NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Investigating posterior contour probabilities using INLA: A case study on recurrence of bladder tumours

by

Rupali Akerkar

PREPRINT
STATISTICS NO. 4/2012

NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2012/S4-2012.pdf
Rupali akerkar has homepage: http://www.math.ntnu.no/∼akerkar
E-mail: akerkar@math.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology, N-7491 Trondheim, Norway.

Investigating posterior contour probabilities using INLA: A case study on recurrence of bladder tumours

Rupali Akerkar
Department of Mathematical Sciences
NTNU, Norway

March 6, 2012

Abstract

In biomedical studies, often there exist large number of covariates. The selection of influential covariates is important. Another natural question is about the appropriate functional form of these covariates. The functional form of covariates cannot be easily judged. The selection of functional form of covariate requires simultaneous probability statements. We carry out an extensive simulation study to investigate the performance of posterior contour probabilities, computed within the INLA frame, to address these questions. Our study evaluates models with five different functional forms, which include a constant effect, a linear effect, a quadratic effect and two nonlinear effect models. We consider three signal to noise levels: low, medium and high for each model. We compare our conclusions based on posterior contour probabilities with other Bayesian model selection tools, such as the deviance information criterion (DIC) and the cross-validated logarithmic scores (log-scores). We include a case study on recurrence of bladder tumour data, to decide about the functional form of the influential covariate. We model the number of tumours occurring to the patients in the data set, using nonhomogeneous Poisson processes.

1 Introduction

The goal of this report is to investigate how appropriate are the posterior contour probabilities computed within the INLA frame in deciding the correct functional form of covariates in case of multiple events data. As a foundation, we first conduct an extensive simulation study to examine the performance of posterior contour probabilities in judging the functional form of covariates in known models.

The statistical analysis of survival data has been extensively studied in the past few decades. In many cases, an outcome of interest is the time to an event, such as death, occurrence of a tumor or even discharge from hospital. Sometimes, instead of one event the subject in the study might experience more events. The idea that an event can occur multiple times in the course of the follow-up of a subject is a conceptually easy extension of the single event model. Multiple events can be of two types, one when identical events are involved and the other when events consider are not identical. We concentrate on multiple events of single type. One example is cancer. Following treatment, the cancer may go in to remission but, at a later point in time, may recur. The defining characteristic of a recurrent event is that we observe the same event in a single subject multiple times during the follow-up period. Various methods have been proposed to discuss such situations, in which subject experiences ordered repeated events, this includes Andersen and Gill (1982), Lawless (1987), Thall (1988), . Cook and Lawless (2007) have reviewed various methods of analysis of repeated events.

In survival data modelling, Bayesian approaches have received much attention due to advances in computational and modelling techniques. Bayesian methods permit a full and exact posterior inference for any parameter or predictive quantity of interest. Inference is done using Markov chain Monte Carlo (MCMC) simulation methods. Sinha (1993) discussed the Bayesian semiparametric model for multiple event-time data, based on proportional intensity model, including a multiplicative frailty component for each individual in the model. He assumed that given the unobserved frailty random effect, the intensity function for an individual does not depend on the number of previous events experienced by the individual.

In biomedical studies, often there exist large number of covariates, not all of them are equally significant but some influence the study. The selection of influential covariates is fundamentally important, as the validity of the fitted model and its inference heavily depends on whether the model is specified correctly or not. In such cases a first step is to select a parsimonious model. Selecting a suitable model from group of reasonable models to explain a given set of data is an important problem. Another natural question is about the appropriate functional form of these covariates. In this report we focus on selecting the functional form of these covariates. This requires simultaneous probability statements. In Bayesian literature there has been limited consideration of simultaneous probability statements in multivariate inferential problems.

Besag et al. (1995) (p.30) proposed a method to calculate simultaneous credible bands based on order statistics. Their approach defines such a region as the product of symmetric univariate posterior credible intervals (of the same univariate level). The confidence intervals are typically constructed based on sample quantiles which are obtain from the Monte Carlo output. The simultaneous credible bands are thus by construction restricted to be hyperrectangular. Held (2004) argued about the risk attached with simultaneous credible bands. He suggested that even though simultaneous credible bands are visually appealing, sometimes they might be misleading about the support of the posterior as these are by construction restricted to be hyperrectangular. As an alternative, Held (2004), by taking into account Box and Tiao (1973), suggested the use of highest density region (HPD) to check the support of the posterior distribution for a certain parameter vector of interest. Consequently Held (2004) proposed an MCMC based algorithm to compute posterior contour probabilities. Using the same approach, Brezger and Lang (2008), developed posterior contour probabilities for Bayesian P-splines to additive models with Gaussian responses. They also discussed pseudo contour probabilities based on simultaneous credible intervals proposed by Besag et al. (1995).

Recently, Sørbye and Rue (2011) used integrated nested Laplace approximation (INLA) methodology to provide simultaneous credible bands for latent Gaussian models. INLA provides fast and accurate deterministic alternative to MCMC. It also computes posterior marginals for each component in model, from which posterior expectation and standard deviations can easily be found.

Sørbye and Rue (2011) presented an algorithm to compute Bayesian simultaneous credible bands analytically for subsets of the latent field. Their algorithm is based on the Gaussian approximations to the joint marginals of subsets of the latent field. When the functional form of marginal density is unknown, estimates of posterior contour probabilities are based on MCMC samples, Rao-Blackwellization approach to estimate density and Monte Carlo estimation (Held (2004)). Sørbye and Rue (2011) suggested calculating these probabilities using Monte Carlo estimation by sampling from the relevant Gaussian mixture. More details about their method are given in section 4.

Our simulation study evaluates five different models, which include a constant effect, a linear effect, a quadratic effect and two nonlinear effect models. We consider three selected noise levels: low, medium and high for each model. We compare the posterior contour probabilities with other Bayesian model selection

tools, such as the deviance information criterion (DIC) and the cross-validated logarithmic scores (log-scores). We also present a case study about multiple events data. To model the multiple events, we use the model and methodology described in Akerkar et al. (2012). We model the multiple events as nonhomogeneous Poisson processes and express it as a latent Gaussian model, which makes it possible to use INLA for Bayesian inference.

There are various approaches available in Bayesian inference for model comparison, but they do not implicitly choose the true model, but rather inform about the preference for the model given the data and other information. These preferences can be used to choose a single "best" model. The deviance information criterion (DIC) introduced by Spiegelhalter et al. (2002), is a popular Bayesian model selection criterion. It is developed for assessing model fit and comparing complex models. DIC is often used to compare different Bayesian models and provide the best model of those tested. A rough measure of selecting model on the basis of differences in DIC of candidate models, is if the differences is more than 10 might definitely rule out the model with higher DIC, differences between 5 and 10 are substantial, but **if** the difference is less than 5, then it could be misleading just to report the model with lowest DIC (Spiegelhalter et al. (2002))

However, DIC has tendency to under-penalize models with many random effects (Plummer (2008)) and "the question of what constitutes a noteworthy difference in DIC between 2 models has not yet received a satisfactory answer ($\cdots$) no credible scale has been proposed for the difference in DIC between 2 models" (Plummer (2008), p. 536).

Another approach to model checking is cross-validation, in which observed data are partitioned, with each part of the data compared to its predictions conditional on the model and the rest of the data. " The concept of such assessment is an old one. The most primitive form consists in the controlled or uncontrolled division of the data sample into two subsamples, the choice of a statistical predictor, including any necessary estimation, on one subsample and then the assessment of its performance by measuring its predictions against the other subsample" (Stone (1974), p. 111). One major difficulty in this approach is the selection of subsamples, different selection of subsamples provide different results. A solution is using the leave-one-out cross-validation. The leave-one-out cross-validation predictive density was first defined by Geisser and Eddy (1979) as a diagnostic to detect observations discrepant from a given model. This is also known as the Conditional Predictive Ordinate or CPO (Gelfand (1996)). Later CPO was discussed by many authors including Pettit (1990), Gelfand et al. (1992) and Gelfand (1996). Gneiting and Raftery (2007) describe the computation of the cross-validated logarithmic scores (log-scores) for model choice.

The DIC are routinely implemented in INLA, and INLA also provides cross validation model checks without rerunning the model for each observation in turn. This report does not discuss the many important practical issues related to use of cross validation and predictive measures in model comparison, but rather focuses on comparing our results. More details about INLA can be easily found from the web site `www.r-inla.org`.

The remainder of the report is organized as follows. Section 2 introduces latent Gaussian models with a short description to the INLA methodology. In section 3, we discuss the nonhomogeneous Poisson process model. In section 4, different model check procedures are briefly described which includes posterior contour probabilities, Bayes factor , DIC, and leave-one-out Cross validation. Section 5 contains a simulation study based on 15 different models. In section 6 we discuss the simulation results. Tumour data by Byar (1980) is re analysed in section 7. Section 8 contains the discussion.

2 Latent Gaussian models and INLA

In general, latent Gaussian models are hierarchical models where the response variables $\boldsymbol{y}$ are non-Gaussian and the latent field $\boldsymbol{x}$ assumed to have Gaussian density conditional on some hyperparameters $\boldsymbol{\theta}_1$, such that, $\boldsymbol{x} \mid \boldsymbol{\theta}_1 \sim \mathcal{N}(0, \boldsymbol{Q}^{-1}(\boldsymbol{\theta}_1))$. $\boldsymbol{x}$ is controlled by vector of hyperparameters $\boldsymbol{\theta}_1$, which are not Gaussian.

Let the likelihood for response variable $\boldsymbol{y} = \{y_i : i = 1, \cdots, N\}$ is denoted by $\pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta}_2)$ and assume that y_is are conditionally independent given $\boldsymbol{x}$ and $\boldsymbol{\theta}_2$. Hence the posterior distribution of the model is given by

$$\begin{aligned} \pi(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) &\propto \pi(\boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta}) \prod_i \pi(y_i|x_i, \boldsymbol{\theta}) \\ &\propto \pi(\boldsymbol{\theta}) \mid \mathbf{Q}(\boldsymbol{\theta}) \mid^{n/2} \exp\Big(-\frac{1}{2}\mathbf{x}^T\mathbf{Q}(\boldsymbol{\theta})\mathbf{x} + \sum_i \log \pi(y_i|x_i, \boldsymbol{\theta})\Big) \end{aligned} \tag{1}$$

Here, $\boldsymbol{\theta} = (\theta_1^{\mathrm{T}}, \theta_2^{\mathrm{T}})^{\mathrm{T}}$ with $\dim((\theta)) = M$ not very large. Except in special cases this posterior density is not analytically tractable, as the likelihood is not Gaussian. The aim is to infer the the posterior marginals of $\pi(x_i|\boldsymbol{y})$ and $\pi(\boldsymbol{\theta}|\boldsymbol{y})$.

Integrated Nested Laplace approximation (INLA) provides a recipe for computing in a fast and accurate way, approximations to marginal posterior densities for the hyperparameters $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ and for the full conditional posterior marginal densities for the latent variables $\tilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, $i = 1, \ldots, n$. Such approximations are based on Laplace or other related analytical approximations. Rue et al. (2009) discussed three different approaches with their features to approximate $\tilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})$, namely a Gaussian, a full Laplace and a simplified Laplace approximation. Posterior marginals for the latent variables $\tilde{\pi}(x_i|\boldsymbol{y})$ are then computed via numerical integration as:

$$\begin{aligned} \tilde{\pi}(x_i|\boldsymbol{y}) &= \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\, d\boldsymbol{\theta} \\ &= \sum_j \tilde{\pi}(x_i|\boldsymbol{\theta}_j, \boldsymbol{y})\tilde{\pi}(\boldsymbol{\theta}_j|\boldsymbol{y})\Delta_j \end{aligned} \tag{2}$$

Where $\boldsymbol{\theta}_j$ are points accurately chosen in the $\boldsymbol{\theta}$ space and Δ_j are integration weights. The posterior marginal $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ of the hyperparameters $\boldsymbol{\theta}$ is approximated using a Laplace approximation

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\tilde{\pi_G}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\Big|_{\boldsymbol{x}=\boldsymbol{x}^*(\boldsymbol{\theta})}$$

where $\tilde{\pi_G}(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation of the full conditional of $\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$, and $\boldsymbol{x}^*(\boldsymbol{\theta})$ is the mode of the full conditional $\pi(\boldsymbol{x} \mid \boldsymbol{\theta}, \boldsymbol{y})$ (Rue and Held, 2005).

In order for the INLA methodology to work in an efficient way, latent Gaussian models have to satisfy some additional properties which will be assumed throughout this report. First, the latent field $\boldsymbol{x}$, often of large dimension, admits conditional independence properties. In other words it is a latent Gaussian Markov random field (GMRF) with a sparse precision matrix $\boldsymbol{Q}$, ((Rue and Held, 2005)). The efficiency of INLA relies, in fact, on algorithms for sparse matrices computation. Many latent Gaussian models in the literature satisfy these conditions. The second condition to be satisfied is that the dimension of the hyperparameter vector $\boldsymbol{\theta}$ should not be too large. This is necessary for the integral in (2) to be computationally feasible. Rue et al. (2009) propose an integration scheme named Central Composite Design (CCD) which allows to explore

the $\boldsymbol{\theta}$ space using a limited number of points. Such scheme, which is the default choice in the INLA package, allows fast computations also for moderate numbers of hyperparameters (say ≤ 15).

Finally to be able to use INLA in survival analysis applications, we assume that each data point y_i depends on the latent Gaussian field only through the predictor $\eta_i = g(u_i)$, where $g(\cdot)$ is a known link function and $\pi(y_i|\boldsymbol{x}, \boldsymbol{\theta}) = \pi(y_i|\eta_i, \boldsymbol{\theta})$. The predictor function η_i for nonhomogeneous Poisson process is being discussed in the next section.

3 Nonhomogeneous Poisson process model

A common model for multiple events data (Lawless (1987); Thall (1988); Lawless and Zhan (1998)) is the nonhomogeneous Poisson process (NHPP) with a multiplicative intensity function

$$h(t|\boldsymbol{x}) = h_0(t) w_i \exp(\boldsymbol{\beta}^T \boldsymbol{z}), \quad t > 0 \tag{3}$$

where $h_0(\cdot)$ is the baseline intensity function, w_i is the unobserved frailty random variable and $\boldsymbol{\beta}$ is the vector of parameters associated with covariates $\boldsymbol{z}$. This is an extension of the proportional hazards model by Cox (1972). A major issue in extending proportional hazards model to this direction is intra-subject correlation. The frailty random effect takes care of heterogeneity among subjects with respect to their trend to develop events.

We consider $h_0(\cdot)$ to be a piecewise constant baseline intensity function, i.e. $h_0(t) = \lambda_k \quad \text{for} \quad t \in I_k = (s_{k-1}, s_k]$, where t is the event time and $0 = s_0 < s_1 < ...s_K = T$ is a pre-specified sequence of constants dividing the time axis.

Let $E_{ik} = E_i(s_k) - E_i(s_{k-1})$ be the number of events occurring to subject i in interval I_k such that $\sum_k E_{ik} = E_i$. We assume that these $\{E_{ik}; i = 1, \cdots N, k = 1, \cdots, K\}$ are independent Poisson distributed random variables (Sinha (1993)) and $\{E_{ik}\}$, the total number of events occurring to subject i by time t is a realization of a conditional nonhomogeneous Poisson process given $\boldsymbol{z}_i$, the covariate vector and unobserved frailty random variable w_i . The conditional proportional intensity function for subject i in interval I_k is given by

$$\begin{aligned} h(t; \boldsymbol{z}_i, \alpha_i) &= h_0(t_k) w_i \exp(\boldsymbol{\beta}^T \boldsymbol{z}_i), \qquad t \in I_k = (s_{k-1}, s_k] \\ &= \exp(\log(h_0(t)) + \log(w_i) + \boldsymbol{\beta}^T \boldsymbol{z}_i) \\ &= \exp(b_k + \alpha_i + \boldsymbol{\beta}^T \boldsymbol{z}_i) \\ &= \exp(\eta_{ik}) \end{aligned} \tag{4}$$

where, $\eta_{ik} = b_k + \alpha_i + \boldsymbol{\beta}^T \boldsymbol{z}_i$ with $b_k = \log(\lambda_k)$ and $\alpha_i = \log(w_i)$. The conditional distributions of the number of events, E_{ik} given w_i and $\boldsymbol{z}_i$, are independent Poisson and can be expressed as

$$E_{ik} \sim P(\exp(\eta_{ik})(s_k - s_{k-1})) \tag{5}$$

And we assume that if $E_{ik} \perp E_{ik'}|w_i, \boldsymbol{z}_i$ for $k \neq k'$. Under non informative censoring, the log-likelihood contribution of subject i is given by

$$l_i \propto \sum_{k=1}^{K} \Big\{ E_{ik} \log \Big(\exp(\eta_{ik})(s_k - s_{k-1}) \Big) - \exp(\eta_{ik})(s_k - s_{k-1}) \Big\} \tag{6}$$

As discussed in Akerkar et al. (2012), we can easily rewrite a nonhomogeneous Poisson process into a latent Gaussian model which allow us to do the Bayesian inference using integrated nested Laplace approximations (INLA).

4 Simultaneous inference

E_{ik} are modelled using nonhomogeneous Poisson processes. However, our purpose is to fit a model with only influential covariates at an appropriate complexity. In general, the structured additive predictor η_i accounts for effects of various covariates in an additive way via

$$\eta_i = f_1(z_{i1}) + \cdots + f_p(z_{ip}) + \varepsilon_i \tag{7}$$

Here, the $\{f_i(\cdot)\}$s are unknown functions of the covariates $\boldsymbol{z}$ and ε_is are unstructured terms. In principle, we are interested in appropriate complexity for parameter vector $\{f_i(\cdot)\}$ in terms of a constant, linear, quadratic or in general, a smooth fit based on simultaneous probability statements. One approach is to consider posterior contour probabilities. In order to avoid complexity with choosing influential covariates, we consider a particular case with a single covariate z with regression parameter $\boldsymbol{\beta}$ in (7). In addition to posterior contour probabilities we will also discuss another popular methods of model selection, namely Bayes factor, DIC and leave-one-out cross-validation, all these criterion are available in INLA frame.

4.1 Posterior contour probabilities

Suppose we are interested in finding posterior contour probability for a particular parameter vector $\boldsymbol{\beta} = \boldsymbol{\beta}_0$. A linear form of $\boldsymbol{\beta}_0$ could be obtained for example, by considering a linear fit to posterior means of covariate z. The definition of contour probabilities is based on highest posterior density (HPD) regions which are constructed such that the posterior density within the region is higher than outside (Held (2004), and Box and Tiao (1973)). The posterior contour probability $P(\boldsymbol{\beta}_0|\boldsymbol{y})$ of $\boldsymbol{\beta}_0$ is defined as 1 minus the content of the HPD region of $p(\boldsymbol{\beta}|y)$ which just cover $\boldsymbol{\beta}_0$ with some predefined credible level $1-\alpha$.

$$p(\boldsymbol{\beta}_0|\boldsymbol{y}) = P(p(\boldsymbol{\beta}|\boldsymbol{y}) \leq p(\boldsymbol{\beta}_0|\boldsymbol{y})|\boldsymbol{y}) \tag{8}$$

Here the posterior density function $p(\boldsymbol{\beta}|\boldsymbol{y})$ is considered as a random variable. The posterior contour probabilities in (8), can be defined using HPD-region as follow

$$p(\boldsymbol{\beta}_0|\boldsymbol{y}) = 1 - P(\boldsymbol{\beta}_0 \in R(c_\alpha)|\boldsymbol{y})$$

where $R(c_\alpha)$ is the HPD-region at predefined α level, and $c_\alpha = p(\boldsymbol{\beta}_0|\boldsymbol{y})$ is a constant, which provide information about the credible level in which the corresponding HPD-region covers the fixed vector $\boldsymbol{\beta}_0$. Theoretically HPD regions should be optimal in the sense of having the smallest volume of all credible regions, but this is hard to achieve as in many cases the posterior density is not analytically available.

For Gaussian distribution, the HPD-regions are ellipsoids and the contour probabilities can can be found analytically by

$$p(\boldsymbol{\beta_0}) = P(\pi_G(\boldsymbol{\beta};\mu,\Sigma) \leq \pi_G(\boldsymbol{\beta_0};\mu,\Sigma)|\boldsymbol{y}) = P(\chi_2^m > (\boldsymbol{\beta_0}-\mu)^T\Sigma^{-1}(\boldsymbol{\beta_0}-\mu)|\boldsymbol{y})$$

where χ_2^m denotes a chi-squared variable with m degrees of freedom.

Unfortunately the functional form of the marginal density is unknown in many cases. To estimate the contour probability, Held (2004) proposed a method based on Rao-Blackwellization and Monte Carlo estimation:

$$\widehat{P(\boldsymbol{\beta}_0|y)} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{\tilde{p}(\boldsymbol{\beta}^{(t)}|y) \leq \tilde{p}(\boldsymbol{\beta}_0|y)\} \tag{9}$$

where, $\boldsymbol{\beta}^{(t)}$, are the posterior samples obtained from MCMC simulation. Then the contour probability is simply the proportion of posterior samples with density smaller or equal to $p(\boldsymbol{\beta}_0|y)$. This method works in both cases when the functional form of posterior distribution is known and when it is unknown.

Instead of using MCMC simulation, we follow the method proposed by Sørbye and Rue (2011) to estimate posterior contour probabilities using INLA. Rue et al. (2009) in section 6.1, discussed details about approximating the posterior marginals of subset of the latent field x, introduced in section 2. ($\boldsymbol{\beta}$ is a component of this latent field). Let x_s be a subset of the latent field x. The the conditional posterior of x_s given $\boldsymbol{y}$ is

$$\pi(x_s|\boldsymbol{y}) = \int \pi(x_x|\boldsymbol{\theta}, \boldsymbol{y})\pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}$$

$\pi(x_s|\boldsymbol{y})$ is then approximated by using numerical integration (Sørbye and Rue (2011))

$$\tilde{\pi}(x_s|\boldsymbol{y}) = \sum_{j=1}^{k} w_j \tilde{\pi}_G(x_s; \mu_j, \Sigma_j|\boldsymbol{y})$$

where $\sum_{j=1}^{k} w_j = \sum_{j=1}^{k} \tilde{\pi}(\theta_j|y)\Delta_j = 1$. Δ_j denotes the area weight corresponding to θ_j. Here $\tilde{\pi}(\cdot; \mu_j, \Sigma_j|\boldsymbol{y})$ denotes the multivariate Gaussian approximation with mean $\mu_j = \mu(\theta_j)$ and covariance matrix $\Sigma_j = \Sigma(\theta_j)$.

For more details on the numerical integration, including the choice of number of terms k, we refer to Rue et al. (2009). INLA provides estimates for all weights, mean and covariance matrices of the posterior marginal for the elements of the latent fields using Gaussian mixtures. These are used to generate independent samples from this distribution which is denoted by $\{\boldsymbol{\beta}^{(t)}\}$, needed in (9).

The estimated contour probabilities are invariant to monotonic transformations of $p(\boldsymbol{\beta}|\boldsymbol{y})$, we replace the marginal densities by the log densities, as discussed (Brezger and Lang (2008)) and thus the contour probabilities can then be calculated using

$$\widehat{P(\boldsymbol{\beta}_0|y)} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{1}\{\log(\tilde{\pi}(\boldsymbol{\beta}^{(t)}|y)) \leq \log(\tilde{\pi}(\boldsymbol{\beta}_0|y))\} \tag{10}$$

4.2 Bayes factor

The Bayes factor for comparing two models may be approximated as the ratio of the marginal likelihood of the data in model 1 and model 2. If we label two competing models as M_1 and M_2 with parameters $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$. Then the Bayes factor is

$$BF = \frac{\pi(M_1|\boldsymbol{y})\pi(M_1)}{\pi(M_2|\boldsymbol{y})\pi(M_2)}$$

with

$$\pi(M_j|\boldsymbol{y}) = \int \pi(\boldsymbol{y}|\boldsymbol{\theta}_j, M_j)\pi(\boldsymbol{\theta}_j|M_j)d\boldsymbol{\theta_j} \quad j = 1, 2$$

if models are apriori equiprobable, ie $\pi(M_1) = \pi(M_2)$, then

$$BF = \frac{\pi(M_1|\boldsymbol{y})}{\pi(M_2|y)}$$

Bayes factors are difficult to compute, and the Bayes factor is only defined when the marginal density of $\boldsymbol{y}$ under each model is proper. In our context, for Bayes factor to be well defined, it is necessary that the prior of the latent field $\pi(\boldsymbol{x}|\boldsymbol{\theta})$ is proper. For intrinsic models, there are an arbitrary missing constants which cannot be determined, see for example Gelfand (1996). For details about how Bayes factor is calculated in INLA frame, we refer to Rue et al. (2009), section 6.1.

4.3 DIC- Deviance information criterion

We assume in general that the distribution of the data $\boldsymbol{y} = \{y_i : i = 1, \cdots, N\}$, depends on parameter vector $\boldsymbol{\theta}$. The Deviance Information Criterion (DIC), which was proposed by Spiegelhalter et al. (2002) as a Bayesian measure consists of two components, a term that measures goodness of fit and a penalty term for model complexity:

$$DIC = \overline{D(\boldsymbol{\theta})} + p_D \tag{11}$$

Which is based on the posterior distribution of the deviance $D(\theta)$,

$$D(\boldsymbol{\theta}) = -2\log f(\boldsymbol{y}\|\boldsymbol{\theta}) + 2\log f(\boldsymbol{y})$$

where $f(\boldsymbol{y} \mid \boldsymbol{\theta})$ is the likelihood function or the conditional joint probability density function of the observation given the unknown parameters, and $f(\boldsymbol{y})$ is some fully specified standardizing term which is function of the data alone. $\overline{D(\boldsymbol{\theta})}$ is the posterior mean deviance, given by,

$$\overline{D(\boldsymbol{\theta})} = E_{\boldsymbol{\theta}}[2\log P(\boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{y}] + 2\log f(\boldsymbol{y}),$$

which is considered as a Bayesian measure of fit. For model comparison, it is assumed without loss of generality that $f(\boldsymbol{y}) = 1$, for all models and thus

$$\overline{D(\boldsymbol{\theta})} = E_{\boldsymbol{\theta}}[2\log P(\boldsymbol{y}|\boldsymbol{\theta})|\boldsymbol{y}]$$

The effective number of parameters p_D, which measures the model complexity, is defined as the difference between the posterior mean of the deviance and the deviance evaluated at the posterior mean of the parameters:

$$p_D = \overline{D(\boldsymbol{\theta})} - D(\tilde{\boldsymbol{\theta}})$$

$\tilde{\boldsymbol{\theta}}$ is an estimate of $\boldsymbol{\theta}$ depending on $\boldsymbol{y}$. A natural choice for $\tilde{\boldsymbol{\theta}}$ is the posterior mean , $\overline{\boldsymbol{\theta}} = E[\boldsymbol{\theta}|\boldsymbol{y}]$. Thus $\overline{D(\boldsymbol{\theta})}$ is the posterior mean log likelihood, i.e., the average of the log likelihood values calculated from the parameters in each sample from the posterior and $D(\tilde{\boldsymbol{\theta}})$ is the log likelihood at the posterior mean, i.e., the log likelihood calculated when all of the parameters are set to their posterior mean. Thus, rearranging terms we get,

$$DIC = D(\tilde{\boldsymbol{\theta}}) + 2p_D \tag{12}$$

The DIC can be estimated with posterior parameter samples from obtained from Markov chain Monte Carlo (MCMC) methods, so that it can also be estimated if the marginal likelihood of a Bayesian model is not analytically available.

The DIC is contained in the standard INLA output, for details we refer to Rue et al. (2009), section 6.4. In the derivation of DIC, it assumed that the specified parametric family of probability distributions that generate future observations encompasses the true model. The observed data are used both to construct the posterior distribution and to evaluate the estimated models. Therefore, DIC tends to select over-fitted models.

4.4 Predictive measures: leave-one-out Cross validation

The cross-validation approach of model selection has the advantage of being able to incorporate improper priors whereas the Bayes factor approach may not always work because marginal densities may not exist or nay be arbitrary under improper prior specification. Several authors (Gelfand et al. (1992), Gelfand (1996)) have proposed the use of cross-validatory predictive densities. Cross Validation is based on splitting the data into two parts: first part is used to fit the model and estimate the posterior distribution of interest, while the other is used for model checking and evaluation. It is difficult to find a best way to split the data. Geisser and Eddy (1979) proposed using the Leave-one-out cross validation predictive density. A method based on the predictive distribution of the ith observation y_i given all the observations except the ith observation $\boldsymbol{y}_{-i}$. This is also known as the conditional predictive ordinate CPO (Gelfand (1996)), CPO_i is

$$\text{CPO}_i = p(y_i|\boldsymbol{y}_{-i}) = \int p(y_i|\boldsymbol{\theta}, \boldsymbol{y}_{-i}) p(\boldsymbol{\theta} \mid \boldsymbol{y}_{-i}) d\boldsymbol{\theta} \tag{13}$$

Where the suffix $-i$ indicates that the ith observation is removed. The values CPO_i give a ranking of the observations, with the most discordant having the smallest value of CPO_i. These values are computed in INLA without rerunning the model. The first term in (13) is

$$p(y_i|\boldsymbol{\theta}, \boldsymbol{y}_{-i}) = \frac{p(x_i \mid \boldsymbol{y}, \boldsymbol{\theta})}{p(y_i \mid x_i, \boldsymbol{\theta})} dx_i \tag{14}$$

and the second term in (13) is

$$p(\boldsymbol{\theta} \mid \boldsymbol{y}_{-i}) = \frac{p(\boldsymbol{\theta} \mid \boldsymbol{y}) p(y_i \mid \boldsymbol{y}_{-i})}{p(y_i \mid \boldsymbol{y}_{-i}, \boldsymbol{\theta})} \tag{15}$$

Both (14) and (15) are computed using numerical integration and so only approximations are known, for more details we refer to Held et al. (2010). The accuracy of CPO values depends on the accuracy of these approximations. Sometimes, some of the CPO values may not be reliable due to numerical problems in evaluating (14). INLA stores a vector which contains a number between zero and one for each observation. If the value of an observation is zero, it indicates that the CPO were calculated without any problem. Otherwise, the value is one , considered unreliable, indicating presence of computational problems.

Extending the use of CPO values, we also compute the predictive logarithmic score or log-score (Gneiting and Raftery (2007)), which gives a measure of the predictive quality of the model. The logarithmic scoring rule is a proper scoring rule, it combines discrepancies like predictive bias and lack of calibration into a single measure. The log-score is defined as:

$$\text{log-score} = -\frac{1}{N} \sum_{i}^{N} \log \text{CPO}_i \tag{16}$$

5 Simulation study

Before using the posterior contour probabilities for nonhomogeneous Poisson processes, we investigate its performance on general known models. We compare the posterior contour probabilities with the DIC and the log-score.

The basic form of model is

$$y_{ijk} = f_j(x_i) + \epsilon_{ijk} \tag{17}$$

where, the ϵ_{ijk} are assumed i.i.d., $\epsilon_{ijk} \sim N(0, \sigma^2_{ijk})$ and $k = 1, 2, 3$ representing low, medium and high noise level. For x_i, we assumed 100 equidistant points in the interval $[-1, 1]$ and thus $n = 100$.

Our study evaluates five different functional forms for $f_j(\cdot), j = 1, \cdots, 5$, which include a constant line, a straight line, a quadratic curve and two nonlinear curves. For each f_j, we consider three selected noise levels: low, medium and high. For investigating the behaviour of posterior contour probabilities, we consider possible models by imposing different prior structures on the vector of parameter, which we hold common for all candidate models.

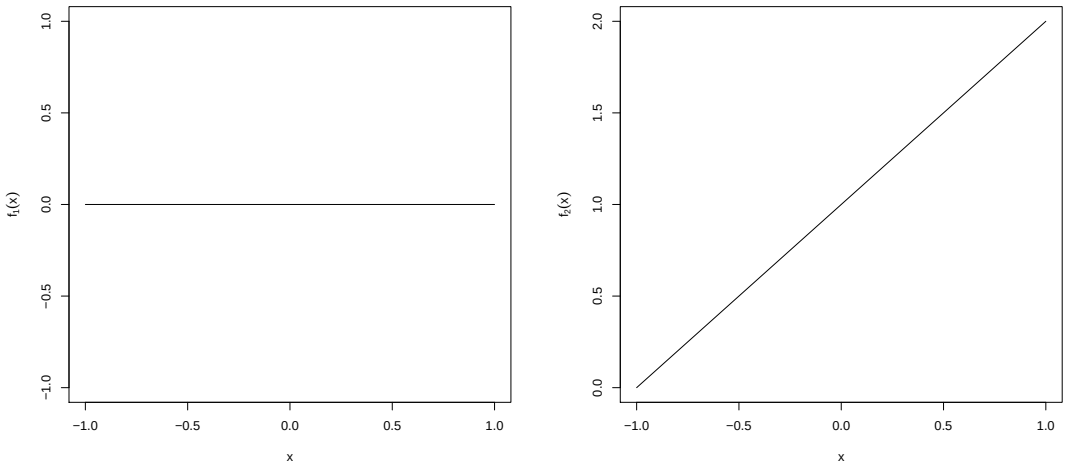
The test functions are presented in Figure 1 and details about investigated models are given in Table 1.

Model no.	Functional form	Model
1	**constant** (Figure 1(a))	$y = 0 * x + \text{rnorm}(n, 0, 0.05)$
2		$y = 0 * x + \text{rnorm}(n, 0, 0.15)$
3		$y = 0 * x + \text{rnorm}(n, 0, 0.45)$
4	**linear** (Figure 1(b))	$y = 1 + x + \text{rnorm}(n, 0, 0.1)$
5		$y = 1 + x + \text{rnorm}(n, 0, 0.2)$
6		$y = 1 + x + \text{rnorm}(n, 0, 0.4)$
7	**quadratic** (Figure 1(c))	$y = 1 + x^2 + \text{rnorm}(n, 0, 0.08)$
8		$y = 1 + x^2 + \text{rnorm}(n, 0, 0.16)$
9		$y = 1 + x^2 + \text{rnorm}(n, 0, 0.24)$
10	**nonlinear 1** (Figure 1(d))	$y = \underbrace{1 + 0.5x}_{-1 \le x \le 0} + \underbrace{1 + 2x^2}_{0 < x \le 1} + \text{rnorm}(n, 0, 0.03)$
11		$y = \underbrace{1 + 0.5x}_{-1 \le x \le 0} + \underbrace{1 + 2x^2}_{0 < x \le 1} + \text{rnorm}(n, 0, 0.12)$
12		$y = \underbrace{1 + 0.5x}_{-1 \le x \le 0} + \underbrace{1 + 2x^2}_{0 < x \le 1} + \text{rnorm}(n, 0, 0.0.48)$
13	**nonlinear 2** (Figure 1(e))	$y = 1 + sin(2\pi x) + \text{rnorm}(n, 0, 0.15)$
14		$y = 1 + sin(2\pi x) + \text{rnorm}(n, 0, 0.30)$
15		$y = 1 + sin(2\pi x) + \text{rnorm}(n, 0, 0.60)$

Table 1: The functions used as test functions in (17).

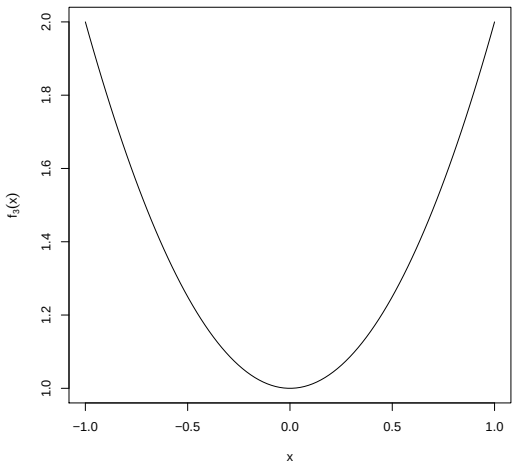
For each model we proceed in the following way:

1. We generate 250 data sets from each model in Table 1.
2. We fit each of the 250 data sets to models with three different functional forms: (i) smooth , (ii) linear and (iii) constant.
3. We calculate the DIC and the log-score values considering same priors.
4. We calculate posterior contour probabilities using (10).
5. We plot the simulation results of the DIC, log-scores and posterior contour probabilities.

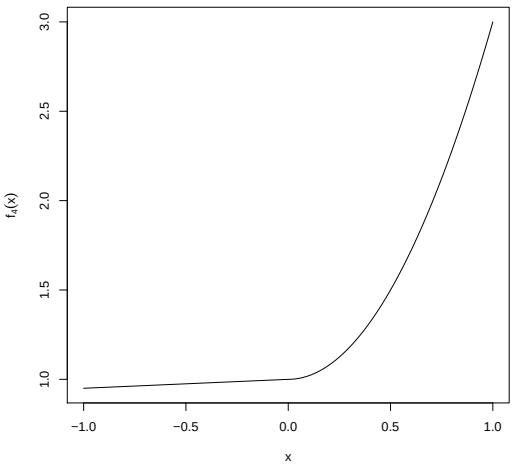


(a)

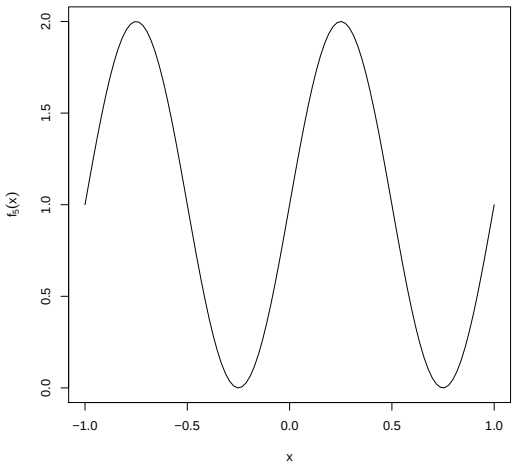
(b)



(c)



(d)



(e)

Figure 1: The test functions used in simulation study: (a) Constant, (b) Linear, (c) Quadratic, (d) nonlinear 1, and (e) nonlinear 2 .

Computational details about posterior contour probabilities

For calculating posterior contour probabilities, we fix $\boldsymbol{\beta}_0$ in (10), corresponding to different functional forms, smooth, quadratic, linear and constant. The details about how we fix the vector $\boldsymbol{\beta}_0$ for these functional forms are given in Table 2. A smooth fit corresponds to fitting a spline to the posterior means of the covariate. Which is the most general case. A quadratic curve and a line are special cases of nonlinear fit.

For calculating the posterior contour probabilities, we model $f(.)$ as smooth effect (RW2) and use the default integration strategy, giving k = 9 terms in the mixture (2).

In cases when the contour probabilities are either equal or one for more than one functional form, we select the least complex functional form.

Functional forms	$\boldsymbol{\beta}_0$
constant	zero vector
linear	fit a straight line to the posterior means of x.
quadratic	fit a quadratic curve to the posterior means of x.
non-linear	fit a spline to the posterior means of x.

Table 2: The fixed vector $\boldsymbol{\beta}_0$.

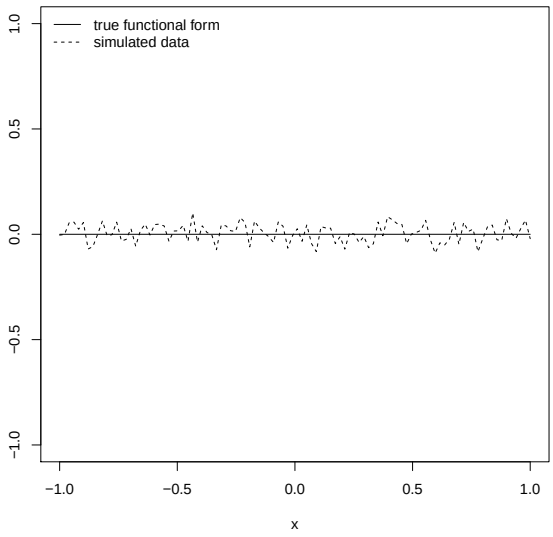


Figure 2: The function $f_1(x)$ and one simulated data set from model-1.

5.1 Model-1 : Constant effect model with low noise level

The first model is:

$$y_{i11} = f_1(x_i) + \epsilon_{i11} \tag{18}$$

where, $f_1(x) = 0 * x$ and $\epsilon_{i11} \sim \mathcal{N}(0, 0.05)$. Figure 2 shows the true functional form $f_1(x)$ and one simulated data set from model (18).

The DIC along with effective number of parameters, log-scores and posterior contour probabilities of one simulated sample from model (18), considering different functional forms for x are given in Table 3. The DIC and log-score of the model with constant effect are the lowest. The lowest DIC and log-score are highlighted. The posterior contour probability is one for all functional forms.

Figure 17(a) and 18(a) show the boxplots of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of all simulated samples from model (18). The boxplots of posterior contour probabilities of simulated samples from model (18), considering various functional forms are shown in Figure 19(a).

Prior for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-318.36	4.31	-1.59	1
Quadratic	-	-	-	1
Linear	-319.38	3.35	-1.6	1
Constant	-321.20	2.36	-1.61	1

Table 3: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-1. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

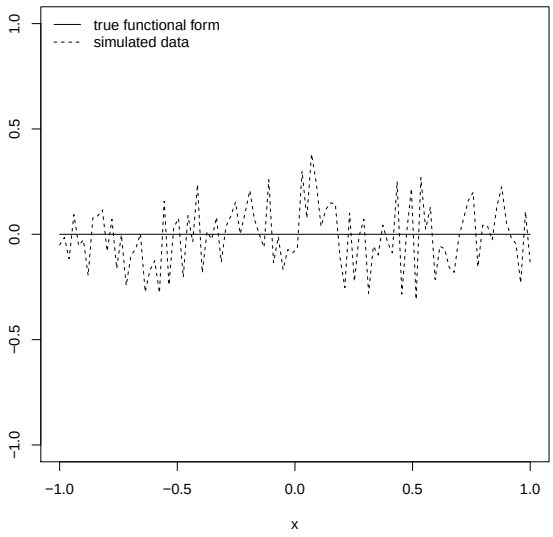


Figure 3: The function $f_1(x)$ and one simulated data set from model-2.

5.2 Model-2 : Constant effect model with medium noise level

The model is:

$$y_{i12} = f_1(x_i) + \epsilon_{i12} \tag{19}$$

where, $f_1(x) = 0 * x$ and $\epsilon_{i12} \sim \mathcal{N}(0, 0.15)$. Figure 3 shows the true functional form $f_1(x)$ and one simulated data set from model (19).

The DIC along with the effective number of parameters, log-scores and posterior contour probabilities of one simulated sample from model (19), considering different functional forms for x are given in Table 4. The DIC and log-score of the model with constant effect are the lowest. The posterior contour probability is one for all functional forms.

Figure 20(a) and 21(a) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of all simulated samples from model (19). The boxplots of posterior contour probabilities of simulated samples from model (19), considering various functional forms are shown in Figure 22(a).

Model for x	DIC	effective number of parameters	log-score probabilities	posterior contour
Nonlinear(RW2)	-90.86	3.16	-0.45	1
Quadratic	-	-	-	1
Linear	-91.74	2.72	-0.46	1
Constant	-93.48	1.72	-0.47	1

Table 4: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-2. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted..

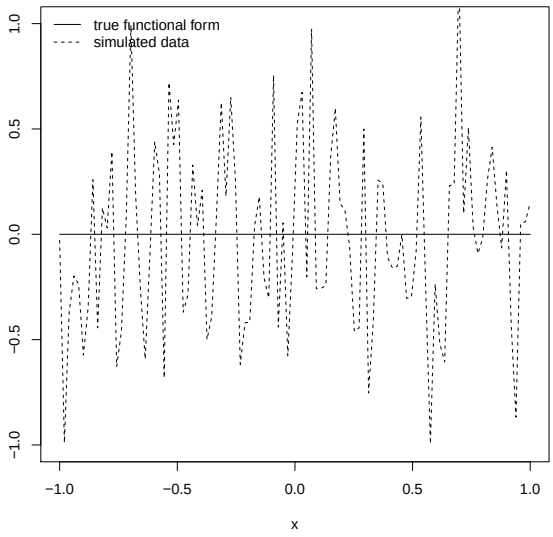


Figure 4: The function $f_1(x)$ and one simulated data set from model-3.

5.3 Model-3 : Constant effect model with high noise level

The model is:

$$y_{i13} = f_1(x_i) + \epsilon_{i13} \tag{20}$$

where, $f_1(x) = 0 * x$ and $\epsilon_{i13} \sim \mathcal{N}(0, 0.45)$. Figure 4 shows the true functional form $f_1(x)$ and one simulated data set from model (20).

The DIC, along with the effective number of parameters, log-scores and posterior contour probabilities of one simulated sample from model (20), considering different functional forms for x are given in Table 5. The DIC and log-score of the model with linear effect are the lowest. The posterior contour probability is one for all functional forms.

Figure 23(a) and 24(a) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of all simulated samples from model (20). The boxplots of posterior contour probabilities of all simulated samples from model (20), considering various functional forms are shown in Figure 25(a).

Model for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	105.49	3	0.53	1
Quadratic	-	-	-	1
Linear	104.89	2.65	0.52	1
Constant	106.79	1.65	0.53	1

Table 5: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-3. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

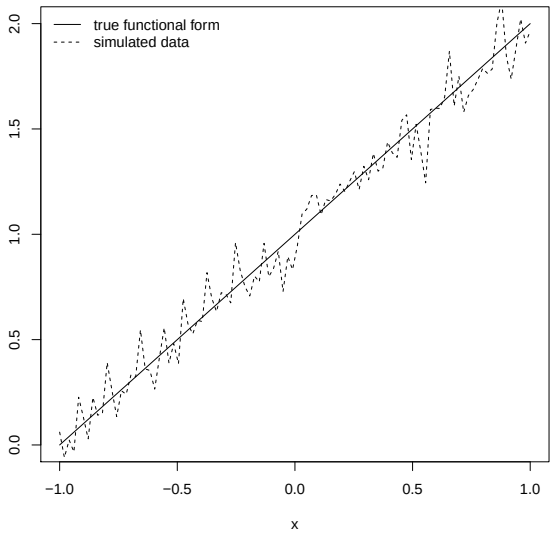


Figure 5: The function $f_2(x)$ and one simulated data set from model-4.

5.4 Model-4 : Linear effect model with low noise level

The model is:

$$y_{i21} = f_2(x_i) + \epsilon_{i21} \tag{21}$$

where, $f_2(x) = 1 + x$ and $\epsilon_{i21} \sim \mathcal{N}(0, 0.1)$. Figure 5 shows the true functional form $f_2(x)$ and one simulated data set from model (21).

The DIC, log-scores and posterior contour probabilities for one simulated sample from model (21), considering different functional forms for x are given in Table 6. The DIC of the model with linear effect is the least. The log-score is the least of the models with either smooth or linear effect. The posterior contour probability is zero for constant functional and is one for the rest of functional forms.

Figure 17(b) and 18(b) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (21). The boxplots of posterior contour probabilities of simulated samples from model (21), considering various functional forms are shown in Figure 19(b).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-159.942	3.4	-0.80	1
Quadratic	-	-	-	1
Linear	-160.72	2.79	-0.80	1
Constant	185.84	1.64	0.93	0

Table 6: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-4.The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

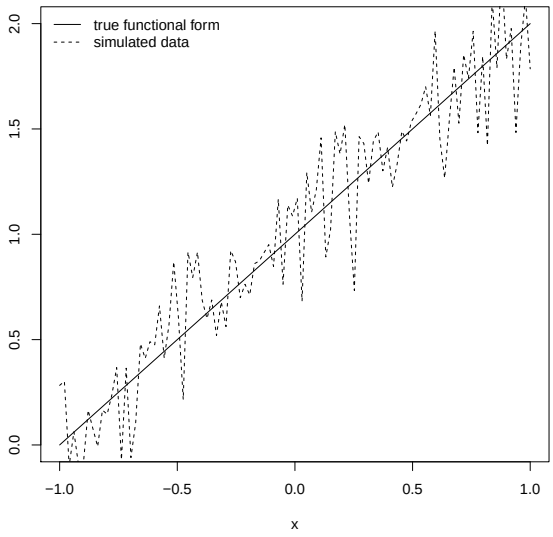


Figure 6: The function $f_2(x)$ and one simulated data set from model-5.

5.5 Model-5 : Linear effect model with medium noise level

The model is:

$$y_{i22} = f_2(x_i) + \epsilon_{i22} \tag{22}$$

where, $f_2(x) = 1 + x$ and $\epsilon_{i22} \sim \mathcal{N}(0, 0.2)$ Figure 6 shows the true functional form $f_2(x)$ and one simulated data set from model (22).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (22), considering different functional forms for x are given in Table 7. The DIC of the model with linear effect is the least. The log-score is least of the model with either smooth or linear effect. The posterior contour probability is zero for constant functional form and is one for the rest of functional forms.

Figure 20(b) and 21(b) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (22). The boxplots of posterior contour probabilities of simulated samples from model (22), considering various functional forms are shown in Figure 22(b).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Smooth	-11.78	3.1	-0.06	1
Quadratic	-	-	-	1
Linear	-12.34	2.68	-0.06	1
Constant	194.82	1.65	0.97	0

Table 7: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-5. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

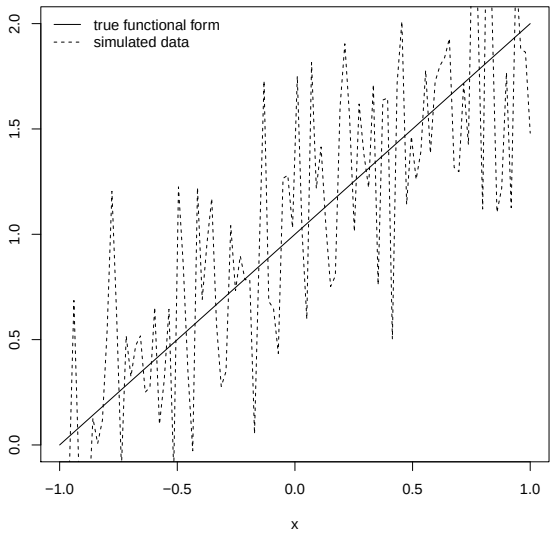


Figure 7: The function $f_2(x)$ and one simulated data set from model-6.

5.6 Model-6 : Linear effect model with high noise level

The model is:

$$y_{i23} = f_2(x_i) + \epsilon_{i23} \tag{23}$$

where, $f_2(x) = 1 + x$ and $\epsilon_{i23} \sim \mathcal{N}(0, 0.2)$. Figure 7 shows the true functional form $f_2(x)$ and one simulated data set from model (23).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (23), considering different functional forms for x are given in Table 8. The DIC of the model with linear effect is the least. The log-score is least of the model with either smooth or linear effect. The posterior contour probability is zero for constant effect and is one for the rest of functional forms.

Figure 23(b) and 24(b) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (23). The boxplots of posterior contour probabilities of simulated samples from model (23), considering various functional forms are shown in Figure 25(b).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Smooth	102.72	3	0.51	1
Quadratic	-	-	-	1
Linear	102.06	2.65	0.51	1
Constant	223.35	11.65	1.12	0

Table 8: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-6. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

5.7 Model-7: Quadratic effect model with low noise level

The model is:

$$y_{i31} = f_3(x_i) + \epsilon_{i31} \tag{24}$$

where, $f_3(x) = 1 + x^2$ and $\epsilon_{i31} \sim \mathcal{N}(0, 0.08)$. Figure 8 shows the true functional form $f_3(x)$ and one simulated data set from model (24).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (24), considering different functional forms for x are given in Table 9. The DIC and log score of the model with smooth effect is minimum. They support smooth effect for this model. The posterior contour probability is zero for constant and linear functional forms, and is one for quadratic and nonlinear functional forms.

Figure 17(c) and 18(c) shows the boxplot of the DIC and the log-score values Figure 23(b) and 24(b) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (23). The boxplots of posterior contour probabilities of simulated samples from model (24), considering various functional forms are shown in Figure 19(c).

Model for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-197.3	9	-0.98	1
Quadratic	-	-	-	1
Linear	62.21	2.66	0.31	0
Constant	60.25	1.66	0.30	0

Table 9: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-7. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

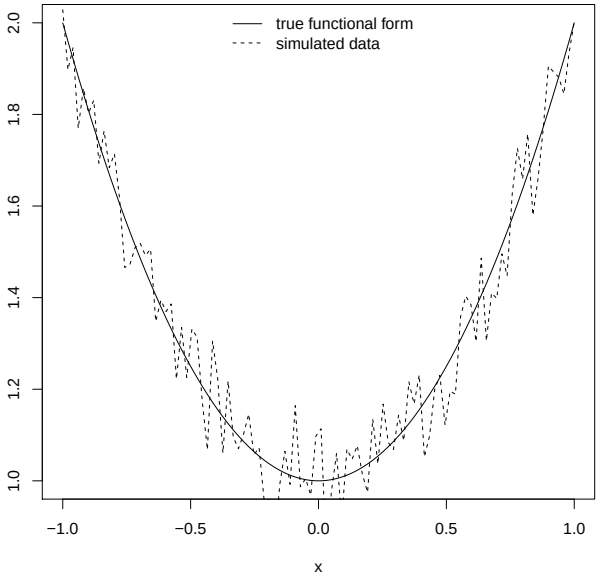


Figure 8: The function $f_3(x)$ and one simulated data set from model-7.

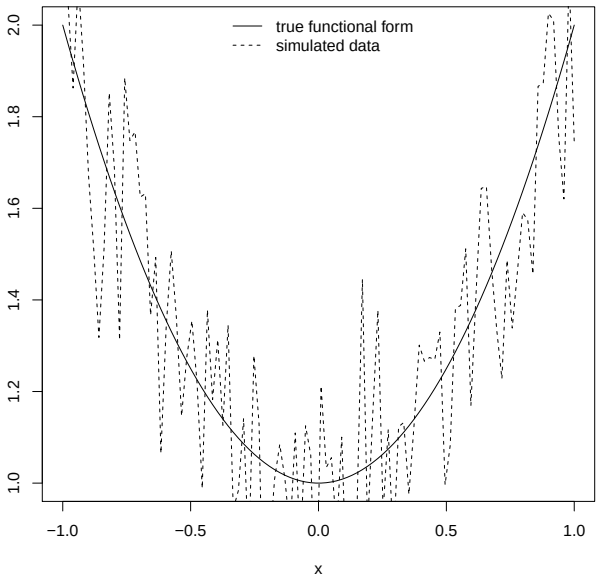


Figure 9: The function $f_3(x)$ and one simulated data set from model-8.

5.8 Model-8 : Quadratic effect model with medium noise level

The model is:

$$y_{i32} = f_3(x_i) + \epsilon_{i32} \tag{25}$$

where, $f_3(x) = 1 + x^2$ and $\epsilon_{i32} \sim \mathcal{N}(0, 0.16)$. Figure 9 shows the true functional form $f_3(x)$ and one simulated data set from model (25).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (25), considering different functional forms for x are given in Table 10. The DIC and log score of the model with smooth effect is minimum. They support smooth effect for this model. The posterior contour probability is zero for constant and linear functional forms, and is one for quadratic and nonlinear functional forms.

Figure 20(c) and 21(c) shows the boxplot of the DIC and the log-score values Figure 23(b) and 24(b) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (23). The boxplots of posterior contour probabilities of simulated samples from model (25), considering various functional forms are shown in Figure 22(c).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-88.9	7.55	-0.44	1
Quadratic	-	-	-	1
Linear	78.02	2.66	0.39	0
Constant	76.17	1.66	0.38	0

Table 10: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-8. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

5.9 Model-9 : Quadratic effect model with high noise level

The model is:

$$y_{i3} = f_3(x_i) + \epsilon_{i33} \tag{26}$$

where, $f_3(x) = 1 + x^2$ and $\epsilon_{i33} \sim \mathcal{N}(0, 0.24)$. Figure 10 shows the true functional form $f_3(x)$ and one simulated data set from model (26).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (25), considering different functional forms for x are given in Table 11. The DIC and log score of the model with smooth effect is minimum. They support smooth effect for this model. The posterior contour probability is zero for constant and linear functional forms, and is one for quadratic and nonlinear functional forms.

Figure 23(c) and 24(c) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (26). The boxplots of posterior contour probabilities of simulated samples from model (26), considering various functional forms are shown in Figure 25(c).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	37.87	6.2	0.19	1
Quadratic	-	-	-	1
Linear	128.60	2.65	0.64	0
Constant	127.18	1.65	0.65	0

Table 11: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-9. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

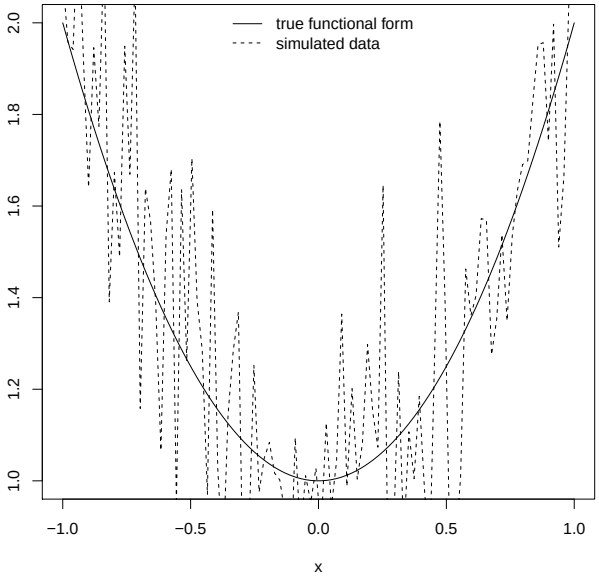


Figure 10: The function $f_3(x)$ and one simulated data set from model-9.

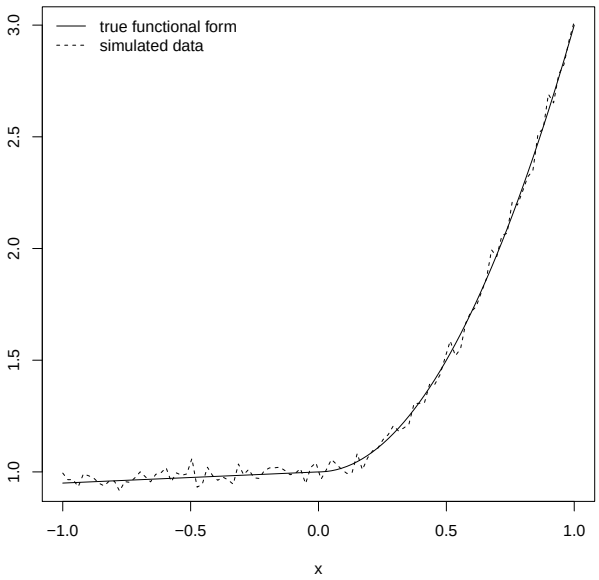


Figure 11: The function $f_4(x)$ and one simulated data set from model-10.

5.10 Model-10 : Nonlinear effect model 1 with low noise level

The model is:

$$y_{i41} = f_4(x_i) + \epsilon_{i41} \tag{27}$$

where, $f_4(x) = \underbrace{1 + 0.5x}_{-1 \leq x \leq 0} + \underbrace{1 + 2x^2}_{0 < x \leq 1}$ and $\epsilon_{i41} \sim \mathcal{N}(0, 0.03)$. Figure 11 shows the true functional form $f_4(x)$ and one simulated data set from model (27).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (27), considering different functional forms for x are given in Table 12. The DIC and log score of the model with smooth effect is minimum. The posterior contour probability is zero for constant, linear and quadratic functional form, and one for nonlinear functional form. Thus DIC, log-score and posterior contour probability support the true functional form.

Figure 17(d) and 18(d) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (27). The boxplots of posterior contour probabilities of simulated samples from model (27), considering various functional forms are shown in Figure 19(d).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-394.38	15	-1.97	1
Quadratic	-	-	-	0
Linear	61.08	2.66	0.31	0
Constant	170.31	1.65	0.85	0

Table 12: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-10. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

5.11 Model-11 : Nonlinear effect model 1 with medium noise level

The model is:

$$y_{i42} = f_4(x_i) + \epsilon_{i42} \tag{28}$$

where, $f_4(x) = \underbrace{1 + 0.5x}_{-1 \leq x \leq 0} + \underbrace{1 + 2x^2}_{0 < x \leq 1}$ and $\epsilon_{i42} \sim \mathcal{N}(0, 0.12)$. Figure 12 shows the true functional form $f_4(x)$ and one simulated data set from model (28).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (28), considering different functional forms for x are given in Table 13. The DIC and log score of the model with smooth effect is minimum . The results show clear evidence for nonlinear effect. The posterior contour probability is zero for constant, linear and quadratic functional form but is one for nonlinear functional form.

Figure 20(d) and 21(d) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (28). The boxplots of posterior contour probabilities of simulated samples from model (28), considering various functional forms are shown in Figure 22(d).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-140.5	9.14	-0.7	1
Quadratic	-	-	-	0
Linear	65.88	2.66	0.33	0
Constant	172.81	1.65	0.87	0

Table 13: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-11. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

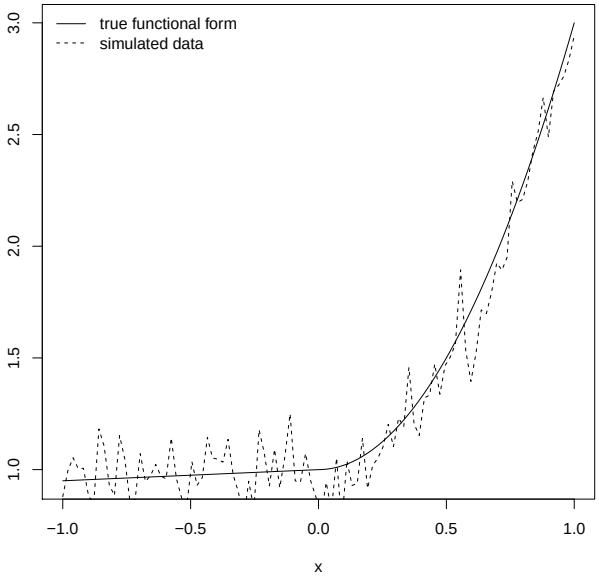


Figure 12: The function $f_4(x)$ and one simulated data set from model-11.

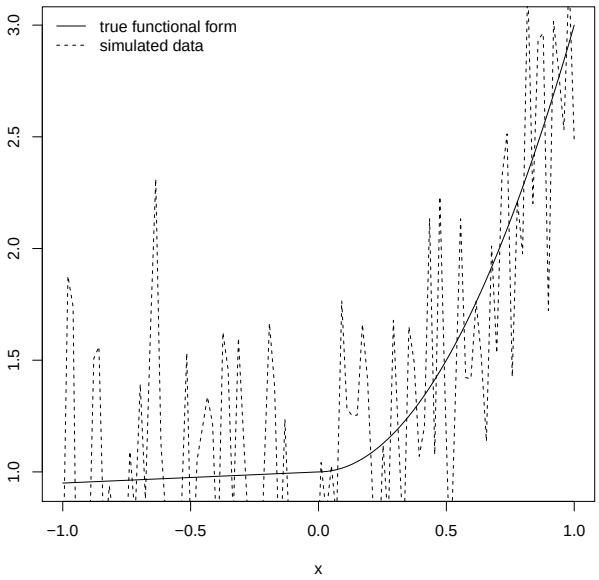


Figure 13: The function $f_4(x)$ and one simulated data set from model-12.

5.12 Model-12 : Nonlinear effect model 1 with high noise level

The model is:

$$y_{i43} = f_4(x_i) + \epsilon_{i43} \tag{29}$$

where, $f_4(x) = \underbrace{1 + 0.5x}_{-1 \leq x \leq 0} + \underbrace{1 + 2x^2}_{0 < x \leq 1}$ and $\epsilon_{i43} \sim \mathcal{N}(0, 0.48)$. Figure 13 shows the true functional form $f_4(x)$ and one simulated data set from model (29).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (29), considering different functional forms for x are given in Table 14. The DIC and log score of the model with smooth effect is minimum. The results show clear evidence for nonlinear effect. The posterior contour probability is zero for constant, linear and quadratic functional form but is one for nonlinear functional.

Figure 23(d) and 24(d) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (29). The boxplots of posterior contour probabilities of simulated samples from model (29), considering various functional forms are shown in Figure 25(d).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	174.18	5.62	0.88	1
Quadratic	-	-	-	0
Linear	192.76	2.648	0.96	0
Constant	237.82	1.65	1.19	0

Table 14: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-12. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

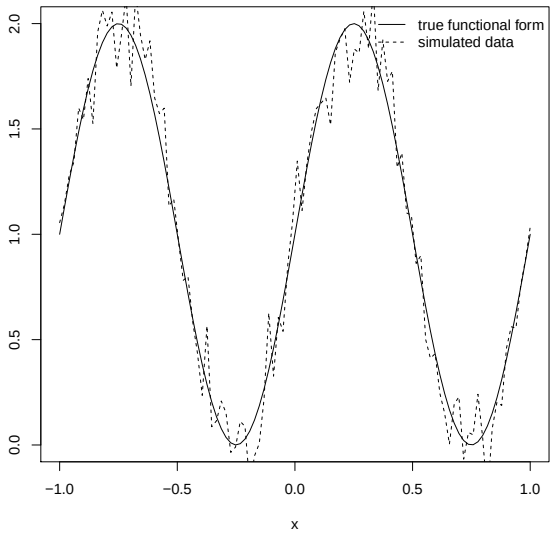


Figure 14: The function $f_5(x)$ and one simulated data set from model-13.

5.13 Model-13 : Nonlinear effect model 2 with low noise level

The model is:

$$y_{i51} = f_5(x_i) + \epsilon_{i51} \tag{30}$$

where, $f_5(x) = 1 + sin(2\pi x)$ and $\epsilon_{i51} \sim \mathcal{N}(0, 0.15)$. Figure 14 shows the true functional form $f_5(x)$ and one simulated data set from model (30).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (30), considering different functional forms for x are given in Table 15. The DIC and log score of the model with smooth effect is minimum. The results show clear evidence for nonlinear effect. The posterior contour probability is zero for constant, linear and quadratic functional but is one for nonlinear functional form.

Figure 17(e) and 18(e) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (30). The boxplots of posterior contour probabilities of simulated samples from model (30), considering various functional forms are shown in Figure 19(e).

Priors for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	-54.56	17.52	-0.26	1
Quadratic	-	-	-	0
Linear	207.11	2.65	1.03	0
Constant	220.57	1.65	1.19	0

Table 15: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-13. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

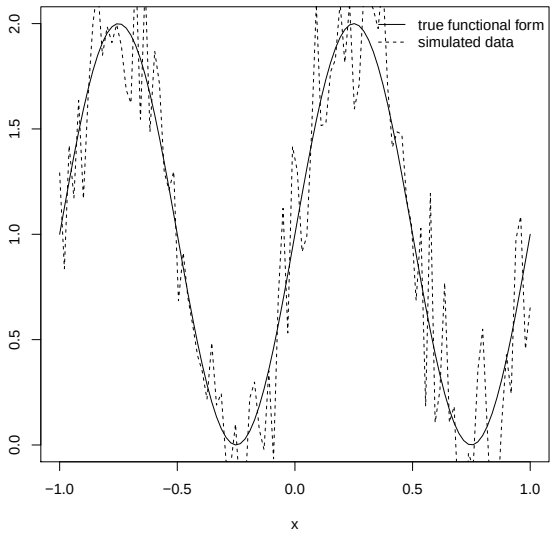


Figure 15: The function $f_5(x)$ and one simulated data set of model-14.

5.14 Mode-14 : Nonlinear effect model 2 with medium noise level

The model is:

$$y_{i52} = f_5(x_i) + \epsilon_{i52} \tag{31}$$

where, $f_5(x) = 1 + sin(2\pi x)$ and $\epsilon_{i52} \sim \mathcal{N}(0, 0.30)$. Figure 15 shows the true functional form $f_5(x)$ and one simulated data set from model (31).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (31), considering different functional forms for x are given in Table 16. The DIC and log score of the model with smooth effect is minimum. The results show clear evidence for nonlinear effect. The posterior contour probability is zero for constant, linear and quadratic effect but is one for nonlinear functional forms.

Figure 20(e) and 21(e) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect. The boxplots of posterior contour probabilities of simulated samples from model (31), considering various functional forms are shown in Figure 22(e).

Model for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear(RW2)	57.48	14.34	0.19	1
Quadratic	-	-	-	0
Linear	225.18	2.65	1.1	0
Constant	239.52	1.65	1.1	0

Table 16: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-14. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

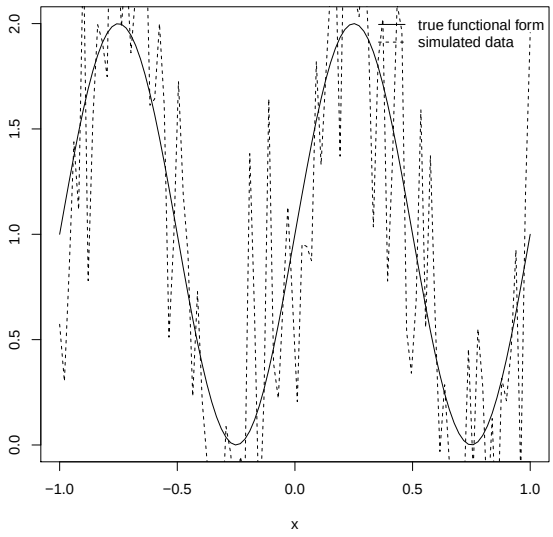


Figure 16: The function $f_5(x)$ and one simulated data set of model-15.

5.15 Model-15 : Nonlinear effect model 2 with high noise level

The model is:

$$y_{i53} = f_5(x_i) + \epsilon_{i53} \tag{32}$$

where, $f_5(x) = 1 + sin(2\pi x)$ and $\epsilon_{i53} \sim \mathcal{N}(0, 0.60)$. Figure 16 shows the true functional form $f_5(x)$ and one simulated data set from model (32).

The DIC, log-scores and posterior contour probabilities of one simulated sample from model (32), considering different functional forms for x are given in Table 17. The DIC and log score of the model with smooth effect is minimum. The results show clear evidence for nonlinear effect. The posterior contour probability is zero for constant, linear effect and quadratic effect but is one for nonlinear effect.

Figure 23(e) and 24(e) shows the boxplot of the DIC and the log-score values of all 250 data sets, after considering smooth effect, linear effect and constant effect of simulated samples from model (32). The boxplots of posterior contour probabilities of simulated samples from model (32), considering various functional forms are shown in Figure 25(e).

Prior for x	DIC	effective number of parameters	log-score	posterior contour probabilities
Nonlinear	181.18	10.64	0.91	1
Quadratic	-	-	-	0
Linear	245.95	2.65	1.276847	0
Constant	257.40	1.65	1.300819	0

Table 17: The DIC, log-scores and posterior contour probabilities for different functional forms of x for model-15. The minimum value of DIC and log-score, and the selected posterior contour probability are highlighted.

6 Simulation results

15 models have been studied. We generate data sets with 250 replications of each of the 15 models. We calculate the posterior contour probabilities corresponding to four functional forms: constant, linear, quadratic and nonlinear for x, and the DIC and log-scores considering three priors: smooth effect (RW2), linear and constant for x of all 250 replications of 15 models. We compare the results of the DIC, log-scores and posterior contour probabilities for all 15 models. The results are summarized as follows:

6.1 Models with low noise

In Table 18, we present the preferences by the DIC, log-score and posterior contour probabilities in model selection, when noise level is low.

True Model	DIC			log-score			posterior contour probabilities			
	smooth	linear	constant	smooth	linear	constant	smooth	quadratic	linear	constant
Constant	4	39	207	4	35	211	-	-	-	250
Linear	5	245	0	3	247	0	-	-	250	0
Quadratic	250	0	0	250	0	0	-	250	0	0
Nonlinear 1	250	0	0	250	0	0	250	0	0	0
Nonlinear 2	250	0	0	250	0	0	250	0	0	0

Table 18: Model selection by the DIC, log-score and posterior contour probabilities when noise level is low. ”-” denotes that the posterior contour probabilities are 1 for more than one functional forms.

In around 19% of the cases the DIC detect linear or smooth effect when there is none. In case when the true effect is linear the DIC detect smooth effect in 5 cases out of 250. For quadratic model, nonlinear model 1 and nonlinear model 2, the DIC work perfectly.

In case when the true effect is constant, log-score support true effect in 111 cases out of 250. In around 16% of the cases log-score detect linear or smooth effect when there is none. For linear, quadratic, nonlinear 1 and nonlinear 2 models, log-score always support the true model.

Posterior contour probabilities support the true and more complex functional forms. We select the least complex forms.

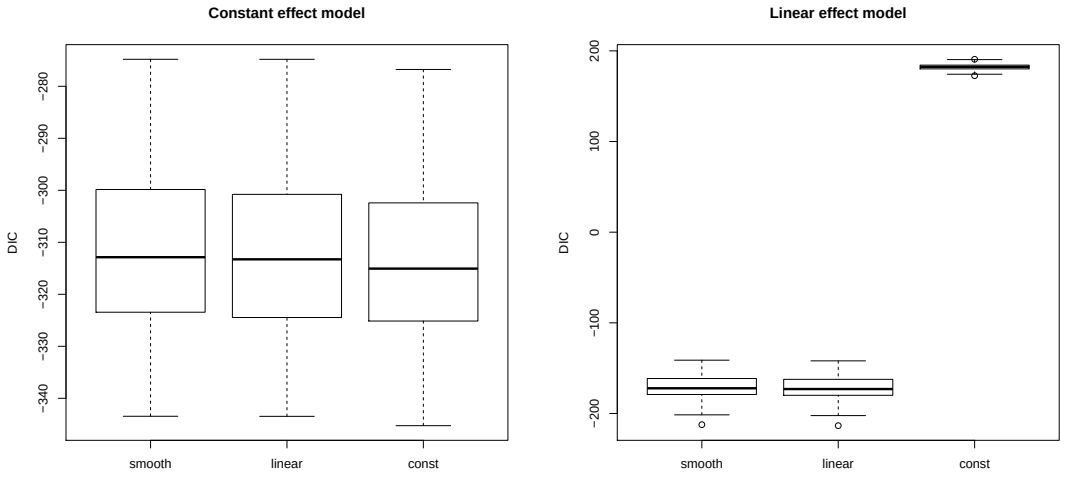
Figure 17 shows the boxplots of the DIC of the studied models when noise level is low. For constant effect model, the DIC support the true effect but the DIC of linear effect and smooth effect are very close to the DIC of constant effect (Figure 17(a)). For linear effect model, the DIC support the true effect but the DIC of smooth effect are very close to DIC of linear effect (Figure 17(b)). In rest of the three models: quadratic, nonlinear 1 and nonlinear 2, the DIC clearly support the true effect (Figure 17(c), Figure 17(d), Figure 17(e)).

Figure 18 shows the boxplots of log scores of the studied models when noise level is low. For constant effect model, log-scores support the true effect but log-scores of linear effect and smooth effect are very close to log-scores of constant effect (Figure 18(a)). For linear effect model, the log-score support the the true effect but the log-score of smooth effect are very close to linear effect (Figure 18(b)). In rest of the three models: quadratic, nonlinear 1 and nonlinear 2, the log-scores clearly support the true model (Figure 18(c), Figure 18(d), Figure 18(e)).

Figure 19 shows the boxplots of the posterior contour probabilities for studied models: constant, linear,

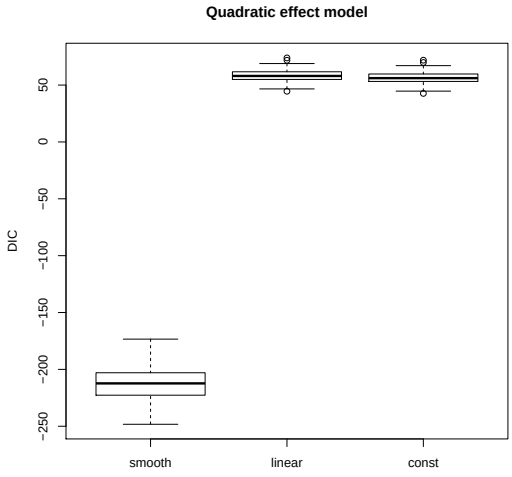
quadratic, nonlinear 1 and nonlinear 2 when noise level is low. For low noise, posterior contour probabilities give evidence for the true and the more complex models.

For constant effect model, posterior contour probabilities are one for all functional forms. Thus say nothing about the influence of x, but we choose the least complex functional form , which is constant. For linear effect model, the contour probabilities are zero for constant effect, and one for linear, quadratic and nonlinear effects (Figure 19(b)). For quadratic, the contour probabilities are zero for constant and linear but one for the quadratic and nonlinear effects (Figure 19(c)). For nonlinear effect model 1 and nonlinear effect model 2, the contour probabilities support only the nonlinear effect (Figure 19(d) & Figure 19(e)).

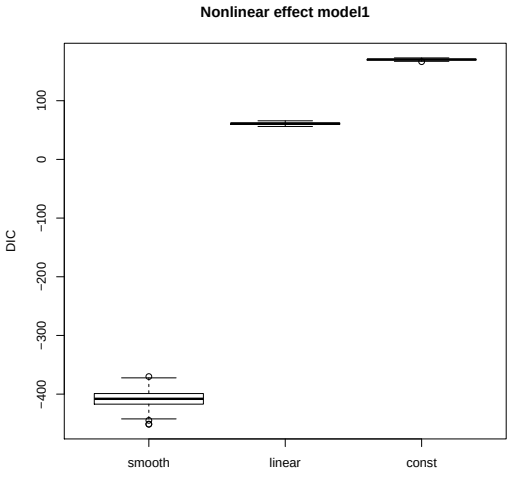


(a)

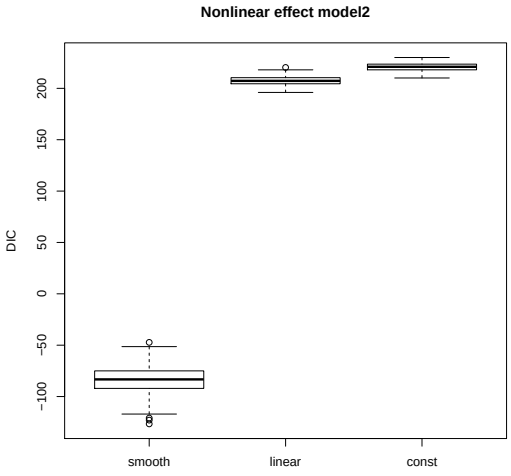
(b)



(c)

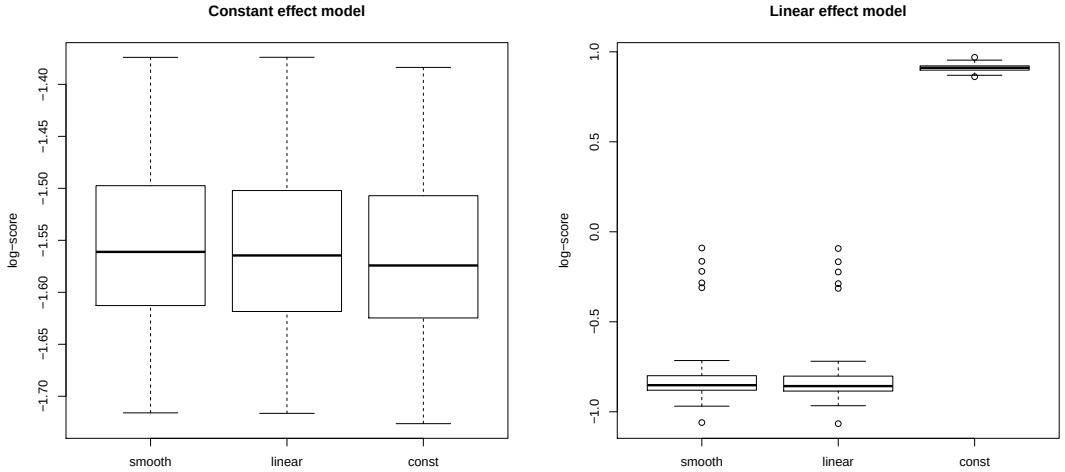


(d)



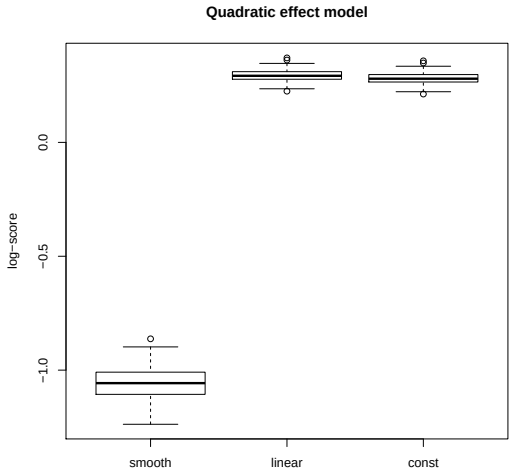
(e)

Figure 17: Boxplots of the DIC of constant effect model, linear effect model, quadratic effect model , nonlinear effect model 1 and nonlinear effect model 2 when noise level is low.

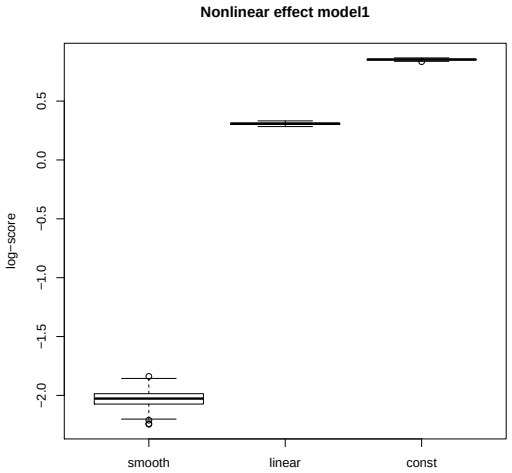


(a)

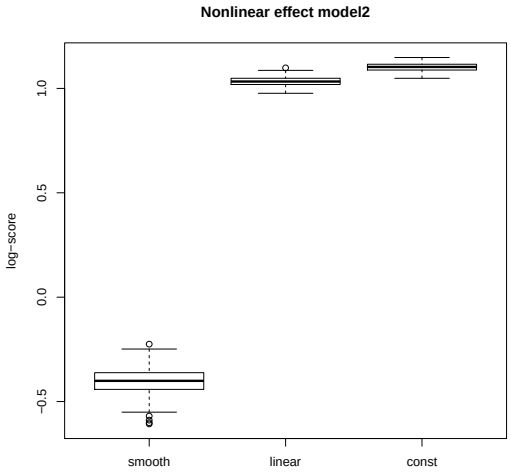
(b)



(c)

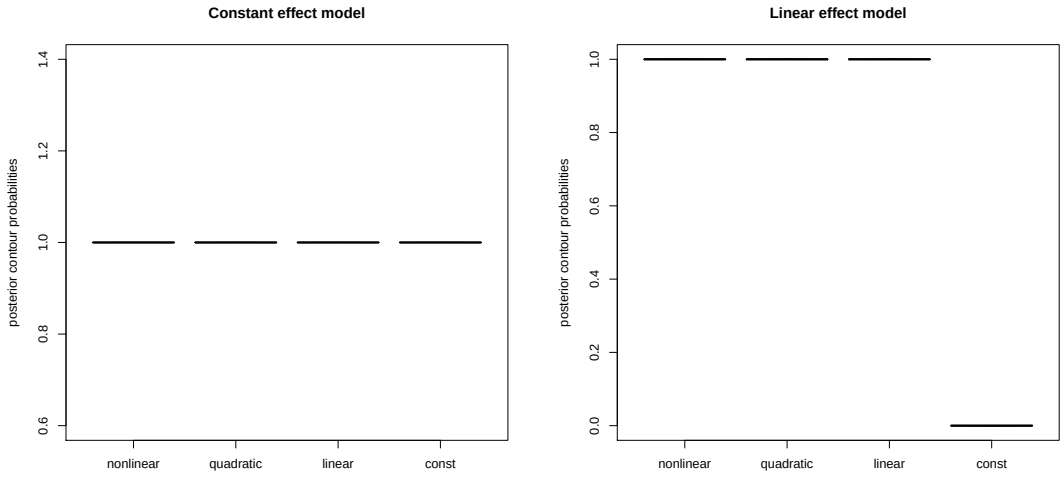


(d)



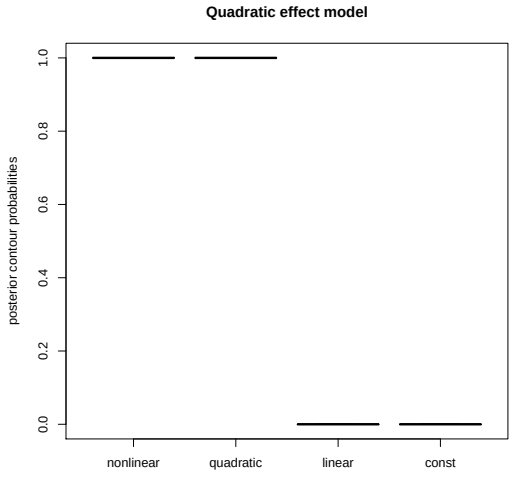
(e)

Figure 18: Boxplots of the log-scores of constant effect model, linear effect model, quadratic effect model , nonlinear effect model 1 and nonlinear effect model 2 when noise level is low.

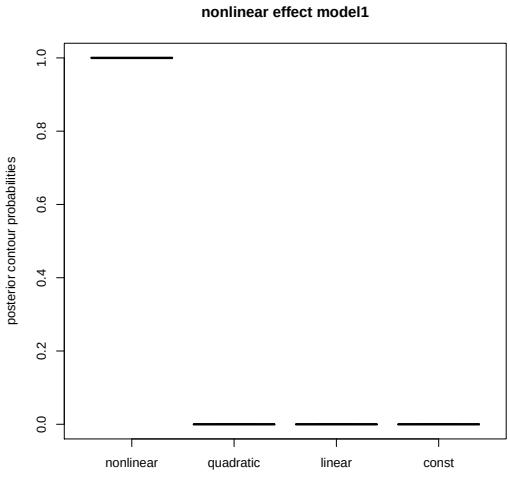


(a)

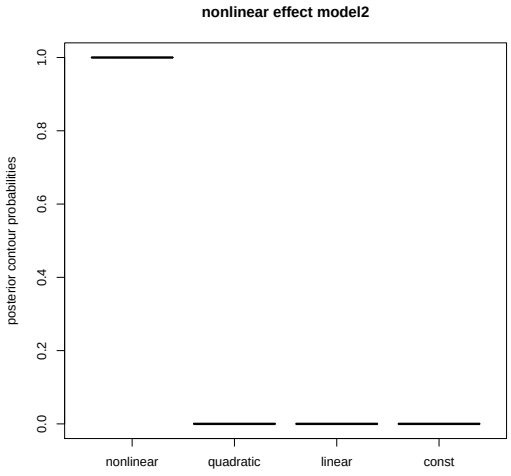
(b)



(c)



(d)



(e)

Figure 19: Boxplots for estimated posterior contour probabilities for constant effect model, linear effect model, quadratic effect model , nonlinear effect model 1 and nonlinear effect model 2 when noise level is low.

6.2 Models with medium noise

In Table 19, we present the preferences by the DIC, log-score and posterior contour probabilities in model selection, for models when noise level is medium.

For constant effect model, the DIC select the true effect in most of the cases, but in about 3 % of the cases, the DIC select a linear or smooth effect. For linear, quadratic, nonlinear 1 and nonlinear 2 effect models, the DIC support the true effect.

In case when the true effect is constant, the log-score support true effect in 245 cases, and only in 5 cases it support the linear effect. For linear, quadratic, nonlinear 1 and nonlinear 2 effect models, the log-score always support the true effect.

The posterior contour probabilities for the considered models with medium noise are similar to the posterior contour probabilities for the models with low noise. The posterior contour probabilities prefer the true and more complex functional forms.

True Model	DIC			log-score			posterior contour probabilities			
	smooth	linear	constant	smooth	linear	constant	smooth	quadratic	linear	constant
Constant	1	5	244	4	35	211	-	-	-	250
Linear	0	250	0	0	250	0	-	-	250	0
Quadratic	250	0	0	250	0	0	-	250	0	0
Nonlinear 1	250	0	0	250	0	0	250	0	0	0
Nonlinear 2	hl250	0	0	250	0	0	250	0	0	0

Table 19: Model selection by the DIC, log-score and posterior contour probabilities when noise level is medium. ”-” denotes that the posterior contour probabilities are 1 for more than one functional forms.

Figure 20 shows the boxplots of the DIC of the studied models when noise level is medium. For constant effect models, the DIC support the true model but the DIC for linear effect and smooth effect are very close to the DIC for constant effect (Figure 20(a)). Again, for linear effect model, the DIC support the the true effect but the DIC for smooth effect are very close to linear effect (Figure 20(b)). For rest of the three models with quadratic, nonlinear 1 and nonlinear 2 effects, the DIC clearly support the true effect (Figure 20(c), Figure 20(d), Figure 20(e)).

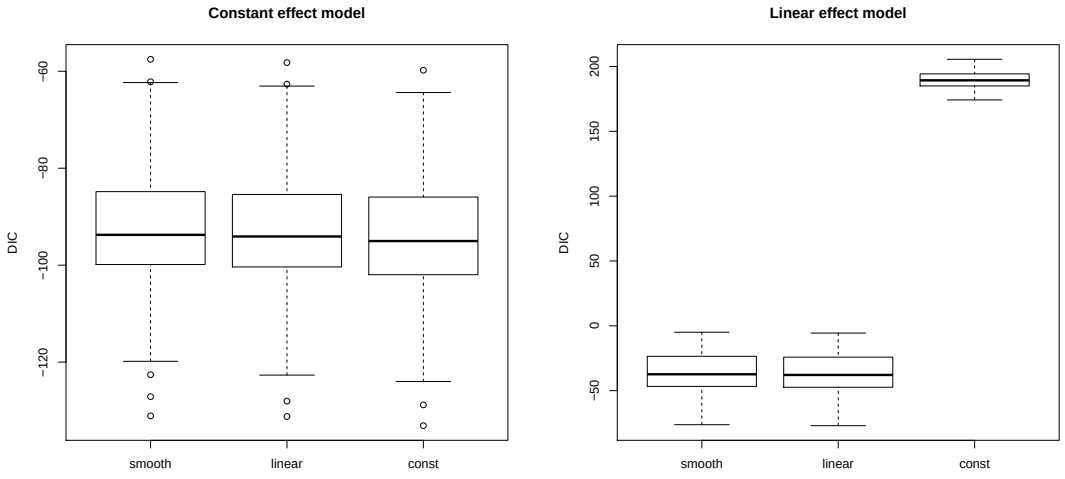
Figure 21 shows the boxplots of the log-scores for studied models when noise level is medium. For constant effect model, the log-scores support the true effect but the log-scores for linear effect and smooth effect are very close to the log-scores for constant effect (Figure 21(a)). For linear effect model, the log-scores support the the linear effect but the log-scores for smooth effect are very close to linear effect (Figure 21(b)). For rest of the three models with quadratic, nonlinear 1 and nonlinear 2 effects, the log-scores clearly support the true model (Figure 21(c), Figure 21(d), Figure 21(e)). The behaviour of the log-score in all five models with medium noise is very similar to its behaviour for same models with low noise.

Figure 22 shows the boxplots of the posterior contour probabilities for studied models when noise level is medium. For constant effect model, the posterior contour probabilities are one for all functional forms (Figure 22(a)). For linear effect model, the contour probabilities are zero for constant effect, and one for linear, quadratic and nonlinear effects (Figure 22(b)). For quadratic effect model, posterior contour probabilities support both the quadratic and nonlinear effects.

For nonlinear effect model 1, the contour probabilities show notable difference for quadratic effect (Figure

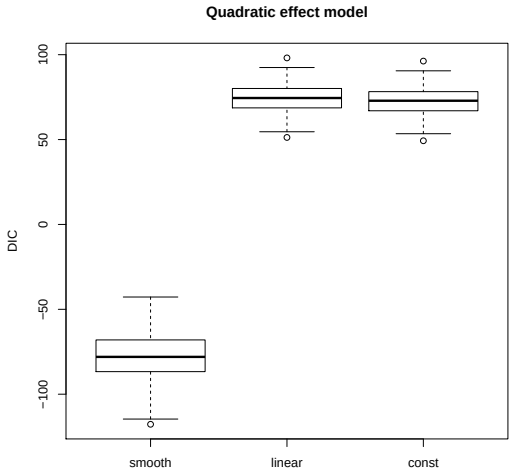
22(d)). For quadratic effect, the median of the posterior contour probabilities is around 0.62, which was zero for the same model when the noise level was low.

For nonlinear effect model 2, the contour probabilities are one only for nonlinear effect (Figure 22(e)).

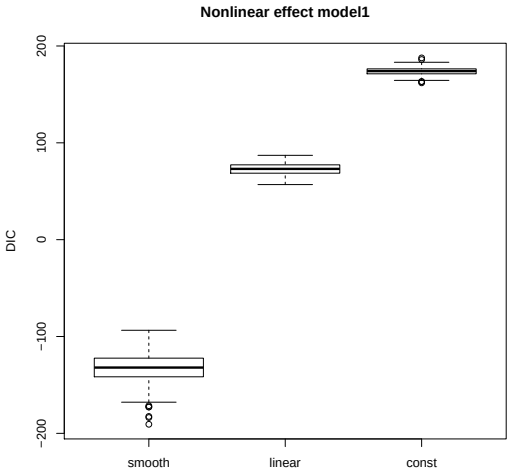


(a)

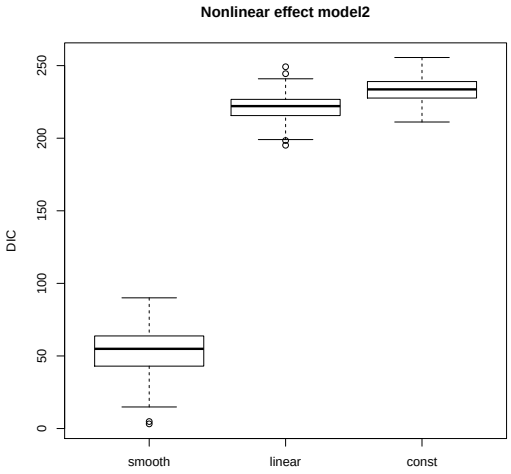
(b)



(c)

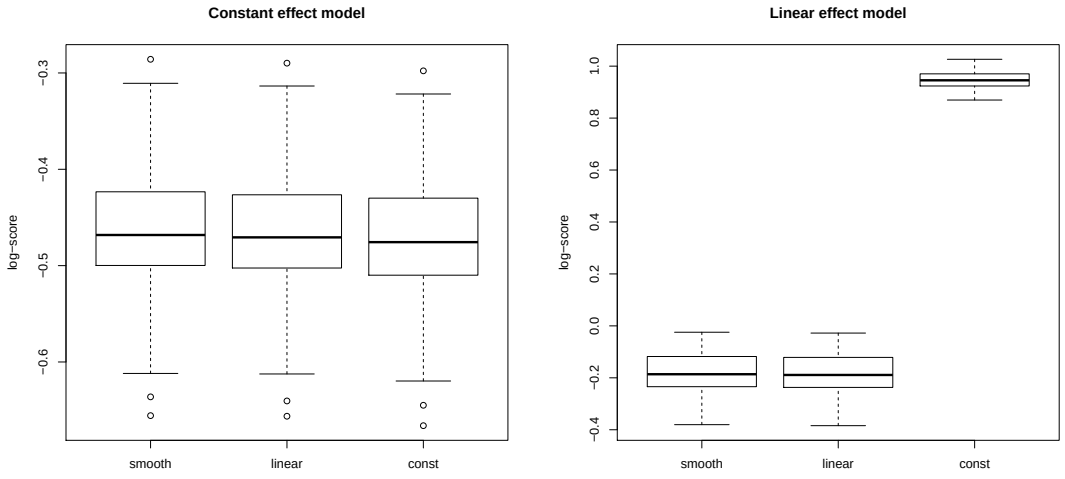


(d)



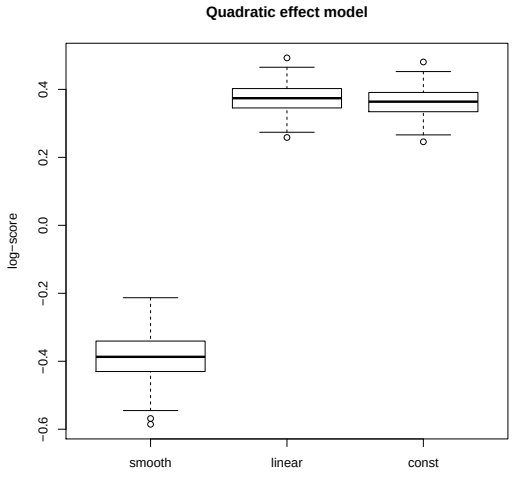
(e)

Figure 20: Boxplots of the DIC for constant effect model, linear effect model, quadratic effect model , nonlinear effect model1 and nonlinear effect model2 when noise level is medium.

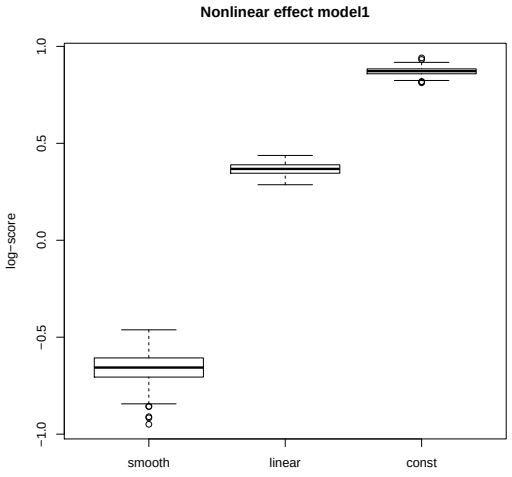


(a)

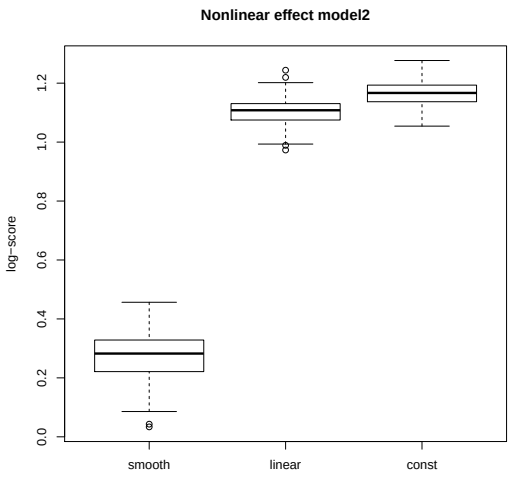
(b)



(c)

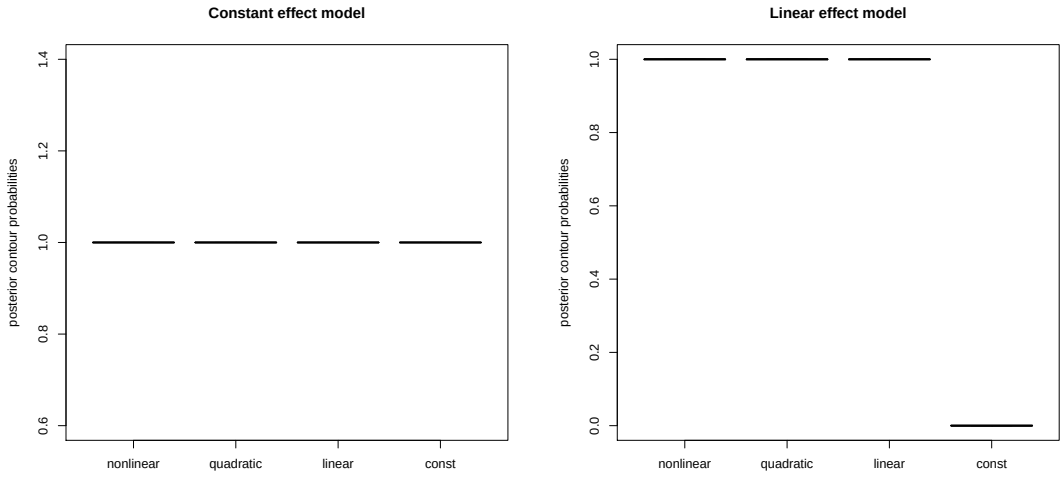


(d)



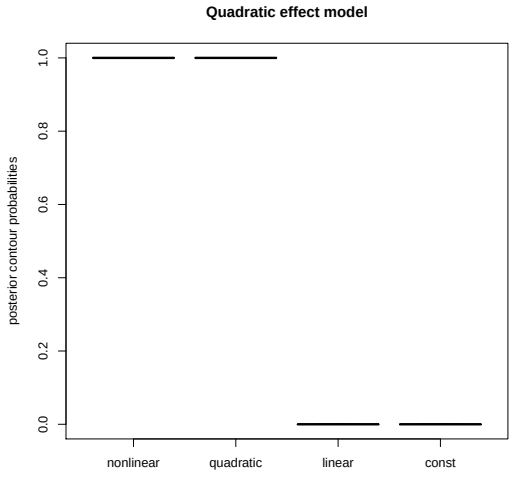
(e)

Figure 21: Boxplots of the log-scores for constant effect model, linear effect model, quadratic effect model , nonlinear effect model1 and nonlinear effect model2 when noise level is medium.

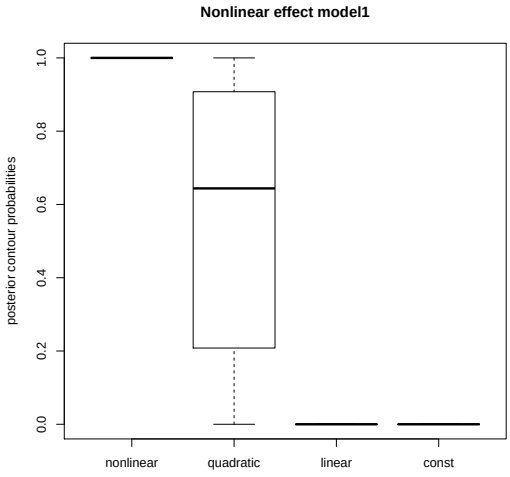


(a)

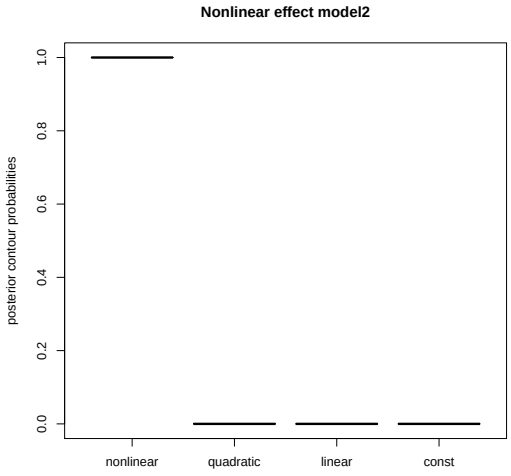
(b)



(c)



(d)



(e)

Figure 22: Boxplots for estimated posterior contour probabilities for constant effect model, linear effect model, quadratic effect model , nonlinear effect model1 and nonlinear effect model2 when noise level is medium.

6.3 Models with high noise

In Table 20, we present the preferences by the DIC, log-score and posterior contour probabilities in model selection.

For constant effect model, in 20 % of the cases the DIC and log-score detect linear or smooth effect when there is none. For linear, quadratic, nonlinear 1 and nonlinear 2 effect models, the DIC and log-score support the true effect. For linear effect model, only in one case the log-score supports the smooth effect but otherwise it support the true effect.

Like in case of low and medium noise levels, the posterior contour probabilities support the true and more complex functional forms (effects).

True Model	DIC			log-score			posterior contour probabilities			
	smooth	linear	constant	smooth	linear	constant	smooth	quadratic	linear	constant
Constant	3	47	200	3	48	199	-	-	-	250
Linear	0	250	0	1	249	0	-	-	250	0
Quadratic	250	0	0	250	0	0	-	250	0	0
Nonlinear 1	250	0	0	250	0	0	250	0	0	0
Nonlinear 2	243	7	0	243	7	0	250	0	0	0

Table 20: Model selection by the DIC, log-score and posterior contour probabilities when noise level is high. ”-” denotes that the posterior contour probabilities are 1 for more than one functional forms.

Figure 23 shows the boxplots of the DIC for studied models when noise level is high. For constant effect model, the performance of the DIC is similar to its performance for the same model with low and medium noise. Thus, the DIC for constant effect is minimum but the DIC for linear and smooth effects are close to it (Figure 23(a)). For linear effect model, the DIC for linear effect is minimum. But the DIC for smooth effect is very near to it (sometimes the difference is even less than 1). For rest of the three models with quadratic, nonlinear 1 and nonlinear 2 effects, the DIC clearly support the true effect (Figure 23(c), Figure 23(d), figure 23(e)).

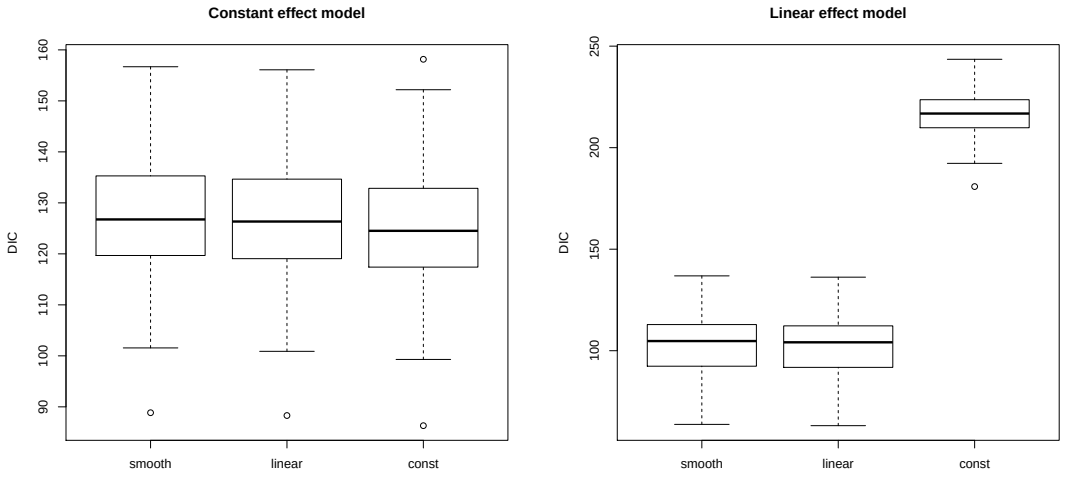
Figure 24 shows the boxplots of the log-scores for studied models when noise level is high. The log-score support the true effect, for all 5 models (Figure 24(a), Figure 24(b), Figure 24(c), Figure 24(d), Figure 24(e)). But for the constant effect model and linear effect models, the log-score for true effect is nearly equal to the log-score of complex models. The behaviour of the log-scores in all five models with high noise is similar to their behaviour for same models with low and medium noise.

Figure 25 shows the boxplots of the posterior contour probabilities for studied models when noise level is high. The contour probabilities for constant effect model and linear effect model follow the same pattern as for the respective models with low and medium noise (Figure 25(a) & Figure 25(b)). For quadratic effect model, the posterior contour probabilities are one for quadratic and nonlinear effects. But in very few cases, the contour probabilities for constant and linear functional forms show an increase in the values, though for majority of times the posterior contour probabilities are nearly zero.

For nonlinear effect model 1, the posterior contour probabilities show an increase in values from zero, for null, constant and linear effects. Surprisingly, the posterior contour probabilities are one for the linear and quadratic vectors (Figure 25(c) and Figure 25(d)). For nonlinear effect model 1, increasing noise in the model influence the contour probabilities. Posterior contour probabilities show support for nonlinear vector (Figure

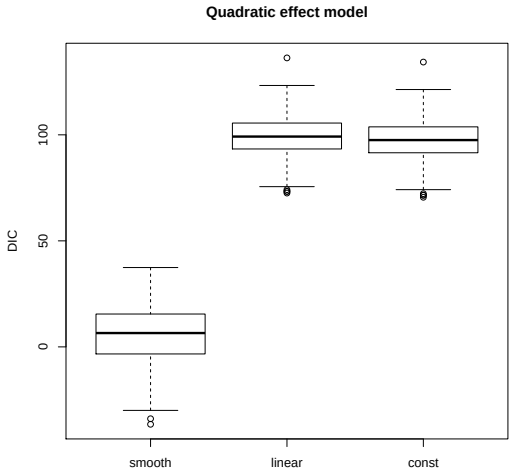
25(e)). Which is quite natural as the true model consists of nonlinear effect.

For nonlinear effect model 2, there is also rise in posterior contour probabilities for constant, linear and quadratic vectors, but still posterior contour probabilities clearly supports the nonlinear vector or functional form.

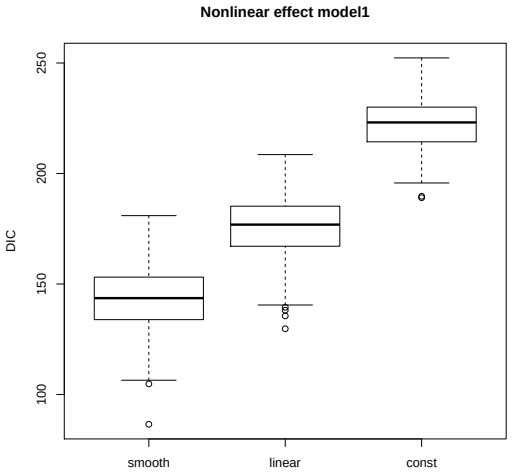


(a)

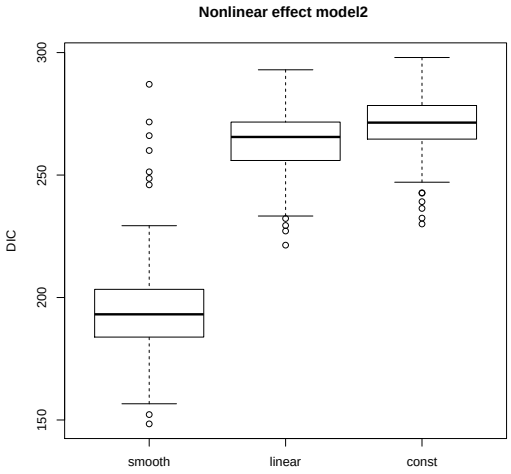
(b)



(c)

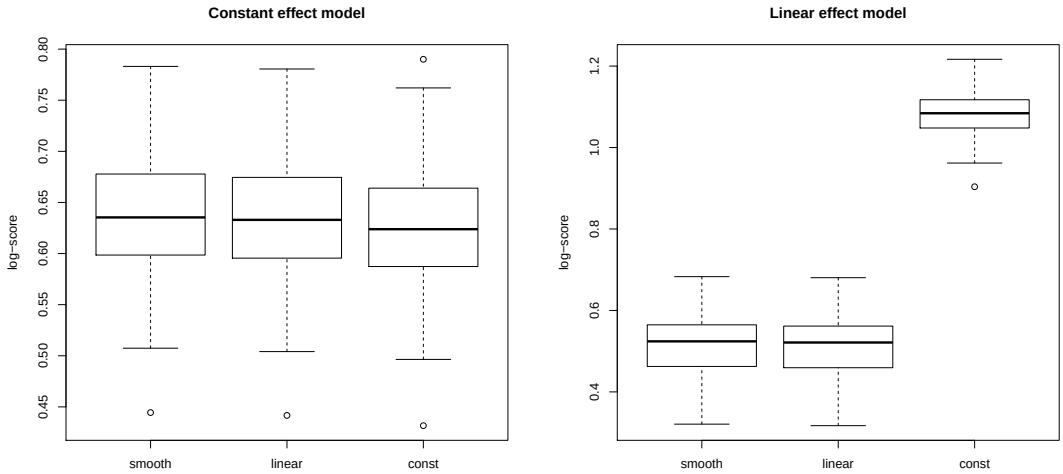


(d)



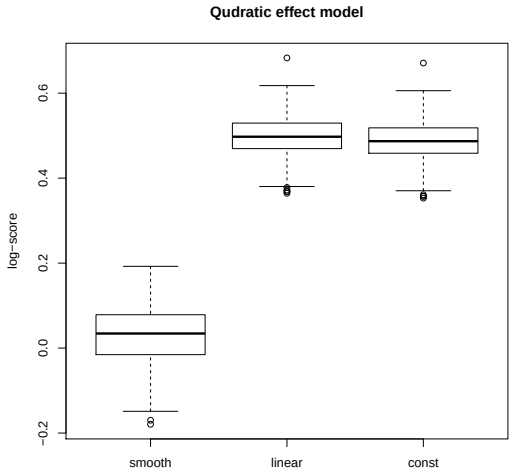
(e)

Figure 23: Boxplots of the DIC for constant effect model, linear effect model, quadratic effect model , nonlinear effect model1 and nonlinear effect model2 when noise level is high.

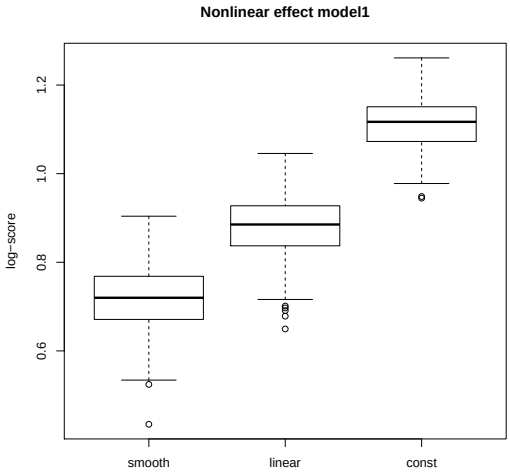


(a)

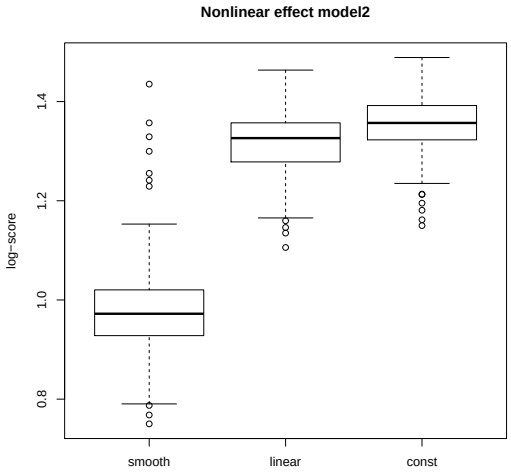
(b)



(c)

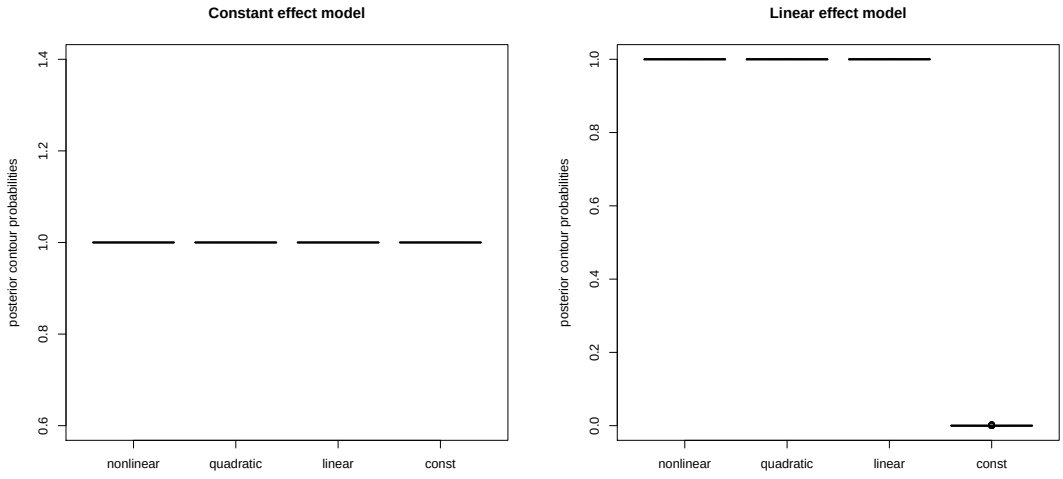


(d)



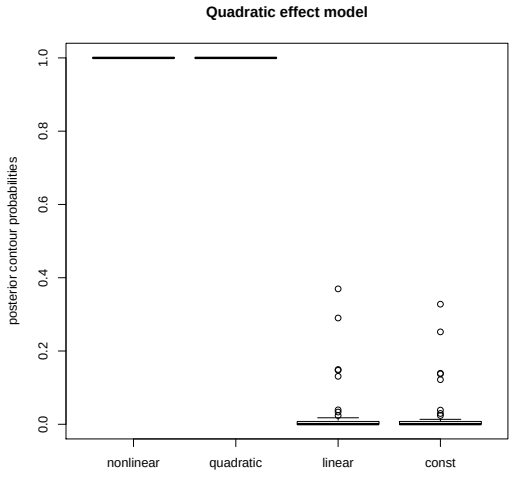
(e)

Figure 24: Boxplots of the log-scores for constant effect model, linear effect model, quadratic effect model , nonlinear effect model1 and nonlinear effect model2 when noise level is high.

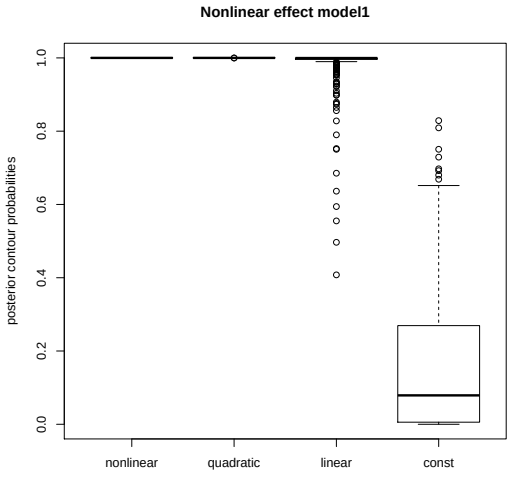


(a)

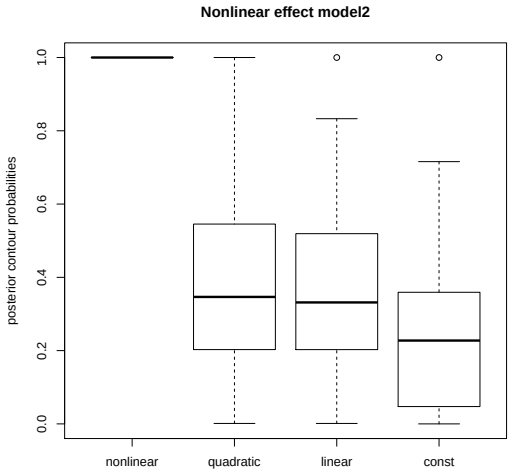
(b)



(c)



(d)



(e)

Figure 25: Boxplots for estimated posterior contour probabilities for constant effect model, linear effect model, quadratic effect model , nonlinear effect model 1 and nonlinear effect model 2 when noise level is high.

6.4 The posterior contour probabilities Versus the DIC and the log-score

The posterior contour probabilities support the true effect and more complex effects in case of four models: constant, linear, quadratic and nonlinear 2 irrespective of noise levels.

The performance of posterior contour probabilities is not accurate in case of nonlinear effect model 1. Nonlinear effect model 1 is a special model in the sense that part of it is constant line and part of it is a curve but more closer to linearity. So increasing noise makes it difficult for posterior contour probabilities to judge the correct form of the effect in the model. When noise level is medium posterior contour probabilities show some support for quadratic functional form, when actually the model is not quadratic. When we increase the noise even more, then posterior contour probabilities do not detect proper effect. In such a case, posterior contour probabilities even support the linear and the quadratic functional forms along with nonlinear form.

The DIC and the log-score support the same functional forms, they go parallel irrespective of noise level. In case of constant effect model, the DIC and log-scores support the true effect but the results are quite close to the more complex effects irrespective of noise level.

A rough measure of selecting model on the basis of DIC of the candidate models is, if the difference is more than 10 might definitely rule out the model with higher DIC, differences between 5 and 10 are substantial, but **if** the difference is less than 5, then it could be misleading just to report the model with lowest DIC (Spiegelhalter et al. (2002)).

For constant and linear effect models, the DIC of true models are very close to the DIC of models with more complex effects, irrespective of noise levels. Which makes model selection questionable.

It has always been argued, whether or not DIC support more complex models. It is true at least in our experience that DIC sometimes favour more complex models. As for constant effect models, in some cases the DIC detect linear or more complex effect when actually there was none. Even the log-scores behaves in similar manner of detecting linear or more complex effect when actually there was none.

In most of the cases discussed, the DIC, log-scores and the posterior contour probabilities perform equally good. For posterior contour probabilities, the only exception is the nonlinear effect model 1 with high noise level.

7 Application : Example - Bladder cancer data

In this section we re-analyse the bladder cancer data from Byar (1980) using nonhomogeneous Poisson processes. We calculate the posterior contour probabilities and compare them with the DIC and log-score values to explore the correct functional form of the covariates.

This bladder cancer data is discussed by many including Byar (1980) and Lawless (1987). The data was collected as event-time data from a clinical trial to analyze the efficiency of three treatments for recurring bladder cancer. All patients had superficial bladder tumours when they enter the trial. Tumours were removed and then patients were randomly assigned to one of the treatments: placebo pills (this refers to control group), or either thiotepa or pyridoxine (this refers to treatment group). Many had multiple recurrences of tumours during the study, these tumours were removed at each visit. The data record up to four recurrence times (in months) of tumours for the patients, where each recurrence time of a patient was measured from the beginning of his treatment. In the following analysis we consider the patients in the control or treatment groups. One of the subjects (the one with no follow-up after 0th recurrence) has no event and 0 months of follow-up. We removed this subject from the data set since it adds nothing to the likelihood. The time of study for each patient

is about 64 months after entering into the system. The covariates are the number of initial tumours present (intno), the initial size (the diameter of the largest initial tumour) (intsize), and treatment group (group).

We perform the analysis in the following manner:

1. We model the number of tumours in each patient according to nonhomogeneous Poisson processes as described in section 3.
2. We then calculate posterior contour probabilities in order to know the correct functional form of the covariates.
3. For model section we calculate the DIC and log-scores for models under consideration.
4. We perform a simulation study to investigate the behaviour of contour probabilities.
5. We discuss our results.

7.1 Model specification:

check about # of new tumors as conditional poisson. We consider the same model as defined in section 3. In the beginning we partitioned the follow-up period (time) into 8 equal intervals and assume that baseline intensity for these intervals is constant, $\{\lambda_k, k = 1, \cdots, 8\}$. The conditional proportional intensity function for patient $i, i = 1, \cdots, 85$ in the interval I_k is considered according to equation (4). The additive predictor function η_{ik}, which accounts for the effect of various covariates of ith patient in interval I_k is follows

$$\eta_{ik} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{intsize}_i + f^{(\text{intno})}(\text{intno}_i) + \alpha_i + b_k \tag{33}$$

We let the initial size of tumours have a linear effect while for the initial number of tumours we assume a smooth effect. The number of initial tumours are $n_{\text{intno}} = 1, \cdots, 8$. Where 8 denotes eight or more initial tumours. The vector $\boldsymbol{f}^{(\text{intno})} = \{f_1^{(\text{intno})}, \cdots, f_8^{(\text{intno})}\}$ is assumed to follow a second order random walk (Rue and Held (2005), Ch. 3) defined as

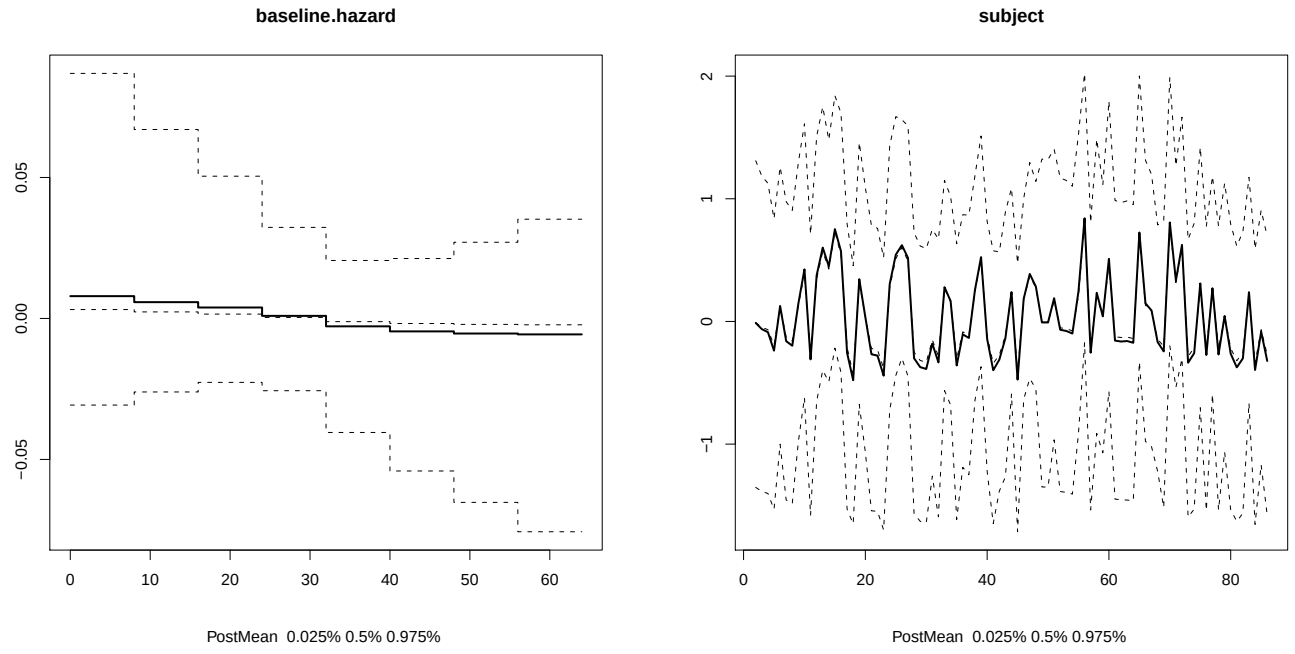
$$\pi(\boldsymbol{f}^{(\text{intno})} \mid \tau_{\text{intno}}) \propto \tau_{\boldsymbol{f}}^{(n_{\text{intno}}-2)/2} \exp\Big\{ -\frac{1}{2}\tau_{\text{intno}} \sum_{i=3}^{8} \big(f_i^{(\text{intno})} - 2f_{i-1}^{(\text{intno})} + f_{i-2}^{(\text{intno})}\big)^2 \Big\}$$

The priors assume for $\boldsymbol{\beta} = \{\beta_0, \beta_1, \beta_2\}$ are $\mathcal{N}(0, 10^{-3}\mathbf{I})$. We assume independent gamma priors $\Gamma(a, b)$ with mean a/b and variance a/b^2 for hyperparameters τ_b and τ_{intno}. More specifically, $\tau_{\text{intno}} \sim \Gamma(1, 0.001)$ and $\tau_\alpha \sim \Gamma(0.001, 0.001)$. We choose a non informative prior for precision of frailty term α as we want to check our results with the frequentist analysis.

7.2 INLA results

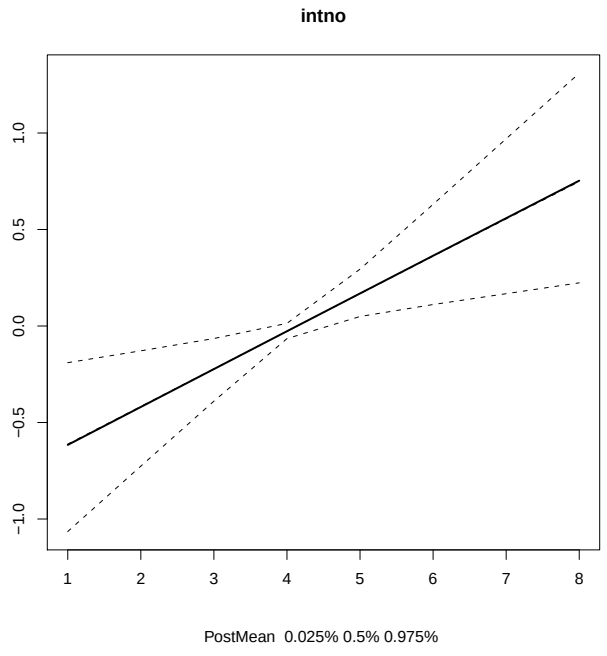
Estimates of the log-baseline intensity, log frailty and initial number of tumours are shown in Figure 26. The log-baseline intensity decreases slowly over the whole study period (Figure 26(a)). (Figure 26(b) provides the evidence of the existence of subject random frailty. Which suggests that some patients have higher chances of developing tumours as compared to the others in the same group.

The effect $f^{(\text{intno})}$ of initial number of tumour is linear (Figure 26(c)). In other words, the patients with a large number of initial tumours have higher rate of tumour recurrence if all other conditions are same. We



(a)

(b)



(c)

Figure 26: Posterior means of the model in (33) (a) log baseline intensity, (b) log frailty and (c) initial number of tumours.

check the linearity of this effect by using posterior contour probabilities. It is clear from Figure 26(c) that the 95% credible intervals are wide at the boundaries where there are less observations.

The estimate of group effect is -0.490 with s.d. 0.275, indicates that the patients in the first group (thiotepa) have lower rate of recurrence than the patients in second (control) group (Figure 27(a)). The effect of initial size of tumour is constant (Figure 27(b)).

7.3 Model selection

Even though the effect of initial number of tumours looks linear (Figure 26(c)), we will investigate whether the effect is linear or smooth using the posterior contour probabilities, the DIC and the log-score. We calculate the posterior contour probabilities for fixed vector $\boldsymbol{\beta}_0$ as discussed in section 5 corresponding to smooth, quadratic, linear and constant functional forms for the initial number of tumours. The posterior contour probabilities, given in Table 21 suggest that the linear effect of initial number of tumours is sufficient.

vector	posterior contour probability
constant	0.34
linear	0.9971
nonlinear(quadratic)	0.9971
nonlinear(spline)	0.9971

Table 21: Posterior contour probabilities for initial number of tumours

We comparing the DIC and the log-score for two models, complex and reduced. Model is complex (given in (33)), when we assume smooth effect for the initial number of tumours and model is reduced, when we assume the initial number of tumours as linear effect. The additive predictor function for ith patient in kth interval for reduced model is

$$\eta_{ik} = \beta_0 + \beta_1 \text{group}_i + \beta_2 \text{intsize}_i + \beta_3 \text{intno}_i + \alpha_i + b_k \tag{34}$$

The estimated mean and s.d. of the initial number of tumours for the reduced model are 0.2 and 0.074. The DIC and log-score values for complex model and the reduced models are given in Table 22. The DIC and the log-score values both support the reduced model.

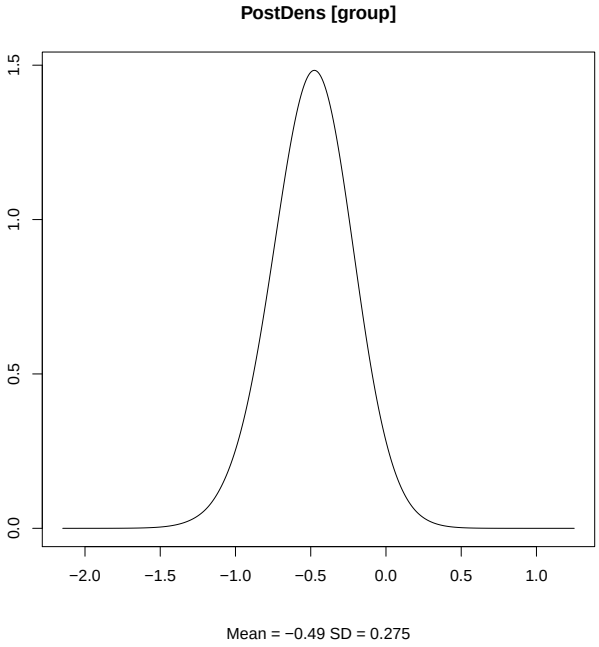
Model	effective number of parameters	DIC	log-score
model:intno=smooth	30.96	502.52	0.6677
model: intno=linear	31.37	502.41	0.6675

Table 22: Model selection

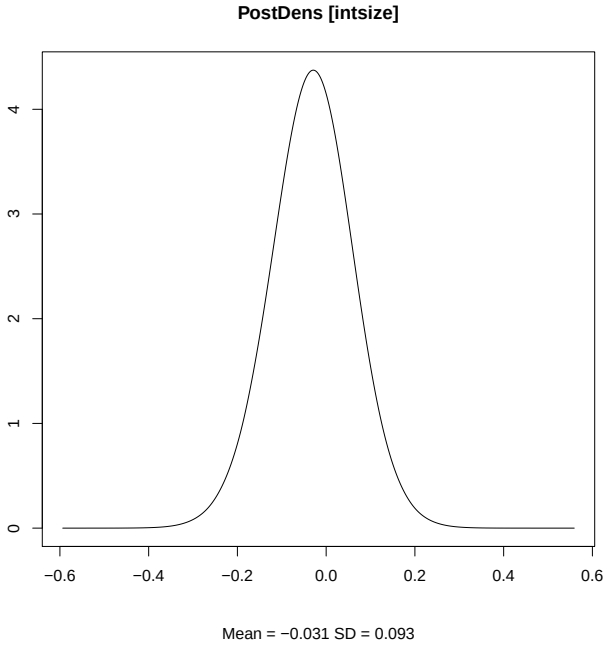
Thus the posterior contour probabilities, the DIC and the log-score are in favour of reduced model. Based on the results, we conclude that effect of the initial number of tumours is linear.

7.4 Simulation

To verify the performance of the posterior contour probabilities for real data set. We generated data sets with 100 replications from the model (33), i.e. by considering smooth effect for the initial number of tumours and



(a)



(b)

Figure 27: Posterior marginals of the model in (33) (a) group (treatment) effect β_1 and (b) initial size β_2.

calculate the posterior contour probabilities corresponding to smooth (nonlinear), linear and constant functional forms for the initial number of tumours (the parameter of interest is $f^{(\text{intno})}$). Histograms of posterior contour probabilities for constant, linear and nonlinear vectors are shown in Figure 28. The posterior contour probabilities for null vector are always near 0.28 and for linear and nonlinear vector they are near 0.997, indicating that the initial number of tumours has a linear effect on recurrences of tumour.

8 Discussion

The purpose of this report is to examine the performance of the posterior contour probabilities. Posterior contour probabilities are used in order to decide about the posterior support for a particular vector of interest. These help in deciding not only the correct functional form of a covariate but they even help to get a more parsimonious model. We use the algorithm suggested by Sørbye and Rue (2011) using INLA methodology to compute posterior contour probabilities.

There are various approaches available of model choice in Bayesian inference. Thus besides posterior contour probabilities, we also computed the DIC and the leave-one-out cross validated logarithmic score or log-score. Both of these can be computed in INLA.

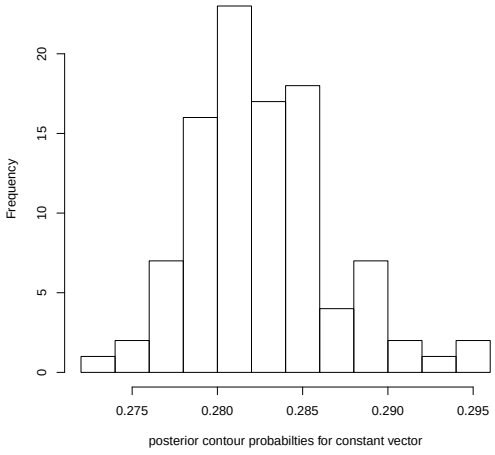
An extensive simulation study has been performed. We studied five different models with different effects: a constant, a linear, a quadratic, and two nonlinear effects. For each model we studied three different noise levels. According to our simulation study results, the contour probabilities performed reasonably well in coordination with the DIC and the log-score values. The DIC and the log-score support the same functional forms, they go parallel irrespective of noise levels.

Though posterior contour probabilities did not work that well for one model we studied. For the nonlinear effect model 1 with high noise, posterior contour probabilities even supported linear and quadratic vectors along with nonlinear vector. The test function used in this model is nonlinear in a special way, as, a part of it is constant and remaining part is a curve but more close to a straight line. Increasing noise makes it difficult for posterior contour probabilities to judge the correct form. For the same model, the DIC and log-score support smooth effects. For this case the DIC and log-score perform well.

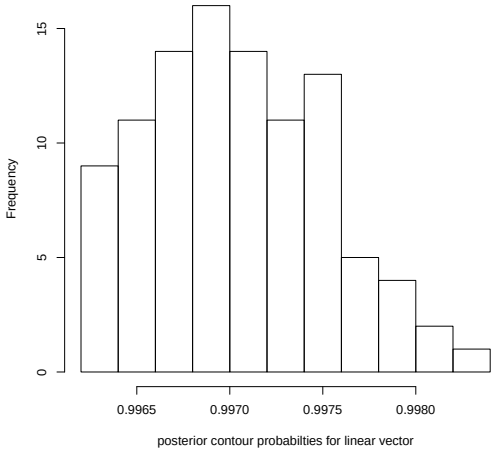
The efficiency of the DIC is argued by many authors and it is clearly mentioned that the DIC penalizes for the number of parameters required to be estimated in the model (Plummer (2008)) or in other words, the DIC works fine only when the effective number of parameters in the model is much smaller than the number of independent observations. In our simulation study and in the application, the effective number of parameters are always less than the number of independent observations.

Depending solely on DIC for model selection, when a true effect in a model is constant or linear is questionable. In our experience, the DIC even support more complex effects along with the true effect irrespective of the noise level. For such models, the differences in DIC is less than 5. Also DIC might detect effect when there is none. It is true at least in our experience, for constant effect models, in some cases the DIC detected linear or more complex effect when actually there was none.

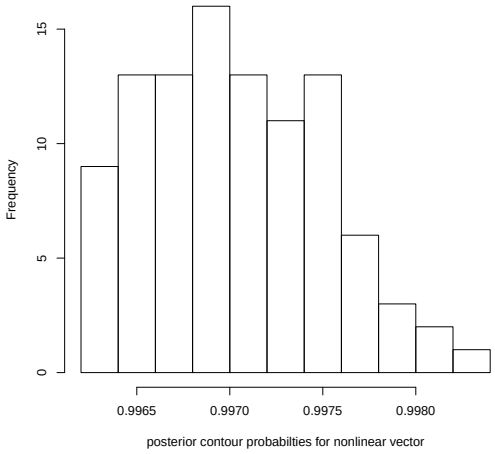
In this report, we consider a cross validation approach of model selection using CPO because it has an advantage over the Bayes factor approach. As cross validation approach can be used even for improper priors whereas the Bayes factor may not always work due to the reason that the marginal densities may not exist or may be arbitrary for improper priors. We excluded Bayes factors from the study as in our simulation examples, the prior we have considered for x is a RW2 (intrinsic) model and in our application, on bladder tumour data,



(a)



(b)



(c)

Figure 28: Histograms of posterior contour probabilities obtained for (a) constant vector, (b) linear vector and (c) nonlinear vector, for the initial number of tumours

the prior we have considered for the log baseline intensity is RW1 model. Which violates the condition for a meaningful Bayes factor.

We also explore the performance of posterior contour probabilities for nonhomogeneous Poisson processes using bladder tumour data. Posterior contour probabilities perform well together with other model selection criterion. Posterior contour probabilities can really be beneficial to the investigators in deciding about the functional forms of covariates in the semiparametric survival models, especially when the differences in DIC for the candidate models are small .

References

Akerkar, R., Martino, S., and Rue, H. (2012). Approximate Bayesian inference for nonhomogeneous Poisson processes with application to survival analysis. Technical report 3, Department of mathematical sciences, Norwegian University of Science and Technology.

Andersen, P. K. and Gill, R. D. (1982). Cox's regression models for counting processes: A large sample study. *The Annals of Statistics*, 10:1100–1120.

Besag, J., Green, P., Higdon, D., and Mengersen, K. (1995). Bayesian computation and stochastic systems. *Statistical Science*, pages 3–41.

Box, G. and Tiao, G. (1973). *Bayesian inference in statistical analysis*. Wiley.

Brezger, A. and Lang, S. (2008). Simultaneous probability statements for Bayesian p-spline. *Statistical Modelling*, 8(2):141.

Byar, D. P. (1980). The veterans administration study of chemoprophylaxis for recurrent stage i bladder tumours: comparisons of placebo, pyridoxine and topical thiotepa. *"in Bladder tumors and other topics in urological oncology, eds. M. Pavone-Macaluso, P.H. Smith, and F. Edsmyn, New York"*, pages 363–370.

Cook, R. and Lawless, J. (2007). *The statistical analysis of recurrent events*. Springer Verlag.

Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society, Series B*, 34:187–220.

Geisser, S. and Eddy, W. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, pages 153–160.

Gelfand, A., Dey, D., and Chang, H. (1992). Model determination using predictive distributions with implementation via sampling-based methods (with discussion). *Bayesian Statistics 4*, pages 147–167.

Gelfand, A. E. (1996). Model determination using sampling-based methods. In Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., editors, *Markov Chain Monte Carlo in Practice*, pages 145–161. Chapman & Hall, London.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378.

Held, L. (2004). Simultaneous posterior probability statements from Monte Carlo output. *Journal of Computational and Graphical Statistics*, 13:20–35.

Held, L., Schrödle, B., and Rue, H. (2010). Posterior and cross-validatory predictive checks: A comparison of MCMC and INLA. *Statistical Modelling and Regression Structures*, pages 91–110.

Lawless, J. F. (1987). Regression methods for Poisson process data. *Journal of the American Statistical Association*, 82:808–815.

Lawless, J. F. and Zhan, M. (1998). Analysis of interval-grouped recurrent-event data using piecewise constant rate function. *The Canadian Journal of Statistics*, 26:549–565.

Manda, S. and Meyer, R. (2005). Bayesian inference for recurrent events data using time-dependent frailty. *Statistics in medicine*, 24(8):1263–1274.

Pettit, L. I. (1990). The conditional predictive ordinate for the normal distribution. *Journal of the Royal Statistical Society, Series B*, 52(1):175–184.

Plummer, M. (2008). Penalized loss functions for Bayesian model comparison. *Biostatistics*, 9(3):523–539.

Rue, H. and Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London.

Rue, H., Martino, S., and Chopin, N. (2009). Approximate Bayesian inference for latent Gaussian models using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319–392.

Sinha, D. (1993). Semiparametric Bayesian analysis of multiple event time data. *Journal of the American Statistical Association*, 88:979–983.

Sørbye, S. H. and Rue, H. (2011). Simultaneous credible bands for latent Gaussian models. *Scandinavian Journal of Statistics*, 38 (4):712–725.

Spiegelhalter, D., Best, N., Carlin, B., and Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64(4):583–639.

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 111–147.

Thall, P. F. (1988). Mixed Poisson likelihood regression model for longitudinal interval count data. *Biometrics*, 44:197–209.