



Equivariance and Invariance for Optimal Designs in Generalized Linear Models Exemplified by a Class of Gamma Models

Osama Idais¹ · Rainer Schwabe¹

Accepted: 30 July 2021 / Published online: 21 December 2021
© The Author(s) 2021

Abstract

The main intention of the present work is to outline the concept of equivariance and invariance in the design of experiments for generalized linear models and to demonstrate its usefulness. In contrast with linear models, pairs of transformations have to be employed for generalized linear models. These transformations act simultaneously on the experimental settings and on the location parameters in the linear component. Then, the concept of equivariance provides a tool to transfer locally optimal designs from one experimental region to another when the nominal values of the parameters are changed accordingly. The stronger concept of invariance requires a whole group of equivariant transformations. It can be used to characterize optimal designs which reflect the symmetries resulting from the group actions. The general concepts are illustrated by models with gamma distributed response and a canonical link. There, for a given transformation of the experimental settings, the transformation of the parameters is not unique and may be chosen to be nonlinear in order to fully exploit the model structure. In this case, we can derive invariant maximin efficient designs for the D - and the $IMSE$ -criterion.

Keywords Optimal design · Invariance · Equivariance · Maximin efficiency · Generalized linear model

This article is part of the topical collection “Special Issue: State of the art in research on design and analysis of experiments” guest edited by John Stufken, Abhyuday Mandal, and Rakhi Singh.

✉ Osama Idais
osama.idais@ovgu.de
Rainer Schwabe
rainer.schwabe@ovgu.de

¹ Institute for Mathematical Stochastics, Otto-von-Guericke University Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

1 Introduction

Generalized linear models are a powerful tool to analyze data for which the standard linear model approach is not adequate. The idea of generalized linear models goes back to Nelder and Wedderburn [28], and their concept is comprehensively presented in the monograph by McCullagh and Nelder [27]. The statistical analysis is well developed in generalized linear models, and there is also a considerable amount of literature on optimal design in this situation (see, e.g., Atkinson and Woods [3] and the literature cited therein).

In generalized linear models, the performance of a design depends not only on the experimental settings but, in contrast with linear models, also on the values of the underlying parameters. Even more crucially, the solutions for optimal designs also depend on the parameters. As those are commonly unknown at the design step, nominal values of these parameters have to be specified prior to the experiment which leads to the concept of locally optimal designs (see Chernoff [5]). This approach has been frequently employed (see [1, 13, 14, 36, 41–43] among others), and it provides at least a benchmark for the quality of a design.

To overcome the parameter dependence, robust criteria have been proposed which either impose a prior weight on the parameters (Bayesian design, see, e.g., Atkinson et al. [2], ch. 18) or choose a minimax approach over a parameter region of interest (maximin efficiency, see, e.g., [8, 10, 17, 22]).

The construction of optimal designs for generalized linear models is difficult, and often numerical algorithms are employed to find a solution. To reduce the complexity of the search for a good design, one can make use of symmetries (“invariance”) in the design problem which can be described by transformations of the experimental settings and the location parameters in the linear component. The concept of invariance or, more specifically, equivariance with respect to transformations has been used for long in statistical analysis and dates back to Pitman [30] (see, e.g., Lehmann [24, ch. 6], for a comprehensive description). The underlying idea of transformations which are conformable with the model (“reparameterization”) has been successfully adapted to optimal design theory in linear models. In contrast with equivariance in statistical analysis, the parameter values do not play a role in optimal designs for linear models. Therefore, only transformations of the experimental settings have to be considered there. With these transformations, optimal designs may be first determined on a standardized experimental region and then transferred to more general regions as long as the transformation is order preserving with respect to the design criterion (see, e.g., Heiligers and Schneider [19]). This covers, for example, the situation of D -optimal designs for polynomial regression on an arbitrary (multivariate) interval.

The stronger concept of invariance requires a whole group of equivariant transformations. In linear models, invariance has been widely used to characterize optimal designs which reflect the symmetries resulting from the group actions (see Pukelsheim [33, ch. 5], or Schwabe [38, ch. 3]). These groups of transformations may cover reflections and rotations for quantitative variables as well as permutations of levels and factors for categorical variables and combinations thereof. In the context of generalized linear models, however, the concept of invariance is not well established. This seems to be mainly due to the fact that local optimality criteria lack symmetries, in

general, because they depend on the parameter values. Therefore, we have to account for this dependence by also transforming the parameters similar to the situation in statistical analysis.

For the underlying concept of equivariance, we thus need a pair of transformations which acts simultaneously on the experimental settings and on the parameter values. The most famous representative of this concept is known as the canonical transformation defined in Ford et al. [13]. But the motivation there is different from our approach exhibited in Radloff and Schwabe [34]. The canonical transformation starts with a standardization of the nominal value of the parameters. This standardization is compensated by an associated transformation of the experimental settings which leaves the value of the linear component unchanged. In contrast with that, we start with a transformation of the experimental settings as in linear models. This transformation is conformable (“linearly equivariant,” see Schwabe [38, ch. 3]) with the regression functions in the sense that it results in a linear reparameterization of the linear component. This linear reparameterization might be the associated action on the parameter values as in the canonical transformation, but we allow for more general, even nonlinear, transformations of the parameters. Moreover, in its standard formulation, the canonical transformation deals with one quantitative explanatory variable with a straight line relationship for the linear component. A generalization of the canonical transformation to multiple explanatory variables is given in Sitter and Torsney [40]. In our approach, there is no restriction on the number of explanatory variables, on their impact on the linear component described by the regression functions, or on whether they are quantitative or categorical.

For the concept of invariance in generalized linear models, symmetries are also required in the parameters which concur with the symmetries in the experimental settings. This requirement is hardly met in the case of local optimality, but Bayesian or maximin efficiency criteria can incorporate symmetries in their prior or in their parameter region of interest (see Radloff and Schwabe [34]).

Based on this approach, we develop in the following the concept of equivariance and invariance in generalized linear models and their application to optimal designs step-by-step and illustrate each step by a running example of gamma models with canonical link functions. This kind of gamma model is chosen because it exhibits an additional scaling property which provides a more complex, nonlinear symmetry structure.

The paper is organized as follows. In Sect. 2, we introduce the model assumptions and the design criteria. In Sect. 3, we discuss the concept of equivariance under standard linear transformations of the parameters and show how optimal designs can be transferred from one experimental region to another. In Sect. 4, the concept of equivariance is extended to nonlinear transformations of the parameters. In Sect. 5, the general concept of invariance is introduced and optimal designs are obtained for various situations. Finally, Sect. 6 concludes the paper with a short discussion and an outlook.

2 Basics: Model Specification, Information, and Design

We consider a response variable Y for which the dependence on a (potentially multi-dimensional) covariate \mathbf{x} can be described by a generalized linear model. This means that the distribution of Y comes from a given exponential family and the mean $\mu = E(Y)$ is related to the linear component $\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} = \sum_{j=0}^{p-1} \beta_j f_j(\mathbf{x})$ by a one-to-one link function. In the linear component, $\mathbf{f}(\mathbf{x}) = (f_0(\mathbf{x}), \dots, f_{p-1}(\mathbf{x}))^T$ is a p -dimensional vector of given regression functions $f_0(\mathbf{x}), \dots, f_{p-1}(\mathbf{x})$ and $\boldsymbol{\beta} = (\beta_0, \dots, \beta_{p-1})^T$ is a p -dimensional vector of parameters $\beta_0, \dots, \beta_{p-1}$ to be estimated. Traditionally the link function maps the mean to the linear component (see McCullagh and Nelder [27, ch. 2]). For analytical purposes, however, it is more convenient to describe the dependence of the mean on the linear component,

$$\mu = \mu(\mathbf{x}; \boldsymbol{\beta}) = \eta(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}), \quad (1)$$

where η is the inverse of the link function. For example, for the log link η is the exponential and for the inverse link η is the reciprocal function.

As a particular case and for illustrative purposes, we consider gamma models. Such models are frequently used in engineering applications. For example, in Dette et al. [9] a gamma model is considered in a thermal spraying process. Further applications in the fields of ecology, medicine, and psychology can be found in Gea-Izquierdo and Cañellas [15], Grover et al. [18], and Ng and Cribbie [29]. In a gamma model, the response Y is gamma distributed. One possibility to parameterize its density is given by $f_Y(y) = y^{\kappa-1} \exp(-y/\theta) / (\theta^\kappa \Gamma(\kappa))$, where $\kappa > 0$ and $\theta > 0$ denote the shape and scale parameters, respectively. In this case, the expectation of Y is given by $\mu = \kappa\theta$. In order to end up with a one-parametric exponential family, we suppose that the shape parameter κ is a fixed nuisance parameter (see Atkinson and Woods [3]). For example, $\kappa = 1$ gives the family of exponential distributions, or for fixed integer κ one obtains a family of certain Erlang distributions.

For the link function, we assume the inverse link $\kappa/\mu = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$. Alternatively, the log link is frequently used in gamma models (see, e.g., Ford et al. [13]). However, the inverse link appears to be more suitable for illustrative purposes. Moreover, the inverse link is equal to the canonical link $-\kappa/\mu = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$ up to the minus sign (see McCullagh and Nelder [27, ch. 3]). This means that all subsequent results will also be valid for the canonical link, and the minus sign is suppressed for notational reasons.

The inverse η of the inverse link is given by

$$\eta(z) = \kappa/z, \quad (2)$$

which itself is equal to the inverse $-\kappa/z$ of the canonical link up to the minus sign. Then, the responses Y_i of a sample Y_1, \dots, Y_n with covariates $\mathbf{x}_1, \dots, \mathbf{x}_n$ are gamma distributed with means $\mu_i = \eta(\mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta})$ and common shape parameter κ .

In an experimental design setup, the covariates \mathbf{x}_i may be chosen by the experimenter from an experimental region \mathcal{X} over which the model under consideration is assumed to be valid. For gamma distributed responses, as an additional side condition,

the means μ_i have to be positive ($\mu(\mathbf{x}_i; \boldsymbol{\beta}) > 0$). This implies the natural restriction on the parameter region \mathcal{B} of potential values for the parameter vector $\boldsymbol{\beta}$ that for every $\boldsymbol{\beta} \in \mathcal{B}$ the linear component has to be positive ($\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta} > 0$) for all $\mathbf{x} \in \mathcal{X}$. Further note that for reasons of parameter identifiability the regression functions f_0, \dots, f_{p-1} are assumed to be linearly independent on the experimental region \mathcal{X} .

The aim of experimental designs is to optimize the performance of statistical analysis. The contribution of an observation Y_i to the performance is measured in terms of its information. In the present generalized linear models framework, for a single observation at an experimental setting \mathbf{x} the elemental information matrix is given by

$$\mathbf{M}(\mathbf{x}; \boldsymbol{\beta}) = \lambda(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}) \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T \quad (3)$$

(see Fedorov and Leonov [12] or Atkinson and Woods [3]) where λ is a positive valued function which is called the intensity function. Note that through the intensity function the elemental information depends on the parameter vector $\boldsymbol{\beta}$.

In generalized linear models, the intensity is given by

$$\lambda(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}) = \eta'(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})^2 / \text{Var}(Y). \quad (4)$$

In the case of a canonical link, we have $\text{Var}(Y) = \eta'(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})$ and the intensity reduces to the variance. In particular, in the gamma model with inverse link the intensity function is

$$\lambda(z) = \kappa / z^2, \quad (5)$$

because the minus sign in the inverse of the link function does not affect the intensity (cf. Gaffke et al. [14]). The (per experiment) Fisher information of n independent observations Y_i at experimental settings \mathbf{x}_i is then given by

$$\mathbf{M}(\mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\beta}) = \sum_{i=1}^n \mathbf{M}(\mathbf{x}_i; \boldsymbol{\beta}) = \sum_{i=1}^n \lambda(\mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}) \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T. \quad (6)$$

The aim of finding an exact optimal design $\mathbf{x}_1^*, \dots, \mathbf{x}_n^*$ is to optimize the Fisher information in a certain sense because the inverse is proportional to the asymptotic covariance matrix of the maximum likelihood estimator for $\boldsymbol{\beta}$ (see Fahrmeir and Kaufmann [11]).

As this discrete optimization problem is too difficult, in general, we will deal with approximate (continuous) designs ξ in the spirit of Kiefer [23] (see also Silvey [39, p. 15]) throughout the remainder of the present paper. An approximate design ξ is defined on the experimental region \mathcal{X} by mutually distinct support points $\mathbf{x}_1, \dots, \mathbf{x}_m$ and corresponding weights $w_1, \dots, w_m > 0$ such that $\sum_{i=1}^m w_i = 1$. In terms of an exact design, the support points \mathbf{x}_i may be interpreted as the distinct experimental settings and the weights w_i as their corresponding relative frequencies in the sample. The relaxation of an approximate design is then that the weights w_i may be chosen continuously and need not be multiples of $1/n$. The standardized (per observation)

information matrix of a design ξ is defined by

$$\mathbf{M}(\xi; \boldsymbol{\beta}) = \sum_{i=1}^m w_i \mathbf{M}(\mathbf{x}_i; \boldsymbol{\beta}) = \sum_{i=1}^m w_i \lambda(\mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}) \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T. \tag{7}$$

Design optimization is now concerned with finding an approximate design ξ^* which minimizes a convex real-valued criterion function Φ of the Fisher information $\mathbf{M}(\xi; \boldsymbol{\beta})$ of the design ξ . A design ξ^* will then be called Φ -optimal when it minimizes $\Phi(\xi)$, $\Phi(\xi^*) = \min \Phi(\xi)$. As the information matrix depends on the parameter vector $\boldsymbol{\beta}$, the obtained design ξ^* is locally Φ -optimal at a given parameter value $\boldsymbol{\beta}$ [5] and may change with $\boldsymbol{\beta}$. To avoid the parameter dependence, so-called robust versions of the criteria can be considered like ‘‘Bayesian’’ criteria which involve a weighting measure (‘‘prior’’) on the parameters (see Atkinson et al. [2, ch. 18]) or ‘‘minimax’’ criteria which aim at minimizing the worst case scenario for the parameter settings (see the ‘‘standardized minimax’’ criteria in [8]). In the following, we will focus on the local D - and $IMSE$ -criteria and the corresponding maximin efficiency (‘‘standardized maximin’’) criteria.

The D -criterion is the most commonly used design criterion. It is related to the estimation of the model parameters $\boldsymbol{\beta}$ and aims at minimizing the determinant of the asymptotic covariance matrix, $\Phi(\mathbf{M}) = \det(\mathbf{M}^{-1})$ for positive definite information matrix \mathbf{M} , and $\Phi(\mathbf{M}) = \infty$ for singular \mathbf{M} . A design ξ^* is then called locally D -optimal at $\boldsymbol{\beta}$ when $\det(\mathbf{M}(\xi^*; \boldsymbol{\beta})^{-1}) = \min \det(\mathbf{M}(\xi; \boldsymbol{\beta})^{-1})$. The D -criterion can be motivated by the fact that it measures the (squared) volume of the asymptotic confidence ellipsoid of the maximum likelihood estimator for $\boldsymbol{\beta}$. However, its popularity predominantly stems from its nice analytic properties.

Note that in the present situation the property of $\mathbf{M}(\xi; \boldsymbol{\beta})$ being nonsingular does not depend on the value of the parameter vector $\boldsymbol{\beta}$ because the intensity $\lambda(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})$ is greater than zero for all $\mathbf{x} \in \mathcal{X}$ and all $\boldsymbol{\beta} \in \mathcal{B}$.

The definition of the $IMSE$ -criterion (alternatively, also called I -, V - or Q -optimality in the literature) is based on the estimation (prediction) of the mean response $\mu(\mathbf{x}; \boldsymbol{\beta})$. It aims at minimizing the average asymptotic variance of the predicted mean response $\hat{\mu}(\mathbf{x}) = \mu(\mathbf{x}; \hat{\boldsymbol{\beta}})$, where averaging is taken with respect to a standardized measure ν on \mathcal{X} (see Li and Deng [25,26]). For a generalized linear model, the asymptotic variance is given by

$$\text{asVar}(\hat{\mu}(\mathbf{x}); \xi, \boldsymbol{\beta}) = \eta'(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})^2 \mathbf{f}(\mathbf{x})^T \mathbf{M}(\xi; \boldsymbol{\beta})^{-1} \mathbf{f}(\mathbf{x}), \tag{8}$$

for all $\mathbf{x} \in \mathcal{X}$. For a canonical link, we have $\lambda = \eta'$ and hence

$$\text{asVar}(\hat{\mu}(\mathbf{x}); \xi, \boldsymbol{\beta}) = \lambda(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})^2 \mathbf{f}(\mathbf{x})^T \mathbf{M}(\xi; \boldsymbol{\beta})^{-1} \mathbf{f}(\mathbf{x}). \tag{9}$$

The integrated mean-squared error (IMSE) is then defined as the average prediction variance

$$\begin{aligned} \text{IMSE}(\xi; \beta, \nu) &= \int \text{asVar}(\hat{\mu}(\mathbf{x}); \xi, \beta) \nu(d\mathbf{x}) \\ &= \int \lambda(\mathbf{f}(\mathbf{x})^T \beta)^2 \mathbf{f}(\mathbf{x})^T \mathbf{M}(\xi; \beta)^{-1} \mathbf{f}(\mathbf{x}) \nu(d\mathbf{x}) \end{aligned} \tag{10}$$

with respect to a given standardized measure ν on the experimental region \mathcal{X} ($\nu(\mathcal{X}) = 1$).

By a standard method to express the *IMSE*-criterion (see, e.g., Li and Deng [26]), the asymptotic variance can be rewritten as

$$\text{asVar}(\hat{\mu}(\mathbf{x}); \xi, \beta) = \text{trace}(\lambda(\mathbf{f}(\mathbf{x})^T \beta)^2 \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T \mathbf{M}(\xi; \beta)^{-1}). \tag{11}$$

Hence, the IMSE is given by

$$\text{IMSE}(\xi; \beta, \nu) = \text{trace}(\mathbf{V}(\beta; \nu) \mathbf{M}(\xi; \beta)^{-1}), \tag{12}$$

where

$$\mathbf{V}(\beta; \nu) = \int \lambda(\mathbf{f}(\mathbf{x})^T \beta)^2 \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T \nu(d\mathbf{x}) \tag{13}$$

denotes a weighted ‘‘moment’’ matrix with respect to the measure ν . Note that the leading term under the integral in $\mathbf{V}(\beta; \nu)$ differs from that in the virtual information matrix $\mathbf{M}(\nu; \beta)$ by replacing the intensity λ by λ^2 . Moreover, in contrast with the *D*-criterion, the *IMSE*-criterion does not solely depend on the information matrix $\mathbf{M}(\xi; \beta)$, but also depends through the weighting matrix $\mathbf{V}(\beta; \nu)$ explicitly on the parameter vector β and additionally on the measure ν as a supplementary argument. The *IMSE*-criterion is thus defined by $\Phi(\mathbf{M}; \beta, \nu) = \text{trace}(\mathbf{V}(\beta; \nu) \mathbf{M}^{-1})$. A design ξ^* is then called locally *IMSE*-optimal with respect to ν at β when $\text{trace}(\mathbf{V}(\beta; \nu) \mathbf{M}(\xi^*; \beta)^{-1}) = \min \text{trace}(\mathbf{V}(\beta; \nu) \mathbf{M}(\xi; \beta)^{-1})$.

To avoid the parameter dependence of an optimal design under local criteria, we will also consider as a ‘‘robust’’ alternatives maximin efficiency criteria which are also called standardized optimality criteria (see Dette et al. [10]). For this, we first have to introduce the concept of efficiency. Let the local criterion Φ_β at β depend homogeneously on the information matrix, i.e., $\Phi_\beta(\xi) = \phi(\mathbf{M}(\xi; \beta))$ for some function ϕ on the set of positive definite matrices satisfying $\phi(c\mathbf{M}) = c^{-1}\phi(\mathbf{M})$ for $c > 0$ (cf Pukelsheim [33], ch. 5, for the related concept of information functions). Then, the efficiency of a design ξ (locally at β) is defined by

$$\text{eff}(\xi; \beta) = \frac{\Phi_\beta(\xi_\beta^*)}{\Phi_\beta(\xi)},$$

where ξ_β^* is the Φ -optimal design (locally at β). Maximin efficiency then aims at maximizing the worst efficiency $\inf_{\beta \in \mathcal{B}'} \text{eff}_\phi(\xi; \beta)$ over a given subset \mathcal{B}' of interest

of the parameter region \mathcal{B} . In order to arrive at a minimization problem, we define the maximin efficiency criterion by the inverse relation

$$\Phi(\xi) = \sup_{\beta \in \mathcal{B}'} \frac{\Phi_{\beta}(\xi)}{\Phi_{\beta}(\xi_{\beta}^*)}. \tag{14}$$

Note that Φ is convex if, for all β , the local criteria Φ_{β} are convex.

For maximin D -efficiency, we have to choose the homogeneous version $\Phi_{\beta}(\xi) = (\det(\mathbf{M}(\xi; \beta)))^{-1/p}$ of the local D -criterion (see [33, ch. 6]) to get the maximin D -efficiency criterion

$$\Phi_{D-ME}(\xi) = \sup_{\beta \in \mathcal{B}'} \left(\frac{\det(\mathbf{M}(\xi; \beta))}{\det(\mathbf{M}(\xi_{\beta}^*; \beta))} \right)^{-1/p},$$

where ξ_{β}^* denotes the locally D -optimal design at β . The D -efficiency can then be interpreted as the proportion of observations required under the D -optimal design ξ_{β}^* to obtain the same value of the determinant as for design ξ . For example, an efficiency of 0.5 means that with a D -optimum design ξ_{β}^* only half as many observations as for ξ are necessary to get the same precision. A design ξ^* is then called maximin D -efficient on \mathcal{B}' when $\Phi_{D-ME}(\xi^*) = \min \Phi_{D-ME}(\xi)$.

The local $IMSE$ -criterion is already homogeneous because it is a linear criterion. Thus, the maximin $IMSE$ -efficiency criterion can be defined directly as

$$\Phi_{IMSE-ME}(\xi; \nu) = \sup_{\beta \in \mathcal{B}'} \frac{IMSE(\xi, \beta, \nu)}{IMSE(\xi_{\beta}^*, \beta, \nu)},$$

where ξ_{β}^* denotes the locally $IMSE$ -optimal design at β . A design ξ^* is then called maximin $IMSE$ -efficient with respect to ν on \mathcal{B}' when $\Phi_{IMSE-ME}(\xi^*; \nu) = \min \Phi_{IMSE-ME}(\xi; \nu)$.

In particular, for the gamma model with inverse link we have $\lambda(z) = \kappa/z^2$ (see (5)) which implies that

$$\mathbf{M}(\xi; \beta) = \sum_{i=1}^m w_i \kappa (\mathbf{f}(\mathbf{x}_i)^T \beta)^{-2} \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T \tag{15}$$

and

$$\mathbf{V}(\beta; \nu) = \int \kappa^2 (\mathbf{f}(\mathbf{x})^T \beta)^{-4} \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T \nu(d\mathbf{x}). \tag{16}$$

Hence, in both the D - and the $IMSE$ -criterion the shape parameter κ occurs only as a factor which does not affect the optimization problem. Without loss of generality, we may thus assume $\kappa = 1$ in the remainder of the text.

3 Equivariance

Invariance and equivariance play an important role for optimal design in linear models. However, these concepts can also be applied in the context of generalized linear models as established in Radloff and Schwabe [34].

The essential idea of equivariance in the design setup is to transfer an already known optimal design on a given (standardized) experimental region to another experimental region of interest by a suitable transformation while keeping the model structure unchanged. The most prominent approach of this kind is the method of canonical transformation propagated by Ford et al. [13].

Throughout we accompany each conceptual step by a simple running example (Example 1). We start with a one-to-one transformation $g : \mathcal{X} \rightarrow \mathcal{Z}$ which maps the experimental region \mathcal{X} onto a potentially different region \mathcal{Z} .

Example 1 Let $\mathcal{X} = [0, 1]$ be the one-dimensional standard unit interval and $\mathcal{Z} = [a, b]$ another non-degenerate interval, $b > a$. Then, the shift and scale transformation $g(x) = a + cx$, where $c = b - a$, maps \mathcal{X} onto \mathcal{Z} . □

The next ingredient connects the transformation g with the vector of regression functions: \mathbf{f} is said to be linearly equivariant with respect to g if there exists a (nonsingular) matrix \mathbf{Q}_g such that $\mathbf{f}(g(\mathbf{x})) = \mathbf{Q}_g \mathbf{f}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$, which will be assumed to hold throughout the remainder of this text.

Example (Example 1 continued) Let $\mathbf{f}(x) = (1, x)^T$ be the vector of regression functions for a simple one-dimensional linear regression, $p = 2$, such that the linear component is $\mathbf{f}(x)^T \boldsymbol{\beta} = \beta_0 + \beta_1 x$. Then, for $g(x) = a + cx$ the transformation matrix \mathbf{Q}_g is given by

$$\mathbf{Q}_g = \begin{pmatrix} 1 & 0 \\ a & c \end{pmatrix}.$$

□

In contrast with the situation in linear models, additionally a transformation $\tilde{g} : \mathcal{B} \rightarrow \tilde{\mathcal{B}}$ of the parameter vector $\boldsymbol{\beta}$ is required in the present setup of generalized linear models. This approach of equivariance with respect to a pair (g, \tilde{g}) of transformations of the settings \mathbf{x} and the parameters $\boldsymbol{\beta}$, respectively, is in accordance with the general concept of equivariance in statistical analysis (see, e.g., Lehmann [24, ch. 6]).

A natural choice for the transformation \tilde{g} is a reparameterization which leaves the value of the linear component unchanged, $\mathbf{f}(g(\mathbf{x}))^T \tilde{g}(\boldsymbol{\beta}) = \mathbf{f}(\mathbf{x})^T \boldsymbol{\beta}$ for all $\mathbf{x} \in \mathcal{X}$. This is accomplished by setting $\tilde{g}(\boldsymbol{\beta}) = \mathbf{Q}_g^{-T} \boldsymbol{\beta}$, where “ \cdot^{-T} ” denotes the inverse of a transposed matrix, and $\tilde{\mathcal{B}} = \tilde{g}(\mathcal{B})$. For convenience, we denote $\tilde{g}(\boldsymbol{\beta})$ by $\tilde{\boldsymbol{\beta}} = (\tilde{\beta}_0, \dots, \tilde{\beta}_{p-1})^T$.

Example (Example 1 continued) For $g(x) = a + cx$ and simple linear regression $\mathbf{f}(x) = (1, x)^T$, the transformation matrix for the parameter vector is

$$\mathbf{Q}_g^{-T} = \begin{pmatrix} 1 & -a/c \\ 0 & 1/c \end{pmatrix},$$

and the transformation $\tilde{g}(\boldsymbol{\beta}) = \mathbf{Q}_g^{-T} \boldsymbol{\beta}$ results in $\tilde{\beta}_0 = \beta_0 - a\beta_1/c$ and $\tilde{\beta}_1 = \beta_1/c$. If for given value of $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$, the pair (g, \tilde{g}) is chosen in such a way that $\tilde{\boldsymbol{\beta}} = (0, 1)^T$, i.e., $c = \beta_1$ and $a = \beta_0$, then g represents essentially the canonical transformation used in Ford et al. [13].

For the gamma model, the parameter region \mathcal{B} is restricted by the constraint that the linear component $\mathbf{f}(x)^T \boldsymbol{\beta} = \beta_0 + \beta_1 x$ is positive for all $x \in \mathcal{X} = [0, 1]$. Hence, the maximal parameter region is $\mathcal{B} = \{\boldsymbol{\beta}; \beta_0 > 0, \beta_1 > -\beta_0\}$ which is displayed in Fig. 2 in Sect. 4. The transformed parameter region is then $\tilde{\mathcal{B}} = \tilde{g}(\mathcal{B}) = \{\boldsymbol{\beta}; \tilde{\beta}_0 + \tilde{\beta}_1 a > 0, \tilde{\beta}_0 + \tilde{\beta}_1 b > 0\}$. In particular, to obtain the symmetric unit interval $\mathcal{Z} = [-1, 1]$ as secondary experimental region, the transformation $g(x) = 2x - 1$ is to be chosen with $a = -1$ and $c = 2$, and the transformed parameter region becomes $\tilde{\mathcal{B}} = \{\boldsymbol{\beta}; |\tilde{\beta}_1| < \tilde{\beta}_0\}$. \square

Note that for each pair (g, \tilde{g}) of transformations the mean response and the intensity remain unchanged, $\mu(g(\mathbf{x}); \tilde{g}(\boldsymbol{\beta})) = \mu(\mathbf{x}; \boldsymbol{\beta})$ and $\lambda(\mathbf{f}(g(\mathbf{x}))^T \tilde{g}(\boldsymbol{\beta})) = \lambda(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})$. Having this in mind, we study how these transformations act on a design and its information matrix: For a design ξ with support points \mathbf{x}_i and corresponding weights $w_i, i = 1, \dots, m$, we denote by ξ^g its image under the transformation g , i.e., ξ^g has support points $\mathbf{z}_i = g(\mathbf{x}_i)$ with weights $w_i, i = 1, \dots, m$, respectively, and is hence a design on \mathcal{Z} . Then, for the associated information matrices we obtain

$$\begin{aligned} \mathbf{M}(\xi^g; \tilde{g}(\boldsymbol{\beta})) &= \sum_{i=1}^m w_i \lambda(\mathbf{f}(g(\mathbf{x}_i))^T \tilde{g}(\boldsymbol{\beta})) \mathbf{f}(g(\mathbf{x}_i)) \mathbf{f}(g(\mathbf{x}_i))^T \\ &= \sum_{i=1}^m w_i \lambda(\mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}) \mathbf{Q}_g \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T \mathbf{Q}_g^T \\ &= \mathbf{Q}_g \left(\sum_{i=1}^m w_i \lambda(\mathbf{f}(\mathbf{x}_i)^T \boldsymbol{\beta}) \mathbf{f}(\mathbf{x}_i) \mathbf{f}(\mathbf{x}_i)^T \right) \mathbf{Q}_g^T = \mathbf{Q}_g \mathbf{M}(\xi; \boldsymbol{\beta}) \mathbf{Q}_g^T \end{aligned} \tag{17}$$

(see Radloff and Schwabe [34]). In short, the pair (g, \tilde{g}) of simultaneous transformations induces the transformation $\mathbf{M}(\xi; \boldsymbol{\beta}) \rightarrow \mathbf{Q}_g \mathbf{M}(\xi; \boldsymbol{\beta}) \mathbf{Q}_g^T$ of the information matrix.

Example (Example 1 continued) Let ξ be supported on the endpoints $x_1 = 0$ and $x_2 = 1$ of the experimental region $\mathcal{X} = [0, 1]$ with corresponding weights $w_1 = 1 - w$ and $w_2 = w$, respectively. For the gamma model with simple linear regression, $\mathbf{f}(x) = (1, x)^T$, denote by $\lambda_0 = \lambda(\beta_0)$ and $\lambda_1 = \lambda(\beta_0 + \beta_1)$ the intensities at the support points 0 and 1. The information matrix of ξ is given by

$$\mathbf{M}(\xi; \boldsymbol{\beta}) = \begin{pmatrix} (1 - w)\lambda_0 + w\lambda_1 & w\lambda_1 \\ w\lambda_1 & w\lambda_1 \end{pmatrix}.$$

For $g(x) = a + cx$, the induced design ξ^g is supported on the endpoints $z_1 = a$ and $z_2 = b$ of the induced experimental region $\mathcal{Z} = [a, b]$ with weights $1 - w$ at a and w

at b . Under $\tilde{\boldsymbol{\beta}} = \tilde{g}(\boldsymbol{\beta})$, the intensities at a and b are λ_0 and λ_1 , respectively, and the information matrix of ξ^g is

$$\mathbf{M}(\xi^g; \tilde{g}(\boldsymbol{\beta})) = \begin{pmatrix} (1-w)\lambda_0 + w\lambda_1 & (1-w)\lambda_0a + w\lambda_1b \\ (1-w)\lambda_0a + w\lambda_1b & (1-w)\lambda_0a^2 + w\lambda_1b^2 \end{pmatrix} = \mathbf{Q}_g \mathbf{M}(\xi; \boldsymbol{\beta}) \mathbf{Q}_g^T.$$

□

The final step is the equivariance of the criterion Φ . In analogy to the terminology in Heiligers and Schneider [19] for linear models, we will call a convex optimality criterion Φ equivariant with respect to a transformation g , if Φ preserves the ordering under the transformation g , i.e., for any two designs ξ_1 and ξ_2 the relation $\Phi(\xi_1) \leq \Phi(\xi_2)$ implies $\Phi(\xi_1^g) \leq \Phi(\xi_2^g)$.

In the present situation of generalized linear models, more care has to be taken, since in addition the parameter vector $\boldsymbol{\beta}$ and eventually some supplementary arguments have to be changed in the criterion during the transformation. We therefore introduce a second criterion function $\Phi' = \Phi_{g, \tilde{g}}$ for the designs on \mathcal{Z} which may depend on the transformations g and \tilde{g} . Then, we will call a pair of criteria Φ and Φ' equivariant with respect to the pair (g, \tilde{g}) of transformations, when the ordering is preserved, i.e., the relation $\Phi(\xi_1) \leq \Phi(\xi_2)$ implies $\Phi'(\xi_1^g) \leq \Phi'(\xi_2^g)$.

With these definitions, we obtain the following result that in the case of equivariance the optimality of designs is preserved under transformations.

Theorem 1 *Let the pair of criteria Φ and Φ' be equivariant with respect to the pair (g, \tilde{g}) of transformations. If ξ^* is Φ -optimal, then its image $(\xi^*)^g$ is Φ' -optimal.*

We will now establish that the D - and $IMSE$ -criteria are equivariant, if simultaneously the parameter vector $\boldsymbol{\beta}$ and potential supplementary arguments are transformed. By (17), we obtain for the D -criterion

$$\det(\mathbf{M}(\xi^g; \tilde{g}(\boldsymbol{\beta}))^{-1}) = \det(\mathbf{Q}_g)^{-2} \det(\mathbf{M}(\xi; \boldsymbol{\beta})^{-1}). \tag{18}$$

Let Φ be the local D -criterion at $\boldsymbol{\beta}$ and Φ' be the local D -criterion at $\tilde{g}(\boldsymbol{\beta})$, then the D -criterion is equivariant under simultaneous transformation of $\boldsymbol{\beta}$, and by Theorem 1 the locally D -optimal design can be transferred.

Corollary 1 *If ξ^* is locally D -optimal on \mathcal{X} at $\boldsymbol{\beta}$, then $(\xi^*)^g$ is locally D -optimal on \mathcal{Z} at $\tilde{\boldsymbol{\beta}} = \tilde{g}(\boldsymbol{\beta})$.*

Example (Example 1 continued) For the gamma model with simple linear regression, $\mathbf{f}(x) = (1, x)^T$, the locally D -optimal design ξ^* on the unit interval $\mathcal{X} = [0, 1]$ is supported by the endpoints $x_1 = 0$ and $x_2 = 1$ and assigns equal weights $w^* = 1/2$ to these endpoints for any value of the parameter vector $\boldsymbol{\beta} \in \mathcal{B}$ (see Gaffke et al. [14]). Then, for any other interval $\mathcal{Z} = [a, b]$ as the experimental region we may consider the transformation $g(x) = a + cx$, $c = b - a$, together with $\tilde{g}(\boldsymbol{\beta}) = \mathbf{Q}_g^{-T} \boldsymbol{\beta}$. By Corollary 1, the design $(\xi^*)^g$ which assigns equal weights $w^* = 1/2$ to the endpoints $z_1 = a$ and $z_2 = b$ of the experimental region \mathcal{Z} is locally D -optimal for any value of the parameter vector $\tilde{\boldsymbol{\beta}} = \tilde{g}(\boldsymbol{\beta}) \in \tilde{\mathcal{B}} = \tilde{g}(\mathcal{B})$. □

In the situation of Example 1, the locally D -optimal design does not depend on the parameter β . This will typically not hold true, if the underlying model for the linear component becomes more complex.

Example 2 We consider the gamma model with the linear component $\mathbf{f}(\mathbf{x})^T \beta = \beta_0 + \beta_1 x_1 + \beta_2 x_2$, that is multiple linear regression of two covariates, $\mathbf{x} = (x_1, x_2)^T$, where $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2)^T$, $p = 3$, with the unit square $\mathcal{X} = [0, 1]^2$ as the experimental region. Denote by $\mathbf{x}_1 = (0, 0)^T$, $\mathbf{x}_2 = (1, 0)^T$, $\mathbf{x}_3 = (0, 1)^T$ and $\mathbf{x}_4 = (1, 1)^T$ the vertices of \mathcal{X} . The parameter region \mathcal{B} is the set of all parameter vectors $\beta = (\beta_0, \beta_1, \beta_2)^T$ such that the linear component at $\mathbf{x}_1, \dots, \mathbf{x}_4$ is positive, i.e., $\beta_0 > 0$, $\beta_0 + \beta_1 > 0$, $\beta_0 + \beta_2 > 0$, and $\beta_0 + \beta_1 + \beta_2 > 0$. This region, depicted in the left panel of Fig. 1, constitutes a cone in the three-dimensional Euclidean space.

According to Burrige and Sebastiani [4], the minimally supported design ξ^* which assigns equal weights $w_i^* = 1/3$ to the support points \mathbf{x}_i , $i = 1, 2, 3$, is locally D -optimal at β , when β satisfies $\beta_0^2 - \beta_1 \beta_2 \leq 0$. The subset \mathcal{B}_1 of these β in \mathcal{B} is shown in the right panel of Fig. 1.

Now equivariance can be used to find D -optimal designs for other parameter values different from those in \mathcal{B}_1 . For this, we use transformations which map the experimental region onto itself, $\mathcal{Z} = \mathcal{X}$:

$$g_2(\mathbf{x}) = (1 - x_1, 1 - x_2)^T, \quad g_3(\mathbf{x}) = (1 - x_1, x_2)^T \quad \text{and} \quad g_4(\mathbf{x}) = (x_1, 1 - x_2)^T. \tag{19}$$

Here g_3 and g_4 represent the reflection with respect to the first and second covariate x_1 and x_2 , respectively, and g_2 is the simultaneous reflection with respect to both covariates. Alternatively, g_2 can also be described as a rotation by 180 degree. We also introduce $g_1 = \text{id}$ as the identity mapping.

The regression function $\mathbf{f}(\mathbf{x}) = (1, x_1, x_2)^T$ is linearly equivariant with respect to these transformations with corresponding matrices

$$\mathbf{Q}_{g_2} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{pmatrix}, \quad \mathbf{Q}_{g_3} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Q}_{g_4} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & -1 \end{pmatrix}.$$

For each g_k , $k = 2, 3, 4$, the corresponding parameter transformation is given by $\tilde{g}_k(\beta) = \mathbf{Q}_{g_k}^T \beta$. Because g_k maps the experimental region \mathcal{X} onto itself, also the related transformation \tilde{g}_k maps the parameter regions onto itself, $\tilde{\mathcal{B}} = \mathcal{B}$.

Starting from the parameter subregion \mathcal{B}_1 , where the design ξ^* is locally D -optimal, we can define parameter subregions $\mathcal{B}_k = \tilde{g}_k(\mathcal{B}_1)$ induced by the transformations g_k , $k = 2, 3, 4$. These subregions are characterized explicitly in Table 1 by the inequalities in the last column, and they are also shown in the right panel of Fig. 1. All these subregions constitute cones. Now by equivariance we can conclude that the designs $\xi_k^* = (\xi^*)^{g_k}$ are locally D -optimal at β for $\beta \in \mathcal{B}_k$. The results are explicitly stated in Table 1.

Note that the same optimal designs have been obtained before in Idais [20] by a straightforward application of the celebrated Kiefer–Wolfowitz equivalence theorem

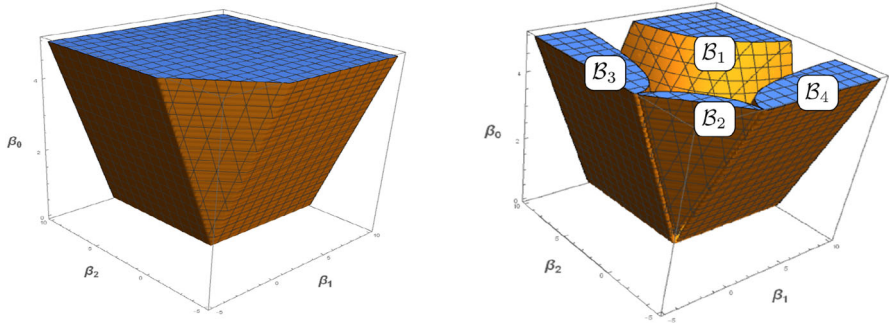


Fig. 1 Parameter region \mathcal{B} in the two-factor gamma model on $[0, 1]^2$ (left panel); subregions $\mathcal{B}_1, \dots, \mathcal{B}_4$ of parameters for minimally supported locally D -optimal designs ξ_1^*, \dots, ξ_4^* (right panel)

(see, e.g., Silvey [39]). Further note that the interior region shown in the right panel of Fig. 1 contains those values for the parameter vector β for which locally D -optimal designs are supported on all four vertices and the corresponding weights depend on the values of β (see Idais [20]). \square

Next we investigate equivariance for the $IMSE$ -criterion. There also the supplementary argument of the weighting measure ν has to be transformed. Similar to the information matrix in (17), the weighting matrix \mathbf{V} is equivariant under the transformations g and \tilde{g} ,

$$\begin{aligned}
 \mathbf{V}(\tilde{g}(\beta); \nu^g) &= \int \lambda(\mathbf{f}(\mathbf{z})^T \tilde{g}(\beta))^2 \mathbf{f}(\mathbf{z}) \mathbf{f}(\mathbf{z})^T \nu^g(d\mathbf{z}) \\
 &= \int \lambda(\mathbf{f}(g(\mathbf{x}))^T \tilde{g}(\beta))^2 \mathbf{f}(g(\mathbf{x})) \mathbf{f}(g(\mathbf{x}))^T \nu(d\mathbf{x}) \\
 &= \int \lambda(\mathbf{f}(g(\mathbf{x}))^T \tilde{g}(\beta))^2 \mathbf{Q}_g \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T \mathbf{Q}_g^T \nu(d\mathbf{x}) \\
 &= \mathbf{Q}_g \left(\int \lambda(\mathbf{f}(g(\mathbf{x}))^T \tilde{g}(\beta))^2 \mathbf{f}(\mathbf{x}) \mathbf{f}(\mathbf{x})^T \nu(d\mathbf{x}) \right) \mathbf{Q}_g^T = \mathbf{Q}_g \mathbf{V}(\beta; \nu) \mathbf{Q}_g^T,
 \end{aligned}
 \tag{20}$$

in the case of a generalized linear model with canonical link. This implies

$$\begin{aligned}
 IMSE(\xi^g; \tilde{g}(\beta), \nu^g) &= \text{trace}(\mathbf{Q}_g \mathbf{V}(\beta; \nu) \mathbf{Q}_g^T (\mathbf{Q}_g \mathbf{M}(\xi; \beta) \mathbf{Q}_g^T)^{-1}) \\
 &= \text{trace}(\mathbf{V}(\beta; \nu) \mathbf{M}(\xi; \beta)^{-1}) = IMSE(\xi; \beta, \nu).
 \end{aligned}
 \tag{21}$$

Let Φ be the local $IMSE$ -criterion at β with respect to ν and Φ' be the local $IMSE$ -criterion at $\tilde{g}(\beta)$ with respect to ν^g , then the $IMSE$ -criterion is equivariant under simultaneous transformation of β and the supplementary argument ν , and by Theorem 1 the locally $IMSE$ -optimal design can be transferred.

Corollary 2 *If ξ^* is locally $IMSE$ -optimal on \mathcal{X} at β with respect to ν , then $(\xi^*)^g$ is locally $IMSE$ -optimal on \mathcal{Z} at $\tilde{\beta} = \tilde{g}(\beta)$ with respect to ν^g .*

Table 1 Minimally supported locally D -optimal designs and optimality regions for the two-factor gamma model on $[0, 1]^2$

Transformation	Induced design	Induced support	Optimality region	Characterizing inequalities
$g_1(\mathbf{x}) = \mathbf{x}$	$\xi_1^* = \xi^*$	$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3$	\mathcal{B}_1	$\beta_0^2 - \beta_1\beta_2 \leq 0$
$g_2(\mathbf{x}) = (1 - x_1, 1 - x_2)^T$	$\xi_2^* = (\xi^*)^{g_2}$	$\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4$	$\mathcal{B}_2 = \tilde{g}_2(\mathcal{B}_1)$	$(\beta_0 + \beta_1 + \beta_2)^2 - \beta_1\beta_2 \leq 0$
$g_3(\mathbf{x}) = (1 - x_1, x_2)^T$	$\xi_3^* = (\xi^*)^{g_3}$	$\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_4$	$\mathcal{B}_3 = \tilde{g}_3(\mathcal{B}_1)$	$(\beta_0 + \beta_1)^2 + \beta_1\beta_2 \leq 0$
$g_4(\mathbf{x}) = (x_1, 1 - x_2)^T$	$\xi_4^* = (\xi^*)^{g_4}$	$\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_4$	$\mathcal{B}_4 = \tilde{g}_4(\mathcal{B}_1)$	$(\beta_0 + \beta_2)^2 + \beta_1\beta_2 \leq 0$

Note that the results of Corollaries 1 and 2 hold not only for any generalized linear model, but also, more generally, for all models, where the elemental information matrix is of the form (3) (see, e.g., Schmidt and Schwabe [37], for further examples).

Example (Example 1 continued) In order to apply the equivariance result of Corollary 2 to the gamma model with simple linear regression, $\mathbf{f}(x) = (1, x)^T$, the locally *IMSE*-optimal design ξ^* on the unit interval $\mathcal{X} = [0, 1]$ has to be determined first.

Proposition 1 For the one-factor gamma model with simple linear regression $\mathbf{f}(x)^T \boldsymbol{\beta} = \beta_0 + \beta_1 x$ on the experimental region $\mathcal{X} = [0, 1]$ locally *IMSE*-optimal designs can be found which are supported on the endpoints 0 and 1 of the experimental region.

Locally optimal weights $1 - w^*$ at 0 and w^* at 1, respectively, are given by

- (a) $1 - w^* = w^* = 1/2$ for ν the uniform (Lebesgue) measure on the interval $[0, 1]$,
- (b) $1 - w^* = (\beta_0 + \beta_1)/(2\beta_0 + \beta_1)$ and $w^* = \beta_0/(2\beta_0 + \beta_1)$ for ν the (discrete) uniform measure on the endpoints $\{0, 1\}$, and
- (c) $1 - w^* = \beta_0/(2\beta_0 + \beta_1)$ and $w^* = (\beta_0 + \beta_1)/(2\beta_0 + \beta_1)$ for ν the one-point measure on the midpoint $1/2$ of the design region.

The proof of Proposition 1 is given in ‘‘Appendix.’’ Note that in Proposition 1 the locally optimal weights may depend on the weighting measure ν used. In particular, for the two measures in Proposition 1 (b) and (c) which are concentrated on the endpoints and the midpoint, respectively, the locally optimal weights at 0 and 1 are interchanged. For the continuous uniform measure (Proposition 1 (a)), equal weights, $w^* = 1/2$, are assigned to both endpoints, and the (locally) *IMSE*-optimal design does not depend on the value of the parameter vector $\boldsymbol{\beta}$.

Now equivariance can be employed to obtain locally *IMSE*-optimal designs for any other interval $\mathcal{Z} = [a, b]$ as the experimental region. We again use the transformation $g(x) = a + cx$, $c = b - a$, together with $\tilde{g}(\boldsymbol{\beta}) = \mathbf{Q}_g^{-T} \boldsymbol{\beta}$. Let ξ^* be the locally *IMSE*-optimal design of Proposition 1 at $\boldsymbol{\beta}$ with respect to one of the given weighting measures ν . Then, by Corollary 2, the design $(\xi^*)^g$ is the locally *IMSE*-optimal design at $\tilde{\boldsymbol{\beta}} = \tilde{g}(\boldsymbol{\beta})$ with respect to ν^g . \square

In order to obtain locally optimal designs at a given value of $\tilde{\boldsymbol{\beta}}$ on the transformed design region \mathcal{Z} , the inverse transformations $g^{-1}(\mathbf{z}) = \mathbf{Q}_g^{-1} \mathbf{z}$ and $\tilde{g}^{-1}(\tilde{\boldsymbol{\beta}}) = \mathbf{Q}_g \tilde{\boldsymbol{\beta}}$ of g and \tilde{g} , respectively, have to be used. We give this general result only for the case of the *D*- and the *IMSE*-criterion.

Corollary 3 Let the equivariance conditions be fulfilled.

- (a) The design $(\xi^*)^g$ is locally *D*-optimal on \mathcal{Z} at $\tilde{\boldsymbol{\beta}}$ if ξ^* is locally *D*-optimal on \mathcal{X} at $\boldsymbol{\beta} = \tilde{g}^{-1}(\tilde{\boldsymbol{\beta}})$.
- (b) The design $(\xi^*)^g$ is locally *IMSE*-optimal on \mathcal{Z} at $\tilde{\boldsymbol{\beta}}$ with respect to ν if ξ^* is locally *IMSE*-optimal on \mathcal{X} at $\boldsymbol{\beta} = \tilde{g}^{-1}(\tilde{\boldsymbol{\beta}})$ with respect to ν^g .

Example (Example 1 continued) By Corollary 3, we can obtain locally *IMSE*-optimal designs for the one-factor gamma model with simple linear regression $\mathbf{f}(x)^T \tilde{\boldsymbol{\beta}} = \tilde{\beta}_0 + \tilde{\beta}_1 x$ on a given interval $\mathcal{Z} = [a, b]$ with respect to suitably specified weighting measures $\nu_{\mathcal{Z}}$. The inversely transformed parameter vector $\boldsymbol{\beta} = \tilde{g}^{-1}(\tilde{\boldsymbol{\beta}})$ is given by

$\beta = (\tilde{\beta}_0 + a\tilde{\beta}_1, (b - a)\tilde{\beta}_1)^T$. By Corollary 3 and Proposition 1, the optimal designs are supported on the endpoints a and b of the interval and the optimal weights $1 - w^*$ at a and w^* at b , respectively, can be obtained as

- (a) $1 - w^* = w^* = 1/2$ for $\nu_{\mathcal{Z}}$ the uniform (Lebesgue) measure on the interval $[a, b]$,
- (b) $1 - w^* = (\tilde{\beta}_0 + b\tilde{\beta}_1)/(2\tilde{\beta}_0 + (a + b)\tilde{\beta}_1) = \mathbf{f}(b)^T \tilde{\beta} / (\mathbf{f}(a)^T \tilde{\beta} + \mathbf{f}(b)^T \tilde{\beta})$ and $w^* = (\tilde{\beta}_0 + a\tilde{\beta}_1)/(2\tilde{\beta}_0 + (a + b)\tilde{\beta}_1) = \mathbf{f}(a)^T \tilde{\beta} / (\mathbf{f}(a)^T \tilde{\beta} + \mathbf{f}(b)^T \tilde{\beta})$ for $\nu_{\mathcal{Z}}$ the (discrete) uniform measure on the endpoints $\{a, b\}$, and
- (c) $1 - w^* = (\tilde{\beta}_0 + a\tilde{\beta}_1)/(2\tilde{\beta}_0 + (a + b)\tilde{\beta}_1) = \mathbf{f}(a)^T \tilde{\beta} / (\mathbf{f}(a)^T \tilde{\beta} + \mathbf{f}(b)^T \tilde{\beta})$ and $w^* = (\tilde{\beta}_0 + b\tilde{\beta}_1)/(2\tilde{\beta}_0 + (a + b)\tilde{\beta}_1) = \mathbf{f}(b)^T \tilde{\beta} / (\mathbf{f}(a)^T \tilde{\beta} + \mathbf{f}(b)^T \tilde{\beta})$ for $\nu_{\mathcal{Z}}$ the one-point measure on the midpoint $(a + b)/2$ of the experimental region.

The continuous uniform measure in (a) is the common choice for the *IMSE*-criterion. The discrete uniform measure in (b) lays equal interest in the extreme values of the experimental region and may also be applied for the restricted experimental region $\mathcal{X} = \{a, b\}$ which can be used to describe two groups “ a ” and “ b .” In that case, the *IMSE*-optimal weights are inverse proportional to the standard deviations $\lambda_x = 1/(\mathbf{f}(x)^T \tilde{\beta})^2$, $x = a, b$, in the groups in accordance with known results on *A*-optimality for group means. The one-point measure in (c) coincides with the *c*-criterion for estimating the mean response at the midpoint of the interval. □

Note that the *D*- and *IMSE*-criteria are equivariant with respect to any transformation g of \mathbf{x} for which the regression function \mathbf{f} is linearly equivariant, $\mathbf{f}(g(\mathbf{x})) = \mathbf{Q}_g \mathbf{f}(\mathbf{x})$, and the corresponding transformation $\tilde{g}(\beta) = \mathbf{Q}_g^T \beta$ of β . For other criteria, additional requirements may have to be fulfilled by the transformations to obtain equivariance results. For example, in the case of Kiefer’s class of Φ_q -criteria (including the *A*-criterion) the transformation matrix \mathbf{Q}_g should be orthogonal or, at least, satisfy that $\mathbf{Q}_g^T \mathbf{Q}_g$ is a multiple of the $p \times p$ identity matrix.

For the equivariance of maximin efficiency criteria, we require additionally that the underlying local criteria are multiplicatively equivariant with respect to (g, \tilde{g}) , which means that for every $\beta \in \mathcal{B}'$ there is a constant $c > 0$ such that $\Phi_{\tilde{g}(\beta)}(\xi^g) = c\Phi_{\beta}(\xi)$ uniformly in ξ . Then, for the corresponding maximin efficiency criterion we get

$$\begin{aligned} \Phi(\xi^g) &= \sup_{\tilde{\beta} \in \tilde{\mathcal{B}}'} \frac{\Phi_{\tilde{\beta}}(\xi^g)}{\Phi_{\tilde{\beta}}(\xi_{\tilde{\beta}}^*)} \\ &= \sup_{\beta \in \mathcal{B}'} \frac{\Phi_{\tilde{g}(\beta)}(\xi^g)}{\Phi_{\tilde{g}(\beta)}((\xi_{\beta}^*)^g)} \\ &= \sup_{\beta \in \mathcal{B}'} \frac{c\Phi_{\beta}(\xi)}{c\Phi_{\beta}(\xi_{\beta}^*)} = \Phi(\xi), \end{aligned} \tag{22}$$

where in the second equality it is used that by Theorem 1 the image of the locally optimal design at β under g is locally optimal at $\tilde{g}(\beta)$. Hence, the resulting maximin efficiency criterion Φ is equivariant.

By (18), the homogeneous version $\Phi_{\beta}(\xi) = (\det(\mathbf{M}(\xi; \beta)))^{-1/p}$ of the local *D*-criterion is multiplicatively equivariant with $c = \det(\mathbf{Q}_g)^{-2/p} > 0$. Accordingly, the

local *IMSE*-criterion is multiplicatively equivariant with $c = 1$ by (21). Hence, both the maximin *D*-efficiency criterion and the maximin *IMSE*-efficiency criterion retain their value under the transformation and are thus equivariant.

Corollary 4 (a) *If ξ^* is maximin *D*-efficient on \mathcal{B}' , then $(\xi^*)^g$ is maximin *D*-efficient on $\tilde{\mathcal{B}}' = \tilde{g}(\mathcal{B}')$.*

(b) *If ξ^* is maximin *IMSE*-efficient with respect to ν on \mathcal{B}' , then $(\xi^*)^g$ is maximin *IMSE*-efficient with respect to ν^g on $\tilde{\mathcal{B}}' = \tilde{g}(\mathcal{B}')$.*

Example (Example 1 continued) In the gamma model with simple linear regression on $[0, 1]$ the design ξ^* which assigns equal weights $1/2$ to both endpoints 0 and 1 is both locally *D*-optimal and by Proposition 1 locally *IMSE*-optimal with respect to the uniform measure ν on $[0, 1]$ for any $\beta \in \mathcal{B}$. Hence, ξ^* is obviously both maximin *D*-efficient and maximin *IMSE*-efficient with respect to ν on \mathcal{B} on $[0, 1]$. Then, with $g(x) = a + cx$, $c = b - a$, by Corollary 4 the design $(\xi^*)^g$ which assigns equal weights $1/2$ to a and b is maximin *D*-efficient and maximin *IMSE*-efficient with respect to the uniform measure ν^g on $\tilde{\mathcal{B}} = \tilde{g}(\mathcal{B})$ on $[a, b]$. \square

Further maximin *D*- and *IMSE*-efficient designs are derived in Sect. 5.

4 Extended Equivariance

The concept of equivariance can be extended when the structure of the intensity function is compatible with some transformation of the parameters. More specifically, we will consider situations where the intensity function λ is multiplicatively equivariant with respect to a transformation \tilde{g}_0 of β , i.e., there exists a constant $c_0 > 0$ such that $\lambda(\mathbf{f}(\mathbf{x})^T \tilde{g}_0(\beta)) = c_0 \lambda(\mathbf{f}(\mathbf{x})^T \beta)$ for all $\mathbf{x} \in \mathcal{X}$.

For example, in the gamma model with inverse link we have $\lambda(\mathbf{f}(\mathbf{x})^T \tilde{c}\beta) = \tilde{c}^{-2} \lambda(\mathbf{f}(\mathbf{x})^T \beta)$ for any scaling factor \tilde{c} (see Idais and Schwabe [21], for some specific models). Hence, the intensity function is multiplicatively equivariant with respect to any transformation $\tilde{g}_0(\beta) = \tilde{c}\beta$ which scales all components of the parameter vector β simultaneously by the same factor $\tilde{c} > 0$, and the multiplicative factor is $c_0 = \tilde{c}^{-2} > 0$. Note that the scaling \tilde{g}_0 retains the positivity of the linear component $\mathbf{f}(\mathbf{x})^T \tilde{g}_0(\beta) = \tilde{c} \mathbf{f}(\mathbf{x})^T \beta > 0$ for the scaled vector $\tilde{g}_0(\beta) = \tilde{c}\beta$, $\tilde{c} > 0$. Thus, the maximal region \mathcal{B} of parameter values β such that the linear component $\mathbf{f}(\mathbf{x})^T \beta$ is positive constitutes a cone in the p -dimensional Euclidean space, i.e., for each vector $\beta \in \mathcal{B}$ and every positive scale factor $\tilde{c} > 0$ the scaled vector $\tilde{c}\beta$ lies also in \mathcal{B} .

Another, more basic example arises in Poisson regression with canonical log link when the value β_0 of the intercept parameter is changed to $\tilde{\beta}_0$. The corresponding transformation of β can be described by the (affine) linear mapping $\tilde{g}_0(\beta) = \beta + (\tilde{\beta}_0 - \beta_0)\mathbf{e}_1$, where \mathbf{e}_1 denotes the first unit vector of appropriate length p . Then, the intensity function $\lambda(z) = \exp(z)$ is multiplicatively equivariant with respect to \tilde{g}_0 with multiplicative factor $c_0 = \exp(\tilde{\beta}_0 - \beta_0) > 0$. This has been implicitly applied in the literature when concluding that optimal designs do not depend on the value β_0 of the intercept parameter (see, e.g., Russell et al. [36]).

To embed these transformations \tilde{g}_0 into the concept of equivariance of Sect. 3, we combine them with the identity mapping $g = \text{id}$ on the experimental region \mathcal{X} . Then, the multiplicative equivariance obviously carries over from the intensity to the information matrix.

Lemma 1 *If the intensity function λ is multiplicatively equivariant with respect to \tilde{g}_0 with multiplicative factor $c_0 > 0$, then the information matrix is (multiplicatively) equivariant with respect to $g = \text{id}$ and \tilde{g}_0 , $\mathbf{M}(\xi; \tilde{g}_0(\boldsymbol{\beta})) = c_0 \mathbf{M}(\xi; \boldsymbol{\beta})$.*

To transfer optimal designs by Theorem 1, it remains to show that the criteria under consideration are order preserving with respect to transformations which act multiplicatively on the intensity. By Lemma 1, we directly get $\det(\mathbf{M}(\xi; \tilde{g}_0(\boldsymbol{\beta}))) = c_0^p \det(\mathbf{M}(\xi; \boldsymbol{\beta}))$ and, hence, the equivariance of the D -criterion.

For the $IMSE$ -criterion, we additionally require multiplicative equivariance of the function η' , i.e., $\eta'(\mathbf{f}(\mathbf{x})^T \tilde{g}_0(\boldsymbol{\beta})) = c_\eta \eta'(\mathbf{f}(\mathbf{x})^T \boldsymbol{\beta})$ uniformly in \mathbf{x} for some constant $c_\eta > 0$. This condition is fulfilled for generalized linear models with canonical link because of $\eta' = \lambda$. Then, the equivariance property $\mathbf{V}(\tilde{g}_0(\boldsymbol{\beta}); \nu) = c_\eta^2 \mathbf{V}(\boldsymbol{\beta}; \nu)$ of the weighting matrix yields equivariance, $IMSE(\xi; \tilde{g}_0(\boldsymbol{\beta}), \nu) = c_0^{-1} c_\eta^2 IMSE(\xi; \boldsymbol{\beta}, \nu)$.

Corollary 5 *If the intensity function λ is multiplicatively equivariant with respect to \tilde{g}_0 , then*

- (a) *A locally D -optimal design ξ^* at $\boldsymbol{\beta}$ is also locally D -optimal at $\tilde{\boldsymbol{\beta}} = \tilde{g}_0(\boldsymbol{\beta})$,*
- (b) *If additionally η' is multiplicatively equivariant, a locally $IMSE$ -optimal design ξ^* at $\boldsymbol{\beta}$ with respect to ν is also locally $IMSE$ -optimal at $\tilde{\boldsymbol{\beta}} = \tilde{g}_0(\boldsymbol{\beta})$ with respect to ν .*

When a whole family of such transformations \tilde{g}_0 is available, as scaling by $\tilde{c} > 0$ in the gamma model with inverse link or shifting the intercept in Poisson regression, then we can use the result of Corollary 5 to reduce the number of parameters in the optimization problem. Therefore, we first solve the optimization problem for a standardized parameter setting and then transfer the obtained optimal design to a general parameter vector by a suitable choice of the transformation \tilde{g}_0 . For example, in the gamma model one component of the parameter vector can be set equal to 1 for standardization. Then, a general parameter vector can be obtained by choosing \tilde{c} equal to the nominal value of the component of the parameter vector used for standardization. Similarly, in Poisson regression, we can first set the intercept parameter equal to 0 and then transfer the optimal design to the parameter vector with given nominal value β_0 .

Example (Example 1 continued) For the one-factor gamma model with simple linear regression $\mathbf{f}(x)^T \boldsymbol{\beta} = \beta_0 + \beta_1 x$ on $\mathcal{X} = [0, 1]$, the locally $IMSE$ -optimal design at $(1, \gamma)^T$ with respect to the discrete uniform weighting measure ν on the endpoints $\{0, 1\}$ assigns weights $1 - w^* = 1/(2 + \gamma)$ and $w^* = (1 + \gamma)/(2 + \gamma)$ to 0 and 1, respectively. Under scaling with $\tilde{c} = \beta_0 > 0$, these weights remain locally optimal for any parameter vector $\boldsymbol{\beta} = (\beta_0, \beta_1)^T$ with $\beta_1/\beta_0 = \gamma$ by Corollary 5. The corresponding reduced parameter region for γ is given by $\mathcal{C} = \{\gamma; \gamma > -1\}$ which is displayed in Fig. 2 as the vertical dashed line at $\beta_0 = 1$. There the diagonal dotted line represents one ray $\{\boldsymbol{\beta}; \beta_1 = \gamma \beta_0\}$ of values of $\boldsymbol{\beta}$ which are reduced to one specific value of γ indicated by the intersection of the ray with the vertical line at $\beta_0 = 1$. \square

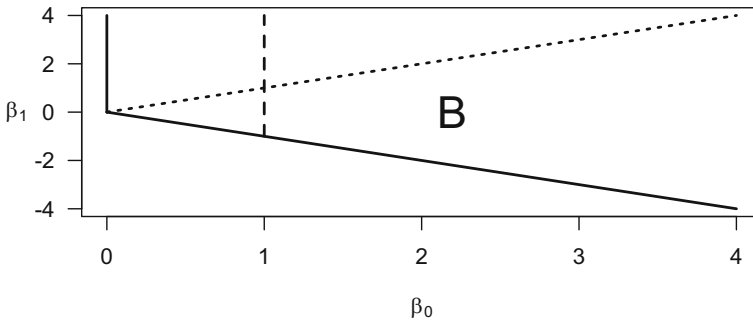


Fig. 2 Parameter region \mathcal{B} for the one-factor gamma model on $[0, 1]$; reduced parameter region \mathcal{C} (vertical dashed line); values of β reduced to $\gamma = 1$ (diagonal dotted line)

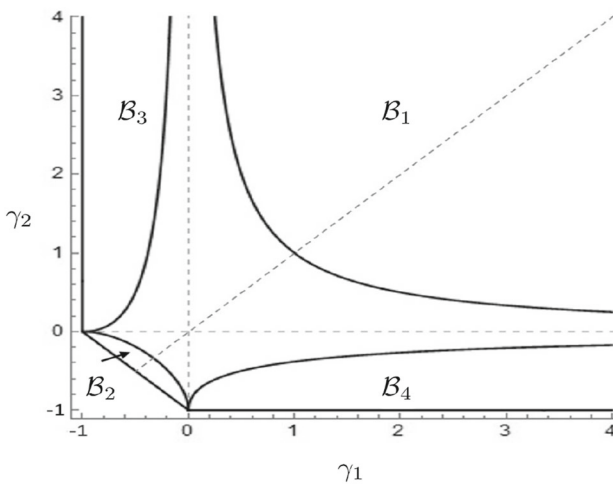


Fig. 3 Scaled parameter region of the two-factor gamma model on $[0, 1]^2$; the diagonal dashed line represents $\gamma_1 = \gamma_2$

Example (Example 2 continued) Similarly, in the two-factor gamma model on $[0, 1]^2$, the three-dimensional parameter vector $\beta = (\beta_0, \beta_1, \beta_2)^T$ can be reduced to $\tilde{\beta} = (1, \gamma_1, \gamma_2)^T$, where $\gamma_1 = \beta_1/\beta_0$ and $\gamma_2 = \beta_2/\beta_0$, by setting the value of the intercept parameter β_0 equal to 1. As a consequence, the three-dimensional parameter region \mathcal{B} in Fig. 1 is reduced to the two-dimensional region \mathcal{C} for $\gamma = (\gamma_1, \gamma_2)^T$ in Fig. 3 which is characterized by the linear constraints $\gamma_1 > -1$, $\gamma_2 > -1$, and $\gamma_1 + \gamma_2 > -1$. The optimality regions \mathcal{B}_k , $k = 1, \dots, 4$ of Table 1 can now be described in terms of the ratios $\gamma_j = \beta_j/\beta_0$, $j = 1, 2$, resulting in $\mathcal{B}_1 \equiv 1 - \gamma_1\gamma_2 \leq 0$, $\mathcal{B}_2 \equiv (1 + \gamma_1 + \gamma_2)^2 - \gamma_1\gamma_2 \leq 0$, $\mathcal{B}_3 \equiv (1 + \gamma_1)^2 + \gamma_1\gamma_2 \leq 0$ and $\mathcal{B}_4 \equiv (1 + \gamma_2)^2 + \gamma_1\gamma_2 \leq 0$, as exhibited in Fig. 3. The scaling property explains why the subregions in Fig. 1 constitute cones in the three-dimensional Euclidean space. \square

By combination of the transformation \tilde{g}_0 with the linear transformations of the preceding Sect. 3, we get an extension of Corollaries 1 and 2 by Theorem 1.

Corollary 6 *If the intensity function λ is multiplicatively equivariant with respect to \tilde{g}_0 , then:*

- (a) *If ξ^* is locally D-optimal on \mathcal{X} at β , then $(\xi^*)^g$ is locally D-optimal on \mathcal{Z} at $\tilde{\beta} = \tilde{g}_0(\mathbf{Q}_g^{-T}\beta)$,*
- (b) *If ξ^* is locally IMSE-optimal on \mathcal{X} at β with respect to ν and if, additionally, η' is multiplicatively equivariant, then $(\xi^*)^g$ is locally IMSE-optimal on \mathcal{Z} at $\tilde{\beta} = \tilde{g}_0(\mathbf{Q}_g^{-T}\beta)$ with respect to ν^g .*

This result indicates that for a given transformation g of \mathbf{x} the associated transformation $\tilde{g}(\beta) = \tilde{g}_0(\mathbf{Q}_g^{-T}\beta)$ of β needs not be unique. Moreover, we may let the transformation $\tilde{g}_0 = \tilde{g}_{0,\beta}$ depend on the parameter vector β , where the intensity function λ is multiplicatively equivariant with respect to $\tilde{g}_{0,\beta}$ for any β . Then, also the multiplicative factor $c_0 = c_{0,\beta}$ will depend on β so that $\lambda(\mathbf{f}(\mathbf{x})^T \tilde{g}_{0,\beta}(\beta)) = c_{0,\beta} \lambda(\mathbf{f}(\mathbf{x})^T \beta)$. In combination with the linear transformation of Sect. 3, this leads to a nonlinear transformation $\tilde{g}(\beta) = \tilde{g}_{0,\beta}(\mathbf{Q}_g^{-T}\beta)$ of the parameter vector β so that the information matrix is equivariant with respect to the pair (g, \tilde{g}) of transformations. For the gamma model with inverse link, this can be accomplished by choosing the scaling factor $\tilde{c} = \tilde{c}_\beta$ in dependence on β .

Example (Example 1 continued) For the one-factor gamma model with simple linear regression on the unit interval $[0, 1]$, consider the reflection $g(x) = 1 - x$ that maps $[0, 1]$ onto itself. The corresponding linear transformation of the parameter vector $\beta = (\beta_0, \beta_1)^T$ is given by $\mathbf{Q}_g^{-T}\beta = (\beta_0 + \beta_1, -\beta_1)^T$. In particular, for a scaled reduced parameter vector $\beta = (1, \gamma)^T, \gamma = \beta_1/\beta_0$, we have $\mathbf{Q}_g^{-T}\beta = (1 + \gamma, -\gamma)^T$. In order to obtain a transformed parameter vector $\tilde{\beta} = \tilde{g}(\beta)$ in reduced form, $\tilde{\beta}_0 = 1$ the linear transformation has to be rescaled by $\tilde{c}_\beta = 1/(1 + \gamma)$. This results in the nonlinear transformation $\tilde{g}((1, \gamma)^T) = (1, -\gamma/(1 + \gamma))^T$. Note that this transformation is a one-to-one mapping of the maximal region $\mathcal{C} = (-1, \infty)$ for the reduced parameter γ .

For general values of the first entry β_0 in the parameter vector β , the value of β_0 can be preserved by the nonlinear transformation $\tilde{g}(\beta) = \beta_0/(\beta_0 + \beta_1)\mathbf{Q}_g^{-T}\beta$, where the scaling factor $\tilde{c}_\beta = \beta_0/(\beta_0 + \beta_1) = 1/(1 + \gamma)$ only depends on $\gamma = \beta_1/\beta_0$. □

The result of Corollary 6 carries over directly also for the nonlinear transformation, when \tilde{g}_0 is replaced by $\tilde{g}_{0,\beta}$.

Example (Example 1 continued) For the one-factor gamma model with simple linear regression on $[0, 1]$ and reflection $g(x) = 1 - x$, the weighting measures ν specified in Proposition 1 are all invariant with respect to g , i.e., $\nu^g = \nu$. The corresponding locally IMSE-optimal designs ξ^* on $[0, 1]$ with respect to ν are supported by the endpoints with optimal weights $1 - w^*$ and w^* at 0 and 1, respectively. Then, the designs $(\xi^*)^g$ which assign the interchanged weights w^* to 0 and $1 - w^*$ to 1 are locally IMSE-optimal on $[0, 1]$ with respect to ν at $\tilde{\beta} = \tilde{c}_\beta \mathbf{Q}_g^{-T}\beta = (\beta_0, -\beta_1/(1 + \gamma))^T$.

The standardization with respect to the intercept can be extended to more complex models.

Example (Example 2 continued) In the two-factor gamma model on $[0, 1]^2$, we consider the transformation $g_2(\mathbf{x}) = (1 - x_1, 1 - x_2)^T$ of simultaneous reflection of both explanatory variables and the corresponding rescaled transformation $\tilde{g}_2(\boldsymbol{\beta}) = \tilde{c}_\beta \mathbf{Q}_{g_2}^{-T} \boldsymbol{\beta}$ of $\boldsymbol{\beta}$ which leaves the intercept β_0 unchanged, i.e., $\tilde{c}_\beta = \beta_0 / (\beta_0 + \beta_1 + \beta_2)$ and, hence, $\tilde{g}_2(\boldsymbol{\beta}) = \beta_0(\beta_0, -\beta_1, -\beta_2)^T / (\beta_0 + \beta_1 + \beta_2)$. The scaling factor $\tilde{c}_\beta = 1 / (1 + \gamma_1 + \gamma_2)$ only depends on the reduced parameters $\gamma_j = \beta_j / \beta_0, j = 1, 2$, so that the one-to-one transformation $\boldsymbol{\gamma} \rightarrow -(1 / (1 + \gamma_1 + \gamma_2)) \boldsymbol{\gamma}$ is induced on the reduced parameter region \mathcal{C} onto itself. Hence, if a design ξ^* is locally D -optimal at $\boldsymbol{\gamma}$ which assigns weights w_i^* to $\mathbf{x}_i, i = 1, \dots, 4$, then the design $(\xi^*)^{g_2}$ which assigns weights w_4^*, w_3^*, w_2^* and w_1^* to $\mathbf{x}_1, \dots, \mathbf{x}_4$, respectively, is locally D -optimal at $-(1 / (1 + \gamma_1 + \gamma_2)) \boldsymbol{\gamma}$.

Similar results hold for *IMSE*-optimality. \square

For maximin efficiency criteria, we additionally allow here that the multiplicative factor in the equivariance of the underlying local criteria may depend on the parameter $\boldsymbol{\beta}, c = c_\beta$. This does not affect the arguments in (22), and hence, the resulting maximin efficiency criteria remain equivariant. The homogeneous version of the local D -criterion and the local *IMSE*-criterion is multiplicatively equivariant with $c_\beta = c_{0,\beta}^{-1} \det(\mathbf{Q}_g)^{-2/p} > 0$ and $c_\beta = c_{0,\beta} > 0$, respectively. Hence, for both the maximin D -efficiency and the maximin *IMSE*-efficiency criterion their value is not changed under the transformation. These criteria are thus equivariant, and the result of Corollary 4 remains valid so that maximin efficient designs can be transferred also for nonlinear transformations $\tilde{g}(\boldsymbol{\beta}) = \tilde{g}_{0,\beta}(\mathbf{Q}_g^{-T} \boldsymbol{\beta})$ when the intensity function is multiplicatively equivariant with respect to $\tilde{g}_{0,\beta}$ for all $\boldsymbol{\beta}$.

5 Invariance

While equivariance can be used to transfer optimal designs, the concept of invariance allows reduction in the complexity of finding optimal designs by exploiting symmetries (see, e.g., Schwabe [38, ch. 3], in the case of linear models). As in linear models, we need a (finite) group G of transformations g which map the experimental region \mathcal{X} onto itself. For each of these transformations g , the regression functions \mathbf{f} are assumed to be linearly equivariant, $\mathbf{f}(g(\mathbf{x})) = \mathbf{Q}_g \mathbf{f}(\mathbf{x})$. For generalized linear models, we require additionally that the corresponding transformations \tilde{g} of $\boldsymbol{\beta}$ also constitute a group \tilde{G} such that the set (G, \tilde{G}) of pairs (g, \tilde{g}) of transformations shares the group structure. This requirement is automatically fulfilled for the linear transformations $\tilde{g}(\boldsymbol{\beta}) = \mathbf{Q}_g^{-T} \boldsymbol{\beta}$, because the transformation matrices $\mathbf{Q}_g, g \in G$, constitute a group with respect to matrix multiplication. For extended equivariance (Sect. 4), also the factor $c_{0,\beta}$ has to share the group property. This holds in the gamma model for rescaling by \tilde{c}_β which leaves the value for the standardized component of the parameter vector unchanged. Similarly, for Poisson regression, standardization of the intercept to 0 preserves the group structure.

Example (Example 1 continued) For the one-factor gamma model with simple linear regression on $[0, 1]$, the reflection $g(x) = 1 - x$ maps $[0, 1]$ onto itself and is self-inverse, i.e., $g^{-1} = g$. Hence, g together with the identity id constitute a

group $G = \{\text{id}, g\}$ of transformations. For g the associated transformation of the parameter vector β is $\tilde{g}(\beta) = \mathbf{Q}_g^{-T}\beta = (\beta_0 + \beta_1, -\beta_1)^T$ in the linear case and $\tilde{g}(\beta) = \tilde{c}_\beta \mathbf{Q}_g^{-T}\beta = (\beta_0, -\beta_0\beta_1/(\beta_0 + \beta_1))^T$ in the extended case. As always, the identity on \mathcal{B} is associated with the identity id on \mathcal{X} . Because of $\tilde{c}_{\tilde{g}(\beta)} = 1 + \gamma = 1/\tilde{c}_\beta$ also \tilde{g} is self-inverse, and the group structure is retained. \square

The final ingredient for invariance is that the optimality criterion Φ is invariant with respect to the group G of transformations, i.e., $\Phi(\xi^g) = \Phi(\xi)$ for all $g \in G$ and any design ξ . Then, we can make use of convexity arguments to improve designs by symmetrization. For this, define by $\bar{\xi} = (1/|G|) \sum_{g \in G} \xi^g$ the symmetrized version of a design ξ with respect to the group G , where $|G|$ denotes the number of elements in the (finite) group G . Note that $\bar{\xi}$ is itself a design and that $\bar{\xi}$ is invariant with respect to G , i.e., $\bar{\xi}^g = \bar{\xi}$ for all $g \in G$. If Φ is invariant and convex, we obtain

$$\Phi(\bar{\xi}) \leq \frac{1}{|G|} \sum_{g \in G} \Phi(\xi^g) = \Phi(\xi), \tag{23}$$

where the inequality follows from convexity and the equation from invariance. From this majorization property, we can conclude that the designs which are invariant with respect to G constitute an essentially complete class with regard to Φ . This means that we can confine the search for a Φ -optimal design to the class of invariant designs.

Theorem 2 *If Φ is invariant (with respect to G) and convex, then there exists an invariant design ξ^* (with respect to G) which is Φ -optimal over all designs.*

The class of invariant designs is often much smaller than the class of all designs, and optimization can be simplified. Invariant designs are uniform on orbits $\mathcal{O}_x = \{g(\mathbf{x}); g \in G\} \subset \mathcal{X}$, i.e., all \mathbf{x} in the same orbit have the same weight. In particular, for an invariant design, either all \mathbf{x} in an orbit \mathcal{O} are included with weight $w_{\mathcal{O}}$ or the whole orbit is not in the support of the design. For optimization in the class of invariant designs, it remains to find the optimal orbits and the corresponding optimal weights which is often a much easier task than to optimize over all possible designs.

Example (Example 1 continued) For the reflection group $G = \{\text{id}, g\}$, $g(x) = 1 - x$ on $[0, 1]$, the orbits are all of the form $\{x, 1 - x\}$ for $x < 1/2$ and $\{1/2\}$ for $x = 1/2$, respectively. In the one-factor gamma model with simple linear regression, it is known that the optimal designs are supported at the endpoints 0 and 1 (see Gaffke et al. [14]). Thus, the only remaining orbit for an optimal design is $\{0, 1\}$, and hence, there is only one invariant design which assigns equal weights $1/2$ to each endpoint. This design is optimal with respect to each convex invariant criterion. \square

In the case of local optimality criteria, the requirement of invariance is rather restrictive. In particular, the local parameter β has to be invariant under all transformations, i.e., $\tilde{g}(\beta) = \beta$ for all $g \in G$. This condition typically holds only for a few values of β .

Example (Example 1 continued) For the one-factor gamma model with simple linear regression under the reflection $g(x) = 1 - x$, the parameter β is only invariant if

$\beta_1 = 0$, i.e., there is no effect of the covariate x . The invariant design which assigns equal weights $1/2$ to the endpoints is locally optimal at β for $\beta_1 = 0$.

Example (Example 2 continued) For the two-factor gamma model with multiple linear regression on $[0, 1]^2$, the reflections g_2 , g_3 , and g_4 are all self-inverse and the composition of any two reflections yields the third one. Together with the identity $g_1 = \text{id}$, the reflections constitute a group $G = \{g_1, g_2, g_3, g_4\}$ of transformations. Locally optimal designs are supported at the vertices of $[0, 1]^2$ (see Gaffke et al. [14]). The vertices lie all on one orbit, and hence, the unique invariant design on the vertices assigns equal weights $1/4$ to each of the vertices. Under the group G , the parameter vector β is only invariant if $\beta_1 = \beta_2 = 0$, i.e., there is no effect for both covariates x_1 and x_2 . Thus, the invariant design which assigns equal weights $1/4$ to the vertices is locally optimal at β only in the case $\beta_1 = \beta_2 = 0$. \square

Note that in both examples above locally optimal designs are obtained for the situation of constant intensity λ . In that case, the information matrix is proportional to that in the corresponding linear model with the same linear component. Hence, the locally optimal design coincides with the optimal design in the linear model (see Cox [6, Section 4]).

In more complex situations, however, invariance may be helpful for local optimality at certain parameter values which are invariant with respect to \tilde{g} for all $g \in G$. To this end, first note that in the case of a finite group G of transformations g the corresponding transformation matrices \mathbf{Q}_g are unimodal, i.e., $|\det(\mathbf{Q}_g)| = 1$ (see Schwabe [38, ch. 3]). For the *IMSE*-criterion, we additionally require that the weighting measure ν is invariant with respect to G , i.e., $\nu^g = \nu$ for all $g \in G$.

Corollary 7 *If $\tilde{g}(\beta) = \beta$ for all $g \in G$, then there exists a locally D -optimal design ξ^* at β which is invariant with respect to G .*

*If additionally ν is invariant with respect to G , then there exists a locally *IMSE*-optimal design ξ^* at β with respect to ν which is invariant with respect to G .*

Example (Example 2 continued) In the two-factor gamma model on $[0, 1]^2$, we consider nominal parameter values β with $\beta_1 = 0$, i.e., where the first covariate x_1 has no effect. Such parameter vectors are invariant with respect to the linear transformation $\tilde{g}_3(\beta) = (\beta_0 + \beta_1, -\beta_1, \beta_2)^T$ associated with the reflection $g_3(\mathbf{x}) = (1 - x_1, x_2)^T$ of the first covariate x_1 . As the transformation g_3 is self-inverse, together with the identity id it constitutes a group $G_3 = \{\text{id}, g_3\}$. Then, the local D -criterion at such β with $\beta_1 = 0$ is invariant with respect to G_3 . By Corollary 7 a locally D -optimal design can be found in the class of designs which are invariant with respect to G_3 . Moreover, here we can also restrict attention to designs supported by the vertices. With respect to G_3 , the relevant orbits are then $(\mathbf{x}_1, \mathbf{x}_2)$ and $(\mathbf{x}_3, \mathbf{x}_4)$, and invariant designs on the vertices have equal weights w at \mathbf{x}_1 and \mathbf{x}_2 and equal weights $1/2 - w$ at \mathbf{x}_3 and \mathbf{x}_4 , respectively. We will denote such designs by $\tilde{\xi}_w$. The optimization problem for a local D -optimal design reduces to finding the optimal weight w^* . Note that for $\beta_1 = 0$ the intensities on the orbits are constant, i.e., $\lambda_1 = \lambda_2$ and $\lambda_3 = \lambda_4$, where again λ_i denotes the intensity at \mathbf{x}_i . For the designs $\tilde{\xi}_w$, the determinant of the information

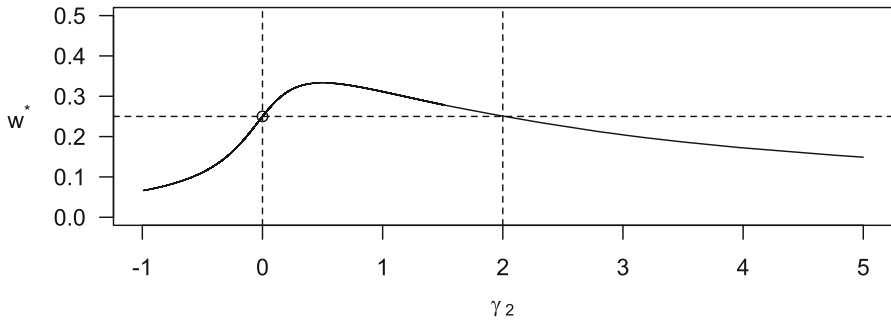


Fig. 4 Optimal weights w^* in the two-factor gamma model when $\beta_1 = 0$. The vertical dashed lines indicate $\gamma_2 = 0$ and $\gamma_2 = 2$. The horizontal dashed line indicates $w^* = 0.25$

matrix becomes

$$\det(\mathbf{M}(\bar{\xi}_w; \boldsymbol{\beta})) = 2(\lambda_1^2 \lambda_3 w^2 (1/2 - w) + \lambda_1 \lambda_3^2 w (1/2 - w)^2)$$

in the case $\beta_1 = 0$. The optimal weight w^* can be determined by straightforward computations as

$$w^* = \frac{3\gamma_2 - 1 + \sqrt{12\gamma_2^2 + 1}}{6\gamma_2(\gamma_2 + 2)}, \tag{24}$$

for $\beta_2 \neq 0$, and $w^* = 1/4$ for $\beta_2 = 0$ or $\beta_2 = 2\beta_0$. The dependence of the optimal weight w^* on γ_2 is shown in Fig. 4. The resulting invariant design $\bar{\xi}_{w^*}$ is locally D -optimal at $\boldsymbol{\beta}$ with $\beta_1 = 0$.

An analogous result holds for $\beta_2 = 0$, when the reflection g_4 of the second covariate x_2 is used instead of g_3 . □

Example 3 In the two-factor gamma model on $[0, 1]^2$, there are further symmetries which can be employed. In particular, we may consider parameter vectors $\boldsymbol{\beta}$ with equal slopes, i.e., $\beta_1 = \beta_2 = \beta$ for some β when both covariates x_1 and x_2 have an effect of the same size. These values for the parameter vector $\boldsymbol{\beta}$ are invariant with respect to the linear transformation $\tilde{g}_5(\boldsymbol{\beta}) = (\beta_0, \beta_2, \beta_1)^T$ associated with the permutation $g_5(\mathbf{x}) = (x_2, x_1)^T$ of the covariates. The transformation g_5 is self-inverse and constitutes together with the identity id a group $G_5 = \{\text{id}, g_5\}$. Because locally optimal designs are supported by the vertices of $[0, 1]^2$, there are only three relevant orbits $\{\mathbf{x}_1\}$, $\{\mathbf{x}_2, \mathbf{x}_3\}$, and $\{\mathbf{x}_4\}$. Optimal invariant designs can thus be characterized by two weights w_1^* assigned to \mathbf{x}_1 and w_4^* assigned to \mathbf{x}_4 , while the remaining equal weights $w_2^* = w_3^* = (1 - w_1^* - w_4^*)/2$ are assigned to each of \mathbf{x}_2 and \mathbf{x}_3 .

For the local D -criterion, optimal weights have been obtained (see Gaffke et al. [14, Theorem 4.3]). There it was shown that minimally supported designs are locally D -optimal at $\boldsymbol{\beta}$ with $\beta_1 = \beta_2 = \beta$ for $\beta > \beta_0$ or $-\beta_0/2 < \beta \leq \beta_0/3$ with weights $w_2^* = w_3^* = 1/3$ and $w_1^* = 1/3$ or $w_4^* = 1/3$, respectively. In the intermediate case

Table 2 Locally *IMSE*-optimal weights in the two-factor gamma model on $[0, 1]^2$

γ	0	1	2	3	10	-3/7
w_1^*	0.250	0.250	0.242	0.236	0.214	0.000
$w_2^* = w_3^*$	0.250	0.300	0.362	0.382	0.393	0.382
w_4^*	0.250	0.150	0.034	0.000	0.000	0.236

$-\beta_0/3 < \beta \leq \beta_0$ the locally *D*-optimal designs are supported on all four vertices with weights

$$w_1^* = \frac{3\gamma + 1}{4(2\gamma + 1)}, w_2^* = w_3^* = \frac{(\gamma + 1)^2}{4(2\gamma + 1)} \text{ and } w_4^* = \frac{1 - \gamma}{4},$$

where $\gamma = \beta/\beta_0$. In particular, uniform weights $w_i^* = 1/4$ are again seen to be optimal in the case $\beta = 0$ of constant intensity, as indicated in Fig. 4 by the vertical and horizontal dashed lines at $\gamma_2 = \beta_2 = 0$ and $w^* = 0.25$, respectively.

Also for *IMSE*-optimality, the optimal weights depend only on $\gamma = \beta/\beta_0$ by the scaling property of Sect. 4. The locally optimal weights can only be determined numerically. For selected values of γ , we present some numerical solutions in Table 2 in the case of the uniform weighting measure ν on $[0, 1]^2$ which is invariant with respect to g_5 . These results were obtained by the method of augmented Lagrange multipliers implemented in the R package `Rsolnp` [35]. Similar to the *D*-criterion, the locally *IMSE*-optimal designs are seen to be minimally supported on $\mathbf{x}_1, \mathbf{x}_2$, and \mathbf{x}_3 when the standardized effect $\gamma = \beta/\beta_0$ is sufficiently large, but the optimal weights vary in contrast to the local *D*-criterion. All four vertices are required for smaller values of $\gamma \geq 0$. Note that this parameter region is considerably larger here than for the *D*-criterion. In the case $\gamma = 0$, the optimal weights are again uniform on all vertices. Moreover, by the additional reflection $g_2(\mathbf{x}) = (1 - x_1, 1 - x_2)^T$ the optimal weights can be transferred from $\gamma > 0$ to $\gamma < 0$ by the nonlinear transformation \tilde{g}_2 described in Sect. 4. For example, in the last column of Table 2, the locally *IMSE*-optimal design at $\tilde{\boldsymbol{\beta}} = (1, -3/7, -3/7)^T$ is obtained from the locally *IMSE*-optimal design at $\boldsymbol{\beta} = (1, 3, 3)^T$ by g_2 and the corresponding (nonlinear) transformation $\tilde{g}_2(\boldsymbol{\beta})$ which results in $\tilde{\gamma} = -\gamma/(1 + 2\gamma)$.

We now turn to maximin efficiency where invariance can become a powerful tool. For this, we additionally require that the subregion \mathcal{B}' of interest is also invariant with respect to the pair (g, \tilde{g}) of transformations, i.e., $\tilde{g}(\mathcal{B}') = \mathcal{B}'$ for all $g \in G$. As mentioned in Sect. 3, the homogeneous version of the *D*-criterion and the *IMSE*-criterion is multiplicatively equivariant. Hence, by (22) both maximin *D*- and *IMSE*-efficiency are invariant with respect to any group G of transformations satisfying the conditions of this section.

Corollary 8 *If $\tilde{g}(\mathcal{B}') = \mathcal{B}'$ for all $g \in G$, then there exists a maximin *D*-efficient design ξ^* on \mathcal{B}' which is invariant with respect to G .*

*If additionally ν is invariant with respect to G , then there exists a maximin *IMSE*-efficient design ξ^* on \mathcal{B}' with respect to ν which is invariant with respect to G .*

Example (Example 1 continued) In the one-factor gamma model with simple linear regression on $[0, 1]$, the invariant design $\bar{\xi}$ which assigns equal weights $1/2$ to the endpoints is both maximin D -efficient and maximin $IMSE$ -efficient with respect to the uniform measure ν on $[0, 1]$ on \mathcal{B} as has already been pointed out at the end of Sect. 3.

However, in contrast with the local criteria, there is no general majorization argument available for maximin efficiency criteria which allows restriction of the support of an optimal design to the extremal points of the experimental region. Therefore, to keep argumentation simple and to concentrate on the concept of invariance, we deliberately confine the support of the designs under consideration to these endpoints. Then, with respect to reflection, $g(x) = 1 - x$, the only invariant design which assigns equal weights $1/2$ to the endpoints is maximin efficient for any invariant criterion. In particular, this design is maximin $IMSE$ -efficient on \mathcal{B} with respect to any invariant weighting measure ν as specified in Proposition 1. \square

Although, generally, the determination of the efficiencies requires the knowledge of all locally optimal designs, the maximin efficient design may be constructed without this information as the above example shows. This result can be extended to more complex models.

Example (Example 2 continued) In the two-factor gamma model on $[0, 1]^2$ with multiple regression, we first consider maximin efficiency on the region \mathcal{B} of all possible values for the parameter vector. This region is invariant under the transformations associated with the group $G = \{g_1, \dots, g_4\}$ of reflections of the covariates. As in the case of the one-factor gamma model, we deliberately confine the support of the designs to the vertices $\mathbf{x}_1, \dots, \mathbf{x}_4$ of the experimental region to keep argumentation simple. Then, there is only one orbit which contains all vertices, and the only invariant design with respect to G is the uniform design $\bar{\xi}$ on the vertices which assigns equal weights $1/4$ to each vertex. Hence, the design $\bar{\xi}$ is maximin efficient on \mathcal{B} for any invariant criterion with respect to G .

This result carries over to any parameter subregion \mathcal{B}' which is invariant with respect to G . For example, if the intercept β_0 is restricted to a subset, $\beta_0 \in \mathcal{B}_0$, of its marginal region $(0, \infty)$ or, more specifically, set to a fixed value ($\mathcal{B}_0 = \{\beta_0\}$), while the slopes may vary across their corresponding (conditional) regions, then the resulting subregion $\mathcal{B}' = \{\beta \in \mathcal{B}; \beta_0 \in \mathcal{B}_0\}$ is invariant with respect to the rescaled transformations \tilde{g} associated with the transformations $g \in G$. Hence, the uniform design $\bar{\xi}$ is also maximin efficient on \mathcal{B}' for any invariant criterion with respect to G . In particular, this holds for the reduced parameter region \mathcal{C} displayed in Fig. 3 when $\beta_0 = 1$ is fixed. \square

Invariance can also be employed in cases where there are fewer symmetries, and thus, there is more than one orbit, so that the weights of the orbits still have to be optimized.

Example (Example 3 continued) In the two-factor gamma model on $[0, 1]^2$, we are now only interested in the parameter vectors β with equal slopes, i.e., $\beta_1 = \beta_2 = \beta$. We thus consider the parameter subregion $\mathcal{B}' = \{(\beta_0, \beta, \beta)^T; \beta > -\beta_0/2, \beta_0 > 0\}$. In terms of the reduced parameter $\gamma = (\gamma_1, \gamma_2)^T$, $\gamma_j = \beta_j/\beta_0$, the subset \mathcal{B}' reduces

to $C' = \{(\gamma, \gamma)^T : \gamma > -1/2\}$ which is exhibited in Fig. 3 by the diagonal dashed line.

For the transformation $g_2(\mathbf{x}) = (1 - x_1, 1 - x_2)^T$ of simultaneous reflection of both explanatory variables, we use the rescaled transformation $\tilde{g}_2(\boldsymbol{\beta}) = \tilde{c}_\beta \mathbf{Q}_{g_2}^{-T} \boldsymbol{\beta}$ of the parameter vector $\boldsymbol{\beta}$ which leaves the intercept β_0 unchanged, i.e., $\tilde{c}_\beta = \beta_0 / (\beta_0 + \beta_1 + \beta_2) = 1 / (\gamma_0 + \gamma_1 + \gamma_2)$ which specializes to $\tilde{c}_\beta = \beta_0 / (\beta_0 + 2\beta) = 1 / (1 + 2\gamma)$ on the subsets B' and C' , respectively. In particular, the relevant reduced slope parameter $\gamma > -1/2$ is mapped to $-\gamma / (1 + 2\gamma) > -1/2$ as mentioned before. Obviously, both B' and C' are invariant with respect to \tilde{g}_2 .

To make use of the symmetries with respect to the transformations g_2 and g_5 jointly, we consider the group $G' = \{\text{id}, g_2, g_5, g_6\}$ generated by them, where the composition g_6 of g_2 and g_5 is the reflection at the secondary diagonal of the unit square \mathcal{X} , $g_6(\mathbf{x}) = (1 - x_2, 1 - x_1)^T$. We again restrict attention to the vertices of the experimental region. Then there are just two orbits $\{\mathbf{x}_1, \mathbf{x}_4\}$ and $\{\mathbf{x}_2, \mathbf{x}_3\}$ of the group G' . The invariant designs $\tilde{\xi}_w$ can thus be characterized by the weight w assigned to each of the settings \mathbf{x}_1 and \mathbf{x}_4 in the first orbit, $0 < w < 1/2$, while weight $1/2 - w$ is assigned to each of the settings \mathbf{x}_2 and \mathbf{x}_3 in the second orbit. Design optimization is then reduced to determining the optimal weight w .

The determinant of the invariant design $\tilde{\xi}_w$ is given by

$$\det(\mathbf{M}(\tilde{\xi}_w; \boldsymbol{\beta})) = \frac{w(1 - 2w) ((1 + \gamma)^2 + \gamma^2(1 - 2w))}{2\beta_0^6(1 + \gamma)^4(1 + 2\gamma)^2}$$

locally at $\boldsymbol{\beta} = (\beta_0, \beta, \beta)^T$, where $\gamma = \beta / \beta_0$. To determine the maximin D -efficient design over B' , we can confine the analysis to the reduced parameter region C' , i.e., $\gamma > -1/2$. For $\gamma \geq 1$ the minimally supported design ξ_β^* with equal weights $1/3$ on $\mathbf{x}_1, \mathbf{x}_2$, and \mathbf{x}_3 is locally D -optimal and has $\det(\mathbf{M}(\xi_\beta^*; \boldsymbol{\beta})) = \beta_0^6(1 + \gamma)^4/27$. Hence, for the D -efficiency of $\tilde{\xi}_w$ at $\gamma \geq 1$, we get

$$\text{eff}_D(\tilde{\xi}_w; \boldsymbol{\beta})^3 = 27 \frac{w(1 - 2w) ((1 + \gamma)^2 + \gamma^2(1 - 2w))}{2(1 + 2\gamma)^2}. \tag{25}$$

The efficiency is decreasing in $\gamma \geq 1$. Therefore, its infimum $\inf_{\gamma \geq 1} \text{eff}_D(\tilde{\xi}_w; \boldsymbol{\beta})^3 = 27w(1 - w)(1 - 2w)/4$ is obtained when γ tends to ∞ . This expression is maximized by $w^* = (3 - \sqrt{3})/6 = 0.2113$, and the design $\tilde{\xi}_{w^*}$ is maximin D -efficient over $\gamma \geq 1$ within the class of invariant designs $\tilde{\xi}_w$. This also holds for the parameter region $-1/2 < \gamma \leq -1/3$ by symmetry considerations with respect to the transformation g_2 (or g_6). For the intermediate region $(-1/3 < \gamma < 1)$, the efficiency has to be computed numerically. The D -efficiency of $\tilde{\xi}_{w^*}$ is displayed in Fig. 5. By inspection of the plot, it can be concluded that $\tilde{\xi}_{w^*}$ is maximin D -efficient on C' and its minimal D -efficiency $3(w^*(1 - w^*)(1 - 2w^*)/4)^{1/3} = 0.8660$ is attained at the boundary of the parameter region. This result carries over to the whole parameter region B' of equal slopes as well as to subregions $\{\boldsymbol{\beta} \in B'; \beta_0 \in \mathcal{B}_0\}$ with constraints on the intercept β_0 .

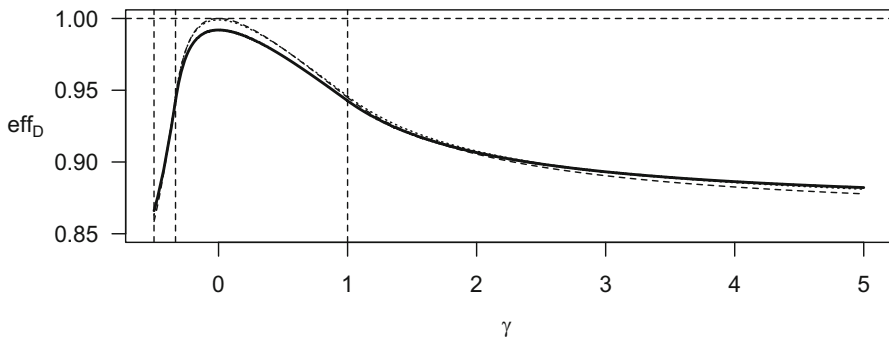


Fig. 5 D -efficiency of the maximin D -efficient design $\bar{\xi}_{w^*}$ (solid line) and the uniform design $\bar{\xi}_{1/4}$ (dashed line); the vertical dashed lines indicate the lower bound ($\gamma = -1/2$) and the thresholds $\gamma = -1/3$ and $\gamma = 1$ between the subregions

For comparison, the D -efficiency is also plotted in Fig. 5 for the uniform design $\bar{\xi}_{1/4}$ which is locally optimal at $\gamma = 0$. The maximin D -efficiency of $\bar{\xi}_{1/4}$ is also attained at the boundary and can be computed to be 0.8585 by (25). Hence, the maximin D -efficiency of $\bar{\xi}_{1/4}$ is slightly worse than for the maximin D -efficient design $\bar{\xi}_{w^*}$. \square

6 Discussion

In this article, we present an outline of the concept of invariance and equivariance in the design of experiments for generalized linear models. In contrast with the well-known results in linear models, where only the experimental settings are transformed, we have to consider pairs of transformations in generalized linear models which act simultaneously on the experimental settings and on the location parameters in the linear component. We focus on local optimality and maximin efficiency for the common D - and $IMSE$ -criteria which allow a wide range of transformations for the experimental settings like scaling, permutations or reflections. As in linear models, the transformation of the experimental settings has to act in a linear way on the regression functions of the linear component. The parameters can then be transformed linearly in such a way that the value of the linear component and, hence, of the intensity is not changed (see Radloff and Schwabe [34]). Besides this natural choice, nonlinear transformations of the parameters may also be employed, if additional properties of the intensity function can be used. We illustrate this feature by the gamma model with inverse link for which the intensity is only scaled by a multiplicative factor based on the parameter. This scaling does not affect standardized design criteria like maximin efficiency, and invariance can also be used here. In Table 3, we exhibit which concepts of equivariance and invariance can be used under the model conditions of linear and generalized linear models and, in particular, for the gamma model with canonical (inverse) link.

The general results on equivariance and invariance in generalized linear models can be extended in a straightforward manner to other model specifications in which the intensity depends only on the linear component, as in censoring (see Schmidt and

Table 3 Transformations for equivariance (Equiv.) and invariance (Inv.); “+” indicates whether this property is required / can be used

Transformation	Linear model		GLM		Gamma (inverse link)	
	Equiv.	Inv.	Equiv.	Inv.	Equiv.	Inv.
g	$\mathcal{X} \rightarrow \mathcal{Z}$	$\mathcal{X} \rightarrow \mathcal{X}$	$\mathcal{X} \rightarrow \mathcal{Z}$	$\mathcal{X} \rightarrow \mathcal{X}$	$\mathcal{X} \rightarrow \mathcal{Z}$	$\mathcal{X} \rightarrow \mathcal{X}$
$\mathbf{f} \circ g = \mathbf{Q}_g \mathbf{f}$	+	+	+	+	+	+
Group G		+		+		+
\tilde{g}			$\mathcal{B} \rightarrow \tilde{\mathcal{B}}$	$\mathcal{B} \rightarrow \mathcal{B}$	$\mathcal{B} \rightarrow \tilde{\mathcal{B}}$	$\mathcal{B} \rightarrow \mathcal{B}$
Linear			\mathbf{Q}_g^{-T}	\mathbf{Q}_g^{-T}	\mathbf{Q}_g^{-T}	\mathbf{Q}_g^{-T}
Group (G, \tilde{G})				+		+
Nonlinear					$\tilde{c} \cdot \mathbf{Q}_g^{-T}$	$\tilde{c} \cdot \mathbf{Q}_g^{-T}$

Schwabe [37], for examples). How far the results on nonlinear transformations can be extended, however, depends on the structure of the intensity function.

For other optimality criteria, such as A -, E - or, more generally, Kiefer’s Φ_q -criteria which are based on the eigenvalues of the information matrix, the use of equivariance and invariance is limited, because additional structures of the transformations would be required, like orthogonality of the transformation matrices \mathbf{Q}_g .

For the case of maximin efficiency in the gamma model, it would be also desirable to obtain the majorization results as we have found for local optimality. These would allow us to restrict the search for the optimal experimental settings to the extremal points of the experimental region. However, the findings in Gaffke et al. [14] do not carry over, because the arguments used there are of a local nature and do not work uniformly on the parameter region. Alternatively, equivalence theorems could be employed for establishing maximin efficiency (see Pronzato and Pázman [31, ch. 8]), but in their formulation these theorems require that the minimal efficiency is attained inside the parameter region which is violated in our examples. It thus remains an open problem whether the restriction to the extremal points of the experimental region can be justified.

The concepts of equivariance and invariance can further be extended to models with random effects (see Graßhoff et al. [16] and Debusho and Haines [7] for the estimation of population parameters, and Prus and Schwabe [32] for individual prediction in linear mixed models).

Appendix

Proof of Proposition 1 First note that by the majorization argument in Gaffke et al. [14] there is an optimal design of the form ξ_w which assigns weight w to 1 and weight $1 - w$ to 0. The information matrix for ξ_w and its inverse are

$$\mathbf{M}(\xi_w; \boldsymbol{\beta}) = \begin{pmatrix} (1-w)/a^2 + w/b^2 & w/b^2 \\ w/b^2 & w/b^2 \end{pmatrix} \quad \text{and}$$

$$\mathbf{M}(\xi_w; \boldsymbol{\beta})^{-1} = \begin{pmatrix} a^2/(1-w) & -a^2/(1-w) \\ -a^2/(1-w) & a^2/(1-w) + b^2/w \end{pmatrix},$$

where $a = \beta_0$ and $b = \beta_0 + \beta_1$ for simplification.

For given measure ν , denote by $v_k = \int x^k / (\beta_0 + \beta_1 x)^4 \nu(dx)$, $k = 0, 1, 2$ the entries in the weighting matrix $\mathbf{V}(\boldsymbol{\beta}; \nu) = \begin{pmatrix} v_0 & v_1 \\ v_1 & v_2 \end{pmatrix}$. Then, the *IMSE*-criterion becomes

$$\begin{aligned} \text{IMSE}(\xi_w; \boldsymbol{\beta}, \nu) &= \text{trace}(\mathbf{V}(\boldsymbol{\beta}; \nu)\mathbf{M}(\xi_w; \boldsymbol{\beta})^{-1}) \\ &= (1/(1-w))a^2(v_0 - 2v_1 + v_2) + (1/w)b^2v_2. \end{aligned}$$

To obtain the v_k , we introduce the auxiliary terms $c_k = \int 1/(\beta_0 + \beta_1 x)^k \nu(dx)$, $k = 2, 3, 4$. Then, the v_k can be represented as

$$v_0 = c_4, \quad v_1 = (c_3 - ac_4)/(b - a) \quad \text{and} \quad v_2 = (c_2 - 2ac_3 + a^2c_4)/(b - a)^2,$$

if $\beta_1 = b - a \neq 0$. The case $\beta_1 = 0$, when the intensity is constant, follows directly from the corresponding linear model.

(a) For the (continuous) uniform measure ν on $[0, 1]$ we have $c_k = (a^{-k+1} - b^{-k+1})/((k - 1)(b - a))$ and, hence,

$$v_0 = \frac{a^2 + ab + b^2}{3a^3b^3}, \quad v_1 = \frac{2a + b}{6a^2b^3} \quad \text{and} \quad v_2 = \frac{1}{3ab^3}.$$

From this, we get

$$\text{IMSE}(\xi_w; \boldsymbol{\beta}, \nu) = \frac{1}{3ab} \left(\frac{1}{w} + \frac{1}{1-w} \right)$$

which is optimized for $w^* = 1/2$.

(b) For the (discrete) uniform measure ν on $\{0, 1\}$, we have $c_k = (a^{-k} + b^{-k})/2$ and, hence,

$$v_0 = (a^{-4} + b^{-4})/2, \quad \text{and} \quad v_1 = v_2 = b^{-4}/2.$$

From this, we get

$$\text{IMSE}(\xi_w; \boldsymbol{\beta}, \nu) = \frac{1}{2} \left(\frac{1}{b^2w} + \frac{1}{a^2(1-w)} \right)$$

which is optimized for $w^* = a/(a + b)$.

(c) For the one-point measure ν at $1/2$, we have $c_k = \frac{2^k}{(a+b)^k}$ and, hence, $v_k = 2^{4-k}/(a + b)^4$. From this, we get

$$\text{IMSE}(\xi_w; \boldsymbol{\beta}, \nu) = \frac{4}{(a + b)^4} \left(\frac{a^2}{1-w} + \frac{b^2}{w} \right)$$

which is optimized for $w^* = b/(a + b)$.

Acknowledgements The first author acknowledges support by the fund of the Graduate Academy, Otto-von-Guericke-University Magdeburg.

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Atkinson A, Haines L (1996) Designs for nonlinear and generalized linear models. In: Ghosh S, Rao C (eds) Design and analysis of experiments. Handbook of statistics, vol 13. Elsevier, Amsterdam, pp 437–475
2. Atkinson AC, Donev AN, Tobias RD (2007) Optimum experimental designs, with SAS. Oxford University Press, Oxford
3. Atkinson AC, Woods DC (2015) Designs for generalized linear models. In: Dean A, Morris M, Stufken J, Bingham D (eds) Handbook of design and analysis of experiments. Chapman & Hall/CRC Press, Boca Raton, pp 471–514
4. Burrige J, Sebastiani P (1994) D-optimal designs for generalised linear models with variance proportional to the square of the mean. *Biometrika* 81:295–304
5. Chernoff H (1953) Locally optimal designs for estimating parameters. *Ann Math Stat* 24:586–602
6. Cox DR (1988) A note on design when response has an exponential family distribution. *Biometrika* 75:161–164
7. Debusho LK, Haines LM (2010) D-optimal population designs for the simple linear random coefficients regression model. *Annu Proc S Afr Stat Assoc Conf* 2010:52–59
8. Dette H (1997) Designing experiments with respect to ‘standardized’ optimality criteria. *J R Stat Soc Ser B* 59:97–110
9. Dette H, Hoyden L, Kuhnt S, Schorning K (2017) Optimal designs for thermal spraying. *J R Stat Soc Ser C* 66:53–72
10. Dette H, Martinez Lopez I, Ortiz Rodriguez IM, Pepelyshev A (2006) Maximin efficient design of experiment for exponential regression models. *J Stat Plan Inference* 136:4397–4418
11. Fahrmeir L, Kaufmann H (1985) Consistency and asymptotic normality of the maximum likelihood estimator in generalized linear models. *Ann Stat* 13:342–368
12. Fedorov VV, Leonov SL (2013) Optimal design for nonlinear response models. CRC Press, Boca Raton
13. Ford I, Torsney B, Wu CFJ (1992) The use of a canonical form in the construction of locally optimal designs for non-linear problems. *J R Stat Soc B* 54:569–583
14. Gaffke N, Idais O, Schwabe R (2019) Locally optimal designs for gamma models. *J Stat Plan Inference* 203:199–214
15. Gea-Izquierdo G, Cañellas I (2009) Analysis of holm oak intraspecific competition using gamma regression. *For Sci* 55:310–322
16. Graßhoff, U, Holling, H, Schwabe, R (2009). On optimal design for a heteroscedastic model arising from random coefficients, in: Proceedings of the 6th St. Petersburg Workshop on Simulation, pp. 387–392.

17. Graßhoff U, Schwabe R (2008) Optimal design for the Bradley-Terry paired comparison model. *Stat Methods Appl* 17:275–289
18. Grover G, Sabharwal A, Mittal J (2013) An application of gamma generalized linear model for estimation of survival function of diabetic nephropathy patients. *Int J Stat Med Res* 2:209–219
19. Heiligers B, Schneider K (1992) Invariant admissible and optimal designs in cubic regression on the v-ball. *J Stat Plan Inference* 31:113–125
20. Idais, O (2021) On local optimality of vertex type designs in generalized linear models. *Stat Pap* 62: 1871–1898.
21. Idais O, Schwabe R (2021) Analytic solutions for locally optimal designs for gamma models having linear predictors without intercept. *Metrika* 84:1–26
22. Imhof L, Wong WK (2000) A graphical method for finding maximin efficiency designs. *Biometrics* 56:113–117
23. Kiefer J (1959) Optimum experimental designs. *J R Stat Soc B* 21:272–304
24. Lehmann E (1959) *Testing statistical hypotheses*. Wiley, New York
25. Li Y, Deng X (2018) On I-optimal designs for generalized linear models: an efficient algorithm via general equivalence theory. *arXiv preprint arXiv:1801.05861*
26. Li, Y. and Deng, X. (2021), An efficient algorithm for Elastic I-optimal design of generalized linear models. *Can J Statistics*, 49:438–470.
27. McCullagh P, Nelder JA (1989) *Generalized linear models*, 2nd edn. Chapman & Hall, London
28. Nelder JA, Wedderburn RWM (1972) *Generalized linear models*. *J R Stat Soc Ser A* 135:370–384
29. Ng VK, Cribbie RA (2017) Using the gamma generalized linear model for modeling continuous, skewed and heteroscedastic outcomes in psychology. *Curr Psychol* 36:225–235
30. Pitman E (1939) The estimation of the location and scale parameters of a continuous population of any given form. *Biometrika* 30(3–4):391–421
31. Pronzato L, Pázman A (2013) *Design of experiments in nonlinear models: asymptotic normality. Optimality criteria and small-sample properties*. Springer, New York
32. Prus M, Schwabe R (2016) Optimal designs for the prediction of individual parameters in hierarchical models. *J R Stat Soc B* 78:175–191
33. Pukelsheim F (1993) *Optimal design of experiments*. Wiley, New York
34. Radloff M, Schwabe R (2016) Invariance and equivariance in experimental design for nonlinear models. In: Kunert J, Müller CH, Atkinson AC (eds) *mODA 11—advances in model-oriented design and analysis*. Springer, New York, pp 217–224
35. R Core Team (2020) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria
36. Russell KG, Woods DC, Lewis SM, Eccleston JA (2009) D-optimal designs for Poisson regression models. *Stat Sin* 19:721–730
37. Schmidt D, Schwabe R (2017) Optimal design for multiple regression with information driven by the linear predictor. *Stat Sin* 27:1371–1384
38. Schwabe R (1996) *Optimum designs for multi-factor models*. Springer, New York
39. Silvey SD (1980) *Optimal design*. Chapman & Hall, London
40. Sitter RR, Torsney B (1995) D-optimal designs for generalized linear models. In: Kitsos CP, Müller WG (eds) *MODA4—advances in model-oriented data analysis*. Physica-Verlag, Heidelberg, pp 87–102
41. Tong L, Volkmer HW, Yang J (2014) Analytic solutions for D-optimal factorial designs under generalized linear models. *Electron J Stat* 8:1322–1344
42. Yang M, Stufken J (2009) Support points of locally optimal designs for nonlinear models with two parameters. *Ann Stat* 37:518–541
43. Yang M, Zhang B, Huang S (2011) Optimal designs for generalized linear models with multiple design variables. *Stat Sin* 21:1415–1430

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.