma} J_{v-1}. \end{aligned}$$

By using the special form of $Q^T M_\tau^- Q$ we can easily calculate $N_A = (Q^T M_\tau^- Q)^{-1}$. Let us assume that N_A has form $aI_{v-1} + bJ_{v-1}$, then it must satisfy

$$\big(aI_{v-1} + bJ_{v-1}\big) \left(\frac{v-1}{1-\gamma} I_{v-1} + \frac{1}{\gamma} J_{v-1}\right) = I_{v-1},$$

resulting in

$$a\frac{v-1}{1-\gamma} = 1 \quad \text{and} \quad a\frac{1}{\gamma} + b\frac{v-1}{1-\gamma} + b(v-1)\frac{1}{\gamma} = 0.$$

Thus

$$a = \frac{1-\gamma}{v-1},\ b = -\left(\frac{1-\gamma}{v-1}\right)^2,\ \text{hence } N_A = \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}.$$

Earlier in this proof, we defined D and noted that using Lemma A.5, D has eigenvalues

$$\lambda_1 = \frac{1-\gamma}{v-1} \quad \text{with multiplicity } v-1$$

and

$$\lambda_2 = \frac{\gamma(1-\gamma)}{v-1} \quad \text{with multiplicity 1.}$$

Furthermore, we proved that $N_A = D$. Therefore N_A has the aforementioned eigenvalues. □

2.5 ϕ_p-optimal designs

The aim of this section is to provide ϕ_p-optimal designs for the experiment defined by model (2.2). We will show that for a suitably chosen γ, the γ-designs are ϕ_p-optimal.

Before formulating the main results, we examine a particular equation.

Lemma 2.13. *For any $p \in (-\infty, 0]$ and $v > 2$ there exists a unique solution γ of the equation*

$$(v-2)\gamma^{1-p} + 2\gamma - 1 = 0 \tag{2.11}$$

in the interval $(0, 1/2)$. Furthermore the solution satisfies $\lim_{p\to-\infty} \gamma(p) = 1/2$.

For $v = 2$, the unique solution satisfies $\gamma(p) = 1/2$ for any $p \in (-\infty, 0]$.

Proof. Let $g(\gamma)$ be the function of the left-hand side of the equation (2.11), i.e. $g(\gamma) = (v-2)\gamma^{1-p} + 2\gamma - 1$. The solution of the equation is then the root of $g(\gamma)$. Let us calculate the derivative

$$g'(\gamma) = (1-p)(v-2)\gamma^{-p} + 2,$$

which is positive on $\gamma \in (0, 1/2)$. Thus there exists at most one solution in $(0, 1/2)$. The values of $g(\gamma)$ on the boundaries of the interval in consideration are $g(0) = -1 < 0$ and $g(1/2) = (v-2)2^{p-1} + 1 - 1 = (v-2)2^{p-1} > 0$. Hence there exists a solution on $(0, 1/2)$.

Together this means that there always exists a unique solution of (2.11) on the interval $(0, 1/2)$.

Let $\{p_n\}_{n=1}^{\infty}$ be any sequence such that $p_n \to -\infty$. Then $\lim_{n\to\infty}(v-2)\gamma^{1-p_n} = 0$ for any $\gamma \in (0, 1/2)$. Thus the sequence of solutions $\{\gamma(p_n)\}_{n=1}^{\infty}$ implied by (2.11) converges to $1/2$, because it must satisfy $0 + 2\gamma - 1 = 0$.

When $v = 2$, we get $2\gamma - 1 = 0$ and hence the solution is $\gamma(p) = 1/2$ for any p. □

We examine further the function given by the left-hand side of the equation (2.11). If the function has 'nice' properties, it will be easy to numerically find its root.

Lemma 2.14. *The function $g(\gamma) := (v-2)\gamma^{1-p} + 2\gamma - 1$ is an increasing convex function for $\gamma \in (0, 1/2)$.*

Proof. In the proof of Lemma 2.13 we already calculated the first derivative of $g(\gamma)$

$$g'(\gamma) = (1-p)(v-2)\gamma^{-p} + 2.$$

Hence, $g'(\gamma) > 0$ for $\gamma \in (0, 1/2)$ and therefore $g(\gamma)$ is increasing in $(0, 1/2)$. We can easily calculate the second derivative

$$g''(\gamma) = -p(1-p)(v-2)\gamma^{-p} \text{ for } p < 0$$

and $g''(\gamma) = 0$ for $p = 0$. Thus $g''(\gamma) \leq 0$ for $\gamma \in (0, 1/2)$ and we get that $g(\gamma)$ is convex in $(0, 1/2)$. □

The main theorem we will use to prove the optimality of the chosen designs, is the Equivalence Theorem, as stated in [20].

Theorem 2.15 (from [20])**.** *(i) For $p \in (-\infty, 0]$ a design ξ^* is ϕ_p-optimal if and only if there exists a generalized inverse G of $M(\xi^*)$ that satisfies the normality inequality*

$$f^T(x)GAN_A^{p+1}(\xi^*)A^TG^Tf(x) \leq \text{trace}N_A^p(\xi^*) \text{ for all } x \in \mathcal{X}. \tag{2.12}$$

Furthermore in case of optimality, in (2.12) we get equality if for x we insert any support point x_i of any design that is ϕ_p-optimal.
(ii) A design ξ^* is $\phi_{-\infty}$-optimal if and only if there exists a generalized inverse G of the moment matrix $M(\xi^*)$ and a nonnegative definite matrix E with $\text{trace}(E) = 1$ *such that they satisfy the normality inequality*

$$f^T(x)GAN_A(\xi^*)EN_A^T(\xi^*)G^Tf(x) \leq \lambda_{\min}(N_A(\xi^*)) \text{ for all } x \in \mathcal{X}. \tag{2.13}$$

Furthermore, in case of optimality, in (2.13) we get equality if for x we insert any support point of any $\phi_{-\infty}$-optimal design.

Proof. See [20]. □

Now we may formulate the main theorem which summarizes the key results of this work.

Theorem 2.16. *Let $A = \left(Q^T, 0_{v-1\times d}\right)^T$, where $Q = \left(-1_{v-1}, I_{v-1}\right)^T$. Let $p \in [-\infty, 0]$. If $p > -\infty$, let γ be the unique solution of the equation (2.11)*

$$(v-2)\gamma^{1-p} + 2\gamma - 1 = 0$$

in the interval $(0, 1/2]$. If $p = -\infty$, let $\gamma = 1/2$. Then any γ-design is ϕ_p-optimal for the estimation of the system $A^T\beta$ in model (2.1).

Proof. Let $p \in [-\infty, 0]$. For $p < 0$ we set γ as the solution of (2.11). We note that from Lemma 2.13 it follows that γ is well defined. For $p = 0$ we set $\gamma = 1/2$.

In this proof we will follow these steps:

1. We will reparametrize the nuisance time trend, so that the computations will be easier.

2. For $p \in (-\infty, 0]$ we will prove that one of the elementary γ-designs is ϕ_p-optimal.

3. We will prove that from the step 2. it follows that all γ-designs are ϕ_p-optimal.

4. We will repeat the steps 2. and 3. for $p = -\infty$.

5. Finally, we will prove optimality of γ-designs in a 'degenerate' situation with no time trend present.

1. Let there be j such that $f_\theta(j) \neq 0$. From now on, this j will be fixed. Then there exists a non-singular matrix R such that $Rf_\theta(j) = e_1$, where e_1 is a $d \times 1$ elementary unit vector.

Let us consider a new model created by a regular reparametrization of nuisance parameters of the original model $f_\theta(u) = Rf_\theta(u)$ for $u \in \{1, \ldots, n\}$. As we proved in Lemma 2.4, the information matrix does not change under a regular reparametrization of the nuisance parameters $f_\theta \mapsto Rf_\theta$. Thus when we consider the new nuisance vectors, all the designs have the same information matrices as previously. Since the criterial values ϕ_p depend only on the information matrices, the new model has for any p the same set of ϕ_p-optimal designs as the original one. This allows us to use the new model where $\tilde{f}_\theta(j) = e_1$ without any loss of information.

2. Let us denote the j-th elementary γ-design as ξ^*. Now we will prove that ξ^* is ϕ_p-optimal, using Theorem 2.15(i). That is, we will prove that ξ^* satisfies the normality inequality (2.12)

$$f(x)^T G A N_A^{p+1}(\xi^*) A^T G^T f(x) \leq \text{trace} N_A^p(\xi^*) \text{ for all } x \in \mathcal{X},$$

where G is some generalized inverse of $M(\xi^*)$.

For that we will need to know the moment matrix M for ξ^*. So let us calculate the blocks of M.

$$M_{11} = \operatorname{diag}\left(\xi(1,j), \ldots, \xi(v,j)\right),$$
$$M_{12} = \begin{bmatrix} \xi(1,j) & 0 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ \xi(v,j) & 0 & \ldots & 0 \end{bmatrix},$$
$$M_{22} = \big(1, 0 \ldots, 0\big)^T \big(1, 0, \ldots, 0\big) = \operatorname{diag}(1,\, 0,\, \ldots,\, 0)$$

Using Gaussian elimination we easily get from M to a matrix

$$\begin{bmatrix} M_{11} & M_{12} \\ 0_{d\times v} & 0_{d\times d} \end{bmatrix}$$

by subtracting the first v rows from the $(v+1)$-st row. Note that we used the fact that $\sum_t \xi(t,j) = 1$. This means that M has rank v. M_{11} is a nonsingular matrix, because none of the values $\xi(t,j)$ is equal to zero. Thus, we can use Lemma A.10(i) and therefore the matrix

$$G = \operatorname{diag}\Big(\frac{1}{\xi(1,j)}, \ldots, \frac{1}{\xi(v,j)}, 0, \ldots, 0\Big)$$

is a generalized inverse of M.

From Proposition 2.7 it follows that $\xi(1,j) = \gamma$ and $\xi(r,j) = \frac{1-\gamma}{v-1}$ for $r > 1$. Hence the matrix G may be expressed as

$$G = \operatorname{diag}\Big(\frac{1}{\gamma}, \frac{v-1}{1-\gamma} \ldots, \frac{v-1}{1-\gamma}, 0, \ldots, 0\Big). \tag{2.14}$$

Using Lemma 2.12 we get that ξ has the information matrix $N_A = \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}$ with eigenvalues $\lambda_1 = \frac{1-\gamma}{v-1}$ with multiplicity $v-2$ and $\lambda_2 = \frac{\gamma(1-\gamma)}{v-1}$ with multiplicity 1.

A set of eigenvectors corresponding to λ_1 is

$$q_1 = \begin{pmatrix} -1 \\ e_1 \end{pmatrix}, \; q_2 = \begin{pmatrix} -1 \\ e_2 \end{pmatrix}, \; \ldots, \; q_{v-2} = \begin{pmatrix} -1 \\ e_{v-2} \end{pmatrix}.$$

An eigenvector corresponding to λ_2 is $q_{v-1} = 1_{v-1}$.

We can easily check that by calculating:

$$N_A - \lambda_1 I = -\left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}, \quad N_A - \lambda_2 I = \frac{(1-\gamma)^2}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}$$

and

$$(N_A - \lambda_1 I)q_k = -\left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}\begin{pmatrix}-1\\ e_k\end{pmatrix} = 0_{v-1}.$$

Furthermore, the vectors $q_1, \ldots, q_{v-2}$ are obviously linearly independent. Thus they indeed are a set of eigenvectors corresponding to λ_1. To prove that $q_{v-1} = 1_{v-1}$ is an eigenvector corresponding to λ_2, we calculate

$$\begin{aligned}(N_A - \lambda_2 I)q_{v-1} &= \left(\frac{(1-\gamma)^2}{v-1}I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}\right)1_{v-1}\\ &= \frac{(1-\gamma)^2}{v-1}1_{v-1} - (v-1)\left(\frac{1-\gamma}{v-1}\right)^2 1_{v-1} = 0_{v-1}.\end{aligned}$$

Then N_A may be expressed as $N_A = S\Lambda S^{-1}$, where S and Λ are $(v-1)\times(v-1)$ matrices, $\Lambda = \operatorname{diag}\big(\lambda_1 1_{v-2}^T, \lambda_2\big)$ and $S = \big(q_1, \ldots, q_{v-1}\big)$. We can express S as

$$S = \begin{bmatrix}-1_{v-2}^T & 1\\ I_{v-2} & 1_{v-2}\end{bmatrix}.$$

Now we may easily calculate S^{-1}

$$S^{-1} = \begin{bmatrix}-\frac{1}{v-1}1_{v-2} & I_{v-2} - \frac{1}{v-1}J_{v-2}\\ \frac{1}{v-1} & \frac{1}{v-1}1_{v-2}^T\end{bmatrix}.$$

We verify that $SS^{-1} = I$

$$SS^{-1} = \begin{bmatrix}\frac{v-2}{v-1} + \frac{1}{v-1} & -1_{v-2}^T + \frac{v-2}{v-1}1_{v-2}^T + \frac{1}{v-1}1_{v-2}^T\\ -\frac{1}{v-1}1_{v-2} + \frac{1}{v-1}1_{v-2} & I_{v-2} - \frac{1}{v-1}J_{v-2} + \frac{1}{v-1}J_{v-2}\end{bmatrix} = I_{v-1}$$

From $N_A = S\Lambda S^{-1}$ it follows that for any $r \in \mathbb{R}$: $N_A^r = S\Lambda^r S^{-1}$, where $\Lambda^r = \operatorname{diag}\big(\lambda_1^r 1_{v-2}^T, \lambda_2^r\big)$. Hence N_A^{p+1} may be calculated as

$$\begin{aligned}N_A^{p+1} &= \begin{bmatrix}-1_{v-2}^T & 1\\ I_{v-2} & 1_{v-2}\end{bmatrix}\begin{bmatrix}\lambda_1^{p+1}I_{v-2} & 0\\ 0 & \lambda_2^{p+1}\end{bmatrix}\begin{bmatrix}-\frac{1}{v-1}1_{v-2} & I_{v-2} - \frac{1}{v-1}J_{v-2}\\ \frac{1}{v-1} & \frac{1}{v-1}1_{v-2}^T\end{bmatrix}\\ &= \begin{bmatrix}-\lambda_1^{p+1}1_{v-2}^T & \lambda_2^{p+1}\\ \lambda_1^{p+1}I_{v-2} & \lambda_2^{p+1}1_{v-2}\end{bmatrix}\begin{bmatrix}-\frac{1}{v-1}1_{v-2} & I_{v-2} - \frac{1}{v-1}J_{v-2}\\ \frac{1}{v-1} & \frac{1}{v-1}1_{v-2}^T\end{bmatrix}\\ &= \begin{bmatrix}\frac{(v-2)\lambda_1^{p+1}+\lambda_2^{p+1}}{v-1} & \frac{-\lambda_1^{p+1}+\lambda_2^{p+1}}{v-1}1_{v-2}^T\\ \frac{-\lambda_1^{p+1}+\lambda_2^{p+1}}{v-1}1_{v-2} & \lambda_1^{p+1}I_{v-2} + \frac{-\lambda_1^{p+1}+\lambda_2^{p+1}}{v-1}J_{v-2}\end{bmatrix},\end{aligned}$$

which can be written as

$$N_A = \lambda_1^{p+1}I_{v-1} + \frac{-\lambda_1^{p+1}+\lambda_2^{p+1}}{v-1}J_{v-1} \tag{2.15}$$

We know that $A = \left(Q^T, 0_{v-1\times d+1}\right)^T$, where $Q = \left(-1_{v-1}, I_{v-1}\right)^T$ and we can finally calculate the matrix product on the left-hand side of (2.12).

$$AN_A^{p+1}A^T = \begin{bmatrix} QN_A^{p+1}Q^T & 0 \\ 0 & 0 \end{bmatrix}, \tag{2.16}$$

$$QN_A^{p+1}Q^T = \begin{bmatrix} 1_{v-1}^T N_A^{p+1} 1_{v-1} & -1_{v-1}^T N_A^{p+1} \\ -N_A^{p+1}1_{v-1} & N_A^{p+1} \end{bmatrix}. \tag{2.17}$$

Moreover

$$N_A^{p+1}1_{v-1} = \Big(\lambda_1^{p+1} + (v-1)\frac{-\lambda_1^{p+1}+\lambda_2^{p+1}}{v-1}\Big)1_{v-1} = \lambda_2^{p+1}1_{v-1} = \Big(\frac{\gamma(1-\gamma)}{v-1}\Big)^{p+1}1_{v-1} \tag{2.18}$$

and

$$1_{v-1}^T N_A^{p+1}1_{v-1} = (v-1)\lambda_2^{p+1} = (v-1)\Big(\frac{\gamma(1-\gamma)}{v-1}\Big)^{p+1}. \tag{2.19}$$

Let us express G in the block form

$$G = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix},$$

where G_{11} is a $v \times v$ matrix and G_{22} is a $d \times d$ matrix. Then, because G is diagonal (see (2.14)), the following holds

$$\begin{aligned} GAN_A^{p+1}A^TG^T &= \begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix}\begin{bmatrix} QN_A^{p+1}Q^T & 0 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} G_{11} & 0 \\ 0 & G_{22} \end{bmatrix} \\ &= \begin{bmatrix} G_{11}QN_A^{p+1}Q^TG_{11} & 0 \\ 0 & 0 \end{bmatrix} \end{aligned}$$

and, because $f(x) = f(t,u) = \left(e_t^T, h_1(u), \ldots, h_d(u)\right)^T$,

$$f^T(x)GAN_A^{p+1}A^TG^Tf(x) = e_t^TG_{11}QN_A^{p+1}Q^TG_{11}e_t$$

for $t \in \{1,\ldots,v\}$ and for any $u \in \{1,\ldots,n\}$. We denote $B := G_{11}QN_A^{p+1}Q^TG_{11}$. Then the left-hand side of (2.12) is equal to the $t-$th diagonal element of B, i.e. $B_{t,t}$. Because of the diagonal form of G_{11}, the elements $B_{t,t}$ can be expressed as

$$B_{t,t} = \left(G_{11}(t,t)\right)^2\left(QN_A^{p+1}Q^T\right)(t,t), \tag{2.20}$$

where the argument (t,t) denotes the t-th diagonal element.

We remind the reader that from (2.14), (2.17), (2.18), (2.19) and (2.15), we get $G_{11} = \mathrm{diag}\Big(\frac{1}{\gamma}, \frac{v-1}{1-\gamma} \ldots, \frac{v-1}{1-\gamma}\Big)$,

$$QN_A^{p+1}Q^T = \begin{bmatrix} 1_{v-1}^T N_A^{p+1} 1_{v-1} & -1_{v-1}^T N_A^{p+1} \\ -N_A^{p+1} 1_{v-1} & N_A^{p+1} \end{bmatrix}$$

and where $1_{v-1}^T N_A^{p+1} 1_{v-1} = (v-1)\lambda_2^{p+1} = (v-1)\Big(\frac{\gamma(1-\gamma)}{v-1}\Big)^{p+1}$ and

$$\begin{aligned} N_A^{p+1}(t,t) &= \lambda_1^{p+1} + \frac{-\lambda_1^{p+1} + \lambda_2^{p+1}}{v-1} = \frac{(v-2)\lambda_1^{p+1} + \lambda_2^{p+1}}{v-1} \\ &= \frac{1}{v-1}\Big((v-2)\Big(\frac{1-\gamma}{v-1}\Big)^{p+1} + \Big(\frac{\gamma(1-\gamma)}{v-1}\Big)^{p+1}\Big) \end{aligned}$$

Therefore the elements $B_{t,t}$ (2.20) attain two distinct values: one for $t = 1$ and another one for $t > 1$. They are

$$\begin{aligned} B_{1,1} &= \frac{1}{\gamma^2}(v-1)\Big(\frac{\gamma(1-\gamma)}{v-1}\Big)^{p+1} \\ &= \gamma^{p-1}(1-\gamma)\Big(\frac{1-\gamma}{v-1}\Big)^{p} \end{aligned}$$

and

$$\begin{aligned} B_{t,t} &= \left(\frac{v-1}{1-\gamma}\right)^2 \frac{1}{v-1}\left((v-2)\Big(\frac{1-\gamma}{v-1}\Big)^{p+1} + \Big(\frac{\gamma(1-\gamma)}{v-1}\Big)^{p+1}\right) \\ &= \frac{v-1}{(1-\gamma)^2}\left(\frac{1-\gamma}{v-1}\right)^{p+1}\Big(v-2+\gamma^{p+1}\Big) \\ &= \frac{1}{1-\gamma}\left(\frac{1-\gamma}{v-1}\right)^{p}\Big(v-2+\gamma^{p+1}\Big) \end{aligned}$$

for $t > 1$.

We need to prove that $B_{t,t} \leq \mathrm{trace}(N_A^p)$, knowing that

$$\begin{aligned} \mathrm{trace}(N_A^p) &= (v-2)\lambda_1^p + \lambda_2^p = (v-2)\big(\frac{1-\gamma}{v-1}\big)^p + \left(\frac{\gamma(1-\gamma)}{v-1}\right)^p \\ &= \big(\frac{1-\gamma}{v-1}\big)^p\big(v-2+\gamma^p\big). \end{aligned}$$

We will show that when γ satisfies (2.11), the inequality we want to prove is equivalent to a simple inequality which always holds. First, let us consider $v > 2$. Hence, when γ satisfies (2.11)

$$(v-2)\gamma^{1-p} + 2\gamma - 1 = 0$$

it also satisfies

$$\gamma^{p-1} = \frac{v-2}{1-2\gamma}, \tag{2.21}$$

$$\gamma^{p} = \big(v-2\big)\frac{\gamma}{1-2\gamma} \tag{2.22}$$

and

$$\gamma^{p+1} = \big(v-2\big)\frac{\gamma^2}{1-2\gamma} \tag{2.23}$$

Moreover, we note that since $\gamma \in (0, 1/2)$, the expressions $(1-\gamma)$ and $(1-2\gamma)$ are both positive. Therefore we may multiply or divide an equation by these expressions, without changing its solutions.

a) When $t = 1$, the inequality to prove is

$$\gamma^{p-1}(1-\gamma)\Big(\frac{1-\gamma}{v-1}\Big)^p \leq (\frac{1-\gamma}{v-1})^p\big(v-2+\gamma^p\big), \tag{2.24}$$

which is equivalent to

$$\gamma^{p-1}(1-\gamma) \leq v-2+\gamma^p.$$

Using (2.21) and (2.22) we get

$$\frac{v-2}{1-2\gamma}(1-\gamma) \leq v-2+(v-2)\frac{\gamma}{1-2\gamma},$$

which is equivalent to

$$1-\gamma \leq 1-2\gamma+\gamma$$

and that is equivalent to $1 \leq 1$, which always holds.

b) When $t > 1$, the inequality to prove is

$$\frac{1}{1-\gamma}\left(\frac{1-\gamma}{v-1}\right)^p\Big(v-2+\gamma^{p+1}\Big) \leq \left(\frac{1-\gamma}{v-1}\right)^p\Big(v-2+\gamma^p\Big), \tag{2.25}$$

which is equivalent to

$$\frac{1}{1-\gamma}\Big(v-2+\gamma^{p+1}\Big) \leq v-2+\gamma^p.$$

Using (2.22) and (2.23) we get

$$\frac{1}{1-\gamma}\Big(v-2+\big(v-2\big)\frac{\gamma^2}{1-2\gamma}\Big) \leq v-2+\big(v-2\big)\frac{\gamma}{1-2\gamma}.,$$

which is equivalent to

$$\frac{1-2\gamma+\gamma^2}{1-\gamma} \leq 1-2\gamma+\gamma$$

and that is equivalent to $(1-\gamma)^2 \leq (1-\gamma)^2$, which always holds.

Now we need to return to the case, where $v = 2$. Thus $\gamma = 1/2$ and the inequalities to prove are (2.24) and (2.25), i.e.

$$(1/2)^{p-1}(1/2)\Big(\frac{1}{2(2-1)}\Big)^p \leq \Big(\frac{1}{2(2-1)}\Big)^p\Big(0+(1/2)^p\Big)$$

for $t = 1$ and

$$(2-1)2^2\left(\frac{1}{2(2-1)}\right)^{p+1}\Big(0+(1/2)^{p+1}\Big) \leq \big(\frac{1}{2(2-1)}\big)^p\big(0+(1/2)^p\big)$$

for $t > 1$. Both of these inequalities apparently hold, because they are equivalent to $(1/2)^{2p} \leq (1/2)^{2p}$.

This proves that the j-th elementary γ-design is ϕ_p-optimal. We may observe, that in the normality inequality we always attained equality. That is in accordance with the second part of Theorem 2.15(i), because we claim that every γ-design is ϕ_p-optimal (for the fixed γ). For every design point $x \in \mathcal{X}$ there is a γ-design which has x as its support point, therefore for every x an equality shall be attained.

3. We note that the value of the criterion ϕ_p for a design ξ depends only on its information matrix $N_A(\xi)$. Since the j-th elementary γ-design is ϕ_p-optimal and all the elementary γ-designs have the same information matrix (we denote it N_A^*), all the elementary γ-designs are ϕ_p-optimal.

Furthermore, from Proposition 2.8 we know that any γ-design ξ_γ can be expressed as a convex combination of the elementary γ-designs. From Remark 1.5 and Lemma 1.10 we know that the function $\xi \mapsto N_A(\xi)$ is a matrix concave function and thus any γ-design ξ satisfies $N_A(\xi) \succeq N_A^*$.

We recall that the real-valued function ϕ_p from the space of $s \times s$ nonnegative definite matrices $N \mapsto \phi_p(N)$ is Loewner isotonic, because it is an information function (see Lemma 1.21). Thus $\phi_p(N_A(\xi_\gamma)) \geq \phi_p(N_A^*)$ for any γ-design ξ_γ. But ϕ_p attains its maximum in N_A^* and therefore $\phi_p(N_A(\xi_\gamma)) = \phi_p(N_A^*)$ and any γ-design is ϕ_p-optimal.

We can make an additional observation. From Lemma 1.21 we also know that for $p > -\infty$, the function ϕ_p is strictly concave on the set of positive definite matrices. Since we are considering only designs feasible for $A^T\beta$ (see Corollary 2.11), the information

matrices $N_A = (A^T M^- A)^{-1}$ are non-singular.

Thus on the subspace of matrices we are considering, the optimality criterion is strictly concave. Hence there is only a single information matrix which attains the maximum value of ϕ_p. We proved that N_A^* is the matrix that attains the optimal value and that all the γ-designs are ϕ_p-optimal. Thus all the γ-designs have the same information matrix N_A^*

4. What remains to prove, is the E-optimality. In this case, the steps are the same up to the point when we started using the equivalence theorem. When considering E-optimality,we will make use of the part (ii) of Theorem 2.15, i.e. we will provide a generalized inverse G of M and a nonnegative definite matrix E with trace trace$(E) = 1$ such that they satisfy the normality inequality

$$f^T(x)GAN_A(\xi^*)EN_A^T(\xi^*)G^T f(x) \leq \lambda_{\min}(N_A(\xi^*)) \text{ for all } x \in \mathcal{X}.$$

We remind the reader that ξ^* is the j-th elementary γ-design. ξ^* has the information matrix $N_A^* = \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}$ with its smallest eigenvalue $\lambda_2 = \frac{\gamma(1-\gamma)}{v-1}$ and we may use the same generalized inverse of $M(\xi^*)$, $G = \text{diag}\Big(\frac{1}{\gamma}, \frac{v-1}{1-\gamma} \ldots, \frac{v-1}{1-\gamma}, 0, \ldots, 0\Big)$.

Finally, we set

$$E = \frac{1}{v-1} J_{v-1}, \tag{2.26}$$

which is obviously nonnegative definite and has trace equal to 1. We recall that $A = \left(Q^T, 0_{v-1\times d+1}\right)^T$, where $Q = \left(-1_{v-1}, I_{v-1}\right)^T$ and $f(x) = f(t,i) = \left(e_t^T, h_1(i), \ldots, h_d(i)\right)^T$. We will prove that these matrices satisfy the normality inequality for E-optimality.

Using $\gamma = 1/2$, we get $N_A = \frac{1}{2(v-1)} I_{v-1} - \frac{1}{4(v-1)^2} J_{v-1}$, $\lambda_2 = \frac{1}{4(v-1)}$ and
$G = \text{diag}\Big(2, 2(v-1)\ldots, 2(v-1), 0, \ldots, 0\Big)$ represented as

$$G = \begin{bmatrix} G_{11} & 0 \\ 0 & 0 \end{bmatrix}.$$

Let us start by calculating $f^T(x)GAN_A$

$$\begin{aligned} f^T(x)GAN_A &= \begin{bmatrix} e_t^T G_{11} & 0 \end{bmatrix} \begin{bmatrix} QN_A \\ 0 \end{bmatrix} = e_t^T G_{11} QN_A = \\ &= e_t^T G_{11} \begin{bmatrix} -1_{v-1}^T \\ I_{v-1} \end{bmatrix} \Big(\frac{1}{2(v-1)} I_{v-1} - \frac{1}{4(v-1)^2} J_{v-1} \Big) \\ &= e_t^T G_{11} \begin{bmatrix} -\frac{1}{2(v-1)} 1_{v-1}^T + (v-1)\frac{1}{4(v-1)^2} 1_{v-1}^T \\ \frac{1}{2(v-1)} I_{v-1} - \frac{1}{4(v-1)^2} J_{v-1} \end{bmatrix} \\ &= g_{tt} e_t^T \begin{bmatrix} -\frac{1}{2(v-1)} 1_{v-1}^T + \frac{1}{4(v-1)} 1_{v-1}^T \\ \frac{1}{2(v-1)} I_{v-1} - \frac{1}{4(v-1)^2} J_{v-1} \end{bmatrix}, \end{aligned}$$

where we denoted g_{tt} the t-th diagonal element of G_{11}. Let us also denote d_t^T the t-th row of the matrix

$$\begin{bmatrix} -\frac{1}{2(v-1)} 1_{v-1}^T + \frac{1}{4(v-1)} 1_{v-1}^T \\ \frac{1}{2(v-1)} I_{v-1} - \frac{1}{4(v-1)^2} J_{v-1} \end{bmatrix}.$$

Then $f^T(x)GAN_A = g_{tt} d_t^T$. Then, using (2.26), the left-hand side of (2.13) is equal to

$$g_{tt} d_t^T E d g_{tt} = g_{tt}^2 \frac{1}{v-1} d^T J_{v-1} d.$$

Again, we have two distinct cases - for $t = 1$ and for $t > 1$.

a) For $t = 1$ the left-hand side of (2.13) is equal to

$$\begin{aligned} \text{LHS} &= \frac{4}{v-1} \Big(-\frac{1}{2(v-1)} + \frac{1}{4(v-1)} \Big) 1_{v-1}^T J_{v-1} \Big(-\frac{1}{2(v-1)} + \frac{1}{4(v-1)} \Big) 1_{v-1} \\ &= \frac{1}{4(v-1)^3} 1_{v-1}^T J_{v-1} 1_{v-1} = \frac{1}{4(v-1)^3} (v-1)^2 \\ &= \frac{1}{4(v-1)} \end{aligned}$$

and that is equal to the right-hand side of (2.13), i.e. λ_2. Thus the inequality holds.

b) For $t > 1$ the left-hand side of (2.13) is equal to

$$\begin{aligned} \text{LHS} &= 4\Big(v-1\Big) \Big(\frac{1}{2(v-1)} e_{t-1} - \frac{1}{4(v-1)^2} 1_{v-1} \Big)^T J_{v-1} \Big(\frac{1}{2(v-1)} e_{t-1} - \frac{1}{4(v-1)^2} 1_{v-1} \Big) \\ &= \frac{1}{(v-1)} \Big(e_{t-1}^T - \frac{1}{2(v-1)} 1_{v-1}^T \Big) J_{v-1} \Big(e_{t-1} - \frac{1}{2(v-1)} 1_{v-1} \Big) \\ &= \frac{1}{(v-1)} \left(e_{t-1}^T J_{v-1} e_{t-1} - \frac{2}{2(v-1)} e_{t-1}^T J_{v-1} 1_{v-1} + \frac{1}{4(v-1)^2} 1_{v-1}^T J_{v-1} 1_{v-1} \right) \\ &= \frac{1}{(v-1)} \left(1 - 1 + \frac{1}{4} \right) = \frac{1}{4(v-1)} \end{aligned}$$

and that is again equal to the right-hand side of (2.13).

Therefore the inequality always holds and moreover it is always attained as an equality. Like the situation with $p \in (0, \infty)$, that is in accordance with the second part of the equivalence theorem.

We can follow the same reasoning as in the case where $p \in (0, \infty)$ (see step 4), because the function $\phi_{-\infty}$ is Loewner isotonic too. Thus for $\gamma = 1/2$, all the γ-designs are $\phi_{-\infty}$-optimal.

6. Finally, we will consider the situation where $h_k(u) = 0$ for each k and for each u, i.e. there is no time trend present. Let us consider any elementary γ-design ξ. Even in this case we may use Lemma 2.12 and we get that $N_A = \frac{1-\gamma}{v-1} I_{v-1} - \left(\frac{1-\gamma}{v-1}\right)^2 J_{v-1}$.

The blocks of the moment matrix of ξ are: $M_{12} = 0$, $M_{22} = 0$ and $M_{11} = \operatorname{diag}\left(\gamma, \frac{1-\gamma}{v-1}, \ldots, \frac{1-\gamma}{v-1}\right)$. Then

$$\begin{aligned} G &= \operatorname{diag}\left(\frac{1}{\xi(1,j)}, \ldots, \frac{1}{\xi(v,j)}, 0, \ldots, 0\right) \\ &= \operatorname{diag}\left(\frac{1}{\gamma}, \frac{v-1}{1-\gamma} \ldots, \frac{v-1}{1-\gamma}, 0, \ldots, 0\right) \end{aligned}$$

is a generalized inverse of M. In the case of $\phi_{-\infty}$-optimality we may again set $E = \frac{1}{v-1} J_{v-1}$. Thus all matrices in normality equations of Equivalence Theorem are the same as previously and following the same reasoning we get that the γ-designs are ϕ_p-optimal.

□

We may observe some interesting properties of the results given by Theorem 2.16.

The theorem gives us an optimal proportion γ of the trials assigned to the first treatment relative to the total number of trials. To better understand the dependence between p and the optimal γ we provide a chart of this dependence for some chosen v (see Figure 1 below).

The designs we provide are invariant to the nuisance time trend present: their optimality does not depend on the degree d of the nuisance time trend nor on the regressors $h_1, \ldots, h_d$. For example, the optimal designs that we give are the same for the linear time trend, the quadratic time trend or even an exponential trend.

The optimality of the designs we provided depends neither on the weights these designs assign to the particular times $1, \ldots, n$. Thus a γ-design is ϕ_p-optimal as long

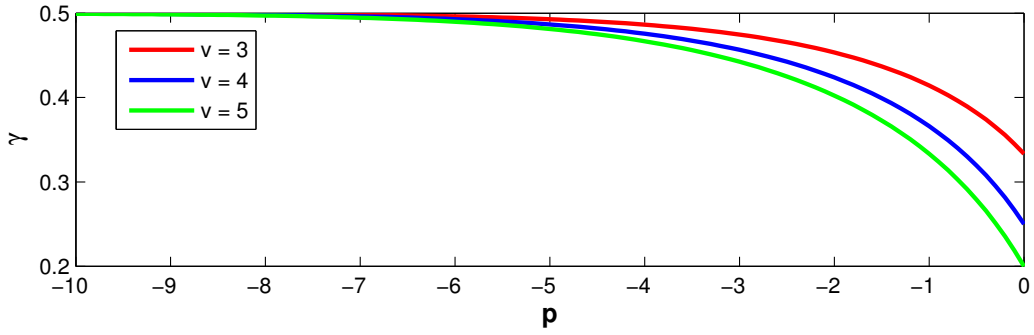

Figure 1: Dependence between p and the optimal γ given by (2.11). Note that, in accordance with Lemma 2.13, with p decreasing from 0 to $-\infty$ the optimal γ increases from $1/v$ to $1/2$. As we decrease p, we increase the weight of the first treatment and proportionally decrease the weight of the rest of the treatments. In D-optimal designs we assign the same weight $1/v$ to all treatments and in E-optimal designs we assign $1/2$ of all trials to the first treatment and the rest of the trials is distributed uniformly among the rest of the treatments.

as it satisfies the equation (2.11) (respectively $\gamma = 1/2$ for $p = -\infty$), even if it assigns all trials to only one time u. Hence, Theorem 2.16 provides an infinite number of ϕ_p-optimal designs for any $p \in [-\infty, 0]$, as we may choose from an infinite number of possible weights for the time moments $1, \ldots, n$.

Directly from Theorem 2.16 we can calculate the optimal values of ϕ_p criterion. For certain p we will provide an explicit formula for the optimal γ and for the optimal value of the ϕ_p criterion.

We note that from Lemma 2.14 we know that the function $g(\gamma)$ of the left-hand side of the equation (2.11) is increasing and convex. Therefore the function is very simple from the numerical point of view and it is easy to numerically calculate its root in $(0, 1/2)$ for any $p \in (-\infty, 0]$.

Corollary 2.17. *For $p \in (-\infty, 0)$ the criterial value $\phi_p(\xi^*)$ of any ϕ_p-optimal design ξ^* is given by the equation*

$$\phi_p(\xi^*) = \frac{1-\gamma_p}{v-1}\left(1 - \frac{1-\gamma_p^p}{v-1}\right)^{1/p}, \tag{2.27}$$

where γ_p is the solution of (2.11).

Furthermore for $v > 2$, the values of γ_p and the optimal criterial values for D-, A-

and E- optimality are

$$\gamma_0 = 1/v, \quad \phi_0(\xi^*) = v^{-v/(v-1)} \quad \textit{(D-optimality)},$$
$$\gamma_{-1} = \frac{\sqrt{v-1}-1}{v-2}, \quad \phi_{-1}(\xi^*) = \Big(\frac{\sqrt{v-1}-1}{v-2}\Big)^2 \quad \textit{(A-optimality)},$$
$$\gamma_{-\infty} = 1/2, \quad \phi_{-\infty}(\xi^*) = \frac{1}{4(v-1)} \quad \textit{(E-optimality)}.$$

Proof. From Theorem 2.16 we know that elementary γ-designs are ϕ_p-optimal as long as γ_p satisfies (2.11)

$$(v-2)\gamma_p^{1-p} + 2\gamma_p - 1 = 0$$

for $p \in (-\infty, 0]$ or $\gamma_p = 1/2$ for $p = -\infty$. Furthermore, from Lemma 2.12 we get that they have information matrix $N_A = \frac{1-\gamma_p}{v-1} I_{v-1} - \Big(\frac{1-\gamma_p}{v-1}\Big)^2 J_{v-1}$ with eigenvalues $\lambda_1 = \frac{1-\gamma_p}{v-1}$ with multiplicity $v-2$ and $\lambda_2 = \frac{\gamma_p(1-\gamma_p)}{v-1}$ with multiplicity 1. Thus for $p \in (-\infty, 0)$, the value of ϕ_p criterion for such designs is

$$\phi_p(N_A) = \left[\frac{1}{v-1}\sum_j \lambda_j^p\right]^{1/p} = \left[\frac{1}{v-1}\left((v-2)\Big(\frac{1-\gamma_p}{v-1}\Big)^p + \Big(\frac{\gamma_p(1-\gamma_p)}{v-1}\Big)^p\right)\right]^{1/p} =$$
$$= \left[\frac{1}{v-1}\Big(\frac{1-\gamma_p}{v-1}\Big)^p\Big(v-2+\gamma_p^p\Big)\right]^{1/p} = \frac{1-\gamma_p}{v-1}\left(1 - \frac{1-\gamma_p^p}{v-1}\right)^{1/p}.$$

For $p = 0$ we have $(v-2)\gamma_0 + 2\gamma_0 - 1 = 0$, thus $\gamma_0 = 1/v$ and we get

$$\phi_0(N_A) = \left(\prod_j \lambda_j\right)^{\frac{1}{v-1}} = \left(\Big(\frac{1-\gamma_0}{v-1}\Big)^{v-2}\frac{\gamma_0(1-\gamma_0)}{v-1}\right)^{\frac{1}{v-1}} =$$
$$= \left(\Big(\frac{1-\gamma_0}{v-1}\Big)^{v-1}\gamma_0\right)^{\frac{1}{v-1}} = \frac{1-\frac{1}{v}}{v-1}\Big(\frac{1}{v}\Big)^{\frac{1}{v-1}} = \frac{1}{v}\Big(\frac{1}{v}\Big)^{\frac{1}{v-1}} = v^{-\frac{v}{v-1}}.$$

For $p = -\infty$ we have $\gamma_{-\infty} = 1/2$ and we get

$$\phi_{-\infty}(N_A) = \lambda_2 = \frac{\gamma_{-\infty}(1-\gamma_{-\infty})}{v-1} = \frac{1}{4(v-1)}.$$

Finally, we will calculate the optimal value of A-optimality criterion, i.e. we set $p = -1$. We note that this is not a special case, we only need to set $p = -1$ in the general formula for $p \in (-\infty, 0)$. First, we will calculate γ_{-1}. We have $(v-2)\gamma_{-1}^2 + 2\gamma_{-1} - 1 = 0$. For $v > 2$, this quadratic equality has two roots

$$\gamma_{1,2} = \frac{-1 \pm \sqrt{v-1}}{v-2},$$

where one of them is obviously negative and thus not permissible. Therefore we get

$$\gamma_{-1} = \frac{\sqrt{v-1}-1}{v-2}.$$

Then we substitute in the general formula:

$$\begin{aligned}\phi_{-1}(N_A) &= \frac{1-\gamma_{-1}}{v-1}\left(1-\frac{1-\gamma_{-1}^{-1}}{v-1}\right)^{-1} = \frac{1-\gamma_{-1}}{v-2+\gamma_{-1}^{-1}} = \\ &= \frac{1-\frac{\sqrt{v-1}-1}{v-2}}{v-2+\frac{v-2}{\sqrt{v-1}-1}} = \frac{\frac{v-1-\sqrt{v-1}}{v-2}}{(v-2)\frac{\sqrt{v-1}}{\sqrt{v-1}-1}} = \\ &= \frac{(v-1-\sqrt{v-1})(\sqrt{v-1}-1)}{(v-2)^2\sqrt{v-1}} = \left(\frac{\sqrt{v-1}-1}{v-2}\right)^2.\end{aligned}$$

Since for a given p the elementary γ-designs are ϕ_p-optimal, every ϕ_p-optimal design needs to have the previously calculated value of ϕ_p criterion. □

To demonstrate the behavior of the function of the optimal criterial values, we display the optimal values of ϕ_p criterion for some chosen values of v.

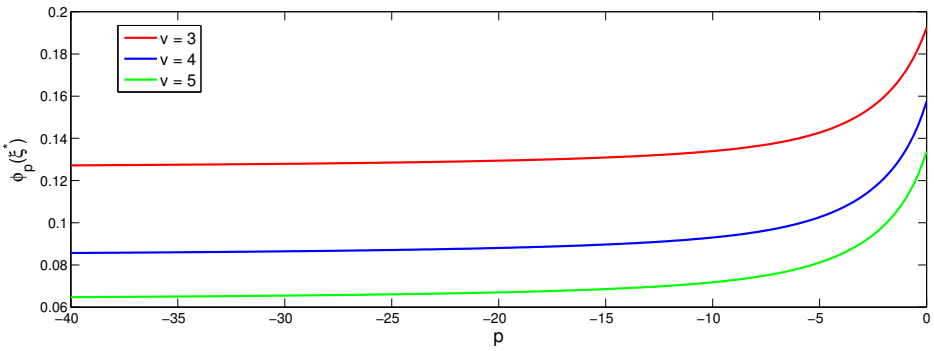

Figure 2: The optimal values $\phi_p(\xi^*)$ for $v = 3$, 4 and 5 treatments.

3 Applications

3.1 Direct Approach

In the previous chapter we provided a class of ϕ_p-optimal approximate designs for model (2.1). However, in order to design a real experiment, the experimenters need an exact design. This limits the application of approximate designs. But although we cannot use them directly to design an experiment, they can still be very useful in practice. In this chapter we will investigate how it is possible to apply the results we provided.

If we were lucky enough, the weights we calculated for different experimental conditions could be directly transformed to the number of trials under each condition; bear in mind that we assume that we have a given total number of trials. Then we can use a direct approach and simply transform the approximate design to the exact one (by multiplying its every value with the same number). Then, since the approximate designs generalize the exact designs, once the approximate design is optimal, so is the exact one. This can be better represented using an example.

Example 3.1. *Consider an experiment described by model (2.1): we have $N = 20$ trials with the choice of $v = 3$ treatments, where the first treatment is the control. We aim to perform these trials in a time sequence of length $n = 5$, where in each time we perform 4 trials. The experimenters wish to use an E-optimal design.*

By using Theorem 2.16 we get an E-optimal approximate design ξ_a which assigns in each time weight $1/2$ to the first treatment and weights $1/4$ to each of the other two treatments. This design can be transformed to an exact design ξ_e, where in each time we perform 2 treatments with the first treatment and 1 treatment with the second and third treatment, respectively. Since ξ_a is an E-optimal approximate design, ξ_e is an E-optimal exact design.

This example not only shows how optimal exact designs can be directly created from optimal exact designs, but also how rare these situations are. If we had $v = 4$ treatments or a longer time sequence with up to 3 trials in each time, we could not get an E-optimal exact design using the direct approach. Especially in a standard situation, where in each time we perform exactly one trial, the direct approach cannot

be used. Thus we need to use this approximation in other ways.

3.2 Efficiency

Efficiency of a design is a measure of how good a design is with respect to a given criterion. It compares its criterial value with the criterial value of an optimal design. The optimal design can be an exact or an approximate one and with either of them we would get a different value of efficiency. In this work we will consider efficiency with respect to an optimal approximate design.

Definition 3.2. *Let ξ be a design and ξ^* be ϕ-optimal approximate design of a given experiment. Then*
(i) if ξ is an approximate design, then the ϕ-efficiency of ξ is

$$\mathrm{eff}_\phi(\xi) = \frac{\phi(\xi)}{\phi(\xi^*)}.$$

(ii) if ξ is an exact design, then the ϕ-efficiency of ξ is

$$\mathrm{eff}_\phi(\xi) = \frac{\phi(\xi_a)}{\phi(\xi^*)},$$

where ξ_a is an approximate design given by ξ, i.e. $\xi_a(x) = \xi(x)/\sum_{z\in\mathfrak{X}}\xi(z)$ for all $x \in \mathfrak{X}$. That is, efficiency of ξ is equal to the efficiency of the approximate design ξ_a given by ξ.

Often finding an optimal exact design is very time consuming. In such cases it is usually sufficient to find an exact design, which has efficiency high enough, e.g. 95%. This is where we may use approximate designs. Knowing the optimal value of ϕ_p-criterion, we can immediately calculate ϕ_p-efficiency of any given design. Therefore we can use some level of efficiency as a stopping rule for calculating nearly ϕ_p-optimal exact designs and thus reducing computation time of algorithms.

We can use optimal approximate designs not only to find efficient exact designs, but also to evaluate the quality of a given exact design. This is especially useful for accepting some candidate exact design. Suppose that a heuristic algorithm has provided us with an exact design which should attain high value of A-optimality criterion.

It is sensible to check whether the candidate design is good enough. Using Corollary 2.17, we may calculate its A-efficiency with respect to the optimal approximate design. Then if the A-efficiency is high enough, we may accept the candidate design.

Moreover, our results may be used for comparing multiple candidates for the 'best' exact design. Often, we choose a criterion of optimality ϕ and execute an algorithm to find a (near-)optimal design with respect to the criterion ϕ. The drawback of such method is that we do not know how good is the design with respect to other optimality criteria. This drawback is emphasized by the fact that none of the optimality criteria is superior to the others, they all represent in some way the amount of information we get from the experiment.

From Theorem 2.16 we know ϕ_p-optimal approximate designs for any p and thus we can for any p calculate ϕ_p-efficiency, i.e. the quality of a design with respect to the ϕ_p criterion, without the need to find optimal exact designs for multiple p. This allows us to choose the candidate that performs the best with respect to other ϕ_p criteria; or even reject the optimal candidate and accept a near-optimal candidate which performs significantly better with respect to other ϕ_p criteria.

Suppose that we were given three competing designs of an experiment given by model (2.1) with a cubic time trend and $N = n = 18$ trials. They might be outputs of a heuristic algorithm performed in order to find D-, A- and E-optimal exact designs. Using Corollary 2.17 we calculated their efficiencies, see Table 1.

ξ	$\mathrm{eff}_D(\xi)$	$\mathrm{eff}_A(\xi)$	$\mathrm{eff}_E(\xi)$
231131232232131132	0.9992	0.9703	0.8875
123311221133112231	0.9613	0.9955	0.9870
213111223123111312	0.8951	0.9508	0.9876

Table 1: Table of D-, A- and E-efficiencies for a model with a cubic time trend and $N = n = 18$. The sequence of numbers in the first column determines the time sequence of treatments (e.g. 23... means that in time 1 the treatment 2 is chosen, in time 2 the treatment 3 is chosen, etc.).

Employing our results, we can guarantee a very high efficiency of the second design with respect to the common optimality criteria (D-, A- and E-optimality). Therefore

it allows us to accept this design without knowing the D-, A- and E-optimal exact designs.

We note that these three designs are in fact D-, A- and E-optimal exact designs, respectively. They were computed using the branch-and-bound algorithm provided by [10]. Notice that the D-optimal design chooses each of the three treatments 6 times; the A-optimal design chooses the first treatment 8 times and the other treatments each 5 times; and the E-optimal design chooses the first treatment 9 times, the second treatment 5 times and the third treatment 4 times. This means that these exact designs (nearly) satisfy the conditions given by Theorem 2.16 and explicitly formulated in Corollary 2.17.

The D-optimal approximate design should allocate the same weight to each treatment (i.e. 6 trials). The A-optimal approximate design should give the first treatment the weight $\gamma_{-1} = (\sqrt{v-1}-1)/(v-2) \cong 0.41$. That represents approximately $18 \cdot \gamma_{-1} \cong 7.5$ trials under the first treatment. The other two treatments should have the same weight. The E-optimal design should assign the first treatment to half of the trials and the other two treatments should be chosen in a quarter of the trials each. Notice that the optimal exact designs approximately satisfy these conditions.

This observation suggests that we may use the weights given by Theorem 2.16 to obtain efficient exact designs. We will examine this hypothesis in the following subsection.

3.3 Randomly Generated Designs

We showed previously that direct approach can be used only rarely. But that is not the only method of obtaining exact designs from our results on approximate designs.

We may get an exact design with high value of ϕ_p-efficiency by generating exact designs that nearly satisfy the conditions of Theorem 2.16. That is, we generate exact designs, that assign (nearly) the same number of trials to treatments $2, \ldots, v$ and the relative number of trials performed under control treatment is given by (2.11). We will show on some examples that these designs tend to have high ϕ_p-efficiency, especially for a higher number of trials.

Let us demonstrate this method on an example. Consider an experiment described by model (2.1), where we have $v = 3$ treatments and in each time u we choose exactly one treatment. The effect of time trend is assumed to be cubic. Thus we may model this effect by a series of polynomials p_0, p_1, p_2, p_3, where p_i is of degree i. We will choose discrete orthogonal polynomials, i.e. $\sum_{u=1}^{n} p_i(u)p_j(u) = 0$ for $i \neq j$. Furthermore we set $p_i(1) = 1$ for each i. These conditions fully determine the polynomials $p_0, \ldots, p_3$.

For $N = n = 20$ and $N = n = 150$ trials we generate exact designs using Theorem 2.16 in order to get high D-efficiency. The weights given by Theorem 2.16 are $1/3$ for each treatment. That translates to $20/3 \approx 6.7$ for $N = 20$ and $150/3 = 50$ trials for $N = 150$ performed under each treatment. In the first case we generated designs that assign 7 trials to two treatments and 6 trials to the remaining treatment (chosen at random). In the second case we generated designs that assign 50 trials to each treatment.

We plotted the histograms of 10^4 randomly generated designs for each case to determine the quality of these designs.

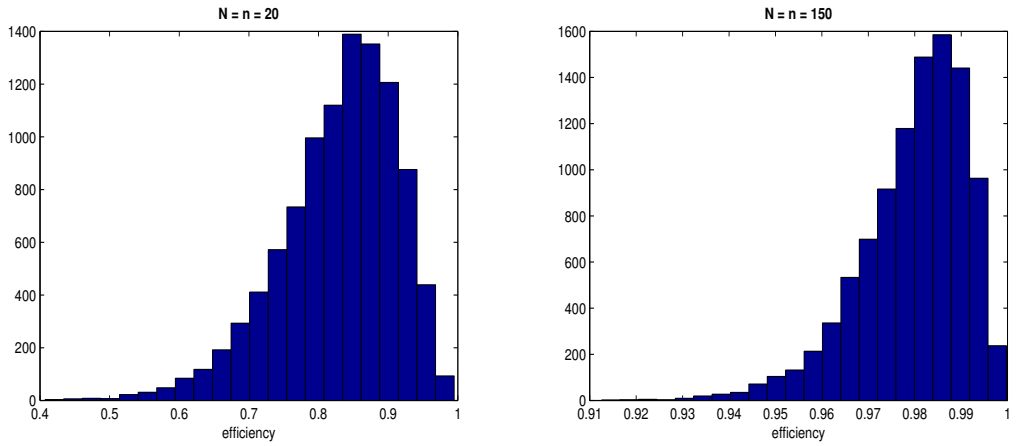

Figure 3: Histograms of efficiencies of 10^4 randomly generated designs using Theorem 2.16, for $N = 20$ and $N = 150$ trials. The designs tend to have high efficiency, especially in the case of $N = 150$ trials.

We observe that in both cases the designs tend to have high efficiency. Moreover the efficiencies of the randomly generated designs in the experiment with $N = 150$ trials are considerably higher than in the former one. This suggests that with the increasing number of trials the designs generated by our method tend to have increasing efficiency.

To support this hypothesis we used our method to generate designs for experiments with number of trials ranging from $N = 10$ to $N = 150$, aiming for high E-efficiency. For each N we used our method to generate 1 000 designs and we calculated their mean E-efficiencies. We used the same model as previously, i.e. we considered cubic time trend modelled by the discrete orthogonal polynomials. The results of these numerical experiments are summarized in Figure 4.

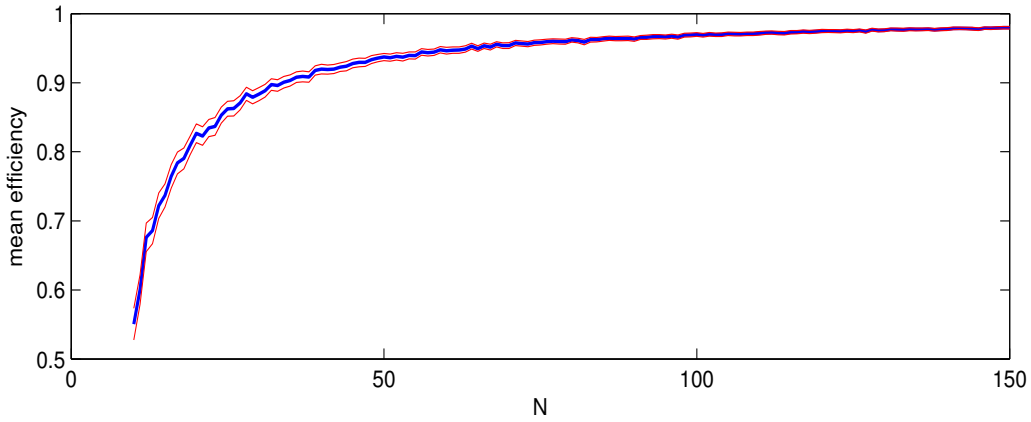

Figure 4: Mean E-efficiency of randomly generated designs using Theorem 2.16 (blue) with calculated confidence belt (red). For each $N = 10, \ldots, 150$ we generated 1000 designs. We note that the mean efficiency of generated designs is increasing with the increasing number of trials N.

In order to better represent the obtained data, we calculated confidence intervals for the mean efficiencies. We considered the confidence intervals for normal data, i.e. for each N the interval was calculated as

$$(\bar{\text{eff}} - \frac{uS}{\sqrt{M}}, \bar{\text{eff}} + \frac{uS}{\sqrt{M}}),$$

where $\bar{\text{eff}}$ is the mean efficiency, S^2 is the sample variance, $M = 1000$ is the number of generated designs and u is the $1 - 0.025/141$ quantile of the standard normal distribution. We estimated 141 efficiencies in total, therefore the value $1 - 0.025/141$ was chosen, so that we would get a 95% simultaneous confidence intervals (using the Bonferroni correction).

We clearly see that the mean efficiency is increasing with the number of trials. Moreover, in experiments with number of trials higher than 50 we obtained mean efficiency higher than 90%.

To summarize, we observed that the designs generated using our method tend to have high efficiency, especially in experiments with a higher number of trials.

This is a useful observation, because in experiments with a higher number of trials the algorithms providing exact optimal designs become time consuming. Using our results we may, in a time-frame of seconds or minutes, generate random designs until we get one with a reasonably high efficiency instead of running algorithms for a considerably higher amount of time.
Another approach would be using heuristics, but they too provide only nearly-optimal results. Moreover they may provide designs not efficient enough or they may be time consuming too.

We used these results to demonstrate the proposed method. We aimed to get efficient designs with respect to the A-optimality criterion ($p = -1$) in experiments with $v = 5$ treatments and exponential time trend. We set $h_1 \equiv 1$, $h_2(u) = u - 1$ and $h_3(u) = 1 + \mathrm{e}^{u/N}$. For multiple choices of N we generated exact designs, until we got one with at least 95% A-efficiency. For each N we calculated the time it took to generate a design efficient enough.
To compare our method we generated random designs from the entire set $\mathfrak{X}$ without using our knowledge on γ-designs, i.e. we generated the designs entirely at random. Table 2 summarizes the results of these two approaches.

We see that our method is considerably faster and as a by-product it provides higher efficiencies for higher number of trials.

	using Theorem 2.16		without Theorem 2.16	
N	time	eff_A	time	eff_A
15	1.5720	0.9512	51.8680	0.9577
20	0.0590	0.9690	0.3720	0.9637
25	0.0090	0.9735	1.9080	0.9582
30	0.0040	0.9564	1.6010	0.9584
40	0.0090	0.9730	1.1120	0.9516
50	0.0080	0.9621	0.0540	0.9640
60	0.0030	0.9823	0.1180	0.9536
70	0.0070	0.9722	0.0420	0.9549
80	0.0040	0.9940	0.3030	0.9606
90	0.0040	0.9771	0.1030	0.9532
100	0.0060	0.9935	0.1700	0.9711
125	0.0050	0.9619	0.0860	0.9545
150	0.0050	0.9895	0.3500	0.9520
175	0.0070	0.9873	0.2900	0.9772
200	0.0070	0.9933	0.8680	0.9523

Table 2: Generating 95% A-efficient designs: comparison of our method and the direct method. The first column contains the number of trials of the experiment we considered in the particular run. The next two columns consist of results from generating random designs using Theorem 2.16 and the last two columns contain results of generating random designs without using Theorem 2.16. The column *time* represents the total time (in seconds) it took to find a design at least 95% efficient and the column eff_A consists of the actual efficiencies of the found designs.

If we aimed for 99% efficiency, the quality of our method would become even more visible as can be seen in Table 3. It is analogous to Table 2, except we demanded to get a design with efficiency 99%.

Here it takes our method to provide a results less than a second and without using Theorem 2.16 it takes even hundreds of seconds.

	using Theorem 2.16		without Theorem 2.16	
N	time	eff_A	time	eff_A
50	0.2440	0.9917	667.8070	0.9935
75	0.0040	0.9925	5.8240	0.9914
100	0.0270	0.9908	207.4280	0.9906
125	0.0190	0.9969	8.3070	0.9914
150	0.0170	0.9906	140.4930	0.9917

Table 3: Generating 99% A-efficient designs: comparison of our method and the direct method. The notation is the same as in the Table 2.

4 Model With a General Nuisance Effect

4.1 Model

In Chapter 2 we examined a model with a nuisance time trend. We will show that our results hold in a more general model with a multidimensional nuisance effect, i.e. we will show that γ-designs that satisfy conditions analogous to the ones of Theorem 2.16 are ϕ_p-optimal for such model.

First, we will provide an example, where a multidimensional nuisance trend is present to show a motivation for the generalized model. Consider an agricultural experiment, where the researchers wish to improve the yields of maize by genetic modifications. From an original crop they created two varieties and they need to determine, whether it is reasonable to introduce these new varieties to the market. Thus they want to determine how much is their yield improved in comparison to the original crop.
The researchers wish to perform the experiment in a large field, which they split to smaller parts by a 8×4 grid. In each of the 32 smaller fields they will sow one of the three varieties: the original, or one of the genetically modified ones. However, the field is not perfectly flat and the experimenters suspect that the skewness of the field might have some unknown impact on the yields. Therefore they consider the effect of the skewness of the field to be a two-dimensional nuisance effect.

To formalize, the model is analogous to the one in Chapter 2, only instead of a time trend $u \in \mathbb{R}$ we have some r-dimensional trend $u \in \mathbb{R}^r$, $r \geq 1$, i.e. we have

$$Y_i = \tau_{t(i)} + \theta_1 h_1(u_1(i), \ldots, u_r(i)) + \ldots + \theta_d h_d(u_1(i), \ldots, u_r(i)) + \varepsilon_i, \quad i = 1, \ldots, N. \tag{4.1}$$

All the assumptions and notations are the same as in model (2.1), except that in addition to the treatment $t(i)$ we also choose other r experimental conditions $u(i) := \big(u_1(i), \ldots, u_r(i)\big)^T$, where the vectors u are assumed to belong to a finite set of all possible nuisance conditions $U \subseteq \mathbb{R}^r$. Then $h_1, \ldots, h_d$ are any regressors of the trend $h_k : U \to \mathbb{R}$.

Therefore, the set of all permissible experimental conditions is $\mathfrak{X} = \{1, \ldots, v\} \times U$ and an (approximate) design ξ is a function $\xi : \{1, \ldots, v\} \times U \to [0, 1]$, such that

$\sum_{t,u} \xi(t,u) = 1$. The value $\xi(t,u), t \in \{1,\dots,v\}, u \in U$, is the relative weight of trials to be performed under the conditions u and with the treatment t. For brevity, we will sometimes call $u \in U$ nuisance points, or simply points, instead of nuisance conditions.

In our example with genetically modified corn we would have $v = 3$ treatments t as the possible varieties of maize with their yields τ_t. The 32 experimental conditions $u = (u_1, u_2)^T$ would be the 32 smaller fields, labelled by their grid row and column number, thus we would have $U = \{1,\dots,8\} \times \{1,\dots,4\}$. The nuisance trend would be the skewness of the large field, approximated by some low-order polynomials.

The model (4.1) can be written as a special case of the general regression model (1.1), where $f(t,u) = \big(e_t^T, h_1(u), \dots, h_d(u)\big)^T$ and $\beta = \big(\tau_1,\dots,\tau_v,\theta_1,\dots,\theta_d\big)^T$. We remind the reader that the general model has the form

$$Y_i = f^T(x_i)\beta + \varepsilon_i, \quad i = 1,\dots,N.$$

We note that the notation $f(t,u)$ with r-dimensional u is not a correct one, it should be $f(t,u^T)$ instead. But the latter notation does not provide a better understanding of the situation, it is only needlessly complicated. Therefore we will use $f(t,u)$ even in case of multidimensional u.

Again, we assume the first treatment to be a control and we are interested in comparison of other treatments to the control, i.e. we aim to estimate the contrasts $\tau_2 - \tau_1, \dots, \tau_v - \tau_1$. Thus we have the same system of contrasts $Q^T\tau$, where Q is a $v \times (v-1)$ matrix

$$Q = (-1_{v-1}, I_{v-1})^T$$

and $\tau = (\tau_1,\dots,\tau_v)^T$. In the terms of Chapter 1, we aim to estimate the system $A^T\beta$, where $A = \big(Q^T, 0_{v-1\times d}\big)^T$ is a $(v+d) \times (v-1)$ matrix.

We may express the model (4.1) as

$$Y_i = \tau_{t(i)} + f_\theta(u(i)) + \varepsilon_i, \quad i = 1,\dots,N, \tag{4.2}$$

where $t \in \{1,\dots,v\}$, $u \in U \subseteq \mathbb{R}^r$ and $f_\theta(u) := \big(h_1(u),\dots,h_d(u)\big)^T$.

4.2 Results

The reader may observe that the model (4.2) is in its notation very similar to the original model (2.2). We will use this fact to provide all the auxiliary results without

proofs, because the proofs are analogous to the ones in Chapter 2: they did not depend on the fact that $u \in \mathbb{R}$; for a general $u \in U$ we get the same results.

Lemma 4.1. *Let ξ be a design for the model (4.2), then its moment matrix is*

$$M(\xi) = \begin{bmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{12}^T(\xi) & M_{22}(\xi) \end{bmatrix},$$

where

$$M_{11}(\xi) = \operatorname{diag}\left(\sum_{u \in U} \xi(1,u), \ldots, \sum_{u \in U} \xi(v,u) \right),$$

$$M_{12}(\xi) = \left(\sum_{u \in U} \xi(1,u) f_\theta(u), \ldots, \sum_{u \in U} \xi(v,u) f_\theta(u) \right)^T,$$

$$M_{22}(\xi) = \sum_{u \in U} \left(\sum_{t=1}^{v} \xi(t,u) \right) f_\theta(u) f_\theta^T(u).$$

We provide a similar characterization of the feasible designs, as in Chapter 2.

Lemma 4.2. *Let $A = \left(Q^T, 0_{s \times d}\right)^T$. Then $A^T\beta$ is estimable in the model (4.2) if and only if $\mathcal{S}(M_\tau) \subseteq \mathcal{S}(Q)$.*

Lemma 4.3. *Let $A = \left(Q^T, 0_{v-1 \times d}\right)^T$ and let $A^T\beta$ be estimable under a given design ξ. Then the information matrix $N_A(\xi)$ can be expressed as $N_A(\xi) = (Q^T M_\tau^-(\xi) Q)^{-1}$, where $M_\tau(\xi) = M_{11}(\xi) - M_{12}(\xi) M_{22}^-(\xi) M_{21}(\xi)$ is the Schur complement for $M(\xi)$.*

Lemma 4.4. *Let $A = \left(Q^T, 0\right)^T$, where Q is a $v \times s$ matrix of full column rank s. Then a regular reparametrization of nuisance parameters by changing f_θ to $\tilde{f}_\theta = R f_\theta$ in model (4.2) does not change the information matrix $N_A(\xi)$ for any design ξ.*

Again, we define elementary designs and γ-designs.

Definition 4.5. *For any $\gamma \in (0,1)$ let* γ-design *be any design ξ_γ that satisfies*

$$\xi_\gamma(1,u) = \gamma \sum_t \xi_\gamma(t,u) \text{ for any } u \in U \tag{4.3}$$

$$\xi_\gamma(2,u) = \ldots = \xi_\gamma(v,u) \text{ for any } u \in U \tag{4.4}$$

Definition 4.6. Elementary design *is any design that for some $j \in U$ satisfies*

$$\xi(t,j) > 0 \text{ for all } t \in \{1, \ldots, v\}, \tag{4.5}$$

$$\xi(t,u) = 0 \text{ for all } u \neq j \text{ and for all } t \in \{1, \ldots, v\}. \tag{4.6}$$

Hence γ-designs are designs, which under each nuisance condition u assign to the first (control) treatment proportion γ of the total weight assigned to the particular nuisance condition u. The rest of the weight is distributed uniformly among the rest of the treatments. Elementary designs are such designs that assign all weight to only one nuisance condition and furthermore they assign non-zero weight to each treatment. We note that an *elementary γ-design* is a design, which is both an elementary design and a γ-design.

Similar characterizations of elementary and γ-designs hold for the model (4.2) as for the model (2.1).

Proposition 4.7. *Let $\gamma \in (0,1)$. Then the elementary γ-design ξ that assigns all trials to a nuisance condition j attains the values*

$$\begin{aligned} &\xi(1,j) = \gamma, \quad \xi(2,j) = \ldots = \xi(v,j) = \frac{1-\gamma}{v-1} \text{ and} \\ &\xi(t,u) = 0 \text{ for } u \neq j \text{ and } t \in \{1,\ldots,v\}. \end{aligned} \tag{4.7}$$

Proposition 4.8. *Let $\gamma \in (0,1)$ and let ξ_γ be a γ-design. Then there exists a set of elementary γ-designs such that ξ_γ can be expressed as their convex combination.*

Lemma 4.9. *Let $f_\theta(j) \neq 0_d$ and let ξ be an elementary design that assigns all its weight to a nuisance point j. Then the Schur complement for its moment matrix is*

$$M_\tau(\xi) = \operatorname{diag}\Big(\xi(1,j), \ldots, \xi(v,j)\Big) - \Big(\xi(1,j),\ldots,\xi(v,j)\Big)^T \Big(\xi(1,j),\ldots,\xi(v,j)\Big).$$

Theorem 4.10. *Let $A = \big(Q^T, 0_{v-1\times d}\big)^T$, where $Q = (-1_{v-1}, I_{v-1})^T$ and let ξ be an elementary design. Then the system $A^T\beta$ is estimable under ξ.*

Corollary 4.11. *Let $\gamma \in (0,1)$ and let $A = \big(Q^T, 0_{v-1\times d}\big)^T$, where $Q = (-1_{v-1}, I_{v-1})^T$. Then any γ-design is feasible for the system $A^T\beta$.*

Lemma 4.12. *Let $\gamma \in (0,1)$ and let ξ be an elementary γ-design. Then its information matrix is $N_A(\xi) = \frac{1-\gamma}{v-1} I_{v-1} - \big(\frac{1-\gamma}{v-1}\big)^2 J_{v-1}$ with eigenvalues*

$$\lambda_1 = \frac{1-\gamma}{v-1} \quad \text{with multiplicity } v-1 \tag{4.8}$$

and

$$\lambda_2 = \frac{1-\gamma}{v-1} - \frac{(1-\gamma)^2}{v-1} = \frac{\gamma(1-\gamma)}{v-1} \quad \text{with multiplicity } 1. \tag{4.9}$$

Using the Equivalence Theorem 2.15 we will prove that these designs are ϕ_p-optimal for a suitably chosen γ.

Theorem 4.13. *Let* $A = \left(Q^T, 0_{v-1\times d}\right)^T$, *where* $Q = \left(-1_{v-1}, I_{v-1}\right)^T$. *Let* $p \in [-\infty, 0]$. *If* $p > -\infty$, *let* γ *be the unique solution of the equation (2.11)*

$$(v-2)\gamma^{1-p} + 2\gamma - 1 = 0$$

in the interval $(0, 1/2]$. *If* $p = -\infty$, *let* $\gamma = 1/2$. *Then any* γ*-design is* ϕ_p*-optimal for the estimation of the system* $A^T\beta$ *in model (4.2).*

Proof. The proof is in fact the same as the proof of Theorem 2.16, only instead of considering times $u \in \{1, \ldots, n\}$, we consider nuisance conditions $u \in U$.

Let $p \in [-\infty, 0]$. For $p < 0$ we set γ as the solution of (2.11) and for $p = 0$ we set $\gamma = 1/2$.

The proof consists of the same steps as the proof of Theorem 2.16:

1. We will reparametrize the nuisance time trend, so that the computations will be easier.

2. For $p \in (-\infty, 0]$ we will prove that one of the elementary γ-designs is ϕ_p-optimal.

3. We will prove that from the step 2. it follows, that all the γ-designs are ϕ_p-optimal.

4. We will repeat the steps 2. and 3. for $p = -\infty$.

5. Finally, we will prove optimality of γ-designs in a 'degenerate' situation with no time trend present.

For brevity, we will refer to proof of Theorem 2.16 simply as P2.16.

We will not provide the full proof, only emphasize that it works the same as P2.16.

1. Let there be j such that $f_\theta(j) \neq 0$. From now on, this j will be fixed. We will prove that the elementary γ-design that assigns all its trials to the nuisance condition j is ϕ_p-optimal. The reparametrization is possible without a loss of information, because of Lemma 4.4.

2. Now there exists an elementary γ-designs for any $u \in U$, $|U|$ in total, instead of n elementary γ-designs. However this obviously does not affect their properties: they have the same moment and information matrix as in P2.16, given by Lemma 4.1 and Lemma 4.12. We will use the same generalized inverse G as in P2.16 and the matrix A is obviously the same as before. Although the mapping $f(x) = f(t, u)$ has

different domain, it still attains the same values as previously. Thus all elements of the normality equality

$$f(x)^T GAN_A^{p+1}(\xi^*)A^T G^T f(x) \leq \text{trace} N_A^p(\xi^*) \text{ for all } x \in \mathcal{X},$$

are the same as in P2.16 and therefore it holds.

3. All elementary γ-designs for the fixed γ are ϕ_p-optimal, because they have the same information matrix and we proved that one of them is ϕ_p-optimal. Similarly as in P2.16, since any γ-design ξ_γ can be expressed as a convex combination of the elementary γ-designs, it is ϕ_p-optimal too.

4. The case of E-optimality is analogous to steps 2 and 3.

5. In the situation where $h_k(u) = 0$ for each k and for each u, i.e. there is no nuisance effect present, we get for any elementary γ-design again the same information matrix from Lemma 4.12. The moment matrix is the same as in the step 5 of P2.16 and following the same reasoning as therein, we get that even in this case, the γ-designs are ϕ_p-optimal.

□

Conclusion

In this work, we studied experimental design for a model of choosing treatments under the presence of a nuisance time trend. We considered a design of experiments for estimating treatment effects compared to the effect of a control treatment. For the studied model we provided a class of ϕ_p-optimal approximate designs for any $p \in [-\infty, 0]$.

In the first chapter, we introduced the theory of experimental design for a linear regression model so that even a reader not familiar with the statistical discipline of experimental design may understand the results that follow in the next chapters. We distinguished between an approximate and an exact design and further examined the former. We studied how the amount of information a design provides is measured: using the optimality criteria, where we have focused on the Kiefer's ϕ_p-optimality criteria. A further aim of this chapter was to clarify the notation that will be used later.

The main results of the work were stated in the second chapter. In this chapter, we examined the model (2.1) with the presence of a nuisance time trend. Then the experimental designs for such model were studied in detail. In Definitions 2.5 and 2.6, we defined two classes of designs - elementary and γ-designs. We used these two defined design classes and some additional lemmas to give a class of ϕ_p-optimal designs for any $p \in [-\infty, 0]$. The results are provided in Theorem 2.16. The key theorem that was used to prove the optimality of the obtained designs was the Equivalence Theorem (Theorem 2.15) as stated in [20].

In the following chapter, we demonstrated how an exact design may be constructed using our results. We showed that exact designs that approximately satisfy the proportions given by Theorem 2.15 tend to be efficient with respect to the ϕ_p-optimality criteria. Therefore we proposed that an efficient exact design may be obtained by generating random designs that approximately satisfy the given proportions, until a design efficient enough is found. We demonstrated on some examples that using such method a highly efficient exact design may be found rapidly. Furthermore, we noted that using the calculated optimal criterial values in Corollary 2.17, we may calculate the ϕ_p-efficiency of any given exact design. This allows us to accept an exact design that is efficient enough without the need to know an optimal exact design.

In the fourth chapter, we showed that our results hold for a generalization of the model (2.1), which allows for a general nuisance effect, e.g. a two-dimensional (space) trend, instead of simply a time trend. The results were summarized in Theorem 4.13 which is an analogy to Theorem 2.16.

The main contribution of this work were the results given by Theorem 2.16 and Theorem 4.13. We found out that designs belonging to a wide class may be ϕ_p-optimal (we denoted them as γ-designs). Then we provided an equation, which the optimal γ must satisfy for the γ-designs to be ϕ_p-optimal. Therefore, for any $p \in [-\infty, 0]$ we provided a class of ϕ_p-optimal approximate designs for estimating the studied treatment contrasts. Unlike designs orthogonal to time trend, our results hold for any nuisance time trend, not only for polynomials of particular degrees. Moreover, we showed that our results may be used to evaluate the quality of given exact designs or to obtain efficient exact designs.

In this thesis, we considered only experiments, where we were interested in comparing treatments to a control treatment. Hence, a possible extension of our work would be to examine ϕ_p-optimal approximate designs for estimating some other treatment contrasts. It would be interesting to know whether some similar results hold in such case.

Even designs for the set of contrasts that we considered may be analysed further. It is possible to extend our results by examining optimal approximate designs with respect to other optimality criteria. The possible future work includes determining whether the set of optimal designs that we provided is complete, i.e. whether there are other ϕ_p-optimal designs for estimating the considered contrasts. Furthermore, additional applications of our results in constructing exact designs may be studied.

Bibliography

[1] Atkinson, A. C., Donev, A. N.: *Experimental design optimally balanced for trend*, Technometrics 38(4) (1996), pp. 333-341

[2] Atkinson, A. C., Donev, A. N.: *Optimum Experimental Designs*, Oxford University Press, Oxford, 1992

[3] Atkinson, A. C., Bailey, R. A.: *One hundred years of the design of experiments on and off the pages of Biometrika*, Biometrika 88 (2001), pp. 53-97

[4] Bailey, R. A., Cheng, C.-S., Kipnis, P.: *Construction of Trend-Resistant Factorial Designs*, Statistica Sinica 2 (1992), 393-411

[5] Cheng, C.-S.: *Construction of Run Orders of Factorial Designs*, Statistical Design and Analysis of Industrial Experiments (1990), Marcel-Dekker, New York, pp. 423-439

[6] Cox, D.R.: *Some systematic experimental designs*, Biometrika 38(3/4) (1951), pp. 312-323

[7] Dey, A., Mukerjee, R.: *Fractional Factorial Plans*, John Wiley & Sons, New York, 1999

[8] Fisher, R.: *The design of experiments*, Oliver and Boyd, Edinburgh, 1935

[9] Gaffke, N.: *Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression*, Annals of Statistics 15 (1987), pp. 942-957

[10] Harman, R.: *Computing D-optimal Experimental Designs for Estimating Linear Combinations of Treatment Effects*, manuscript

[11] Harville, D. A.: *Matrix Algebra From A Statiscian's Perspective*, Springer-Verlag, New York, 1997

[12] Hedayat, A. S., Sloane, N. J. A., Stufken, J.: *Orthogonal Arrays: Theory and Applications*, Springer, New York, 1999

[13] Hotelling, H.: *Some improvements in weighing and other experimental techniques* The Annals of Mathematical Statistics, Volume 15 (1944), pp. 297-306

[14] John, P. W. M.: *Time Trends and Factorial Experiments*, Technometrics 32 (1990), pp. 275-282

[15] Kiefer, J.: *Optimum experimental designs*, Journal of Royal Statistical Society 21 (1959), pp. 272-319

[16] Kiefer, J.: *General equivalence theory for optimum designs (approximate theory)*, Annals of Statistics 2 (1974), pp. 849-879

[17] Mukerjee, R., Wu, C. F. J.: *A Modern Theory of Factorial Designs*, Springer, New York, 2006

[18] Pázman, A., Lacko, V.: *Prednášky z regresných modelov*, Comenius University, Bratislava, 2012

[19] Pázman, A.: *Foundation of Optimum Experimental Design*, Reidel Publ., Dordrecht, 1986

[20] Pukelsheim, F.: *Optimal design of experiments*, Classics in Applied Mathematics, SIAM, Philadelphia, 2006

[21] Raktoe, B.L., Hedayat, A.S., Federer, W.T.: *Factorial Designs*, John Wiley & Sons, New York, 1981

[22] Snell, E. J., Bryan-Jones, J.: *A design balanced for trend*, Biometrika 55 (1968), pp. 535-539

[23] 'Student': *Tables for estimating the probability that the mean of a unique sample of observations lies between -∞ and any given distance of the mean of the population from which the sample is drawn*, Biometrika 11 (1917), 414-417

[24] Zhang, F. (Ed.): *The Schur Complement and Its Applications*, Numerical Methods and Algorithms Vol. 4, Springer-Verlag, New York, 2005

Appendix: Some Selected Parts of Matrix Theory

In this work we use some concepts and results from Matrix Theory, which the reader does not have to be familiar with or which may need to be properly defined. We will introduce these concepts here.

A.1 Basic Matrix Properties

Definition A.1. *Let A be an $n \times n$ matrix. Then A is* symmetric *if it satisfies $A^T = A$. Furthermore if A is a symmetric $n \times n$ matrix, then it is* nonnegative definite *if $x^T Ax \geq 0$ for any $x \in \mathbb{R}^n$, and is* positive definitie *if $x^T Ax > 0$ for any $x \in \mathbb{R}^n, x \neq 0_n$.*

Proposition A.2. *Let A be a symmetric matrix. Then A is nonnegative definite if and only if its eigenvalues are nonnegative, and is positive definite if and only if its eigenvalues are positive.*

Proof. See [11]. □

Lemma A.3. *Let A, B be two $m \times n$ matrices. Then $\mathcal{S}(A) \subseteq \mathcal{S}(B)$ if and only if there exists an $n \times n$ matrix X, such that $A = BX$.*

Proof. Let us denote the columns of A and B as $A = \big(a_1, \ldots, a_n\big)$, $B = \big(b_1, \ldots, b_n\big)$. Then $\mathcal{S}(A) \subseteq \mathcal{S}(B)$ can be written as: for any $i = 1, \ldots, n$ there exist $x_1^{(i)}, \ldots, x_n^{(i)} \in \mathbb{R}$, such that $a_i = x_1^{(i)} b_1 + \ldots + x_n^{(i)} b_n$. That is equivalent to: for any $i = 1, \ldots, n$ there exists $x^{(i)} \in \mathbb{R}^n$, such that $a_i = Bx^{(i)}$. Therefore we get $\mathcal{S}(A) \subseteq \mathcal{S}(B)$ if and only if there exists X, such that $A = BX$. □

Lemma A.4 (by [20])**.** *Let A be a symmetric matrix partitioned as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^T & A_{22} \end{bmatrix}.$$

If A is nonnegative definite, then $\mathcal{S}(A_{12}^T) \subseteq \mathcal{S}(A_{22})$.

Proof. See [20]. □

Lemma A.5. *Let $A = aI_n + bJ_n$ for some $a, b \in \mathbb{R}$. Then A has eigenvalues $\lambda_1 = a$ with multiplicty $n - 1$ and $\lambda_2 = a + nb$ with multiplicty 1.*

Proof. Matrix $B = A - aI_n = bJ_n$ has rank 1 and thus has eigenvalue $\mu_1 = 0$ with multiplicity $n-1$. Then A has eignevalue $\lambda_1 = \mu_1 + a = a$ with multiplicty $n-1$. To find the last eigenvalue, with multiplicity 1, we observe that every column of A has the same sum of its elements: $a+nb$. Therefore the sum of all rows of matrix $A-(a+nb)I_n$ is a row full of zeros. That means, $A-(a+nb)I_n$ is a singular matrix, i.e. $\lambda_2 = a+nb$ is the last eigenvalue of A. □

A.2 Loewner ordering

Definition A.6 (by [20])**.** *We define the* Loewner ordering *as the partial ordering* $\succeq$ *of symmetric matrices*

$$A \succeq B \Leftrightarrow A - B \text{ is a nonnegative definite matrix.}$$

We emphasize that the Loewner ordering is a partial (as opposed to a total) ordering, i.e. there exist symmetric matrices A, B, which satisfy neither $A \succeq B$ nor $B \succeq A$. The book [20] provides a trivial example of such matrices

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \quad B = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

A.3 Generalized Inverse Matrices

Definition A.7 (by [11])**.** Generalized inverse matrix *of an* $m \times n$ *matrix* A *is any* $n \times m$ *matrix* A^- *that satisfies*

$$AA^-A = A.$$

We note that in general the generalized inverse of a given matrix is not uniquely defined. The reader may find more information on generalized inverse matrices in the book [11]. We will use some of their properties listed in [11].

Lemma A.8. *Let* A *be an* $m \times n$ *matrix,* B *a* $k \times m$ *matrix and* C *a* $n \times p$ *matrix. If* $\mathcal{S}(B^T) \subseteq \mathcal{S}(A^T)$ *and* $\mathcal{S}(C) \subseteq \mathcal{S}(A)$*, then* BA^-C *does not depend on the choice of generalized inverse* A^-*.*

Proof (by [11]). If $\mathcal{S}(B^T) \subseteq \mathcal{S}(A^T)$ and $\mathcal{S}(C) \subseteq \mathcal{S}(A)$, then $B = XA$ and $C = AY$ for some matrices X, Y. Then $BA^-C = XAA^-AY = XAY$ and thus it is invariant to the choice of A^-. □

Lemma A.9. *Let A be an $m \times n$ matrix and G be its generalized inverse; let B be an $m \times p$ matrix. Then there exists an $n \times p$ matrix X which satisfies the equation $AX = B$ if and only if $AGB = B$.*

Proof. Let there exist a solution X of $AX = B$. Multiplying the equation by AG we get $AGAX = AGB$, and thus $AX = AGB$. Since $AX = B$, we have $B = AGB$.

Now let $AGB = B$. By setting $X = GB$ we get $AX = B$. □

Lemma A.10. *Let A be an $n \times n$ matrix expressed in the block form*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

where A_{11} is an $q \times q$ matrix and A_{22} is a $r \times r$ matrix, $q + r = n$. Then
(i) if $\mathrm{rank}(A) = \mathrm{rank}(A_{11}) = q$, *then*

$$G_1 := \begin{bmatrix} A_{11}^{-1} & 0 \\ 0 & 0 \end{bmatrix}$$

is a generalized inverse of A.
(ii) if $\mathrm{rank}(A) = \mathrm{rank}(A_{22}) = r$, *then*

$$G_2 := \begin{bmatrix} 0 & 0 \\ 0 & A_{22}^{-1} \end{bmatrix}$$

is a generalized inverse of A.

Proof. The lemma is a special case of Theorem 9.2.3 in the textbook [11]. □

We may state a simple lemma concerning a generalized inverse of a matrix product.

Lemma A.11. *Let A be an $n \times n$ matrix and X and Y be $n \times n$ nonsingular matrices. Then the matrix $Y^{-1}A^{-}X^{-1}$ is a generalized inverse of XAY.*

Proof. Both Y^{-1} and X^{-1} exist, because X and Y are nonsingular. Let us denote $G := Y^{-1}A^{-}X^{-1}$, then

$$XAYGXAY = XAYY^{-1}A^{-}X^{-1}XAY = XAA^{-}AY = XAY,$$

thus $Y^{-1}A^{-}X^{-1}$ is a generalized inverse of XAY. □

A.4 Schur Complement

Definition A.12 (by [20]). *Let A be an arbitrary $m \times n$ matrix partitioned as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Then we call the matrix $A_\tau := A_{11} - A_{12}A_{22}^{-}A_{21}$ the Schur complement *of A_{22} in A.*

For brevity, we will use the notation Schur complement *for* A, instead of Schur complement of A_{22} in A. A detailed analysis of the Schur complement can be found in [24]. The Schur complement does not depend on the choice of generalized inverse A_{22}^{-} (see [20]).

Later, we will use the following lemma, which provides us with a generalized inverse of partitioned matrix, using Schur complement.

Lemma A.13 (by [11]). *Let A be an $m \times n$ matrix partitioned as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}.$$

Then the matrix

$$\begin{bmatrix} A_\tau^{-} & -A_\tau^{-}A_{12}A_{22}^{-} \\ -A_{22}^{-}A_{21}A_\tau^{-} & A_{22}^{-} + A_{22}^{-}A_{21}A_\tau^{-}A_{12}A_{22}^{-} \end{bmatrix}$$

is a generalized inverse of A

Proof. In [11]. □