Technical Report

Unimodal regression using Bernstein-Schoenberg-splines and penalties

Claudia Köllmann[1,*], Björn Bornkamp[2], and Katja Ickstadt[1]

[1]Department of Statistics, TU Dortmund University, Germany

[2]Novartis Pharma AG, Basel, Switzerland

* Email: koellmann@statistik.tu-dortmund.de

6/2012

technische universität dortmund

SFB 876 Verfügbarkeit von Information durch Analyse unter Ressourcenbeschränkung

This work was partly supported by the Research Training School on Statistical Modelling of the German Research Foundation (DFG) and by the German Research Foundation (DFG) within the Collaborative Research Center SFB 876 "Providing Information by Resource-Constrained Analysis", project C4.

Speaker: Prof. Dr. Katharina Morik
Address: TU Dortmund University
Joseph-von-Fraunhofer-Str. 23
D-44227 Dortmund
Web: http://sfb876.tu-dortmund.de

Abstract

Research in the field of nonparametric shape constrained regression has been intensive. However, only few publications explicitly deal with unimodality although there is need for such methods in applications, for example, in dose-response analysis. In this paper we propose unimodal spline regression methods that make use of Bernstein-Schoenberg-splines and their shape preservation property. To achieve unimodal and smooth solutions we use penalized splines, and extend the penalized spline approach towards penalizing against general parametric functions, instead of using just difference penalties. For tuning parameter selection under a unimodality constraint a restricted maximum likelihood and an alternative Bayesian approach for unimodal regression are developed. We compare the proposed methodologies to other common approaches in a simulation study and apply it to a dose-response data set. All results suggest that the unimodality constraint or the combination of unimodality and a penalty can substantially improve estimation of the functional relationship.

1 Introduction

In a variety of applications a dependent variable increases with higher values of an independent variable up to a maximum and then decreases again, *i.e.*, the functional relationship is unimodal. A prominent example is dose-response analysis, where the (beneficial) effect of a substance increases with increasing dose up to a saturation point, after which the effect starts to decrease again, as the substance might cause, for example, interfering toxic effects.

While there exist a variety of parametric approaches to estimate a unimodal relationship, we introduce in this paper a flexible semiparametric method for estimation of a smooth unimodal function, based on spline functions which include polynomials as a special case and which are particularly well-suited for shape-constrained function estimation.

Most common shape constraints used in the context of splines (and polynomials) are monotonicity, convexity or concavity and log-concavity, because finite dimensional constraints on the spline coefficients ensure the desired shape constraint. See, for example, Ramsay (1988), Kelly and Rice (1990), Wood (1994), Hazelton and Turlach (2011) and Wang and Ghosh (2012) for different approaches. Essentially, these shape-constraints induce non-negativity constraints on a derivative which can be ensured using constrained optimization in non-Bayesian approaches (see e.g. Ramsay (1988) or Wang and Ghosh (2012)) and by prior specification in Bayesian approaches (see e.g. Hazelton and Turlach (2011)). The shape constraint of log-concavity (which also guarantees unimodality) can be achieved by imposing concavity on the logarithm of the response, as, for example, described in Eilers (2005). Yet, log-concavity is a stronger requirement and there exist unimodal functions that are not log-concave.

However, the unimodality constraint does not reduce to a single positivity constraint on a derivative of the modelled function and has received less attention in the spline literature so far. In an unpublished manuscript, Woodworth (1999) uses B-splines with a certain hierarchical prior on the coefficients that guarantees unimodality. Unfortunately, no general suggestions on the choice of the prior parameters and the knot sequence are given. The incorporation of a smoothness penalty, which could be used to address the problem of knot placement, is also not straightforward in this model.

Even beyond splines, the literature on nonparametric estimation of a smooth unimodal function is relatively sparse. The first approaches to unimodal regression date back to the 1980s, see, for example, Frisén (1986), where *pointwise* least squares estimates are produced by successively splitting the data into two possible subsets, applying an isotonic and an antitonic regression and choosing the split with the minimal sum of squares. To obtain a smooth function the pointwise estimates have to be interpolated or smoothed in a second step. Even joining two *smooth* monotone estimates will not help, because it yields estimates which are not smooth or even discontinuous at the mode. The problem is that the shape constraint is imposed using two local monotonicity constraints instead of one global constraint.

A different, but very general approach to incorporate shape constraints is data sharpening, where the data points are shifted as little as possible so that the unconstrained estimate fulfils the constraint. See, for example, Braun and Hall (2001) for an approach to unimodal density estimation.

The outline of this paper is as follows: In Section 2 we first introduce a dose-response data set from McLaren et al. (1990) to further motivate the need for unimodal regression in real applications. In Section 3.1 we illustrate the approximation power and the shape-preservation properties of splines using the characteristics of Bernstein-Schoenberg splines. The remainder of Section 3 then describes methods based on maximum likelihood and penalized maximum likelihood with different penalties and a Bayesian approach towards estimation of unimodal splines. Regarding the form of the penalty, which can also be viewed as prior information from the Bayesian perspective, we first extend the finite difference penalties described in Eilers and Marx (1996) to the case of unimodal regression. In a second step, we propose a novel approach to incorporate prior information by penalizing against a parametric model, while still allowing for departures from it, when the data suggest so. All proposed methods are implemented in R and the code is freely available on CRAN (package `uniReg`) or from the authors. In Section 4 the introduced methods are applied to the data example from Section 2 and their characteristics are compared in a simulation study. The last section contains conclusions and an outlook.

2 Unimodality in Dose-Response Data

Characterization of the dose-response relationship for desirable and undesirable effects of a pharmaceutical compound is the central problem of its clinical development. The prespecification of one dose-response model for analysis in the study protocol (before data collection) is difficult, which is why practical methods often rely on specification of a candidate set of parametric dose-response models (see e.g. Bretz et al., 2005) and on model selection or model averaging. Unimodal regression is a nonparametric competitor to these techniques.

A typical assumption in parametric as well as nonparametric dose-response analyses is monotonicity. However, in a variety of cases this assumption can be challenged as the interference of potential saturation or toxicity effects cannot be excluded. When considering a clinical utility index that combines efficacy and safety measures (see e.g. Khan et al., 2009) one explicitly expects a unimodal relationship and a monotonically increasing curve would be surprising to observe. But even when considering efficacy alone unimodality can occur as in the example to follow. A unimodal shape constraint relaxes the assumption of monotonicity and is adequate whenever an umbrella dose-response curve cannot be excluded a priori.

The example data set originates from animal science, where the growth of pigs is evaluated in dependence of an increasing dose of a growth hormone. McLaren et al. (1990) investigated the relationship between administration of porcine somatotropin and several growth variables in 195 pigs. Details on the experimental procedure and data preprocessing can be found in their manuscript. The (aggregated) data used here are the porcine somatotropin dosage levels [mg/pig/day] (PST) and the least squares means and standard deviations of four response variables: Average daily gain of weight [kg/day] (ADG), age at 103.5 kg [days] (Age), gain-to-feed ratio (G/F) and average daily feed consumption [kg/pig/day] (ADF). The five dosage levels are 0, 1.5, 3, 6, 9 mg/pig/day and the means and standard deviations at the respective levels correspond to 29, 29, 57, 58, and 22 pigs. The data are plotted in Figure 1 and the actual data values can be found in Table H.1. While the modes of the means of ADG, Age and ADF are at extreme doses (suggesting monotone relationships), the means of G/F have an interior mode at dose 6. Since mono-

tonicity is a special case of unimodality, it seems reasonable to relax the monotonicity assumption and apply unimodal regression to all four variables (inverse unimodal for the variables Age and ADF). The results are presented in Section 4.1.

3 Methods

3.1 Bernstein-Schoenberg Splines and Shape-Constraints

In this subsection we review some basic properties of Bernstein-Schoenberg splines (see, for example, Goodman, 1995), to illustrate that a particular choice of B-splines is (i) a sufficiently flexible basis and (ii) well-suited for shape-constrained modelling, as a simple constraint on the B-spline coefficients ensures unimodality of a spline.

Let $N_{j,k+1}(x)$ be the (normalized) B-spline basis function of degree $k \geq 1$ with knots $\tau_j, \ldots, \tau_{j+k+1}$, which can be defined by the following recursion formulae: $N_{j,1}(x) = I_{[\tau_j, \tau_{j+1})}(x)$, $N_{j,k+1}(x) = \frac{x-\tau_j}{\tau_{j+k}-\tau_j} N_{j,k}(x) + \frac{\tau_{j+k+1}-x}{\tau_{j+k+1}-\tau_{j+1}} N_{j+1,k}(x)$ for $j = -k, \ldots, g$. Explicit expressions are also available for the B-spline basis functions, but numerically instable to evaluate, which is why only the recursive definition is presented here. See Dierckx (1993) for an introduction to the B-spline basis, or see Appendix I for a graphical representation. The full sequence of knots for all basis splines is denoted by $\mathcal{T} = (\tau_j)_{-k}^{g+k+1}$ and is here restricted to $\tau_{-k} = \ldots = \tau_0 = 0 < \tau_1 \leq \ldots \leq \tau_g < 1 = \tau_{g+1} = \ldots = \tau_{g+k+1}$ (coincident boundary knots) and $\tau_j < \tau_{j+k+1} \; \forall \; j = -k, \ldots, g$, where $g \geq 0$ is the number of inner knots.

Every linear combination of B-spline basis functions, $\sum_{j=-k}^{g} \beta_j N_{j,k+1}(x)$, with $d := g + k + 1$ so-called B-spline coefficients β_j, belongs to the class of spline functions. The Bernstein-Schoenberg (B-S) operator or Bernstein-Schoenberg spline of a function f has specific coefficients and is defined for $x \in [0, 1]$ as $\mathcal{V}_k^{\mathcal{T}} f(x) = \sum_{j=-k}^{g} f(\tau_j^*) N_{j,k+1}(x)$, where $\tau_j^* = \frac{1}{k} \sum_{i=1}^{k} \tau_{j+i}$, $j = -k, \ldots, g$, are the so-called knot averages.

Important properties of the B-S splines are:

(1) For $w = 0, 1$ it holds that, if $f \in \mathcal{C}^w[0,1]$, then $\lim_{k\to\infty} (\mathcal{V}_k^{\mathcal{T}} f)^{(w)} = f^{(w)}$ uniformly on $[0,1]$.

(2) The number of sign changes of the B-S spline of a function f is not larger than the number of sign changes of f itself.

Moreover, it can be shown, that the number of sign changes of a spline $\sum_{j=-k}^{g} \beta_j N_{j,k+1}(x)$ is not larger than the number of sign changes of the coefficient sequence $(\beta_{-k}, \ldots, \beta_g)$. This is called the *variation diminishing property*, which was introduced by Schoenberg (1967). Carnicer and Pena (1994) even found the B-spline basis to be optimally shape preserving for the space of spline functions. This property makes B-S splines very interesting for shape-constrained regression. Concerning unimodality we can derive the following lemma. To our knowledge its contents have not been stated in the literature in this form and thus, we give a proof for it in Appendix A.

LEMMA 1: *Let f be a unimodal function on $[0,1]$ and $\mathcal{T} = (\tau_j)_{-k}^{g+k+1}$ a knot sequence with distinct inner knots and $\tau_{-k} = \ldots = \tau_0 = 0, \quad \tau_{g+1} = \ldots = \tau_{g+k+1} = 1$. Then, the corresponding Bernstein-Schoenberg spline, $\mathcal{V}_k^{\mathcal{T}} f(x) = \sum_{j=-k}^{g} f(\tau_j^*) N_{j,k+1}(x)$, is unimodal, too.*

The reverse implication does not necessarily hold, as can be seen from the counter example in Appendix B.

For $g = 0$ the B-spline basis functions $N_{j,k+1}(x)$ $(j = -k, \ldots, 0)$ reduce to the binomial probabilities $\binom{k}{\ell} x^\ell (1-x)^{k-\ell}$ $(\ell = 0, \ldots, k)$ (see Figures I.1 and I.2) and the B-S splines reduce to the Bernstein polynomials, $\mathcal{B}_k f(x) = \sum_{\ell=0}^{k} f\left(\frac{\ell}{k}\right) \binom{k}{\ell} x^\ell (1-x)^{k-\ell}$, which have been increasingly employed for shape-constrained regression, see, for example, Chang et al. (2005) or more recently Wang and Ghosh (2012). Here, we prefer B-S splines since they are a straightforward, more flexible generalization of Bernstein polynomials and convergence faster than the latter for suitably chosen knot sequences (cp. Marsden, 1970).

In summary, a unimodal function $f \in \mathcal{C}^w[0,1]$ $(w = 0, 1)$ can be approximated by a spline $s(x) = \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x)$ with $\beta_j = f(\tau_j^*)$ with uniform convergence properties, which is unimodal because of the shape-preserving property. The same is true for functions on

arbitrary intervals $[a, b]$ since their support can always be transformed to $[0, 1]$.

B-S splines are a tool in approximation theory problems, but will be used here for a regression problem. Since the functional relationship f between predictor and response is unknown, we cannot choose $\beta_j = f(\tau_j^*)$, but have to estimate the B-spline coefficients $\boldsymbol{\beta} = (\beta_{-k}, \ldots, \beta_g)$ from data. As follows from Lemma 1, it is sufficient to restrict the B-spline coefficients to form a unimodal sequence, i.e., that there exists an index $m \in \{-k, \ldots, g\}$ for which $\beta_{-k} \leq \ldots \leq \beta_{m-1} \leq \beta_m \geq \beta_{m+1} \geq \ldots \geq \beta_g$, to ensure unimodality of the fitted spline. In the following, new fitting procedures are derived for the case when a unimodality constraint is desired.

3.2 Maximum Likelihood

Suppose there are n pairs of observations (x_i, y_i), $x_i \in [a, b]$ $\forall i$, and the underlying model is $Y_i = s(x_i) + \epsilon_i = \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x_i) + \epsilon_i$ with $\epsilon_i \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$.

When using the maximum likelihood approach the aim is to find a spline function s that minimizes the sum of squared residuals $\sum_{i=1}^{n}(y_i - s(x_i))^2 = \|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}\|_2^2$, where $\boldsymbol{y} = (y_1, \ldots, y_n)'$ and $\boldsymbol{B} = (N_{j,k+1}(x_i))_{i=1,\ldots,n, j=-k,\ldots,g}$ is the matrix of B-spline basis functions evaluated at the observation points $x_1, \ldots, x_n$. The minimization problem $\arg\min_{\boldsymbol{\beta}} \|\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}\|_2^2$ is quadratic in $\boldsymbol{\beta}$, if the spline degree, the number of knots and the positions of the knots is fixed. Without imposing shape restrictions and if the number of distinct x-values is greater or equal d, $\boldsymbol{B}'\boldsymbol{B}$ is invertible and the unique solution is simply $\hat{\boldsymbol{\beta}} = (\boldsymbol{B}'\boldsymbol{B})^{-1}\boldsymbol{B}'\boldsymbol{y}$.

Suppose now that the shape of the spline is constrained to be unimodal with a fixed mode m of the B-spline coefficients. Under this constraint, the minimization has to be done for the set of all vectors $\boldsymbol{\beta}$ satisfying $\beta_j \geq \beta_{j-1}$ $\forall j = -k, \ldots, m$ and $\beta_j \leq \beta_{j-1}$

$\forall j = m+1, \ldots, g$. This can be written as the linear constraint $\boldsymbol{A}'_m\boldsymbol{\beta} \geq \mathbf{0}$ with

$$\boldsymbol{A}'_m = \begin{pmatrix} -1 & 1 & & & & & \\ & \ddots & \ddots & & & \mathbf{0} & \\ & & -1 & 1 & & & \\ & & & 1 & -1 & & \\ & \mathbf{0} & & & \ddots & \ddots & \\ & & & & & 1 & -1 \end{pmatrix} \leftarrow m\text{-th row} \quad \in \mathbb{R}^{(d-1)\times d}.$$

This quadratic programming problem has no explicit solution, but can be solved efficiently and reliably, see for example the R package `quadprog` (Turlach and Weingessel, 2011). Usually the mode of the coefficient vector will be unknown and must been learned from data. We propose to fit a unimodal spline for each possible choice of $m \in \{-k, \ldots, g\}$ and then select the fit with the lowest residual sum of squares. This idea is similar to the one in Frisén (1986), but a global shape constraint is used instead of two local ones. An alternative is to take a (Bayesian) average over different choices of the mode, which is done in Section 3.5.

Another problem arises already before fitting any least squares spline: the choices for the degree of the spline and the number and the positions of the knots. Regarding the spline degree we propose to use splines of degree $k = 3$ as recommended by Dierckx (1993). As for the number and the position of the knots we prefer a penalized spline approach that will be explained in the following subsection.

3.3 Penalized Maximum Likelihood

Eilers and Marx (1996) proposed to use a relatively large number of knot positions (compared to the number of the predictor values and the variation in the data) and a penalty term in the objective function to find a compromise between over- and underfitting. In their approach a finite difference penalty on the B-spline coefficients is used, that is,

$$\left\| \frac{1}{\sigma}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}) \right\|_2^2 + \lambda \sum_{j=-k+q}^{g} (\Delta^q \beta_j)^2 \tag{1}$$

is minimized, where $\Delta\beta_j = \beta_j - \beta_{j-1}$, $\Delta^2\beta_j = \Delta(\Delta\beta_j) = \beta_j - 2\beta_{j-1} + \beta_{j-2}$ and so on. The parameter $\lambda > 0$ enables tuning of the penalization. In matrix notation the objective

function is given by $\left\|\frac{1}{\sigma}(\boldsymbol{y}-\boldsymbol{B}\boldsymbol{\beta})\right\|_2^2 + \lambda \left\|\boldsymbol{D}_q\boldsymbol{\beta}\right\|_2^2$, where $\boldsymbol{D}_q \in \mathbb{R}^{(d-q)\times d}$ is the matrix representation of the finite differences of order q, for example,

$$\boldsymbol{D}_2 = \begin{pmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{pmatrix} \in \mathbb{R}^{(d-2)\times d}.$$

Eilers and Marx (1996) propose to use second order differences ($q = 2$) and a knot sequence that is equidistant even beyond the boundary knots, that is, $\tau_j = a + j\frac{b-a}{g+1} \quad \forall\, j = -k, \ldots, g$. In this case the penalty term is zero for linear in- or decreasing coefficient sequences and thus, one penalizes against linear functions. So for the difference penalty we deviate from the knot sequence definition in Section 3.1.

To enable inclusion of other penalties (see also Section 3.4) we will use a generalized penalty term, i.e., $\lambda \left\|\boldsymbol{\Omega}^{\frac{1}{2}}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\right\|_2^2$, in the following. In the case of the difference penalty we thus have $\boldsymbol{\beta}_0 = \mathbf{0}$ and $\boldsymbol{\Omega} = \boldsymbol{D}_q'\boldsymbol{D}_q$.

For fixed λ and σ and without imposing a shape constraint the penalized objective function is minimized by $\hat{\boldsymbol{\beta}} = \left(\frac{1}{\sigma^2}\boldsymbol{B}'\boldsymbol{B} + \lambda\boldsymbol{\Omega}\right)^{-1}\left(\frac{1}{\sigma^2}\boldsymbol{B}'\boldsymbol{y} + \lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right)$. The constrained problem can again be solved using quadratic programming algorithms and performing model selection over the mode using the residual sum of squares criterion.

In real applications the standard deviation σ is usually unknown and has to be estimated as well. If it is possible to attain an accurate estimate from preceeding trials or from the data itself, as, for example, in dose-response trials, then the above process works fine. In other cases, it is still possible to iterate between estimation of σ given an interim estimate of $\boldsymbol{\beta}$ and estimation of $\boldsymbol{\beta}$ given an interim estimate of σ until some defined convergence.

As mentioned before, we have to choose the tuning parameter λ so that the penalization gives a compromise between overfitting and underfitting. In simple regression problems the method of choice is often leave-one-out cross-validation, where the hat matrix $\boldsymbol{H}$ for which $\hat{\boldsymbol{y}} = \boldsymbol{H}\boldsymbol{y}$ can be used as a shortcut for the estimation of the tuning parameter. But when estimating $\boldsymbol{\beta}$ (and calculating $\hat{\boldsymbol{y}}$) under the unimodality constraint, such a matrix and thus a similar simple way of calculating the cross-validated tuning parameter is not available. In what follows we will describe a restricted maximum-likelihood (REML)

approach towards estimation of λ, as this can relatively straightforwardly be extended to the constrained case. We follow an approach, for example, taken by Wood (2011) in other penalized likelihood problems, where no shape contraint is present, and assume that $\boldsymbol{\beta}$ is distributed according to the following (possibly improper) prior: $\boldsymbol{\beta}|\lambda \sim \mathcal{N}_{\mathcal{S}_m}(\boldsymbol{\beta}_0, \lambda^{-1}\boldsymbol{\Omega}^-)$, where $\boldsymbol{\Omega}^- \in \mathbb{R}^{d\times d}$ is the pseudo-inverse of $\boldsymbol{\Omega}$ if $r := \text{rank}(\boldsymbol{\Omega}) < d$, and the regular inverse otherwise, and $\boldsymbol{\Omega}$ and $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ correspond to the respective components in the penalty term. The notation $\mathcal{N}_M(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ stands for a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ truncated on the set $M \subset \mathbb{R}^d$. For $M = \mathbb{R}^d$ there is no shape constraint. The unimodality constraint can be imposed using the set of unimodal coefficient vectors $\boldsymbol{\beta}$ with mode m, that is $\mathcal{S}_m := \{\boldsymbol{\beta} \in \mathbb{R}^d : \boldsymbol{A}_m'\boldsymbol{\beta} \geq \mathbf{0}\}$. The tuning parameter λ is distributed with $p(\lambda)$, a prior density on $(0, \infty)$.

One can show that the joint posterior of $\boldsymbol{\beta}$ and λ, $p(\boldsymbol{\beta}, \lambda|\boldsymbol{y})$, is proportional to

$$\frac{p(\lambda)\ \lambda^{\frac{r}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right)\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}),$$

where $\boldsymbol{E}_\lambda := \left(\frac{1}{\sigma^2}\boldsymbol{B}'\boldsymbol{B} + \lambda\boldsymbol{\Omega}\right)^{-1}$ and $\boldsymbol{e}_\lambda' := \left(\frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{B} + \lambda\boldsymbol{\beta}_0'\boldsymbol{\Omega}\right)\boldsymbol{E}_\lambda$ and c_λ^{prior} is the normalizing constant of the prior, that is, the prior probability of the truncation set $\mathcal{S}_m$.

Moreover, we obtain the marginal density of λ by integration over $\boldsymbol{\beta}$, i.e.,

$$p(\lambda|\boldsymbol{y}) = \int p(\boldsymbol{\beta}, \lambda|\boldsymbol{y})d\boldsymbol{\beta} \quad \propto p(\lambda)\ \lambda^{\frac{r}{2}}\ |\boldsymbol{E}_\lambda|^{\frac{1}{2}}\frac{c_\lambda^{post}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\lambda\boldsymbol{\beta}_0'\boldsymbol{\Omega}\boldsymbol{\beta}_0\right), \quad (2)$$

where c_λ^{post} is the probability of the set $\mathcal{S}_m$ under the multivariate $\mathcal{N}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$-distribution and can be interpreted as the normalizing constant of the posterior of $\boldsymbol{\beta}$. More detailed derivations are given in Appendix C.

In the unconstrained case both normalizing constants are equal to one since we have to integrate over the whole $\mathbb{R}^d$ instead of over $\mathcal{S}_m$. Under the unimodality shape constraint the normalizing constant c_λ^{post} can be approximated numerically (see, for example, the R package `mvtnorm`, Genz et al. (2012)). When the prior is proper ($\boldsymbol{\Omega}$ is positive definite), c_λ^{prior} can be calculated in the same way. If not, as, for example, when using a difference penalty, we propose to use c_λ^{prior} from the normalizing constant of a slightly modified proper prior. If $\boldsymbol{\Omega}$ is determined by a general difference matrix $\boldsymbol{D}_q$, the modified covariance matrix of the prior is given by $\tilde{\boldsymbol{\Omega}} = \frac{1}{\sigma_v^2}\boldsymbol{I}_d + \lambda\boldsymbol{D}_q'\boldsymbol{D}_q$. This can be interpreted as a combination of ridge and difference penalty, where we propose to keep the "tuning

parameter" σ_v^2 of the ridge penalty fixed, so that the influence of the difference penalty still increases with higher λ.

Finally, an automatic choice of the tuning parameter λ based on REML estimation is $\hat{\lambda} = \arg\max_{\lambda} p(\lambda|\boldsymbol{y})$, which can be found using optimization algorithms. This estimate of λ can be used in the above described estimation procedure for $\boldsymbol{\beta}$.

Again, as the mode m will be unknown, one can estimate λ and $\boldsymbol{\beta}$ for every possible choice of m and choose the pair with minimal residual sum of squares. For an alternative way of estimating the mode see also Section 3.5.

Since the described procedure is not invariant to scaling of the data, we propose to scale the observations $\boldsymbol{y}$ into $[-1, 1]$ and transform the fitted values back. For the applications in Section 4 we choose $\sigma_v^2 = 5$, which can be thought of as uninformative since the β_i approximately also lie in $[-1, 1]$.

3.4 Penalization against Parametric Functions

There are situations in which prior information exists from preceding experiments, as, for example, in dose-response trials. Thus, one might have a particular parametric function in mind to estimate the unimodal relationship, but one would like to safeguard against mis-specification of this function (as discussed in Yuan and Yin (2011), for example). We now present how to integrate this information in form of a penalty.

Suppose we want to penalized against the regression function f. The linear interpolant of the points $(\tau_{-k}^*, \beta_{-k}), \ldots, (\tau_g^*, \beta_g)$, which Dierckx (1993) calls the "control polygon" of a spline, mimicks the form of the spline. Thus, it seems natural to penalize the difference in the B-spline coefficients against the differences in the values of the fitted function $\hat{f}$ at the knot averages. Explicitly, the penalty term is

$$\sum_{j=-k}^{g} \left\{ (\beta_j - \beta_{j-1}) - (\hat{f}(\tau_j^*) - \hat{f}(\tau_{j-1}^*)) \right\}^2 = \sum_{j=-k}^{g} \left\{ \Delta(\beta_j - \hat{f}(\tau_j^*)) \right\}^2 = \left\| \boldsymbol{\Omega}^{\frac{1}{2}} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\|_2^2 \tag{3}$$

with $\boldsymbol{\Omega} = \boldsymbol{D}_1' \boldsymbol{D}_1$ and $\boldsymbol{\beta}_0 = (\hat{f}(\tau_{-k}^*), \ldots, \hat{f}(\tau_g^*))'$.

In dose-response applications a possible choice of the parametric model is the sigmoid E_{max} model (implemented, e.g., in the R package `DoseFinding`, Bornkamp, Pinheiro,

and Bretz (2012)). One characteristic of this model is that it allows for a steep increase in the response. Thus, if prior information suggests such a steep increase, it is possible to incorporate this knowledge by penalizing the spline against the fitted sigmoid E_{max} function. Explicitly, the regression function $f(x) = E_0 + E_1 \frac{x^h}{ED_{50}^h + x^h}$ with parameters E_0, E_1, ED_{50} and h is employed in equation (3).

Since we have already seen in Section 3.3 that a penalty term can be formulated as prior distribution of the regression coefficients, estimation of the penalty from the same data resembles an empirical Bayes approach. When replacing maximization by averaging in the procedures of the last two subsections we arrive at the Bayesian approach that will be discussed in the next subsection.

3.5 Bayesian Approach

In contrast to performing model selection over a grid of possible values for the mode, one can also think of model averaging based on a Bayesian approach. Here, the mode is random, too, leading to the same likelihood model as before, but now the prior distributions are $\boldsymbol{\beta}|\lambda, m \sim \mathcal{N}_{\mathcal{S}_m}(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\Omega}}^{-1})$ and $(\lambda, m) \sim p(\lambda, m)$. We directly use the full-rank precision matrix $\tilde{\boldsymbol{\Omega}} = \frac{1}{\sigma_v^2}\boldsymbol{I}_d + \lambda \boldsymbol{D}_q' \boldsymbol{D}_q$ here. The standard deviation $\sigma > 0$ and the penalty components $\boldsymbol{\beta}_0$ and $\tilde{\boldsymbol{\Omega}}$ are assumed to be fixed, since they are either known from preceeding experiments or can be estimated from the data leading to an empirical Bayes approach.

The joint posterior distribution factorizes as $p(\boldsymbol{\beta}, \lambda, m|\boldsymbol{y}) = p(\boldsymbol{\beta}|\lambda, m, \boldsymbol{y})p(\lambda|m, \boldsymbol{y})p(m|\boldsymbol{y})$. Thus, we can generate a Monte Carlo random sample from the posterior distribution by sampling successively from $p(m|\boldsymbol{y})$, $p(\lambda|m, \boldsymbol{y})$ and $p(\boldsymbol{\beta}|\lambda, m, \boldsymbol{y})$.

The marginal posterior $p(\boldsymbol{\beta}|\lambda, m, \boldsymbol{y})$ is again a multivariate normal density truncated onto the set $\mathcal{S}_m$, with posterior mean vector $\boldsymbol{e}_\lambda' := \left(\frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{B} + \boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\right)\boldsymbol{E}_\lambda$ and covariance matrix $\boldsymbol{E}_\lambda := \left(\frac{1}{\sigma^2}\boldsymbol{B}'\boldsymbol{B} + \tilde{\boldsymbol{\Omega}}\right)^{-1}$, that is $\boldsymbol{\beta}|\lambda, m, \boldsymbol{y} \sim \mathcal{N}_{\mathcal{S}_m}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$. Thus, random sampling from this distribution is possible, for example, using the inverse Bayes formulae sampler proposed by Yu and Tian (2011).

Similar to the derivation of the REML likelihood $p(\lambda|\boldsymbol{y})$ (see equation (2)) in the last subsection, we get the posterior $p(\lambda|m, \boldsymbol{y})$ by integrating $\boldsymbol{\beta}$ out of the joint posterior.

Explicitly, we obtain $p(m|\boldsymbol{y}) = \int p(\lambda, m|\boldsymbol{y})d\lambda$ and $p(\lambda|m, \boldsymbol{y}) = \frac{p(\lambda,m|\boldsymbol{y})}{p(m|\boldsymbol{y})}$ where

$$p(\lambda, m|\boldsymbol{y}) \propto p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}\ |\boldsymbol{E}_\lambda|^{\frac{1}{2}}\ \frac{c_\lambda^{post}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right) =: w(\lambda, m).$$

Thus, $p(\lambda|m, \boldsymbol{y})$ is a univariate density that is known up to a constant factor and sampling is possible using, for example, the slice sampler introduced by Neal (2003).
The distribution of the mode can be normalized by $p(m = m^*|\boldsymbol{y}) = \frac{\int w(\lambda, m^*)d\lambda}{\sum_{j=-k}^{g}\int w(\lambda, j)d\lambda}$.
As it is a discrete distribution on $\{-k, \ldots, g\}$, random sampling is simple.
See Appendix D for a step-by-step derivation of the posterior distributions.
Because of the factorization of the joint posterior and successive sampling from the respective factors, the Bayesian method yields a Monte Carlo random sample with uncorrelated draws from the joint posterior. Markov Chain Monte Carlo (MCMC) methods (inverse Bayes formulae and slice sampler) are only necessary for generating a single draw from $p(\boldsymbol{\beta}|\lambda, m, \boldsymbol{y})$ and $p(\lambda|m, \boldsymbol{y})$ for each joint posterior sample. As for the REML approach, we recommend to transform the observations $\boldsymbol{y}$ onto $[-1, 1]$ before fitting the model, to achieve a scale invariant procedure. For the applications we use independent priors for the tuning parameter and the mode, that is $p(\lambda, m) = p(\lambda)p(m)$, where $p(\lambda) \propto \frac{1}{\lambda}I_{[e^{-3},e^{10}]}(\lambda)$ and $p(m) = \frac{1}{d}\ \forall m \in \{-k, \ldots, g\}$. See Appendix F for details.

4 Applications

In this section the proposed methods are evaluated in dose-response applications and compared to existing methodologies. First, the proposed methods are applied to the growth hormone data set introduced in Section 2. In the second subsection a simulation study is performed, which is motivated by the data situation typically observed in pharmaceutical dose-response clinical trials, where increasing levels of a pharmaceutical compound are administered in parallel to a large number of patients to investigate the dose-response relationship.
All calculations are carried out using R, version 2.15.1 (R Core Team, 2012).

Table 1: Weighted residual sums of squares of the exponential fit by McLaren et al. (1990) and the three fitted unimodal regressions for all four response variables.

	Difference penalty	Bayes (Difference)	Exponential fit	Exponential penalty
ADG	0.471	1.847	0.018	0.023
Age	1.043	1.793	0.341	0.337
G/F	1.003	1.560	1.592	1.005
ADF	0.828	1.149	0.514	0.554

4.1 Growth Hormone Dose-Response Analysis

We use three different unimodal regression approaches to analyse the data presented in Section 2: Maximum likelihood and Bayes, each with difference penalty, and maximum likelihood with penalization against an exponential fit. The exponential fit was found in McLaren et al. (1990) to fit the data best in terms of the coefficient of determination. The estimated exponential relationships in dependence of the dosage x are $ADG(x) = 0.7509 + 0.1523\,(1 - e^{-0.6204x})$, $Age(x) = 183.2 - 12.27\,(1 - e^{-0.5x})$, $G/F(x) = 0.2655 + 0.09982\,(1 - e^{-0.5036x})$ and $ADF(x) = 2.932 - 0.7913\,(1 - e^{-0.2097x})$. Those functions are evaluated at the knot averages to calculate the vector $\boldsymbol{\beta}_0$ for the penalty.

For all three methods, we use $g = 10$ equidistant inner knots on the interval $[0, 9]$ and the given standard deviations are accounted for by weighting with the respective reciprocal variances. For the Bayesian fits 1000 samples from the posterior are drawn and averaged. In addition we retain the pointwise 2.5%- and 97.5%-quantiles of the fitted functions to determine a 95% credible region. In the case of Age and ADF we reflect the data along the x-axis before fitting and reflect the fitted values back. For all four response variables, the three fitted unimodal regressions and the exponential fit are plotted in Figure 1 and the weighted residual sums of squares are given in Table 1.

From the plotted functions we can see that the fits with difference penalty do not differ very much (except for ADG), yet there are clear differences in their weighted residual sums of squares (wRSS). The Bayesian method yields higher wRSS values and seems to oversmooth the data. The credible regions lie close around the fitted curves and stay

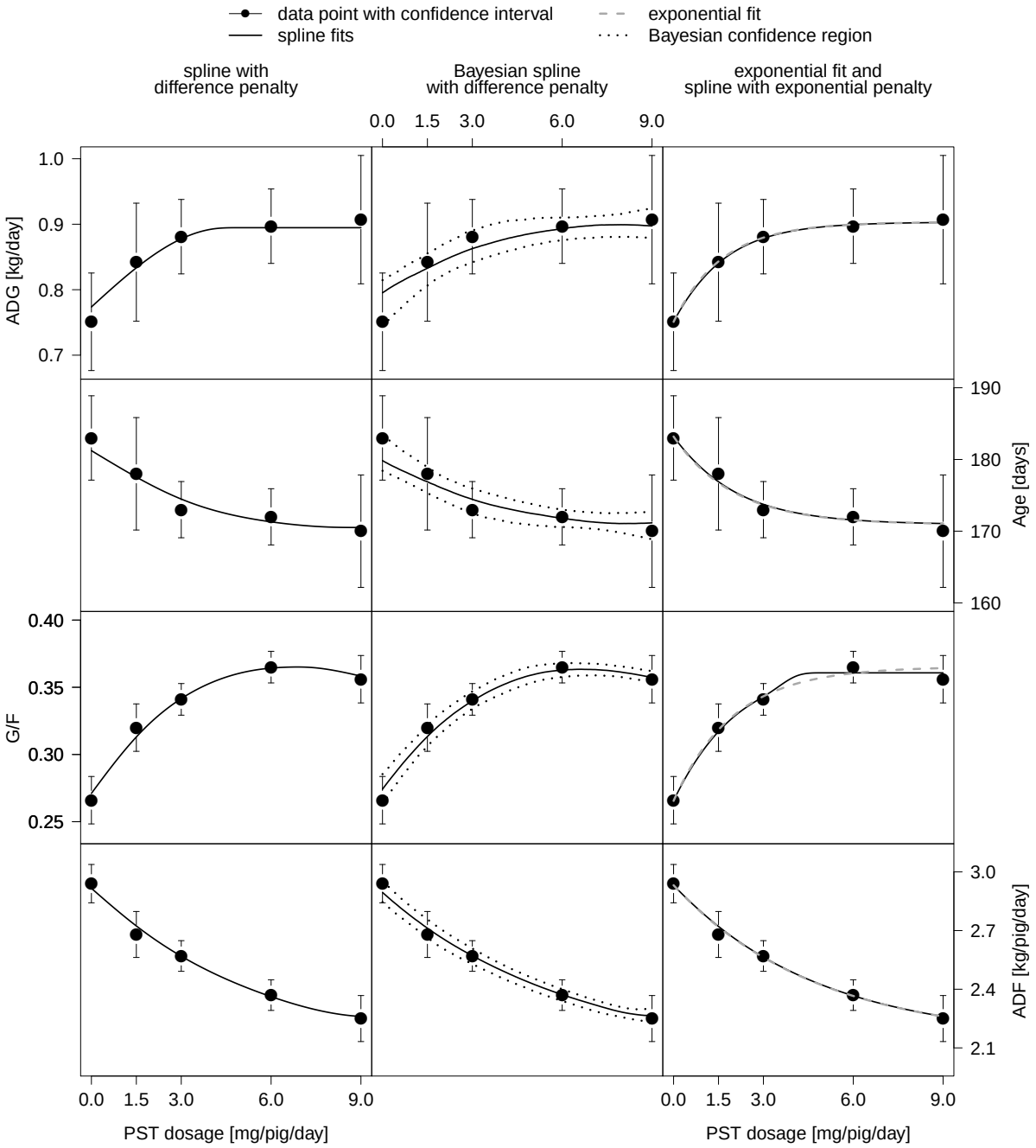

Figure 1: Scatterplots of PST dosage vs. means of the response variables ADG, Age, G/F and ADF. The bars indicate 95% confidence intervals of a normal distribution with the input means and standard deviations (compare Table H.1). The rows correspond to the response variables and the columns to the fitted unimodal functions: Maximum likelihood and Bayesian spline with difference penalty and maximum likelihood spline with exponential penalty. For the Bayesian method the 95% credible regions of the fitted functions are also shown.

within the normal confidence intervals defined by the standard deviations at the observed doses. For the variables ADG, Age and ADF the unimodal regression with exponential penalty and the exponential fit are nearly identical, but in the case of G/F they differ more clearly. For this variable the exponential model seems to be a mis-specification, because all unimodal regressions, also the Bayesian one, have smaller wRSS. McLaren et al. (1990) used the exponential fits to determine optimal dosages. For the G/F variable such subsequent analyses would clearly yield different, presumably more appropiate results using the unimodal fit.

Since we do not know the true underlying relationship we can only asses the different results in terms of wRSS, from which we can derive that the maximum likelihood spline with penalty against the preferred parametric function is more useful than just fitting the parametric model: the difference in wRSS is small, when the parametric model already yields a good fit, but the benefit can be substantial as in the case of the G/F variable.

To provide more insight in the usefulness of the proposed methods in comparison to existing approaches the following subsection describes a simulation study where the true underlying relationship is known and used for the assessment of the results.

4.2 Simulations: Pharmaceutical Dose-Response Trials

4.2.1 Data generation process

The evaluation of the proposed methods primarily follows Bornkamp et al. (2007), where the simulation scenarios were selected so that they are realistic for Phase II trials (for example, in terms of the dose-response shape, the number of doses and the signal to noise ratio).

The data generation process always yields two data vectors: the vector $\boldsymbol{x} = (x_1, \ldots, x_n)'$ of observed predictor values (for example, doses) and the vector $\boldsymbol{y} = (y_1, \ldots, y_n)'$ of observed response values. The vector $\boldsymbol{x}$ is restricted to take values from one of the two equally spaced sequences a) $(0, 2, 4, 6, 8)$ or b) $(0, 1, 2, 3, 4, 5, 6, 7, 8)$. Each value is observed with the same frequency, namely $\frac{n}{5}$ or $\frac{n}{9}$ times. Given the predictor values, the responses are generated according to nine function profiles, that are shown in Figure 2.

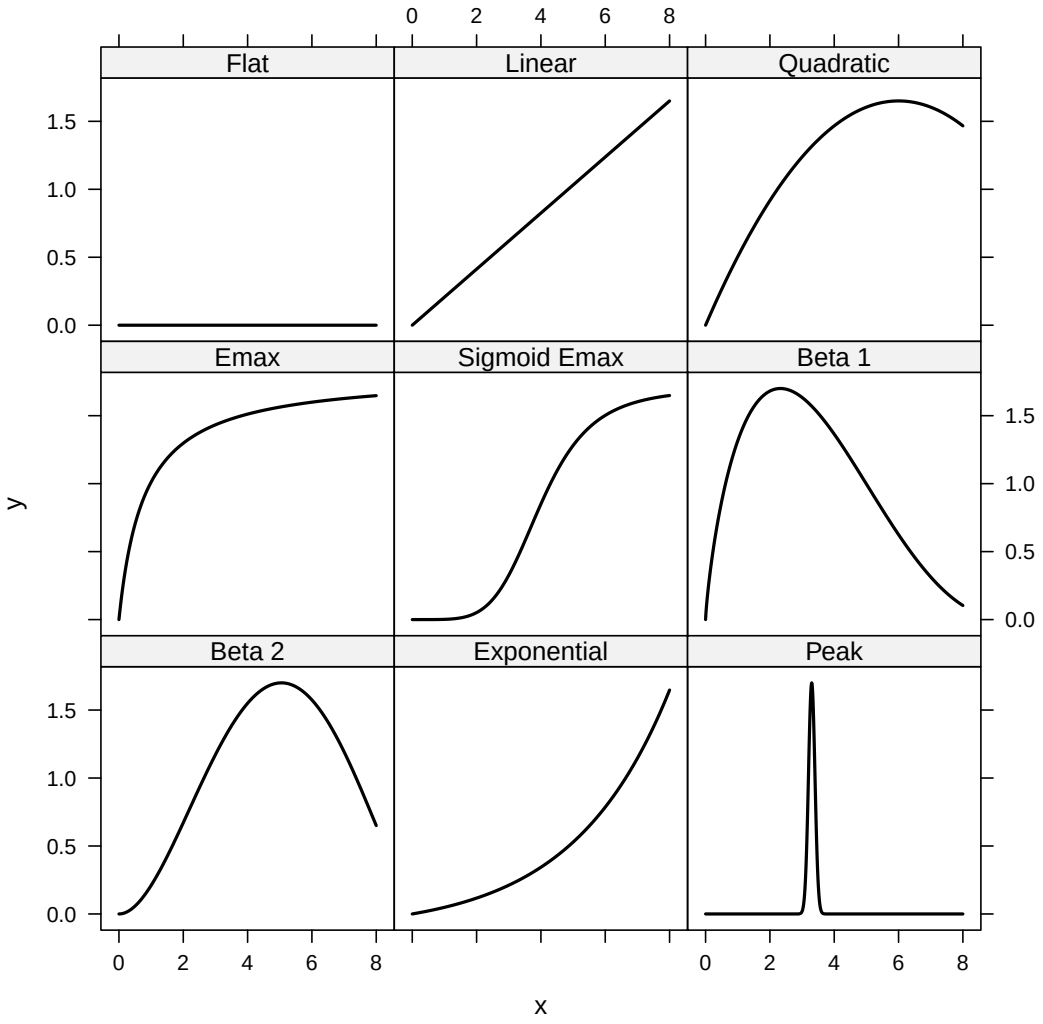

Figure 2: The nine function profiles used in the simulation study.

In comparison with Bornkamp et al. (2007) the logistic profile is left out as it is very similar to sigmoid E_{max}, and four additional profiles are included, three of which have a more pronounced unimodal shape (the two beta profiles and the peak profile) and an exponential function, which is monotone, but convex. The actual values y_i are simulated by $y_i = f(x_i) + \epsilon_i$, where $\epsilon_i \underset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$ and the exact functional forms f are given in Appendix E. The overall sample size and the standard deviation of the errors are set to $n = 250$ and $\sigma = \sqrt{4.5}$, taken from Bornkamp et al. (2007). In the case of repeated measurements at sequence (b), the samples size differs slightly: $n = 252$.

Altogether there are 18 combinations of those settings, see also Table H.2. For each of those 18 scenarios 500 data sets are generated.

4.2.2 Compared methods

Several parametric and nonparametric methods are compared using the above specified simulation scheme. Regarding the proposed frequentist spline methods we compare methods with and without shape constraint and with and without penalization. We use cubic splines and the knot sequence has by default $g = 10$ equidistant inner knots in $[0, 8]$, which is already large compared to the number of distinct x-values (5 or 9 respectively). The second and third order difference penalty (see equation (1)) and the sigmoid E_{max} penalty (see equation (3)) are included. When the unimodality constraint is active, we additionally vary the way of choosing the tuning parameter. We either follow the REML approach as described in Section 3.3 or we choose λ by "approximate REML", which means that the unimodality constraint is not accounted for during tuning parameter optimization, resulting in the same tuning parameter as for the unconstrained models. Thus there are two unpenalized models and three for each penalty.

We also fit four models based on Bernstein polynomials (as used, for example, in Wang and Ghosh (2012)) by applying our penalized maximum likelihood procedure with $g = 0$ inner knots and degree $k = 13$ (which results in the same number of parameters as for the other spline methods). This allows to evaluate the impact of a different selection of the spline basis and to provide a direct comparison of cubic splines and polynomials. Again, an unconstrained and a constrained model is fitted, each combined with the sigmoid E_{max} penalty and, since the difference penalty does not make sense for coincident boundary knots, with the ridge penalty $\sum_{j=-k}^{g} \beta_j^2$.

Altogether there are 15 non-Bayesian procedures, 11 based on B-splines and 4 based on Bernstein polynomials. Moreover, we consider three Bayesian methods applying the proposed methodolody with 2nd and 3rd order difference and sigmoid E_{max} penalty. A fourth Bayesian procedure uses the methodology of Gunn and Dunson (2005), where one samples from the unconstrained model and transforms the coefficients into unimodal sequences. The unconstrained samples are considered as a fifth Bayesian method in the evaluation as well.

In addition to those 20 approaches, we include four models that do not make use of splines: The two parametric methods are the sigmoid E_{max} and the beta model and we

also include model averaging over different parametric models (linear, quadratic, E_{max}, sigmoid E_{max}, beta and exponential model), where the weights for averaging are calculated as described in Buckland et al. (1997) using the AIC criterion. The exact functional forms of those models are given in Appendix F. The nonparametric method by Frisén (1986) is a unimodal transformation of the response means at each dose, connected with straight lines.

A short overview of all 24 methods is given in Table H.3 and a full description and implementation details can be found in Appendix F. All models are fitted to the 500 generated data sets of each scenario. For the Bayesian analysis posterior means were calculated based on 1000 posterior samples.

4.2.3 Evaluation methods

To compare the applied methods, four measures of mean relative loss (MRL), similar to the ones in Morell, Otto, and Fried (2013), are computed. Here the loss of one method in a certain simulation scenario is related to the loss of the best method in that scenario and a mean value of the relative loss is obtained by averaging over a specified subset of scenarios. The closer the MRL is to zero, the less we lose on average by applying this method instead of the respective best methods.

The losses are based on certain performance metrics. MRL_1 is based on the average squared error in function prediction, that is, the mean of the squared differences between the predicted and the true function values at the grid of x-values $(0, 0.01, \ldots, 7.99, 8)$, while MRL_2 and MRL_3 capture the squared error in estimation of the modal value and the mode location, which are interesting in dose-response analysis as estimators of the maximum effect and the dose which yields it. Another characteristic often examined in dose-response studies is the minimum effective dose (MED), the smallest observed x-value that yields a certain response (see also Appendix G). The fourth loss measure MRL_4 is based on the percentage of simulated data sets, in which the MED was incorrectly estimated. A more elaborate description of the performance measures can be found in Appendix G.

4.2.4 Results

The MRL values of the 24 methods are plotted in Figure 3. The subfigures in each row correspond to the scenario subsets with 5 doses, $\mathcal{R}_1 := \{1, 3, \ldots, 17\}$, and with 9 doses, $\mathcal{R}_2 := \{2, 4, \ldots, 18\}$. Means and standard deviations of the four measures of loss are also given in Tables H.4 to H.11.

Except for the mean relative loss in estimation of the modal value (MRL_2) over the scenarios with 5 doses, where the model averaging performs best, always one of the constrained spline methods yields the smallest value of the loss measures MRL_1 to MRL_3. More explicitly, with regard to mean relative loss in function estimation (MRL_1) and in estimation of the mode location (MRL_3) the Bayesian methods with second and third order penalty perform best.

Obviously the inclusion of the constraint mostly reduces the first three loss measures both for unpenalized, penalized and Bayesian splines, and the constrained Bayesian approach always yields smaller values than the frequentists ones. The approximate REML approach is most of the time better than the corresponding unconstrained spline and worse than the exact REML constrained spline.

The Bayesian unimodal transformation method (Gunn and Dunson, 2005) has higher loss MRL_2 and MRL_3 compared to our proposed Bayesian method and comparable loss MRL_1.

With regard to estimation of the MED (MRL_4) the results look different. Here, the beta model performs best (followed by the unpenalized constrained spline) and among the penalized splines the performances of the unconstrained, the constrained (exact and approximate REML) and the Bayesian versions are in reverse order. The Bayesian transformation approach yields smaller values than the proposed Bayesian method. But regarding this fourth measure of loss, we should remark that in some scenarios even the *minimal* relative frequency of incorrect detection (minimal loss) is very high, once as high as 0.79. The methods presumably produce a smooth overall fit that is less beneficial for estimation of the MED.

The best of the non-spline-methods for the performance measures MRL_1 and MRL_2 is the model averaging technique and for MRL_3 and MRL_4 the beta model.

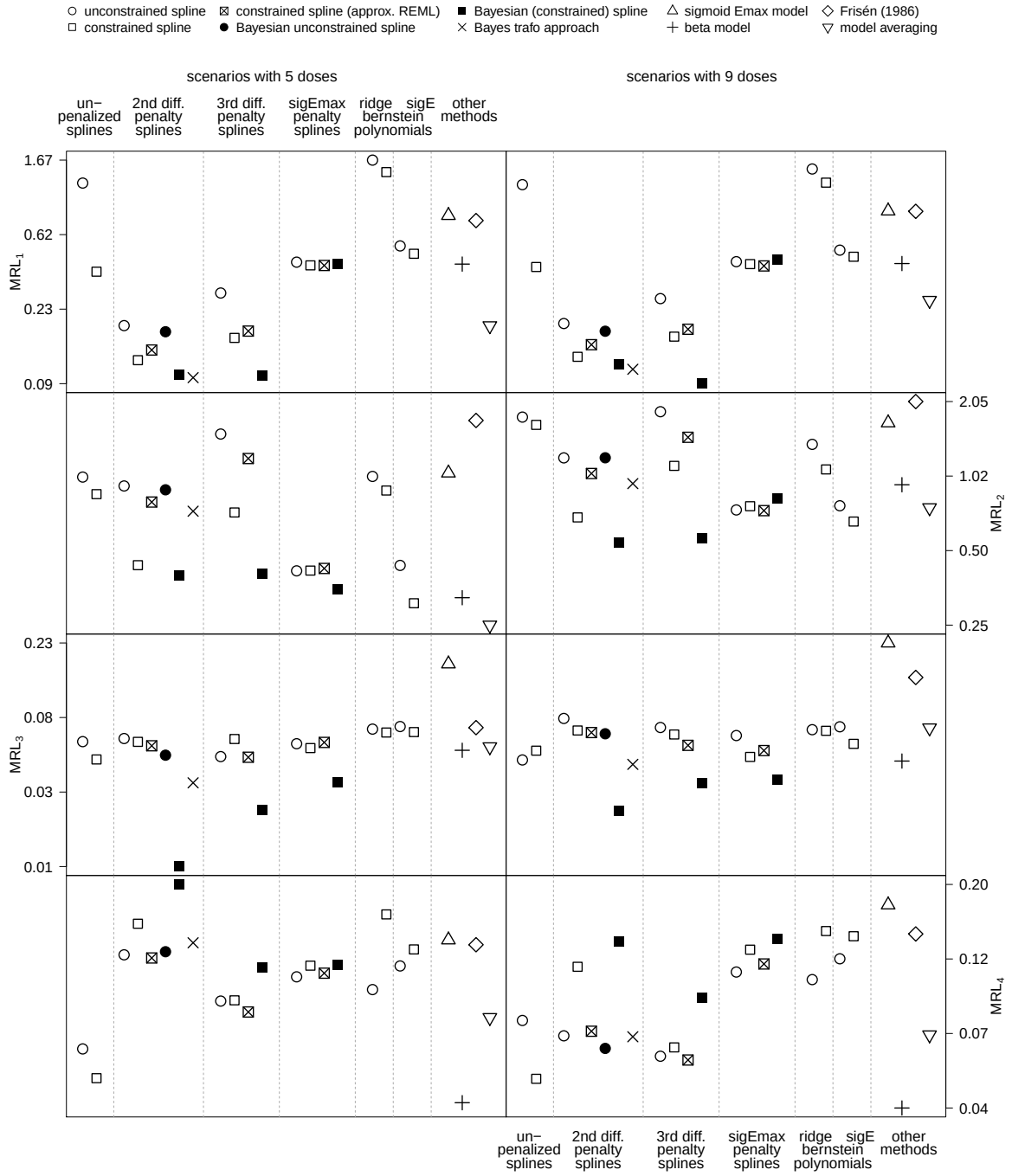

Figure 3: Mean relative loss of the 24 methods in estimation of the function values at the grid $(0, 0.01, 0.02, \ldots, 7.99, 8)$ (MRL_1), of the modal value (MRL_2), of the mode location (MRL_3) and of the minimum effective dose (MRL_4). The rows correspond to the four performance measures and the columns to the two subsets of the scenarios, over which the mean is taken.

Increasing the number of observed doses does not seem to make much difference for the performance of the methods, except that MRL_2 slightly grows for all models and MRL_4 declines for the difference penalized models.

While the sigmoid E_{max} model itself is always one of the worst methods, the sigmoid E_{max} penalized splines have smaller MRL. Regarding the Bernstein polynomials the sigmoid E_{max} penalty performs mostly better than the ridge penalty and comparably to the sigmoid E_{max} penalized B-spline methods, the loss in estimation of the modal value is even slightly smaller.

5 Conclusions and Outlook

In this article we addressed unimodal regression and proposed a range of semiparametric shape-constrained spline regression techniques for this problem. We illustrated the approximation power and the shape-preservation properties of splines using the characteristics of Bernstein-Schoenberg splines and we showed that unimodality of a spline function can easily be guaranteed by placing a linear constraint on the vector of B-spline coefficients. Thus, maximum likelihood estimates can be found with quadratic programming algorithms.

The problem of knot placement was addressed by choosing a large number of knots and putting a penalty on the B-spline coefficients to avoid overfitting. Here the second and third order difference penalties were used, which implicitly penalize against a linear and a quadratic model. We generalized this approach to allow for penalization against arbitrary parametric functions (for example, a sigmoid E_{max} model in dose-response applications). All penalties favour some parametric model, but allow for departures from this model, when the data suggest so. As an automatic choice of the tuning parameter for arbitrary penalities we developed a restricted maximum likelihood (REML) estimate.

The proposed Bayesian approach relies on Monte Carlo random sampling. Since the joint posterior distribution of the B-spline coefficients, the tuning parameter and the mode can be factorized into the marginal posteriors, samples can be drawn by successively sampling from the marginals. MCMC is only required for sampling a B-spline coefficient and

a tuning parameter from their marginal, but the samples from the joint posterior are independent.

When applying our methods to a real dose-response data set of McLaren et al. (1990), we obtained evidence that penalized unimodal spline regression is to some extent able to safeguard against model mis-specification.

A simulation study was conducted to compare the proposed methods among themselves and to other frequently used regression models. The simulations showed that the combination of unimodal shape constraint and a penalty is promising, since adding the constraint improves the fit. Averaging over the posterior mode distribution as in the Bayesian approach can improve the reproduction of the true function characteristics even further.

In this paper the prior for the mode was a uniform distribution on the different mode locations. In many dose-response situations it is also sensible to split the prior probability equally between the case of monotonicity and the remaining mode locations, which is another approach to include prior information, but safeguard against model mis-specification.

The development of the proposed methods was motivated by dose-response analysis, but the methodology is very general and can be used in many other areas of application. We recommend to use penalized unimodal spline regression if a unimodal relationship cannot be excluded a priori. If computation time is available, the Bayesian procedure is preferable.

An interesting aspect for further research are confidence or credible regions. For the Bayesian methods they are straightforward to obtain (and have been included in the analysis of the real data set, see Figure 1), but for the maximum likelihood methods the derivation of the estimator's distribution under the shape constraint is not. Thus, bootstrapping suggests itself, but there is still need to investigate its theoretical properties in the presence of the shape constraint and its computational feasibility.

Furthermore, we expect the unimodal penalized spline regression to be large sample consistent. A proof might be performed along the lines of Meyer (2012), but is beyond the scope of this work. The author shows therein that the convergence rate for monotonicity constrained penalized splines is at least that for unconstrained penalized splines, which were shown to be consistent regression function estimators in Claeskens et al. (2009).

References

Bornkamp, B., Bretz, F., Dmitrienko, A., Enas, G., Gaydos, B., Hsu, C.-H., et al. (2007). Innovative approaches for designing and analyzing adaptive dose-ranging trials. *Journal of Biopharmaceutical Statistics* **17,** 965–995.

Bornkamp, B., Pinheiro, J., and Bretz, F. (2012). *DoseFinding: Planning and Analyzing Dose Finding experiments*. R package version 0.6-2.

Braun, W. J. and Hall, P. (2001). Data sharpening for nonparametric inference subject to constraints. *Journal of Computational and Graphical Statistics* **10,** 786–806.

Bretz, F., Pinheiro, J., and Branson, M. (2005). Combining multiple comparisons and modeling techniques in dose-response studies. *Biometrics* **61,** 738–748.

Buckland, S. T., Burnham, K. P., and Augustin, N. H. (1997). Model selection: an integral part of inference. *Biometrics* **53,** 603–618.

Carnicer, J. M. and Pena, J. M. (1994). Totally positive bases for shape preserving curve design and optimality of b-splines. *Computer Aided Geometric Design* **11,** 633–654.

Chang, I.-S., Hsiung, C. A., Wu, Y.-J., and Yang, C.-C. (2005). Bayesian survival analysis using Bernstein polynomials. *Scandinavian Journal of Statistics* **32,** 447–466.

Claeskens, G., Krivobokova, T., and Opsomer, J. D. (2009). Asymptotic properties of penalized spline estimators. *Biometrika* **96,** 529–544.

Dierckx, P. (1993). *Curve and surface fitting with splines*. Oxford Science Publications. Clarendon Press, Oxford.

Eilers, P. H. C. (2005). Unimodal smoothing. *Journal of Chemometrics* **19,** 317–328.

Eilers, P. H. C. and Marx, B. D. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science* **11,** 89–121.

Frisén, M. (1986). Unimodal regression. *The Statistician* **35,** 479–485.

Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., et al. (2012). *mvtnorm: Multivariate Normal and t Distributions*. R package version 0.9-9993.

Goodman, T. N. T. (1995). Bernstein-Schoenberg operators. In Daehlen, M., Lyche, T., and Schumaker, L., editors, *Mathematical methods for curves and surfaces*. Vanderbilt University Press, Nashville, London.

Gunn, L. and Dunson, D. (2005). A transformation approach for incorporating monotone or unimodal constraints. *Biostatistics* **6,** 434–449.

Hazelton, M. L. and Turlach, B. (2011). Semiparametric regression with shape constrained penalized splines. *Computational Statistics and Data Analysis* **55,** 2871–2879.

Kelly, C. and Rice, J. (1990). Monotone smoothing and its applications to dose-response curves and the assessment of synergy. *Biometrics* **46,** 1071–1085.

Khan, A. A., Perlstein, I., and Krishna, R. (2009). The use of clinical utility assessments in early clinical development. *The AAPS Journal* **11,** 33–38.

Marsden, M. J. (1970). An identity for spline functions with applications to variation-diminishing spline approximation. *Journal of Approximation Theory* **3,** 7–49.

McLaren, D. G., Bechtel, P. J., Grebner, G., Novakofski, J., McKeith, F. K., Jones, R., et al. (1990). Dose response in growth of pigs injected daily with porcine somatotropin from 57 to 103 kilograms. *Journal of Animal Science* **68,** 640–651.

Meyer, M. C. (2012). Constrained penalized splines. *Canadian Journal of Statistics* **40,** 190–206.

Morell, O., Otto, D., and Fried, R. (2013). On robust cross-validation for nonparametric smoothing. *Computational Statistics* **28,** 1617–1637.

Neal, R. (2003). Slice sampling. *Annals of Statistics* **31,** 705–767.

R Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Ramsay, J. O. (1988). Monotone regression splines in action. *Statistical Sciene* **3,** 425–441.

Schoenberg, I. J. (1967). *On spline functions (with a supplement by T. N. E. Greville).* Inequalities I. Academic Press, New York.

Turlach, B. and Weingessel, A. (2011). *quadprog: Functions to solve Quadratic Programming Problems.* R package version 1.5-4, S original by Turlach, B., R port by Weingessel, A.

Turner, R. (2011). *Iso: Functions to perform isotonic regression.* R package version 0.0-10.

Wang, J. and Ghosh, S. K. (2012). Shape restricted nonparametric regression with Bernstein polynomials. *Computational Statistics and Data Analysis* **56,** 2729–2741.

Wood, S. (1994). Monotonic smoothing splines fitted by cross validation. *SIAM Journal on Scientific Computing* **15,** 1126–1133.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society, Series B* **73,** 3–36.

Woodworth, G. (1999). Approaches to bayesian smooth unimodal regression. Technical report, URL: http://www.stat.uiowa.edu/~gwoodwor/unimodal.pdf.

Yu, J.-w. and Tian, G.-l. (2011). Efficient algorithms for generating truncated multivariate normal distributions. *Acta Mathematicae Applicatae Sinica (English Series)* **27,** 601–612.

Yuan, Y. and Yin, G. (2011). Dose-response estimation: A semiparametric mixture approach. *Biometrics* **67,** 1543–1554.

A Proof of Lemma 1

Let f be a unimodal function on $[0,1]$ and $\mathcal{T} = (\tau_j)_{-k}^{g+k+1}$ a knot sequence with distinct inner knots and $\tau_{-k} = \ldots = \tau_0 = 0, \quad \tau_{g+1} = \ldots = \tau_{g+k+1} = 1$.

We need to prove that the Bernstein-Schoenberg spline $\mathcal{V}_k^{\mathcal{T}} f$ is a unimodal function.

Define $\beta_{-k} := f(\tau_{-k}^*), \ldots, \beta_g := f(\tau_g^*)$. Then $\mathcal{V}_k^{\mathcal{T}} f$ is given by $\mathcal{V}_k^{\mathcal{T}} f(x) = \sum_{j=-k}^{g} \beta_j N_{j,k+1}(x)$, which is a spline with B-spline coefficients β_j. Since f is unimodal, the function values at the knot averages and thus, the coefficients β_j form a unimodal sequence, that is, there exists an index $m \in \{-k, \ldots, g\}$ for which $\beta_{-k} \leq \ldots \leq \beta_{m-1} \leq \beta_m \geq \beta_{m+1} \geq \ldots \geq \beta_g$.

The first derivative of $\mathcal{V}_k^{\mathcal{T}} f$ is a spline function of degree $k-1$ with B-spline coefficients $\alpha_j = k \cdot \dfrac{\beta_j - \beta_{j-1}}{\tau_{j+k} - \tau_j}$, $j = -k+1, \ldots, g$ (see, for example, Dierckx (1993), p. 14). As we know that the inner knots are all distinct, we can conclude that $\tau_{j+k} - \tau_j > 0$ for $j = -k+1, \ldots, g$. Thus, the sign of α_j depends only on the sign of $(\beta_j - \beta_{j-1})$.

We have that $\exists\, m \in \{-k, \ldots, g\}$: $\beta_{-k} \leq \ldots \leq \beta_{m-1} \leq \beta_m \geq \beta_{m+1} \geq \ldots \geq \beta_g$

$$\Rightarrow \begin{cases} \alpha_j \geq 0 \;\forall\; j, & m = -k \\ \alpha_j \geq 0 \;\forall\; j \leq m \;\wedge\; \alpha_j \leq 0 \;\forall\; j \geq m+1, & m \notin \{-k, g\} \\ \alpha_j \leq 0 \;\forall\; j, & m = g \end{cases}$$

i.e. the coefficient sequence of the derivative has at most one sign change. Hence, according to the variation diminishing property, the derivative itself has at most one sign change and $\mathcal{V}_k^{\mathcal{T}} f$ is unimodal (including the special cases monotone increasing/decreasing or constant).

B Example: unimodal Bernstein-Schoenberg spline of a non-unimodal function

Let $g = 0$ and $k = 4$, so that the knot sequence is $\tau_{-4} = \ldots = \tau_0 = 0, \tau_1 = \ldots = \tau_4 = 1$ and the knot averages are $\tau^*_{-4} = 0, \tau^*_{-3} = 0.25, \tau^*_{-2} = 0.5, \tau^*_{-1} = 0.75, \tau^*_0 = 1$. Further let $\tilde{f}$ be a function with $\tilde{f}(0) = 0, \tilde{f}(0.25) = 1, \tilde{f}(0.5) = 0, \tilde{f}(0.75) = 1, \tilde{f}(1) = 0$.
Then the Bernstein-Schoenberg spline of $\tilde{f}$ is given by $\mathcal{V}^{\mathcal{T}}_k \tilde{f}(x) = -8(x - 0.5)^4 + 0.5$ for $x \in [0, 1]$. This is a unimodal function as it has only a single maximum at 0.5, but every function $\tilde{f}$ with the specified function values is clearly not unimodal.

C Derivation of equation (2)

Model:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{B} \in \mathbb{R}^{n\times d}, \boldsymbol{\beta} \in \mathbb{R}^d \\
\boldsymbol{y}|\boldsymbol{\beta}, \lambda &\sim \mathcal{N}(\boldsymbol{B}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n), \quad \sigma^2 > 0 \text{ known} \\
\boldsymbol{\beta}|\lambda &\sim \mathcal{N}_{\mathcal{S}_m}(\boldsymbol{\beta}_0, \lambda^{-1}\boldsymbol{\Omega}^-), \quad \boldsymbol{\Omega}^- \in \mathbb{R}^{d\times d}, \mathcal{S}_m := \{\boldsymbol{\beta} \in \mathbb{R}^d : \boldsymbol{A}_m'\boldsymbol{\beta} \geq \boldsymbol{0}\} \\
\lambda &\sim p(\lambda)
\end{aligned}$$

where $\boldsymbol{\Omega}^-$ is the pseudo-inverse of $\boldsymbol{\Omega}$ if $r := \text{rank}(\boldsymbol{\Omega}) < d$, and the regular inverse otherwise.

Thus, the likelihood function is given by $p(\boldsymbol{y}|\boldsymbol{\beta}, \lambda) \propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})\right\}$.

For the (possibly improper) prior density of $\boldsymbol{\beta}$ we adopt the representation in Wood (2011) using the pseudo-determinant $|\cdot|_+$ which is the product of all non-zero eigenvalues of a square matrix, i.e.

$$\begin{aligned}
p(\boldsymbol{\beta}|\lambda) &\propto \frac{|\lambda\boldsymbol{\Omega}|_+^{\frac{1}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\lambda\boldsymbol{\Omega}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}) \\
&= \frac{\lambda^{\frac{r}{2}}|\boldsymbol{\Omega}|_+^{\frac{1}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\lambda\boldsymbol{\Omega}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}) \\
&\propto \frac{\lambda^{\frac{r}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\lambda\boldsymbol{\Omega}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}),
\end{aligned}$$

where c_λ^{prior} is the normalizing constant of the prior, that is, the probability of the truncation set $\mathcal{S}_m$ under the $\mathcal{N}(\boldsymbol{\beta}_0, \lambda^{-1}\boldsymbol{\Omega}^-)$-distribution, and $|\boldsymbol{\Omega}|_+$ denotes the pseudo-determinant if $r < d$, and the regular determinant otherwise, which is independent of $\boldsymbol{\beta}$ and λ.

Then we derive the following for the joint posterior of $\boldsymbol{\beta}$ and λ:

$$
\begin{aligned}
& p(\boldsymbol{\beta}, \lambda | \boldsymbol{y}) \\
= \;& \frac{p(\boldsymbol{\beta}, \lambda, \boldsymbol{y})}{p(\boldsymbol{y})} \\
\propto \;& p(\boldsymbol{\beta}, \lambda, \boldsymbol{y}) \\
= \;& p(\boldsymbol{y} | \boldsymbol{\beta}, \lambda) p(\boldsymbol{\beta} | \lambda) p(\lambda) \\
\propto \;& \exp\left\{ -\frac{1}{2\sigma^2} (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}) \right\} \frac{\lambda^{\frac{r}{2}}}{c_\lambda^{prior}} \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \lambda \boldsymbol{\Omega} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}) \; p(\lambda) \\
= \;& \frac{p(\lambda) \; \lambda^{\frac{r}{2}}}{c_\lambda^{prior}} \exp\left\{ -\frac{1}{2} \left(\frac{1}{\sigma^2} (\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \lambda \boldsymbol{\Omega} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \right) \right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}) \\
= \;& \frac{p(\lambda) \; \lambda^{\frac{r}{2}}}{c_\lambda^{prior}} \exp\left\{ -\frac{1}{2} \left((\boldsymbol{\beta} - \boldsymbol{e}_\lambda)' \boldsymbol{E}_\lambda^{-1} (\boldsymbol{\beta} - \boldsymbol{e}_\lambda) - \boldsymbol{e}_\lambda' \boldsymbol{E}_\lambda^{-1} \boldsymbol{e}_\lambda + \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0 \right) \right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})
\end{aligned}
$$

The last equality holds with $\boldsymbol{E}_\lambda = \left(\frac{1}{\sigma^2} \boldsymbol{B}'\boldsymbol{B} + \lambda \boldsymbol{\Omega} \right)^{-1}$ and $\boldsymbol{e}_\lambda' = (\frac{1}{\sigma^2} \boldsymbol{y}'\boldsymbol{B} + \lambda \boldsymbol{\beta}_0' \boldsymbol{\Omega}) \boldsymbol{E}_\lambda$. This can be seen as follows:

$$
\begin{aligned}
\frac{1}{\sigma^2} (\boldsymbol{y} \;& - \; \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' \lambda \boldsymbol{\Omega} (\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\
&= \frac{1}{\sigma^2} \boldsymbol{y}'\boldsymbol{y} - \frac{2}{\sigma^2} \boldsymbol{y}'\boldsymbol{B}\boldsymbol{\beta} + \frac{1}{\sigma^2} \boldsymbol{\beta}'\boldsymbol{B}'\boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{\beta}' \lambda \boldsymbol{\Omega} \boldsymbol{\beta} - 2 \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta} + \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0 \\
&\propto \boldsymbol{\beta}' \underbrace{\left(\frac{1}{\sigma^2} \boldsymbol{B}'\boldsymbol{B} + \lambda \boldsymbol{\Omega} \right)}_{=: \boldsymbol{E}_\lambda^{-1}} \boldsymbol{\beta} - 2 \left(\frac{1}{\sigma^2} \boldsymbol{y}'\boldsymbol{B} + \lambda \boldsymbol{\beta}_0' \boldsymbol{\Omega} \right) \boldsymbol{\beta} + \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0 \\
&= \boldsymbol{\beta}' \boldsymbol{E}_\lambda^{-1} \boldsymbol{\beta} - 2 \underbrace{\left(\frac{1}{\sigma^2} \boldsymbol{y}'\boldsymbol{B} + \lambda \boldsymbol{\beta}_0' \boldsymbol{\Omega} \right) \boldsymbol{E}_\lambda}_{=: \boldsymbol{e}_\lambda'} \boldsymbol{E}_\lambda^{-1} \boldsymbol{\beta} + \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0 \\
&= \boldsymbol{\beta}' \boldsymbol{E}_\lambda^{-1} \boldsymbol{\beta} - 2 \boldsymbol{e}_\lambda' \boldsymbol{E}_\lambda^{-1} \boldsymbol{\beta} + \boldsymbol{e}_\lambda' \boldsymbol{E}_\lambda^{-1} \boldsymbol{e}_\lambda - \boldsymbol{e}_\lambda' \boldsymbol{E}_\lambda^{-1} \boldsymbol{e}_\lambda + \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0 \\
&= (\boldsymbol{\beta} - \boldsymbol{e}_\lambda)' \boldsymbol{E}_\lambda^{-1} (\boldsymbol{\beta} - \boldsymbol{e}_\lambda) - \boldsymbol{e}_\lambda' \boldsymbol{E}_\lambda^{-1} \boldsymbol{e}_\lambda + \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0
\end{aligned}
$$

Thus we have

$$
\begin{aligned}
& p(\boldsymbol{\beta}, \lambda | \boldsymbol{y}) \\
& \propto \frac{p(\lambda) \; \lambda^{\frac{r}{2}}}{c_\lambda^{prior}} \exp\left(\frac{1}{2} \boldsymbol{e}_\lambda' \boldsymbol{E}_\lambda^{-1} \boldsymbol{e}_\lambda - \frac{1}{2} \boldsymbol{\beta}_0' \lambda \boldsymbol{\Omega} \boldsymbol{\beta}_0 \right) \exp\left\{ -\frac{1}{2} (\boldsymbol{\beta} - \boldsymbol{e}_\lambda)' \boldsymbol{E}_\lambda^{-1} (\boldsymbol{\beta} - \boldsymbol{e}_\lambda) \right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}).
\end{aligned}
$$

Now equation (2) follows by integration over $\boldsymbol{\beta}$:

$$
\begin{aligned}
& p(\lambda|\boldsymbol{y}) \\
&= \int p(\boldsymbol{\beta}, \lambda|\boldsymbol{y}) d\boldsymbol{\beta} \\
&\propto \int_{\mathcal{S}_m} \frac{p(\lambda)\ \lambda^{\frac{r}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right) \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} d\boldsymbol{\beta} \\
&= \frac{p(\lambda)\ \lambda^{\frac{r}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right) \int_{\mathcal{S}_m} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} d\boldsymbol{\beta} \\
&\propto \frac{p(\lambda)\ \lambda^{\frac{r}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right) \\
&\qquad\qquad \times \underbrace{\int_{\mathcal{S}_m} (2\pi)^{-\frac{d}{2}} |\boldsymbol{E}_\lambda|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} d\boldsymbol{\beta}}_{:=c_\lambda^{post}} \\
&= \frac{p(\lambda)\ \lambda^{\frac{r}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right) |\boldsymbol{E}_\lambda|^{\frac{1}{2}}\ c_\lambda^{post} \\
&= p(\lambda)\ \lambda^{\frac{r}{2}}\ |\boldsymbol{E}_\lambda|^{\frac{1}{2}} \frac{c_\lambda^{post}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\lambda\boldsymbol{\beta}_0'\boldsymbol{\Omega}\boldsymbol{\beta}_0\right),
\end{aligned}
$$

where c_λ^{post} is the probability of the truncation set $\mathcal{S}_m$ under the $\mathcal{N}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$-distribution, which is the normalizing constant of the posterior of $\boldsymbol{\beta}$ since

$$
\begin{aligned}
& p(\boldsymbol{\beta}|\lambda, \boldsymbol{y}) \\
&= \frac{p(\boldsymbol{\beta}, \lambda|\boldsymbol{y})}{p(\lambda|\boldsymbol{y})} \\
&\propto \frac{\frac{p(\lambda)\ \lambda^{\frac{r}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\lambda\boldsymbol{\Omega}\boldsymbol{\beta}_0\right) \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})}{p(\lambda)\ \lambda^{\frac{r}{2}}\ |\boldsymbol{E}_\lambda|^{\frac{1}{2}} \frac{c_\lambda^{post}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\lambda\boldsymbol{\beta}_0'\boldsymbol{\Omega}\boldsymbol{\beta}_0\right)} \\
&\propto \frac{|\boldsymbol{E}_\lambda|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})}{c_\lambda^{post}}
\end{aligned}
$$

which is proportional to the density of the $\mathcal{N}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$-distribution truncated onto $\mathcal{S}_m$.

D Derivation of the posterior distributions in the Bayes model

Model:

$$\begin{aligned}
\boldsymbol{y} &= \boldsymbol{B}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad \boldsymbol{B} \in \mathbb{R}^{n\times d}, \boldsymbol{\beta} \in \mathbb{R}^d \\
\boldsymbol{y}|\boldsymbol{\beta}, \lambda, m &\sim \mathcal{N}(\boldsymbol{B}\boldsymbol{\beta}, \sigma^2 \boldsymbol{I}_n), \quad \sigma^2 > 0 \text{ known} \\
\boldsymbol{\beta}|\lambda, m &\sim \mathcal{N}_{\mathcal{S}_m}(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\Omega}}^{-1}), \quad \tilde{\boldsymbol{\Omega}} \in \mathbb{R}^{d\times d}, \mathcal{S}_m := \{\boldsymbol{\beta} \in \mathbb{R}^d : \boldsymbol{A}_m'\boldsymbol{\beta} \geq \boldsymbol{0}\} \\
(\lambda, m) &\sim p(\lambda, m)
\end{aligned}$$

Likelihood and prior density for $\boldsymbol{\beta}$:

$$\begin{aligned}
p(\boldsymbol{y}|\boldsymbol{\beta}, \lambda, m) &\propto \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})\right\} \\
p(\boldsymbol{\beta}|\lambda, m) &\propto \frac{|\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}),
\end{aligned}$$

where c_λ^{prior} is the normalizing constant of the prior, that is, the probability of the truncation set $\mathcal{S}_m$ under the $\mathcal{N}(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\Omega}}^{-1})$-distribution.

Then we derive the following for the joint posterior of $\boldsymbol{\beta}$ and λ and m:

$$\begin{aligned}
& p(\boldsymbol{\beta}, \lambda, m|\boldsymbol{y}) \\
= \;& \frac{p(\boldsymbol{\beta}, \lambda, m, \boldsymbol{y})}{p(\boldsymbol{y})} \\
\propto \;& p(\boldsymbol{\beta}, \lambda, m, \boldsymbol{y}) \\
= \;& p(\boldsymbol{y}|\boldsymbol{\beta}, \lambda, m)p(\boldsymbol{\beta}|\lambda, m)p(\lambda, m) \\
\propto \;& \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})\right\} \frac{|\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})\; p(\lambda, m) \\
= \;& \frac{p(\lambda, m)\; |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}\left(\frac{1}{\sigma^2}(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta})'(\boldsymbol{y} - \boldsymbol{B}\boldsymbol{\beta}) + (\boldsymbol{\beta} - \boldsymbol{\beta}_0)'\tilde{\boldsymbol{\Omega}}(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\right)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}) \\
= \;& \frac{p(\lambda, m)\; |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}} \exp\left\{-\frac{1}{2}\left((\boldsymbol{\beta} - \boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta} - \boldsymbol{e}_\lambda) - \boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda + \boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})
\end{aligned}$$

The last equality holds with $\boldsymbol{E}_\lambda = \left(\frac{1}{\sigma^2}\boldsymbol{B}'\boldsymbol{B} + \tilde{\boldsymbol{\Omega}}\right)^{-1}$ and $\boldsymbol{e}_\lambda' = (\frac{1}{\sigma^2}\boldsymbol{y}'\boldsymbol{B} + \boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}})\boldsymbol{E}_\lambda$ analogous to Appendix C.

Thus we have

$$
\begin{aligned}
&p(\boldsymbol{\beta}, \lambda, m|\boldsymbol{y}) \\
&\propto \frac{p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta}).
\end{aligned}
$$

Now by integration over $\boldsymbol{\beta}$ we obtain:

$$
\begin{aligned}
&p(\lambda, m|\boldsymbol{y}) \\
&= \int p(\boldsymbol{\beta}, \lambda, m|\boldsymbol{y})d\boldsymbol{\beta} \\
&\propto \int_{\mathcal{S}_m} \frac{p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} d\boldsymbol{\beta} \\
&= \frac{p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)\int_{\mathcal{S}_m}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} d\boldsymbol{\beta} \\
&\propto \frac{p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right) \\
&\qquad\qquad \times \int_{\mathcal{S}_m}(2\pi)^{-\frac{d}{2}}|\boldsymbol{E}_\lambda|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} d\boldsymbol{\beta} \\
&= \frac{p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)|\boldsymbol{E}_\lambda|^{\frac{1}{2}}\ c_\lambda^{post} \\
&= p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}\ |\boldsymbol{E}_\lambda|^{\frac{1}{2}}\frac{c_\lambda^{post}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right) =: w(\lambda, m),
\end{aligned}
$$

where c_λ^{post} is again the probability of the truncation set $\mathcal{S}_m$ under the $\mathcal{N}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$-distribution, which is the normalizing constant of the posterior of $\boldsymbol{\beta}$ since

$$
\begin{aligned}
&p(\boldsymbol{\beta}|\lambda, m, \boldsymbol{y}) \\
&= \frac{p(\boldsymbol{\beta}, \lambda, m|\boldsymbol{y})}{p(\lambda, m|\boldsymbol{y})} \\
&\propto \frac{\frac{p(\lambda,m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})}{p(\lambda, m)\ |\tilde{\boldsymbol{\Omega}}|^{\frac{1}{2}}\ |\boldsymbol{E}_\lambda|^{\frac{1}{2}}\frac{c_\lambda^{post}}{c_\lambda^{prior}}\ \exp\left(\frac{1}{2}\boldsymbol{e}_\lambda'\boldsymbol{E}_\lambda^{-1}\boldsymbol{e}_\lambda - \frac{1}{2}\boldsymbol{\beta}_0'\tilde{\boldsymbol{\Omega}}\boldsymbol{\beta}_0\right)} \\
&\propto \frac{|\boldsymbol{E}_\lambda|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)'\boldsymbol{E}_\lambda^{-1}(\boldsymbol{\beta}-\boldsymbol{e}_\lambda)\right\} I_{\mathcal{S}_m}(\boldsymbol{\beta})}{c_\lambda^{post}}
\end{aligned}
$$

which is proportional to the density of the $\mathcal{N}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$-distribution truncated onto $\mathcal{S}_m$ and thus $\boldsymbol{\beta}|\lambda, m, \boldsymbol{y} \sim \mathcal{N}_{\mathcal{S}_m}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$.

Finally, the posterior distribution of the mode is $p(m = m^*|\boldsymbol{y}) = \int p(\lambda, m^*|\boldsymbol{y})d\lambda = \frac{\int w(\lambda, m^*)d\lambda}{\sum_{j=-k}^{g}\int w(\lambda, j)d\lambda}$, where each integral can be approximated using the Riemann sum $\sum_{r=1}^{R}(\lambda_r - \lambda_{r-1})\frac{w(\lambda_r, m^*) + w(\lambda_{r-1}, m^*)}{2}$ over the grid $(\lambda_0, \ldots, \lambda_R)$, and the posterior of the tuning parameter is $p(\lambda|m, \boldsymbol{y}) = \frac{p(\lambda, m|\boldsymbol{y})}{p(m|\boldsymbol{y})}$.

E Functional forms of the nine function profiles used in the simulation study

1. Flat: $f(x) \equiv 0$
2. Linear: $f(x) = \frac{1.65}{8}\ x \cdot \mathcal{I}_{[0,8]}(x)$
3. Umbrella: $f(x) = 1.65\left(\frac{1}{3}x - \frac{1}{36}x^2\right) \cdot \mathcal{I}_{[0,8]}(x)$
4. E_{max}: $f(x) = \dfrac{1.81x}{0.79 + x} \cdot \mathcal{I}_{[0,8]}(x)$
5. sigmoid E_{max}: $f(x) = \dfrac{1.7x^5}{4^5 + x^5} \cdot \mathcal{I}_{[0,8]}(x)$
6. Beta 1: $f(x) = 1.7 \cdot \mathrm{Be}(0.8, 2.5)\left(\frac{x}{9.6}\right)^{0.8}\left(1 - \frac{x}{9.6}\right)^{2.5} \cdot \mathcal{I}_{[0,8]}(x)$
7. Beta 2: $f(x) = 1.7 \cdot \mathrm{Be}(2, 1.8)\left(\frac{x}{9.6}\right)^{2}\left(1 - \frac{x}{9.6}\right)^{1.8} \cdot \mathcal{I}_{[0,8]}(x)$
8. Exponential: $f(x) = 0.123 \cdot (exp(\frac{x}{3}) - 1) \cdot \mathcal{I}_{[0,8]}(x)$
9. Peak: $f(x) = 1.7 \cdot \dfrac{\varphi(x|3.3, 0.25)}{\varphi(0|0, 0.25)} \cdot \mathcal{I}_{[0,8]}(x)$,

where $\mathrm{Be}(\gamma_1, \gamma_2) = \dfrac{(\gamma_1 + \gamma_2)^{\gamma_1+\gamma_2}}{\gamma_1^{\gamma_1}\gamma_2^{\gamma_2}}$ for $\gamma_1, \gamma_2 > 0$ and $\varphi(x|\mu, \sigma) = \dfrac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\dfrac{(x-\mu)^2}{2\sigma^2}\right\}$ for $x, \mu \in \mathbb{R}, \sigma > 0$.

F Details on the fitting procedures

Abbreviations for the 24 methods compared in the simulation study and a short overview of their characteristics are given in Table H.3. A more detailed description is given in this section.

The methods can be divided into 20 semiparametric methods, 16 of which are based on B-splines and 4 on Bernstein polynomials, one nonparametric method and 3 methods based on parametric models.

For all spline-based methods we use cubic splines ($k = 3$) and the knot sequences are chosen as follows: By default, there are $g = 10$ equidistant inner knots in $[0, 8]$ and four coincident knots at each boundary, which yields a knot sequence of length $g+2k+2 = 18$ and a parameter vector of length $d = g + k + 1 = 14$. Only for the difference penalized methods the knots are equidistant also beyond the boundaries. In the case of the unpenalized splines (models "un" and "cn") the number of estimable parameters (including σ) is bounded by the number u of distinct x-values in the respective data set, which is 5 or 9, respectively. Thus, the knot sequence is shorter here, the number of inner knots equals $g = u - 5$, which is 0 or 4, respectively. To avoid data scaling effects, the observations $\boldsymbol{y}$ are projected onto $[-1, 1]$ and the fitted values are transformed back.

The 11 frequentist spline regression methods are fitted using the function `unireg` in R package `unimodal`. We fit two unpenalized maximum likelihood spline models, one unconstrained (model "un") and one constrained (model "cn"). The same is done for second and third order difference (models "ud2", "cd2", "ud3", "cd3") and sigmoid E_{max} penalized splines (models "us" and "cs"). When the unimodality constraint is active, we additionally vary the way of choosing the tuning parameter. The models "cd2", "cd3" and "cs" follow the REML approach as described in Section 2.3, while the models "cda2", "cda3" and "csa" choose λ by "approximate REML", which means that the unimodality constraint is not accounted for during tuning parameter optimization, resulting in the same tuning parameter as for the unconstrained models "ud2", "ud3" and "us".

The Bayesian methods "cdb2", "cdb3" and "csb" follow the procedure described in Section 2.5 with second and third order difference penalty and sigmoid E_{max} penalty. We use independent priors for the tuning parameter and the mode, that is $p(\lambda, m) = p(\lambda)p(m)$.

The prior for λ is the Jeffreys prior $p(\lambda) \propto \frac{1}{\lambda}$, which is restricted to the interval $[e^{-3}, e^{10}] \approx [0.05, 22026.47]$ due to numerical problems that occur for very small or large values of λ and because an improper prior on the interval $(0, \infty)$ might result in an improper posterior. The prior distribution of the mode is simply the uniform distribution on all possible values, $p(m) = \frac{1}{d}\ \forall m \in \{-k, \ldots, g\}$.

The variance σ_v^2 in the proper $\mathcal{N}_{\mathcal{S}_m}(\boldsymbol{\beta}_0, \tilde{\boldsymbol{\Omega}}^{-1})$-prior for $\boldsymbol{\beta}$ is chosen to be 5, which can be thought of as uninformative since the β_i approximately also lie in $[-1, 1]$, the range of the transformed y-values (see the control polygon characteristic in Section 2.4).

The integrals $\int w(\lambda|j)d\lambda$, $j \in \{-k, \ldots, g\}$ required for determining the posterior mode distribution, $p(m|\boldsymbol{y})$, are approximated by a Riemann sum using a grid of 200 values between e^{-3} and e^{10}. A sample from the marginal posterior of λ is drawn using an R implementation of the slice sampler by Neal (2003). For computational efficiency updating was performed on $\log(\lambda)$-scale and by taking samples from an approximation of the likelihood, i.e., a smoothing spline of $w(\lambda|m)$ fitted with the R function `smooth.spline` on same grid as before.

Samples from $p(\boldsymbol{\beta}|\lambda, m, \boldsymbol{y})$, that is from the truncated multivariate normal distribution $\mathcal{N}_{\mathcal{S}_m}(\boldsymbol{e}_\lambda, \boldsymbol{E}_\lambda)$, are obtained using an R implementation of the inverse Bayes formulae sampler by Yu and Tian (2011).

A fourth Bayesian method, "trafo", uses the second order difference penalty and generates samples of the tuning parameter and the spline coefficients from the unconstrained spline model. The sampled coefficient vectors are subsequently mapped onto the space of unimodal vectors in $\mathbb{R}^d$ using the transformation procedure described in Gunn and Dunson (2005), formulae 2.2 and 2.3. The unconstrained samples (before transformation) are also used as the fifth Bayesian method "udb".

For all Bayesian methods the resulting Monte Carlo sample from the joint posterior is of size 1000 and the parameters are estimated by their posterior means.

We also fit constrained and unconstrained Bernstein polynomials by applying our maximum likelihood procedure with $g = 0$ inner knots and degree $k = 13$ (which results in the same number of parameters as for the other spline methods). We penalize with the sigmoid E_{max} penalty (models "uBPs" and "cBPs") and, since the difference penalty does not make sense for coincident boundary knots, with the ridge penalty $\sum_{j=-k}^{g} \beta_j^2$ (models

"uBPr" and "cBPr")).

The model averaging method "modAve" takes a weighted average of the dose-response models `"linear"`, `"quadratic"`, `"emax"`, `"sigEmax"`, `"betaMod"` and `"exponential"`, fitted with the R function `fitDRModel` in R package `DoseFinding` (Bornkamp et al., 2012). The weights for averaging are the models' relative likelihood factors as described in Buckland et al. (1997). The functional forms of those models are as follows:

1. linear: $f(x) = E_0 + \gamma x$
2. quadratic: $f(x) = E_0 + \gamma_1 x + \gamma_2 x^2$
3. E_{max}: $f(x) = E_0 + E_1 \frac{x}{ED_{50} + x}$
4. sigmoid E_{max}: $f(x) = E_0 + E_1 \frac{x^h}{ED_{50}^h + x^h}$
5. beta: $f(x) = E_0 + E_{max} \cdot \text{Be}(\gamma_1, \gamma_2) \left(\frac{x}{scal}\right)^{\gamma_1} \left(1 - \frac{x}{scal}\right)^{\gamma_2}$
6. exponential: $f(x) = E_0 + E_1 \cdot \left(\exp\left(\frac{x}{\gamma}\right) - 1\right)$

where $\text{Be}(\gamma_1, \gamma_2) = \frac{(\gamma_1 + \gamma_2)^{\gamma_1 + \gamma_2}}{\gamma_1^{\gamma_1} \gamma_2^{\gamma_2}}$ for $\gamma_1, \gamma_2 > 0$ and $scal = 9.6$ is a fixed dose scaling parameter.

The models "sigE" and "beta" are the stand-alone parametric sigmoid E_{max} and the beta model and are also fitted using the R function `fitDRModel`.

The nonparametric method by Frisén (1986) (model "frisen") is a unimodal transformation of the response means at each dose, connected with straight lines. It is fitted by calculating the means of y-values with same x-value and applying the R function `ufit()` in R package `Iso` (Turner, 2011) to this sequence of means.

G Measures of mean relative loss computed for evaluation of the simulation study

To compare the methods, that have been applied in the simulation study, several evaluation characteristics have been computed.

First of all, the average squared error (ASE) in function prediction is assessed, that is, the mean of the squared differences between the predicted and the true function values at specified points $\boldsymbol{z} = (z_1, \ldots, z_\eta)$ on the x-axis: $\text{ASE}(\hat{f}_{\rho,\nu}; \boldsymbol{z}) = \frac{1}{\eta} \sum_{i=1}^{\eta} (\hat{f}_{\rho,\nu}(z_i) - f_\rho(z_i))^2$, where f_ρ is the function of scenario $\rho \in \{1, \ldots, 18\}$ and $\hat{f}_{\rho,\nu}$ its estimate using method $\nu \in \{1, \ldots, 24\}$.

We obtain one value $\text{ASE}_\kappa(\hat{f}_{\rho,\nu}; \boldsymbol{z})$ of this measure for each data set $\kappa = 1, \ldots, 500$. To summarize the performance of method ν over all data sets of scenario ρ, we calculate the mean ASE-value and define a measure of loss:

$$\mathcal{L}_1(\rho, \nu) = \frac{1}{500} \sum_{\kappa=1}^{500} \text{ASE}_\kappa(\hat{f}_{\rho,\nu}; \boldsymbol{z} = (0, 0.01, \ldots, 8)), \ \nu = 1, \ldots, 24, \rho = 1, \ldots, 18.$$

Other aspects for comparison of the methods are the modal value $\max_{x \in [0,8]} \hat{f}_{\rho,\nu}(x)$ and the location $\text{mod}(\hat{f}_{\rho,\nu}) = \arg\max_{x \in [0,8]} \hat{f}_{\rho,\nu}(x)$ of the mode, which are interesting, for example, in dose-response analysis as they describe the maximum effect of a drug and the dose level which yields it.

Both values are determined with the help of the function `optimize()` in R package `stats` and for both characteristics we can define a mean performance over all data sets of one scenario as above. In the case of the modal value we execute the above calculation steps of the mean ASE with $\mathbf{z} = (z_1) = \text{mod}(\hat{f}_{\rho,\nu})$ and define the loss $\mathcal{L}_2(\rho, \nu) = \frac{1}{500} \sum_{\kappa=1}^{500} \text{ASE}_\kappa(\hat{f}_{\rho,\nu}; \text{mod}(\hat{f}_{\rho,\nu}))$ for $\nu = 1, \ldots, 24$.

In the case of the mode location we also calculate the mean of a squared error as a measure of loss, namely $\mathcal{L}_3(\rho, \nu) = \frac{1}{500} \sum_{\kappa=1}^{500} (\text{mod}(\hat{f}_{\rho,\nu}) - \text{mod}(f_\rho))^2$. For scenarios $\rho = 1, 2$ this loss is not defined, because a true mode location $\text{mod}(f_\rho)$ does not exist for the flat function profile.

For dose-response applications, another important characteristic is the minimum effective

dose (MED), that is, the smallest (observed) predictor value that yields a response of at least $f(0) + \delta$ for a given $\delta \geq 0$. Here $\delta = 1.3$ as in Bornkamp et al. (2007) is used. The loss $\mathcal{L}_4(\rho, \nu)$ in this context is chosen to be the percentage of data sets of scenario ρ, in which the MED was incorrectly estimated with method ν (which is already a summary of all data sets of one scenario).

Since comparing the 24 methods over 18 data scenarios is still quite unmanageable, we follow Morell, Otto, and Fried (2013) by defining a relative loss (RL) of method ν in scenario ρ:

$$\mathrm{RL}_\ell(\rho, \nu) = \frac{\mathcal{L}_\ell(\rho, \nu) - \mathcal{L}_\ell(\rho, *)}{\psi_\ell(\rho)}, \ \ell = 1, \ldots, 4, \ \rho = 1, \ldots, 18, \ \nu = 1, \ldots, 24,$$

where $\mathcal{L}_\ell(\rho, *) = \min_{\nu=1,\ldots,24} \mathcal{L}_\ell(\rho, \nu)$ is the minimal loss value achieved in scenario ρ by one of the compared methods. The standardization factor $\psi_\ell(\rho)$ equals $\mathcal{L}_\ell(\rho, *)$ for $\ell = 1, 2$. Regarding the losses $\mathcal{L}_3$ and $\mathcal{L}_4$ we face the problem that there are scenarios for which the best method has zero loss. To avoid the problem of division by zero the standardization factors are chosen to be $\psi_3(\rho) = \max\{\mathrm{mod}(f_\rho), 8 - \mathrm{mod}(f_\rho)\}$, which is the maximal possible error since $\mathrm{mod}(\hat{f}_{\rho,\nu})$ and $\mathrm{mod}(f_\rho)$ both lie in $[0, 8]$, and $\psi_4(\rho) = 1$, which is also the maximal possible error.

The closer $\mathrm{RL}_\ell(\rho, \nu)$ is to zero, the less we lose when applying method ν instead of the best method in scenario ρ.

Finally we obtain an even more aggregated performance measure for all four criteria when we take the mean of the relative loss of method ν over all scenarios or a specified subset $\mathcal{R}$ of scenarios:

$$\mathrm{MRL}_\ell(\nu) = \frac{1}{|\mathcal{R}|} \sum_{\rho \in \mathcal{R}} \mathrm{RL}_\ell(\rho, \nu), \quad \ell = 1, \ldots, 4, \ \nu = 1, \ldots, 24.$$

In the main article we show the mean relative loss for the subsets $\mathcal{R}_1 := \{1, 3, \ldots, 17\}$ (scenarios with 5 doses) and $\mathcal{R}_2 := \{2, 4, \ldots, 18\}$ (scenarios with 9 doses). Notice again that the flat scenarios ($\rho = 1, 2$) have to be left out in the case of loss $\mathcal{L}_3$.

H Additional tables

Table H.1: Porcine somatotropin (PST) dosages [mg/pig/day] and least squares means and standard deviations of the four performance variables from McLaren et al. (1990). ADG = average daily gain of weight [kg/d], Age = age at 103.5 kg [d], G/F = gain-to-feed ratio, ADF = average daily feed consumption [kg/pig/d].

	PST dosage				
	0	1.5	3	6	9
ADG	0.751 (0.038)	0.842 (0.046)	0.881 (0.029)	0.897 (0.029)	0.907 (0.050)
Age	183 (3)	178 (4)	173 (2)	172 (2)	170 (4)
G/F	0.266 (0.009)	0.320 (0.009)	0.341 (0.006)	0.365 (0.006)	0.356 (0.009)
ADF	2.940 (0.050)	2.680 (0.060)	2.570 (0.040)	2.370 (0.040)	2.250 (0.060)

Table H.2: The 18 scenarios used for data generation in the simulation study.

no.	x-values	sample size	σ	profile
1	(0,2,4,6,8)	250	$\sqrt{4.5}$	Flat
2	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Flat
3	0,2,4,6,8)	250	$\sqrt{4.5}$	Linear
4	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Linear
5	(0,2,4,6,8)	250	$\sqrt{4.5}$	Quadratic
6	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Quadratic
7	(0,2,4,6,8)	250	$\sqrt{4.5}$	E_{max}
8	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	E_{max}
9	(0,2,4,6,8)	250	$\sqrt{4.5}$	sigmoid E_{max}
10	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	sigmoid E_{max}
11	(0,2,4,6,8)	250	$\sqrt{4.5}$	Beta 1
12	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Beta 1
13	(0,2,4,6,8)	250	$\sqrt{4.5}$	Beta 2
14	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Beta 2
15	(0,2,4,6,8)	250	$\sqrt{4.5}$	Exponential
16	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Exponential
17	(0,2,4,6,8)	250	$\sqrt{4.5}$	Peak
18	(0,1,2,3,4,5,6,7,8)	252	$\sqrt{4.5}$	Peak

Table H.3: The 24 models used in the simulation study. The table gives an abbreviation for each model and a short description of it. Additionally - depending on the model type - information is given about the number of inner knots, if the unimodality constraint was active or not, the type of the penalty, and the way, in which the tuning parameter and the mode location were determined.

Abbr.	Model type	Inner knots	Constraint	Penalty	Estimation of λ	Mode estimation
un	cubic spline	$g = u - 5$	no	no	–	–
cn	cubic spline	$g = u - 5$	yes	no	–	minizing RSS
ud2	cubic spline	$g = 10$	no	difference (2nd)	REML	–
cd2	cubic spline	$g = 10$	yes	difference (2nd)	REML	minizing RSS
cda2	cubic spline	$g = 10$	yes	difference (2nd)	approx. REML	minizing RSS
ud3	cubic spline	$g = 10$	no	difference (3rd)	REML	–
cd3	cubic spline	$g = 10$	yes	difference (3rd)	REML	minizing RSS
cda3	cubic spline	$g = 10$	yes	difference (3rd)	approx. REML	minizing RSS
us	cubic spline	$g = 10$	no	sigmoid E_{max}	REML	–
cs	cubic spline	$g = 10$	yes	sigmoid E_{max}	REML	minizing RSS
csa	cubic spline	$g = 10$	yes	sigmoid E_{max}	approx. REML	minizing RSS
udb	cubic spline (Bayes)	$g = 10$	no	difference (2nd)	posterior mean	–
cdb2	cubic spline (Bayes)	$g = 10$	yes	difference (2nd)	posterior mean	posterior mean
cdb3	cubic spline (Bayes)	$g = 10$	yes	difference (3rd)	posterior mean	posterior mean
csb	cubic spline (Bayes)	$g = 10$	yes	sigmoid E_{max}	posterior mean	posterior mean
trafo	cubic spline (Bayes)	$g = 10$	yes	difference (2nd)	posterior mean	transformation
uBPr	Bernstein polyn. ($k = 13$)	$g = 0$	no	ridge	REML	–
cBPr	Bernstein polyn. ($k = 13$)	$g = 0$	yes	ridge	REML	minizing RSS
uBPs	Bernstein polyn. ($k = 13$)	$g = 0$	no	sigmoid E_{max}	REML	–
cBPs	Bernstein polyn. ($k = 13$)	$g = 0$	yes	sigmoid E_{max}	REML	minizing RSS
sigE	sigmoid E_{max}	–	–	–	–	–
beta	beta	–	–	–	–	–
modAve	model averaging	–	–	–	–	–
frisen	Frisén (1986)	–	yes	–	–	minizing RSS

Table H.4: Mean and standard deviation (in brackets) of the average squared error in function prediction, $\mathrm{ASE}_\kappa(\hat{f}_{\rho,\nu}; \boldsymbol{z} = (0, 0.01, \ldots, 8)$, for all methods $\nu = 1, \ldots, 24$ and for all scenarios with 5 doses ($\varrho \in \{1, 3, \ldots, 17\}$). Mean and standard deviation are calculated over all $\kappa = 1, \ldots, 500$ generated data sets. Thus the mean values are identical to the values of the first measure of loss, $\mathcal{L}_1(\rho, \nu)$, defined in Appendix G.

	Scen1	Scen3	Scen5	Scen7	Scen9	Scen11	Scen13	Scen15	Scen17
un	0.06 (0.05)	0.06 (0.05)	0.06 (0.05)	0.07 (0.05)	0.06 (0.05)	0.07 (0.05)	0.07 (0.05)	0.06 (0.06)	0.21 (0.07)
cn	0.04 (0.04)	0.05 (0.04)	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.06 (0.05)	0.06 (0.05)	0.04 (0.05)	0.18 (0.05)
ud2	0.03 (0.04)	0.04 (0.04)	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.08 (0.06)	0.07 (0.05)	0.05 (0.05)	0.19 (0.05)
cd2	0.03 (0.03)	0.04 (0.04)	0.05 (0.04)	0.06 (0.05)	0.06 (0.04)	0.08 (0.06)	0.07 (0.05)	0.05 (0.04)	0.18 (0.04)
cda2	0.03 (0.03)	0.04 (0.04)	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.08 (0.05)	0.07 (0.05)	0.05 (0.04)	0.18 (0.05)
udb	0.04 (0.03)	0.04 (0.04)	0.05 (0.04)	0.07 (0.05)	0.05 (0.04)	0.08 (0.06)	0.07 (0.05)	0.05 (0.05)	0.19 (0.06)
cdb2	0.03 (0.03)	0.04 (0.04)	0.04 (0.03)	0.05 (0.04)	0.06 (0.04)	0.10 (0.06)	0.08 (0.05)	0.05 (0.05)	0.17 (0.04)
trafo	0.03 (0.03)	0.04 (0.04)	0.05 (0.04)	0.06 (0.05)	0.05 (0.03)	0.08 (0.05)	0.07 (0.05)	0.05 (0.04)	0.18 (0.04)
ud3	0.05 (0.04)	0.05 (0.05)	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.08 (0.06)	0.07 (0.05)	0.06 (0.05)	0.20 (0.07)
cd3	0.04 (0.03)	0.05 (0.04)	0.05 (0.04)	0.07 (0.04)	0.06 (0.04)	0.07 (0.05)	0.07 (0.05)	0.05 (0.05)	0.18 (0.05)
cda3	0.04 (0.04)	0.05 (0.04)	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.07 (0.05)	0.07 (0.05)	0.05 (0.05)	0.19 (0.05)
cdb3	0.04 (0.03)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.07 (0.05)	0.07 (0.05)	0.05 (0.05)	0.18 (0.05)
us	0.04 (0.04)	0.06 (0.05)	0.07 (0.06)	0.10 (0.07)	0.06 (0.04)	0.08 (0.06)	0.09 (0.06)	0.06 (0.05)	0.19 (0.05)
cs	0.04 (0.04)	0.06 (0.05)	0.07 (0.06)	0.10 (0.07)	0.06 (0.04)	0.09 (0.06)	0.09 (0.06)	0.06 (0.05)	0.19 (0.05)
csu	0.04 (0.04)	0.06 (0.05)	0.07 (0.06)	0.10 (0.07)	0.06 (0.04)	0.08 (0.06)	0.09 (0.06)	0.06 (0.05)	0.19 (0.05)
csb	0.04 (0.04)	0.06 (0.05)	0.07 (0.06)	0.10 (0.07)	0.06 (0.04)	0.10 (0.06)	0.09 (0.06)	0.06 (0.05)	0.18 (0.05)
uBPr	0.08 (0.08)	0.14 (0.08)	0.14 (0.09)	0.11 (0.08)	0.12 (0.08)	0.13 (0.09)	0.15 (0.11)	0.13 (0.07)	0.22 (0.09)
cBPr	0.08 (0.08)	0.13 (0.08)	0.13 (0.08)	0.09 (0.07)	0.12 (0.08)	0.13 (0.08)	0.16 (0.10)	0.12 (0.08)	0.23 (0.10)
uBPs	0.04 (0.04)	0.07 (0.06)	0.07 (0.06)	0.07 (0.05)	0.06 (0.05)	0.11 (0.08)	0.12 (0.08)	0.06 (0.06)	0.19 (0.06)
cBPs	0.04 (0.04)	0.07 (0.06)	0.07 (0.06)	0.07 (0.05)	0.06 (0.05)	0.10 (0.06)	0.11 (0.07)	0.06 (0.06)	0.19 (0.05)
sigE	0.04 (0.04)	0.07 (0.06)	0.08 (0.07)	0.12 (0.09)	0.06 (0.04)	0.23 (0.08)	0.14 (0.08)	0.06 (0.05)	0.19 (0.05)
beta	0.07 (0.07)	0.06 (0.05)	0.07 (0.06)	0.07 (0.06)	0.06 (0.05)	0.13 (0.15)	0.07 (0.05)	0.06 (0.07)	0.22 (0.09)
frisen	0.03 (0.03)	0.05 (0.04)	0.05 (0.04)	0.07 (0.05)	0.05 (0.04)	0.38 (0.06)	0.12 (0.05)	0.05 (0.05)	0.19 (0.05)
modAve	0.03 (0.04)	0.04 (0.05)	0.05 (0.05)	0.07 (0.05)	0.06 (0.04)	0.13 (0.15)	0.07 (0.05)	0.05 (0.05)	0.19 (0.06)

Table H.5: Mean and standard deviation (in brackets) of the average squared error in function prediction, $\mathrm{ASE}_\kappa(\hat{f}_{\rho,\nu}; \boldsymbol{z} = (0, 0.01, \ldots, 8)$, for all methods $\nu = 1, \ldots, 24$ and for all scenarios with 9 doses ($\varrho \in \{2, 4, \ldots, 18\}$). Mean and standard deviation are calculated over all $\kappa = 1, \ldots, 500$ generated data sets. Thus the mean values are identical to the values of the first measure of loss, $\mathcal{L}_1(\rho, \nu)$, defined in Appendix G.

	Scen2	Scen4	Scen6	Scen8	Scen10	Scen12	Scen14	Scen16	Scen18
un	0.15 (0.08)	0.16 (0.08)	0.14 (0.08)	0.15 (0.08)	0.15 (0.08)	0.15 (0.08)	0.14 (0.08)	0.15 (0.08)	0.21 (0.08)
cn	0.08 (0.05)	0.07 (0.05)	0.08 (0.05)	0.08 (0.05)	0.08 (0.05)	0.09 (0.05)	0.08 (0.05)	0.06 (0.05)	0.16 (0.06)
ud2	0.04 (0.04)	0.04 (0.04)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.07 (0.05)	0.06 (0.05)	0.05 (0.04)	0.17 (0.04)
cd2	0.04 (0.03)	0.04 (0.04)	0.05 (0.04)	0.05 (0.03)	0.06 (0.04)	0.07 (0.05)	0.06 (0.04)	0.05 (0.04)	0.16 (0.03)
cda2	0.04 (0.04)	0.04 (0.04)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.07 (0.04)	0.06 (0.04)	0.05 (0.04)	0.16 (0.04)
udb	0.04 (0.04)	0.04 (0.04)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.07 (0.05)	0.06 (0.04)	0.05 (0.04)	0.16 (0.04)
cdb2	0.03 (0.03)	0.04 (0.04)	0.05 (0.03)	0.05 (0.03)	0.06 (0.04)	0.08 (0.05)	0.07 (0.04)	0.05 (0.04)	0.16 (0.03)
trafo	0.04 (0.04)	0.04 (0.03)	0.05 (0.03)	0.05 (0.03)	0.05 (0.04)	0.07 (0.04)	0.06 (0.04)	0.04 (0.04)	0.15 (0.03)
ud3	0.05 (0.04)	0.06 (0.04)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.07 (0.04)	0.06 (0.04)	0.05 (0.04)	0.17 (0.05)
cd3	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.04 (0.04)	0.16 (0.04)
cda3	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.04 (0.04)	0.16 (0.04)
cdb3	0.04 (0.04)	0.05 (0.04)	0.05 (0.04)	0.04 (0.03)	0.05 (0.04)	0.06 (0.04)	0.06 (0.04)	0.04 (0.04)	0.16 (0.03)
us	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.06 (0.05)	0.06 (0.04)	0.09 (0.05)	0.09 (0.05)	0.06 (0.05)	0.17 (0.04)
cs	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.09 (0.05)	0.08 (0.04)	0.06 (0.05)	0.17 (0.04)
csu	0.05 (0.04)	0.07 (0.05)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.09 (0.05)	0.09 (0.04)	0.06 (0.05)	0.17 (0.04)
csb	0.06 (0.04)	0.07 (0.05)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.10 (0.05)	0.08 (0.04)	0.06 (0.05)	0.17 (0.04)
uBPr	0.08 (0.08)	0.14 (0.08)	0.13 (0.08)	0.13 (0.08)	0.13 (0.09)	0.15 (0.08)	0.14 (0.09)	0.13 (0.08)	0.20 (0.08)
cBPr	0.08 (0.08)	0.11 (0.08)	0.11 (0.07)	0.09 (0.06)	0.11 (0.08)	0.12 (0.07)	0.12 (0.08)	0.12 (0.07)	0.20 (0.08)
uBPs	0.05 (0.04)	0.07 (0.06)	0.07 (0.05)	0.06 (0.05)	0.06 (0.05)	0.11 (0.07)	0.12 (0.06)	0.06 (0.05)	0.18 (0.06)
cBPs	0.05 (0.04)	0.07 (0.06)	0.07 (0.05)	0.06 (0.04)	0.06 (0.05)	0.10 (0.06)	0.10 (0.06)	0.06 (0.05)	0.17 (0.04)
sigE	0.05 (0.04)	0.07 (0.05)	0.07 (0.05)	0.07 (0.05)	0.06 (0.05)	0.20 (0.10)	0.12 (0.05)	0.06 (0.05)	0.18 (0.04)
beta	0.06 (0.05)	0.06 (0.05)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.07 (0.06)	0.06 (0.05)	0.05 (0.05)	0.18 (0.05)
frisen	0.04 (0.04)	0.07 (0.04)	0.06 (0.04)	0.06 (0.04)	0.06 (0.04)	0.30 (0.04)	0.11 (0.04)	0.06 (0.04)	0.17 (0.03)
modAve	0.04 (0.04)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.06 (0.04)	0.08 (0.06)	0.06 (0.05)	0.05 (0.04)	0.17 (0.04)

Table H.6: Mean and standard deviation (in brackets) of the squared error in estimation of the maximum value, $\text{ASE}_\kappa(\hat{f}_{\rho,\nu}; \boldsymbol{z} = \text{mod}(\hat{f}_{\rho,\nu}))$, for all methods $\nu = 1, \ldots, 24$ and for all scenarios with 5 doses ($\varrho \in \{1, 3, \ldots, 17\}$). Mean and standard deviation are calculated over all $\kappa = 1, \ldots, 500$ generated data sets. Thus the mean values are identical to the values of the second measure of loss, $\mathcal{L}_2(\rho, \nu)$, defined in Appendix G.

	Scen1	Scen3	Scen5	Scen7	Scen9	Scen11	Scen13	Scen15	Scen17
un	0.12 (0.12)	0.07 (0.10)	0.06 (0.09)	0.07 (0.10)	0.08 (0.10)	0.06 (0.09)	0.06 (0.09)	0.09 (0.13)	1.97 (0.55)
cn	0.09 (0.10)	0.07 (0.10)	0.06 (0.08)	0.06 (0.08)	0.07 (0.09)	0.05 (0.08)	0.06 (0.08)	0.09 (0.13)	2.13 (0.55)
ud2	0.09 (0.13)	0.11 (0.17)	0.07 (0.10)	0.10 (0.15)	0.15 (0.17)	0.08 (0.11)	0.08 (0.10)	0.18 (0.27)	2.20 (0.62)
cd2	0.06 (0.09)	0.10 (0.12)	0.05 (0.07)	0.05 (0.09)	0.08 (0.12)	0.10 (0.13)	0.10 (0.12)	0.11 (0.16)	2.33 (0.57)
cda2	0.08 (0.12)	0.12 (0.18)	0.06 (0.09)	0.08 (0.13)	0.14 (0.17)	0.08 (0.11)	0.08 (0.10)	0.18 (0.27)	2.26 (0.61)
udb	0.11 (0.14)	0.13 (0.17)	0.06 (0.10)	0.10 (0.15)	0.13 (0.17)	0.08 (0.11)	0.08 (0.10)	0.18 (0.26)	2.08 (0.61)
cdb2	0.03 (0.05)	0.07 (0.10)	0.04 (0.05)	0.04 (0.06)	0.05 (0.07)	0.13 (0.16)	0.11 (0.12)	0.16 (0.19)	2.49 (0.50)
trafo	0.07 (0.10)	0.12 (0.17)	0.06 (0.09)	0.08 (0.12)	0.13 (0.17)	0.08 (0.11)	0.08 (0.10)	0.18 (0.25)	2.27 (0.57)
ud3	0.19 (0.25)	0.20 (0.32)	0.06 (0.12)	0.14 (0.23)	0.15 (0.23)	0.07 (0.10)	0.07 (0.09)	0.34 (0.49)	1.88 (0.71)
cd3	0.13 (0.17)	0.15 (0.21)	0.06 (0.09)	0.07 (0.12)	0.11 (0.18)	0.07 (0.10)	0.07 (0.09)	0.26 (0.33)	2.04 (0.63)
cda3	0.14 (0.18)	0.18 (0.28)	0.06 (0.10)	0.11 (0.18)	0.13 (0.21)	0.07 (0.10)	0.07 (0.09)	0.32 (0.43)	1.99 (0.65)
cdb3	0.06 (0.09)	0.10 (0.15)	0.05 (0.06)	0.05 (0.07)	0.07 (0.10)	0.07 (0.10)	0.07 (0.09)	0.21 (0.32)	2.33 (0.55)
us	0.08 (0.11)	0.07 (0.11)	0.04 (0.06)	0.04 (0.06)	0.06 (0.08)	0.12 (0.15)	0.13 (0.14)	0.10 (0.14)	2.22 (0.58)
cs	0.08 (0.10)	0.07 (0.10)	0.04 (0.05)	0.04 (0.06)	0.06 (0.08)	0.15 (0.16)	0.13 (0.13)	0.10 (0.14)	2.23 (0.58)
csu	0.08 (0.11)	0.08 (0.11)	0.04 (0.05)	0.04 (0.06)	0.06 (0.08)	0.12 (0.15)	0.13 (0.14)	0.10 (0.14)	2.23 (0.59)
csb	0.09 (0.13)	0.07 (0.10)	0.03 (0.05)	0.03 (0.05)	0.05 (0.07)	0.17 (0.17)	0.12 (0.12)	0.10 (0.14)	2.17 (0.58)
uBPr	0.13 (0.20)	0.09 (0.13)	0.10 (0.18)	0.09 (0.15)	0.07 (0.09)	0.11 (0.17)	0.11 (0.17)	0.10 (0.15)	2.21 (0.80)
cBPr	0.09 (0.13)	0.12 (0.16)	0.08 (0.11)	0.06 (0.11)	0.07 (0.09)	0.13 (0.18)	0.13 (0.17)	0.17 (0.23)	2.51 (0.89)
uBPs	0.08 (0.10)	0.07 (0.10)	0.04 (0.06)	0.04 (0.06)	0.06 (0.08)	0.13 (0.18)	0.13 (0.16)	0.09 (0.14)	2.22 (0.58)
cBPs	0.08 (0.10)	0.07 (0.10)	0.04 (0.05)	0.04 (0.06)	0.06 (0.08)	0.13 (0.15)	0.10 (0.12)	0.10 (0.14)	2.22 (0.58)
sigE	0.08 (0.11)	0.07 (0.11)	0.04 (0.06)	0.04 (0.06)	0.06 (0.08)	0.46 (0.22)	0.22 (0.16)	0.10 (0.14)	2.23 (0.58)
beta	0.15 (0.25)	0.07 (0.10)	0.06 (0.09)	0.05 (0.07)	0.09 (0.11)	0.17 (0.51)	0.07 (0.10)	0.09 (0.14)	1.98 (0.65)
frisen	0.06 (0.09)	0.08 (0.10)	0.04 (0.06)	0.06 (0.09)	0.06 (0.09)	0.60 (0.23)	0.22 (0.16)	0.09 (0.13)	2.45 (0.60)
modAve	0.07 (0.09)	0.06 (0.09)	0.05 (0.06)	0.05 (0.07)	0.06 (0.08)	0.18 (0.49)	0.08 (0.11)	0.10 (0.13)	2.27 (0.57)

Table H.7: Mean and standard deviation (in brackets) of the squared error in estimation of the maximum value, $\mathrm{ASE}_\kappa(\hat{f}_{\rho,\nu}; \boldsymbol{z} = \mathrm{mod}(\hat{f}_{\rho,\nu}))$, for all methods $\nu = 1, \ldots, 24$ and for all scenarios with 9 doses ($\varrho \in \{2, 4, \ldots, 18\}$). Mean and standard deviation are calculated over all $\kappa = 1, \ldots, 500$ generated data sets. Thus the mean values are identical to the values of the second measure of loss, $\mathcal{L}_2(\rho, \nu)$, defined in Appendix G.

	Scen2	Scen4	Scen6	Scen8	Scen10	Scen12	Scen14	Scen16	Scen18
un	0.29 (0.27)	0.13 (0.21)	0.16 (0.21)	0.19 (0.23)	0.18 (0.24)	0.13 (0.17)	0.12 (0.18)	0.15 (0.20)	0.92 (0.53)
cn	0.30 (0.29)	0.14 (0.23)	0.15 (0.19)	0.17 (0.21)	0.18 (0.24)	0.11 (0.15)	0.09 (0.13)	0.14 (0.19)	1.20 (0.49)
ud2	0.11 (0.19)	0.13 (0.20)	0.08 (0.15)	0.12 (0.25)	0.20 (0.25)	0.07 (0.09)	0.06 (0.08)	0.21 (0.37)	1.85 (0.62)
cd2	0.08 (0.12)	0.11 (0.15)	0.05 (0.10)	0.05 (0.10)	0.12 (0.17)	0.09 (0.11)	0.07 (0.09)	0.12 (0.17)	2.09 (0.52)
cda2	0.10 (0.17)	0.14 (0.20)	0.07 (0.13)	0.10 (0.22)	0.19 (0.24)	0.07 (0.09)	0.06 (0.08)	0.21 (0.36)	1.92 (0.60)
udb	0.13 (0.19)	0.14 (0.22)	0.08 (0.14)	0.12 (0.25)	0.19 (0.25)	0.07 (0.09)	0.06 (0.08)	0.21 (0.35)	1.73 (0.57)
cdb2	0.04 (0.06)	0.09 (0.11)	0.04 (0.05)	0.04 (0.06)	0.07 (0.10)	0.10 (0.12)	0.08 (0.09)	0.21 (0.23)	2.15 (0.46)
trafo	0.09 (0.15)	0.14 (0.20)	0.07 (0.12)	0.09 (0.16)	0.18 (0.24)	0.07 (0.08)	0.06 (0.08)	0.21 (0.34)	1.87 (0.53)
ud3	0.24 (0.37)	0.24 (0.52)	0.07 (0.15)	0.19 (0.45)	0.22 (0.34)	0.06 (0.10)	0.05 (0.07)	0.45 (0.74)	1.54 (0.67)
cd3	0.15 (0.24)	0.18 (0.29)	0.06 (0.11)	0.08 (0.18)	0.15 (0.28)	0.06 (0.08)	0.05 (0.07)	0.29 (0.38)	1.80 (0.54)
cda3	0.17 (0.27)	0.22 (0.42)	0.06 (0.13)	0.14 (0.32)	0.19 (0.32)	0.06 (0.08)	0.05 (0.07)	0.41 (0.65)	1.69 (0.59)
cdb3	0.07 (0.12)	0.13 (0.21)	0.04 (0.06)	0.05 (0.10)	0.10 (0.17)	0.06 (0.07)	0.05 (0.07)	0.28 (0.41)	1.95 (0.51)
us	0.12 (0.18)	0.11 (0.13)	0.04 (0.06)	0.04 (0.08)	0.08 (0.12)	0.10 (0.11)	0.10 (0.11)	0.15 (0.21)	1.89 (0.55)
cs	0.12 (0.18)	0.11 (0.13)	0.04 (0.06)	0.04 (0.07)	0.08 (0.12)	0.11 (0.12)	0.09 (0.10)	0.16 (0.22)	1.93 (0.54)
csu	0.12 (0.18)	0.11 (0.13)	0.04 (0.06)	0.04 (0.07)	0.08 (0.12)	0.10 (0.12)	0.11 (0.11)	0.16 (0.21)	1.92 (0.56)
csb	0.15 (0.21)	0.10 (0.12)	0.03 (0.05)	0.03 (0.06)	0.07 (0.10)	0.11 (0.12)	0.09 (0.09)	0.16 (0.21)	1.80 (0.53)
uBPr	0.14 (0.19)	0.11 (0.15)	0.11 (0.17)	0.13 (0.20)	0.13 (0.19)	0.13 (0.18)	0.13 (0.20)	0.15 (0.20)	1.67 (0.84)
cBPr	0.11 (0.14)	0.13 (0.17)	0.08 (0.10)	0.08 (0.11)	0.10 (0.15)	0.11 (0.14)	0.10 (0.14)	0.24 (0.29)	2.03 (0.83)
uBPs	0.12 (0.17)	0.10 (0.13)	0.04 (0.06)	0.04 (0.08)	0.08 (0.11)	0.11 (0.15)	0.11 (0.12)	0.15 (0.20)	1.88 (0.55)
cBPs	0.12 (0.17)	0.11 (0.13)	0.04 (0.06)	0.04 (0.07)	0.07 (0.11)	0.09 (0.11)	0.08 (0.09)	0.15 (0.21)	1.92 (0.54)
sigE	0.12 (0.18)	0.11 (0.14)	0.04 (0.06)	0.04 (0.07)	0.08 (0.12)	0.28 (0.19)	0.15 (0.12)	0.16 (0.22)	1.94 (0.55)
beta	0.15 (0.20)	0.09 (0.12)	0.06 (0.08)	0.05 (0.07)	0.10 (0.13)	0.07 (0.09)	0.06 (0.08)	0.13 (0.18)	1.66 (0.54)
frisen	0.09 (0.15)	0.11 (0.16)	0.06 (0.11)	0.10 (0.16)	0.11 (0.16)	0.49 (0.20)	0.13 (0.12)	0.14 (0.19)	2.10 (0.57)
modAve	0.09 (0.13)	0.08 (0.11)	0.05 (0.08)	0.04 (0.07)	0.09 (0.13)	0.08 (0.10)	0.06 (0.09)	0.13 (0.18)	1.95 (0.54)

Table H.8: Mean and standard deviation (in brackets) of the squared error in mode estimation, $(\text{mod}(\hat{f}_{\rho,\nu}) - \text{mod}(f_\rho))^2$, for all methods $\nu = 1, \ldots, 24$ and for all non-flat scenarios with 5 doses ($\varrho \in \{3, \ldots, 17\}$). Mean and standard deviation are calculated over all $\kappa = 1, \ldots, 500$ generated data sets. Thus the mean values are identical to the values of the third measure of loss, $\mathcal{L}_3(\rho, \nu)$, defined in Appendix G.

	Scen3	Scen5	Scen7	Scen9	Scen11	Scen13	Scen15	Scen17
un	0.42 (1.32)	1.73 (1.93)	7.29 (8.81)	0.73 (0.81)	0.28 (0.51)	0.34 (0.44)	0.08 (0.99)	9.16 (7.74)
cn	0.44 (1.42)	1.34 (1.57)	8.06 (6.36)	0.50 (0.79)	0.33 (0.50)	0.30 (0.38)	0.05 (0.74)	8.74 (8.19)
ud2	0.11 (0.61)	2.12 (1.82)	4.75 (6.91)	0.27 (0.81)	0.41 (0.55)	0.39 (1.17)	0.03 (0.60)	12.76 (8.03)
cd2	0.50 (1.68)	1.52 (1.88)	10.11 (8.89)	0.60 (1.16)	0.36 (0.61)	0.29 (0.42)	0.04 (0.68)	10.86 (8.37)
cda2	0.26 (0.99)	1.82 (1.89)	8.45 (8.99)	0.45 (1.13)	0.34 (0.65)	0.28 (0.51)	0.04 (0.73)	10.78 (8.31)
udb	0.12 (0.60)	2.08 (1.79)	4.35 (6.78)	0.26 (0.79)	0.39 (0.59)	0.35 (0.88)	0.16 (2.93)	10.68 (8.36)
cdb2	0.28 (0.59)	0.96 (0.90)	3.61 (4.19)	0.25 (0.41)	0.61 (0.60)	0.30 (0.61)	0.09 (0.42)	4.88 (5.26)
trafo	0.14 (0.61)	1.70 (1.66)	5.02 (6.97)	0.24 (0.67)	0.40 (0.54)	0.31 (0.71)	0.03 (0.62)	8.04 (7.75)
ud3	0.34 (1.03)	1.30 (1.52)	5.32 (7.10)	0.65 (2.96)	0.45 (0.63)	0.27 (0.55)	0.30 (4.09)	10.30 (8.35)
cd3	0.51 (2.45)	2.03 (2.59)	11.61 (11.69)	0.80 (1.33)	0.31 (0.59)	0.48 (0.88)	0.06 (0.72)	9.38 (8.19)
cda3	0.41 (1.52)	1.71 (1.91)	8.63 (9.86)	0.63 (0.91)	0.37 (0.60)	0.29 (0.57)	0.06 (0.72)	9.06 (8.16)
cdb3	0.32 (0.69)	0.93 (1.13)	4.90 (5.77)	0.41 (0.55)	0.33 (0.52)	0.31 (0.46)	0.05 (0.55)	6.63 (6.14)
us	0.28 (2.41)	3.51 (2.88)	10.28 (15.09)	0.21 (0.77)	0.40 (0.69)	1.18 (2.10)	0.04 (0.77)	9.59 (8.25)
cs	0.54 (2.76)	2.95 (3.07)	10.79 (14.29)	0.71 (1.27)	0.43 (0.65)	0.75 (1.33)	0.07 (0.84)	8.46 (8.02)
csu	0.54 (2.76)	3.29 (3.32)	11.69 (15.32)	0.69 (1.24)	0.40 (0.73)	1.12 (1.94)	0.07 (0.92)	8.60 (7.98)
csb	0.60 (1.52)	1.55 (1.59)	8.37 (8.85)	0.78 (1.04)	0.42 (0.67)	0.58 (0.60)	0.11 (0.81)	6.52 (7.29)
uBPr	0.54 (1.88)	2.12 (2.46)	8.00 (12.84)	1.08 (1.96)	0.77 (0.94)	1.01 (1.41)	0.13 (1.10)	10.61 (7.94)
cBPr	0.55 (1.61)	2.13 (3.19)	8.71 (12.80)	1.10 (1.87)	0.75 (1.21)	1.00 (1.36)	0.12 (0.84)	10.56 (7.68)
uBPs	0.32 (2.76)	4.13 (3.79)	13.37 (17.95)	0.12 (0.63)	0.43 (0.59)	1.57 (2.94)	0.04 (0.86)	10.03 (8.08)
cBPs	0.55 (3.19)	3.76 (4.01)	13.50 (16.60)	0.42 (0.98)	0.40 (0.59)	0.83 (1.75)	0.05 (0.95)	9.05 (8.01)
sigE	0.00 (0.00)	4.00 (0.02)	0.00 (0.00)	0.00 (0.00)	17.99 (13.49)	9.35 (3.13)	0.00 (0.00)	16.25 (5.77)
beta	0.42 (1.16)	1.27 (1.87)	5.96 (8.82)	0.86 (2.93)	1.08 (1.41)	0.27 (0.42)	1.09 (8.04)	9.79 (5.94)
frisen	0.19 (0.50)	1.55 (1.62)	1.72 (2.52)	0.38 (0.58)	7.12 (0.19)	1.18 (1.07)	0.04 (0.24)	11.76 (8.82)
modAve	0.11 (0.68)	1.73 (1.69)	3.98 (5.89)	0.23 (0.61)	1.05 (1.27)	0.37 (0.92)	0.01 (0.17)	12.33 (7.68)

Table H.9: Mean and standard deviation (in brackets) of the squared error in mode estimation, $(\text{mod}(\hat{f}_{\rho,\nu}) - \text{mod}(f_\rho))^2$, for all methods $\nu = 1, \ldots, 24$ and for non-flat all scenarios with 9 doses ($\varrho \in \{4, \ldots, 18\}$). Mean and standard deviation are calculated over all $\kappa = 1, \ldots, 500$ generated data sets. Thus the mean values are identical to the values of the third measure of loss, $\mathcal{L}_3(\rho, \nu)$, defined in Appendix G.

	Scen4	Scen6	Scen8	Scen10	Scen12	Scen14	Scen16	Scen18
un	1.91 (4.51)	2.79 (2.60)	10.25 (10.18)	2.28 (3.44)	1.23 (1.81)	0.89 (1.31)	0.71 (4.36)	2.69 (6.00)
cn	1.10 (2.76)	1.95 (2.08)	8.36 (11.43)	1.61 (2.39)	0.66 (0.91)	0.87 (1.35)	0.20 (0.64)	3.95 (6.42)
ud2	0.10 (0.55)	2.40 (1.75)	4.88 (7.23)	0.48 (1.50)	0.43 (1.54)	0.55 (1.64)	0.01 (0.18)	9.26 (8.48)
cd2	0.63 (2.01)	1.57 (1.74)	9.64 (8.67)	0.87 (1.50)	0.31 (0.56)	0.33 (0.67)	0.04 (0.34)	6.40 (7.65)
cda2	0.28 (1.30)	1.94 (1.86)	8.31 (9.35)	0.73 (1.72)	0.30 (0.49)	0.39 (0.97)	0.02 (0.22)	6.40 (7.72)
udb	0.16 (0.70)	2.21 (1.74)	4.68 (7.24)	0.51 (1.57)	0.38 (0.60)	0.58 (1.55)	0.01 (0.21)	7.17 (8.49)
cdb2	0.30 (0.49)	1.06 (0.89)	3.94 (4.29)	0.34 (0.67)	0.42 (0.44)	0.38 (0.84)	0.12 (0.16)	2.76 (4.00)
trafo	0.19 (0.88)	1.87 (1.67)	5.25 (6.94)	0.42 (1.25)	0.35 (0.50)	0.48 (1.28)	0.02 (0.20)	4.35 (6.80)
ud3	0.67 (4.35)	1.37 (1.54)	5.46 (7.20)	1.15 (5.17)	0.43 (2.07)	0.43 (1.85)	0.42 (4.95)	8.05 (8.84)
cd3	0.60 (2.81)	1.97 (2.38)	13.55 (12.91)	1.03 (1.55)	0.27 (0.37)	0.48 (0.91)	0.11 (1.29)	5.14 (7.29)
cda3	0.40 (1.15)	1.73 (1.79)	9.07 (10.08)	0.95 (1.58)	0.26 (0.34)	0.33 (0.89)	0.04 (0.24)	4.94 (7.23)
cdb3	0.36 (0.76)	1.01 (1.16)	5.34 (6.03)	0.54 (0.82)	0.21 (0.29)	0.33 (0.63)	0.06 (0.21)	3.64 (5.25)
us	0.23 (1.45)	3.02 (1.97)	7.15 (12.58)	0.25 (1.08)	0.54 (0.78)	1.36 (2.30)	0.00 (0.02)	5.07 (7.05)
cs	0.67 (2.56)	2.22 (2.31)	10.02 (13.42)	0.96 (1.74)	0.40 (0.58)	0.63 (1.24)	0.10 (0.67)	3.47 (6.15)
csu	0.67 (2.60)	2.41 (2.50)	10.06 (13.88)	0.98 (1.83)	0.55 (0.80)	0.91 (1.53)	0.09 (0.59)	3.52 (6.13)
csb	0.69 (1.55)	1.22 (1.34)	7.95 (8.50)	0.92 (1.43)	0.45 (0.58)	0.46 (0.68)	0.14 (0.50)	3.10 (5.61)
uBPr	1.06 (3.92)	3.01 (3.10)	9.80 (12.70)	1.65 (3.42)	1.51 (2.14)	1.18 (1.65)	0.17 (1.77)	5.13 (7.42)
cBPr	0.87 (3.00)	2.09 (2.38)	9.32 (13.07)	1.51 (2.43)	1.04 (1.25)	1.21 (1.72)	0.25 (2.44)	5.46 (7.09)
uBPs	0.26 (2.16)	3.49 (2.70)	11.47 (17.24)	0.14 (0.70)	0.66 (0.84)	1.74 (2.66)	0.00 (0.00)	5.15 (6.71)
cBPs	0.64 (3.50)	2.66 (3.02)	13.91 (16.86)	0.56 (1.44)	0.36 (0.57)	0.69 (1.46)	0.06 (0.53)	4.04 (6.46)
sigE	0.00 (0.00)	4.06 (1.35)	0.00 (0.00)	0.00 (0.00)	13.64 (12.53)	8.86 (1.53)	0.00 (0.00)	15.58 (5.77)
beta	2.04 (9.78)	1.16 (1.41)	5.12 (6.40)	2.97 (11.16)	0.51 (1.01)	0.33 (1.31)	2.77 (12.84)	6.38 (6.19)
frisen	0.26 (0.59)	1.65 (1.57)	1.50 (2.39)	0.44 (0.70)	5.48 (2.89)	1.36 (1.78)	0.08 (0.26)	10.71 (8.84)
modAve	0.09 (0.44)	1.94 (1.70)	3.33 (4.67)	0.33 (0.85)	0.53 (0.88)	0.49 (1.40)	0.01 (0.12)	8.40 (7.85)

Table H.10: Values of the fourth measure of loss defined in Appendix G, $\mathcal{L}_4(\rho, \nu)$, for all methods $\nu = 1, \ldots, 24$ and for all scenarios with 5 doses ($\varrho \in \{1, 3, \ldots, 17\}$), that is, the percentage of the 500 data sets, in which the MED was incorrectly estimated.

	Scen1	Scen3	Scen5	Scen7	Scen9	Scen11	Scen13	Scen15	Scen17
un	0.00	0.45	0.35	0.39	0.00	0.81	0.84	0.47	0.68
cn	0.00	0.43	0.37	0.35	0.00	0.75	0.85	0.50	0.64
ud2	0.00	0.54	0.38	0.37	0.00	0.78	0.84	0.88	0.73
cd2	0.00	0.54	0.42	0.45	0.00	0.82	0.88	0.90	0.78
cda2	0.00	0.53	0.39	0.37	0.00	0.79	0.85	0.86	0.71
udb	0.00	0.54	0.41	0.40	0.00	0.77	0.84	0.88	0.71
cdb2	0.00	0.68	0.40	0.53	0.00	0.81	0.94	0.99	0.87
trafo	0.00	0.57	0.36	0.40	0.00	0.79	0.86	0.93	0.70
ud3	0.00	0.42	0.42	0.37	0.00	0.80	0.79	0.77	0.64
cd3	0.00	0.44	0.38	0.39	0.00	0.81	0.86	0.70	0.64
cda3	0.00	0.42	0.38	0.37	0.00	0.79	0.81	0.77	0.63
cdb3	0.00	0.47	0.36	0.39	0.00	0.79	0.80	0.92	0.70
us	0.00	0.42	0.38	0.46	0.00	0.84	0.90	0.75	0.62
cs	0.00	0.42	0.38	0.46	0.00	0.84	0.90	0.80	0.64
csu	0.00	0.42	0.38	0.46	0.00	0.84	0.90	0.75	0.64
csb	0.00	0.41	0.36	0.42	0.00	0.86	0.89	0.81	0.71
uBPr	0.00	0.60	0.51	0.42	0.00	0.90	0.90	0.38	0.57
cBPr	0.00	0.71	0.64	0.56	0.00	0.90	0.94	0.48	0.66
uBPs	0.00	0.58	0.41	0.61	0.00	0.83	0.93	0.57	0.51
cBPs	0.00	0.59	0.41	0.58	0.00	0.83	0.93	0.70	0.52
sigE	0.00	0.46	0.40	0.60	0.00	0.84	0.91	0.82	0.62
beta	0.00	0.47	0.29	0.31	0.00	0.83	0.86	0.43	0.63
frisen	0.00	0.54	0.41	0.60	0.00	0.78	0.86	0.77	0.65
modAve	0.00	0.54	0.36	0.35	0.00	0.80	0.88	0.56	0.64

Table H.11: Values of the fourth measure of loss defined in Appendix G, $\mathcal{L}_4(\rho, \nu)$, for all methods $\nu = 1, \ldots, 24$ and for all scenarios with 9 doses ($\varrho \in \{2, 4, \ldots, 18\}$), that is, the percentage of the 500 data sets, in which the MED was incorrectly estimated.

	Scen2	Scen4	Scen6	Scen8	Scen10	Scen12	Scen14	Scen16	Scen18
un	0.62	0.64	0.22	0.30	0.07	0.83	0.76	0.72	0.14
cn	0.65	0.54	0.28	0.28	0.05	0.78	0.73	0.67	0.08
ud2	0.61	0.62	0.42	0.37	0.00	0.80	0.70	0.69	0.01
cd2	0.67	0.64	0.60	0.45	0.00	0.79	0.75	0.71	0.00
cda2	0.63	0.64	0.42	0.37	0.00	0.79	0.69	0.68	0.01
udb	0.62	0.60	0.45	0.34	0.00	0.77	0.69	0.69	0.01
cdb2	0.70	0.67	0.76	0.47	0.00	0.80	0.66	0.75	0.00
trafo	0.62	0.59	0.52	0.36	0.00	0.78	0.68	0.67	0.01
ud3	0.63	0.54	0.35	0.32	0.01	0.81	0.72	0.75	0.02
cd3	0.65	0.60	0.32	0.31	0.00	0.81	0.73	0.74	0.01
cda3	0.64	0.57	0.35	0.28	0.01	0.81	0.71	0.75	0.02
cdb3	0.69	0.59	0.55	0.35	0.00	0.78	0.72	0.74	0.00
us	0.71	0.61	0.51	0.34	0.00	0.86	0.72	0.81	0.00
cs	0.71	0.61	0.64	0.35	0.00	0.87	0.74	0.81	0.00
csu	0.71	0.62	0.51	0.35	0.00	0.87	0.74	0.82	0.00
csb	0.76	0.62	0.67	0.36	0.00	0.85	0.73	0.83	0.00
uBPr	0.61	0.78	0.25	0.39	0.01	0.86	0.81	0.77	0.03
cBPr	0.65	0.84	0.41	0.49	0.00	0.88	0.82	0.80	0.01
uBPs	0.66	0.76	0.40	0.30	0.00	0.90	0.74	0.89	0.01
cBPs	0.67	0.76	0.60	0.30	0.00	0.90	0.75	0.86	0.00
sigE	0.72	0.69	0.88	0.35	0.00	0.88	0.74	0.89	0.00
beta	0.71	0.68	0.13	0.33	0.00	0.81	0.66	0.63	0.02
frisen	0.62	0.74	0.88	0.28	0.01	0.81	0.73	0.79	0.01
modAve	0.64	0.68	0.30	0.34	0.00	0.84	0.71	0.71	0.00

I Additional figures

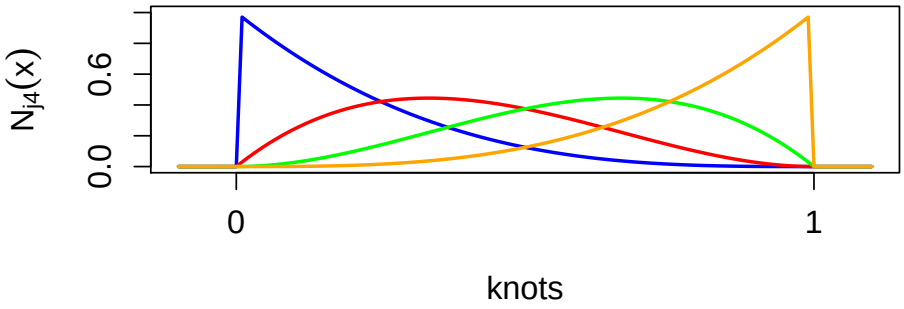

Figure I.1: Binomial probabilities with $k = 3$ (identical with cubic B-splines on the interval $[0, 1]$ without inner knots ($g = 0$) and with coincident outer knots on 0 and 1).

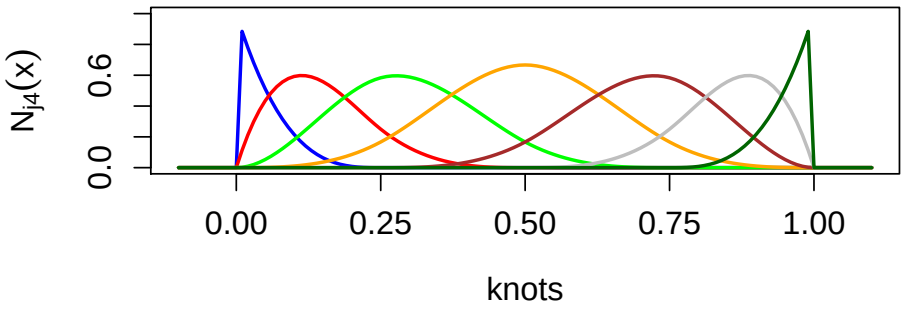

Figure I.2: Cubic B-spline basis functions with $g = 3$ equidistant inner knots in the interval $[0, 1]$ and coincident outer knots on 0 and 1.