



## bsamGP: An R Package for Bayesian Spectral Analysis Models Using Gaussian Process Priors

Seongil Jo  
Chonbuk National  
University

Taeryon Choi  
Korea University

Beomjo Park  
Korea University

Peter Lenk  
University of  
Michigan

---

### Abstract

The Bayesian spectral analysis model (BSAM) is a powerful tool to deal with semiparametric methods in regression and density estimation based on the spectral representation of Gaussian process priors. The **bsamGP** package for R provides a comprehensive set of programs for the implementation of fully Bayesian semiparametric methods based on BSAM. Currently, **bsamGP** includes semiparametric additive models for regression, generalized models and density estimation. In particular, **bsamGP** deals with constrained regression models with monotone, convex/concave, S-shaped and U-shaped functions by modeling derivatives of regression functions as squared Gaussian processes. **bsamGP** also contains Bayesian model selection procedures for testing the adequacy of a parametric model relative to a non-specific semiparametric alternative and the existence of the shape restriction. To maximize computational efficiency, we carry out posterior sampling algorithms of all models using compiled Fortran code. The package is illustrated through Bayesian semiparametric analyses of synthetic data and benchmark data.

*Keywords:* cosine basis, Gaussian process priors, Markov chain Monte Carlo, R, shape restrictions, semiparametric models, spectral representation.

---

## 1. Introduction

Gaussian processes provide a natural method for specifying prior distributions on the space of functions in Bayesian inference. Since the seminal results on Gaussian process priors (Leonard 1978; O'Hagan 1978) in regression and density estimation, Gaussian processes have been widely used as nonparametric priors for unknown random functions. For instance, Gelfand, Kottas, and MacEachern (2005), Duan, Guindani, and Gelfand (2007), and Banerjee, Gelfand, Finley, and Sang (2008) developed models based on Gaussian processes for spatial statistics; Lenk (1999), Shi, Murray-Smith, and Titterton (2005), and Banerjee, Dunson,

and Tokdar (2013) used a Gaussian process as a prior for unknown regression functions; and, Lenk (1988, 1991, 2003) and Tokdar (2007) proposed logistic Gaussian process priors based on the logarithmic transformation of Gaussian processes for nonparametric density estimation. More recently, Gaussian process priors have also been utilized in functional and longitudinal data analyses (see, e.g., Zhu, Vannucci, and Cox 2010, Kaufman and Sain 2010, Wang and Shi 2014, and Yang, Zhu, Choi, and Cox 2016).

With the increasing interest in using Gaussian process priors in Bayesian inference, many easy-to-use software packages have been developed for a wide variety of applications. **GPML** (Rasmussen and Nickisch 2010), and **GPstuff** (Vanhatalo, Riihimäki, Hartikainen, Jylänki, Tolvanen, and Vehtari 2013) are the widely-used toolboxes in MATLAB (The MathWorks Inc. 2019) for Gaussian process models. These toolboxes provide modular functions to specify mean and covariance functions as well as various likelihood functions involving Gaussian process models and inference methods such as Markov chain Monte Carlo (MCMC), expectation propagation, and variational Bayes methods. Further discussions on and comparisons of different features in **GPML** and **GPstuff** can be found in Vanhatalo *et al.* (2013) (see, e.g., Table 1 of Vanhatalo *et al.* 2013). In R (R Core Team 2019), examples of packages for Gaussian process regression models include **tgp** (Gramacy 2007), **spectralGP** (Paciorek 2007), **bigGP** (Paciorek, Lipshitz, Zhuo, Prabhat, Kaufman, and Thomas 2015), and **laGP** (Gramacy 2016). **tgp** implements fully Bayesian inference on treed Gaussian process models including linear models, classification and regression tree models, and single-index models. **spectralGP** implements Bayesian smoothing with a Fourier basis representation of Gaussian processes based on an extension of the work by Wikle (2002) and provides template code for fitting Bayesian models for exponential family data with normal, Poisson, and binomial distributions. With the recent advancement in parallel computing, **bigGP** implements parallel linear algebra operations using threading and message-passing, which is useful for kriging and Gaussian process regression, and **laGP** implements local approximate Gaussian process regression for large-scale modeling and sparse computation with massive data sets. Other software packages written in different programming languages, such as C, C++, and Python, are available at <http://www.gaussianprocess.org/>.

The Bayesian spectral analysis model (BSAM) with Gaussian process priors (e.g, Lenk 1999, Lenk 2003, and Lenk and Choi 2017) provides a powerful tool to deal with semiparametric methods in regression and density estimation. Here, the prior for the unknown function is based on an infinite series expansion with a Karhunen-Loève representation of a second-order Gaussian process. The BSAM has been shown to work well in estimating an unknown semiparametric function and in choosing appropriate functional components for model selection. More importantly, the BSAM framework provides a Bayesian method for shape-restricted regressions with monotone, convex/concave, S-shaped and U-shaped curves by modeling derivatives of regression functions as squared Gaussian processes (Lenk and Choi 2017).

Accordingly, we develop a user-friendly software application called **bsamGP** (Jo, Choi, Park, and Lenk 2019) that not only implements the models in Lenk (1999), Lenk (2003), Jo, Roh, and Choi (2016), and Lenk and Choi (2017) for regression and density estimation with continuous data, but also implements other semiparametric models for discrete data. Thus, we aim to provide practitioners with easy access to Bayesian spectral analysis models based on recent methodologies. **bsamGP** is an R package available from <http://statlab2.korea.ac.kr/software/bsamgp> and from the Comprehensive R Archive Network (CRAN) at <https://CRAN.R-project.org/package=bsamGP>. It is implemented in the R statistical

computing environment and utilizes compiled Fortran 90 code to maximize computational efficiency when generating random samples from posterior distributions.

Although Bayesian semiparametric models are extremely powerful, their applicability depends on the availability of user-friendly software. Because semiparametric models are typically based on complex representations, the tools to fit semiparametric models in regression and density estimation may not be easily accessible to non-experts or practitioners. Currently, there are a few good command-driven software packages in R for semiparametric modeling. Package **DPpackage** (Jara, Hanson, Quintana, Müller, and Rosner 2011) is perhaps the most popular R package that includes many semiparametric models; it mostly uses Dirichlet process mixture models and B-spline functions, and it provides either flexible regression or density estimation for data analysis. Moreover, there are several general-purpose programs available for Bayesian inference, including BUGS (Lunn, Spiegelhalter, Thomas, and Best 2009), JAGS (Plummer 2003), and Stan (Carpenter *et al.* 2016), which can be tailored for fitting semiparametric models for regression and density estimation. For example, BUGS can handle Dirichlet process priors and nonparametric regression with spline and radial basis functions as well as the basic BSAM of Lenk (1999) with cosine series (see, e.g., Congdon 2006, Marley and Wand 2010, and Congdon 2014). Wood (2016) developed the `jagam` function in package **mgcv** (Wood 2017), which automatically generates JAGS code for the generalized additive model (GAM) supported by **mgcv**, utilizing spline-like basis expansions. The **brms** (Bürkner 2017) package implements Bayesian multilevel models in R using Stan, which also includes non-linear models. In addition, Gaussian process models can be built using Stan for regression and classification (Stan Development Team 2017).

Although a number of fully Bayesian programs and user-friendly software packages continue to be developed, as mentioned earlier, these programs and packages generally do not include semiparametric models with shape restrictions, with the exceptions of the monotonicity constraints in **GPstuff** (Vanhatalo *et al.* 2013) for Gaussian process regression, in **bisoreg** (Curtis and Ghosh 2011) with Bernstein polynomials, and in **bnpmr** (Bornkamp and Ickstadt 2009) with mixtures of cumulative distributions, in spite of recent results using Bayesian shape constraints for semiparametric models (see, e.g., Meyer, Hackstadt, and Hoeting 2011, Lin and Dunson 2014, Golchi, Bingham, Chipman, and Campbell 2015, Wang and Berger 2016, Lenk and Choi 2017 and the references therein). To the best of our knowledge, **bsamGP** is the first user-friendly software developed for the purpose of Bayesian shape-restricted function estimation for semiparametric generalized additive models, handling monotonicity, convexity, S-shaped functions and U-shaped functions.

Note that it is possible to incorporate the basic BSAM without shape restrictions into the general programs either through direct coding with BUGS and Stan or by interfacing JAGS and **mgcv** developed in Wood (2016). However, it would be rather unfriendly, time-consuming, and tedious for practitioners to fit the semiparametric models with shape restrictions. In addition, BSAM with the shape restrictions of Lenk and Choi (2017) involves numerical integration and quadratic forms to be evaluated for modeling the derivatives of the unknown function as squared Gaussian processes. This would be computationally inefficient and cumbersome when using the general purpose MCMC sampling implementations, particularly for estimating S-shaped and U-shaped functions.

**bsamGP** currently supports various parametric and semiparametric models with nonparametric components using the Gaussian process priors: linear regression models with Gaussian and non-Gaussian errors, partial linear regression models with Gaussian and non-Gaussian

errors, partial linear quantile regression models based on asymmetric Laplace distributions and Dirichlet mixtures of asymmetric Laplace distributions, generalized partial linear regression models (probit and logistic regression models for binary data and Poisson and negative binomial regression models for count data), and density estimation models using a logistic Gaussian process prior. For non-Gaussian errors, Dirichlet process mixture models are exploited for either non-Gaussian or unknown error distributions. In particular, **bsamGP** models unknown regression functions with monotonic, monotonic convex or concave, S-shaped, or U-shaped shape constraints. As explained previously, both applications use novel squared Gaussian process priors that assume that the derivatives of the functions are the squares of Gaussian processes. In addition, **bsamGP** contains model selection procedures for testing the existence of a shape restriction in a regression model as well as the adequacy of a parametric model relative to a non-specific alternative in a partial regression model based on the computation of marginal likelihoods.

The rest of this paper is organized as follows. Section 2 briefly reviews the approach to partial linear models using BSAM, as implemented in **bsamGP**, and explains how shape constraints are imposed on the unknown function. Then, we present basic procedures for partial linear quantile models, generalized partial linear models, and density estimation using BSAM. Section 3 illustrates the basic usage of the main functions and methods of **bsamGP** using synthetic data sets and demonstrates how shape restrictions are specified on the semiparametric regression models with various options in **bsamGP** using benchmark data sets. Section 4 concludes with a discussion on a possible extension to the **bsamGP** package.

## 2. Bayesian spectral analysis models

In this section, we briefly review a Gaussian process and its spectral representation, and then describe various semiparametric models using a spectral analysis of a Gaussian process as a prior for nonparametric components. Gaussian processes are widely used to define prior distributions over functions with continuous domains (e.g., see [Rasmussen and Williams 2006](#), [Murphy 2012](#), and [Gelman, Carlin, Stern, Dunson, Vehtari, and Rubin 2013](#)). Let  $Z = (Z(x), x \in \mathcal{X})$  and  $\mathcal{X}$  be an arbitrary index set, such as time or space. A stochastic process  $Z$  is called a Gaussian process if the marginal distribution of any finite-dimensional  $Z(x_1), \dots, Z(x_m)$ , where  $x_1, \dots, x_m \in \mathcal{X}$ , is multivariate normal. This process is clearly determined by the mean function  $\mu(x) = \mathbf{E}[Z(x)]$  and the covariance function  $\nu(x, x') = \mathbf{E}[\{Z(x) - \mu(x)\}\{Z(x') - \mu(x')\}]$  for  $x, x' \in \mathcal{X}$ .

A spectral analysis model of a Gaussian process is defined by linearizing its covariance matrix with the Karhunen-Loève representation (see, e.g., [Grenander 1981](#)). Let  $Z$  be a second-order Gaussian process with a mean function equal to zero and the covariance function  $\text{COV}(s, t)$ . Let  $\varphi_0(x), \varphi_1(x), \varphi_2(x), \dots$  be orthonormal basis functions for all  $x \in [0, 1]$ . Then, a spectral analysis model represents the Gaussian process  $Z$  as an infinite series expansion with the Karhunen-Loève representation

$$Z(x) = \sum_{j=0}^{\infty} \theta_j \varphi_j(x), \quad x \in [0, 1], \quad (1)$$

and its covariance function is given by  $\text{COV}(x, x') = \sum_{j=0}^{\infty} \nu_j^2 \varphi_j(x) \varphi_j(x')$ , where  $\sum_{j=0}^{\infty} \nu_j^2 < \infty$  and  $\nu_j^2 = \int_0^1 \int_0^1 \text{COV}(s, t) \varphi_j(s) \varphi_j(t) ds dt$ .

All of the methods implemented in **bsamGP** are based on the spectral representation in (1) of Gaussian processes, in particular using a Fourier series expansion based on the cosine basis functions (Lenk 1999, 2003; Lenk and Choi 2017):

$$\varphi_0(x) = 1 \quad \text{and} \quad \varphi_j(x) = \sqrt{2} \cos(\pi j x), \quad j \geq 1, \quad 0 \leq x \leq 1, \quad (2)$$

$$\theta_j = \int_0^1 Z(x) \varphi_j(x) dx, \quad j \geq 0.$$

Note that the cosine functions in (2) form an orthonormal basis for piece-wise continuous functions on  $[0, 1]$  (see, e.g., Kreider, Kuller, Ostberg, and Perkins 1966), and that the spectral analysis models facilitate the computation of the posterior distribution of the Gaussian process. If the support of  $x$  is  $S$ , then the orthonormal basis can be defined as  $\varphi_0(x) = \sqrt{q(x)}$  and  $\varphi_j(x) = \sqrt{2q(x)} \cos[\pi j Q(x)]$ , where  $Q$  is a cumulative distribution function with support  $S$ , and  $q$  is its density.

## 2.1. The BSAR model

We first consider a partial linear additive model for the mean regression. Let  $y_i$  and  $\mathbf{w}_i$  be the response and the vector of parametric predictors, respectively. Further, let  $x_{i,k}$  be the covariate related to the response through an unknown nonlinear function. The partial linear additive model is given as follows:

$$y_i = \mathbf{w}_i^\top \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{i,k}) + \epsilon_i, \quad i = 1, \dots, n, \quad (3)$$

where  $f_k$  is an unknown nonparametric function of the scalar  $x_{i,k} \in [0, 1]$ , which is estimated using the spectral analysis model of the Gaussian process given in (1) and (2), and the error terms  $\{\epsilon_i\}$  are a random sample from a normal distribution,  $\mathcal{N}(0, \sigma^2)$ . For the spectral coefficients of each function  $f_k$ , we assign the scale-invariant prior (Lenk and Choi 2017):

$$\theta_{jk} \mid \sigma, \tau_k, \gamma_k \sim \mathcal{N}\left(0, \sigma^2 \tau_k^2 \exp[-j\gamma_k]\right), \quad j \geq 1. \quad (4)$$

With regard to the priors of the hyperparameters  $\tau_k^2$  and  $\gamma_k$ , we consider the following two specifications:

- T smoother:  $\tau_k^2 \sim \mathcal{IG}\left(\frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right)$  and  $\gamma_k \sim \mathcal{Exp}(w_0)$ .
- Lasso smoother:  $\tau_k^2 \sim \mathcal{Exp}(u_0)$  and  $\gamma_k \sim \mathcal{Exp}(w_0)$ .

To complete the model specification, we choose the conjugate priors for  $\boldsymbol{\beta}$  and  $\sigma^2$ :

$$\boldsymbol{\beta} \mid \sigma \sim \mathcal{N}\left(m_{0,\boldsymbol{\beta}}, \sigma^2 V_{0,\boldsymbol{\beta}}\right), \quad \sigma^2 \sim \mathcal{IG}\left(\frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right).$$

We refer to this model as the Bayesian spectral analysis regression (BSAR) in the remainder of this paper. The MCMC sampling steps for BSAR are not complicated because all parameters except  $\gamma_k$  have conjugate priors for T smoother, whereas all the parameters other than  $\tau_k^2$  and  $\gamma_k$  have conjugate priors for Lasso smoother. Specifically, we generate the posterior samples

of  $\beta, \sigma^2$  and  $\theta$  from the explicit full conditional distributions, and we use slice sampling methods (e.g., [Damien, Wakefield, and Walker 1999](#) and [Neal 2003](#)) in order to generate  $\gamma_k$  for T smoother and both  $\tau_k^2$  and  $\gamma_k$  for Lasso smoother. The further details of sampling procedures with posterior distributions can be found in [Lenk \(1999\)](#) and [Lenk and Choi \(2017\)](#).

Additionally, we consider modeling the error terms  $\{\epsilon_i\}$  nonparametrically as a random sample from the Dirichlet process mixture models, which can provide a flexible and robust regression analysis of data. Specifically, the error distributions are assumed to follow one of two mixture models as follows:

- Scale mixture:  $\epsilon_i \stackrel{\text{iid}}{\sim} \int \mathcal{N}(\epsilon_i | 0, \sigma^2) dG(\sigma^2)$ ,

$$G \sim DP(M, G_0), \quad M \sim \mathcal{Ga}(a_M, b_M), \quad G_0 = \mathcal{IG}\left(\sigma^{-2}; \frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right).$$

- Location-scale mixture:  $\epsilon_i \stackrel{\text{iid}}{\sim} \int \mathcal{N}(\epsilon_i | \mu, \sigma^2) dG(\mu, \sigma^2)$ ,

$$G \sim DP(M, G_0), \quad M \sim \mathcal{Ga}(a_M, b_M), \quad G_0 = \mathcal{N}(\mu; \mu_0, \kappa\sigma^2) \mathcal{IG}\left(\sigma^{-2}; \frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right).$$

Here  $DP(\cdot)$  denotes the Dirichlet process ([Ferguson 1973](#)),  $M > 0$  is the precision parameter, and  $G_0(\cdot)$  is the centering measure. We implement the Dirichlet process mixture models with the “*no-gaps*” algorithm of [MacEachern and Müller \(1998\)](#).

## 2.2. The BSAR model with shape restrictions

[Lenk and Choi \(2017\)](#) developed the BSAR models for the shape-restricted regression functions by assuming that the derivatives of the functions are the squares of Gaussian processes. That is, the  $q$ th ( $q \geq 1$ ) derivative of the unknown function  $f$ ,  $f^{(q)}$  is modeled as the square of a zero-mean, second-order Gaussian process, depending on the *a priori* shape constraint imposed on the unknown function. Specifically, for monotonicity and convexity/concavity, we consider

- Monotone:

$$\begin{aligned} f^{(1)}(x) &= \delta Z^2(x), \quad \delta \in \{1, -1\}, \\ f(x) &= \delta \left[ \int_0^x Z^2(s) ds - \int_0^1 \int_0^x Z^2(s) ds dx \right], \end{aligned} \tag{5}$$

- Convex/concave:

$$\begin{aligned} f^{(2)}(x) &= \delta Z^2(x), \quad \delta \in \{1, -1\}, \\ f(x) &= \delta \left[ \int_0^x \int_0^s Z^2(t) dt ds - \int_0^1 \int_0^x \int_0^s Z^2(t) dt ds dx \right] + \alpha(x - 0.5), \end{aligned} \tag{6}$$

where  $\delta$  can be either 1 (non-decreasing) or  $-1$  (non-increasing) as given by the user. Note that the  $f$  is assumed to be mean centered and orthogonal to the constant function for identification



and that the constant of integration is included in each of integral representations of  $f$  with  $\alpha$  for mean-centering constraints (see [Lenk and Choi 2017](#) for further detailed discussions).

Because  $Z$  uses a spectral representation  $Z(x) = \sum_{j=1}^{\infty} \theta_j \varphi_j(x)$  with the same cosine basis functions as in (1) and (2), the BSAR model with a monotone restriction (5) is given by

$$f(x) = \delta \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j \theta_k \varphi_{j,k}^a(x), \quad \varphi_{j,k}^a(x) = \int_0^x \varphi_j(s) \varphi_k(s) ds - \int_0^1 \int_0^s \varphi_j(t) \varphi_k(t) dt ds.$$

The infinite series representation of  $Z$  is approximated by a finite sum  $Z_J(x)$ ,  $Z(x) \approx Z_J(x) = \sum_{j=0}^J \theta_j \varphi_j(x)$ , where  $J$  denotes the truncation point, and the BSAR model with monotone restrictions can be written in matrix notation as

$$\mathbf{y} = \mathbf{W}\boldsymbol{\beta} + \delta \boldsymbol{\theta}_J^\top \boldsymbol{\Phi}_J^a(x) \boldsymbol{\theta}_J + \boldsymbol{\epsilon}, \quad (7)$$

where  $\mathbf{y} = (y_1, \dots, y_n)^\top$ ,  $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^\top$ ,  $\mathbf{x} = (x_1, \dots, x_n)^\top$ ,  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^\top$ ,  $\boldsymbol{\theta}_J = (\theta_0, \dots, \theta_J)^\top$  is the  $J+1$  vector of spectral coefficients, and  $\boldsymbol{\Phi}_J^a(x)$  is a  $(J+1) \times (J+1)$  matrix with  $(j, k)$  entry  $\varphi_{j,k}^a(x)$ . Similarly, the BSAR model with monotone convexity or concavity in (6) is given by the spectral representation

$$\begin{aligned} f(x) &= \delta \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} \theta_j \theta_k \varphi_{j,k}^b(x) + \alpha(x - 0.5), \\ \varphi_{j,k}^b(x) &= \int_0^x \int_0^s \varphi_j(t) \varphi_k(t) dt ds - \int_0^1 \int_0^x \int_0^s \varphi_j(t) \varphi_k(t) dt ds dx, \\ f(x) &\approx f_J(x) = \delta \boldsymbol{\theta}_J^\top \boldsymbol{\Phi}_J^b(x) \boldsymbol{\theta}_J + \alpha(x - 0.5), \end{aligned}$$

where  $\boldsymbol{\Phi}_J^b(x)$  is the  $(J+1) \times (J+1)$  matrix with  $(j, k)$  entry  $\varphi_{j,k}^b(x)$ . Thus, the scale-invariant prior distributions of the spectral coefficients of the shape-restricted functions basically inherit those of the BSAR model described in Section 2.1 but are modified as

$$\theta_0 | \sigma \sim \mathcal{N}(m_{\theta_0}, \sigma v_{\theta_0}^2), \quad \theta_j | \sigma, \tau, \gamma \sim \mathcal{N}(m_{\theta_j}, \sigma \tau^2 \exp[-j\gamma]), \quad j \geq 1.$$

The other prior distributions are the same as those for the BSAR model described in Section 2.1 except for the truncated normal prior distribution of  $\alpha$ ,

$$\alpha | \sigma \sim \mathcal{N}(m_{0,\alpha}, \sigma^2 v_0^2 | 0, \alpha) I(\delta \alpha \geq 0),$$

which ensures that  $f'$  is positive or negative.

For the U-shaped and S-shaped restrictions, we additionally include a decreasing logistic function  $h(x)$  scaled between  $-1$  and  $1$  in (5) and (6),

- U-shaped:

$$\begin{aligned} f^{(1)}(x) &= \delta Z^2(x) h(x), \quad \delta \in \{1, -1\}, \\ h(x) &= \frac{1 - \exp[\psi(x - \omega)]}{1 + \exp[\psi(x - \omega)]}, \quad \psi > 0, \quad 0 < \omega < 1, \\ f(x) &= \delta \left[ \int_0^x Z^2(s) h(s) ds - \int_0^1 \int_0^x Z^2(t) h(t) dt ds \right], \end{aligned} \quad (8)$$

- S-shaped:

$$\begin{aligned}
f^{(2)}(x) &= \delta Z^2(x)h(x), \quad \delta \in \{1, -1\}, \\
h(x) &= \frac{1 - \exp[\psi(x - \omega)]}{1 + \exp[\psi(x - \omega)]}, \quad \psi > 0, \quad 0 < \omega < 1, \\
f(x) &= \delta \zeta \left[ \int_0^x \int_0^s Z^2(t)h(t)dt ds - \int_0^1 \int_0^x \int_0^s Z^2(t)h(t)dt ds dx \right] + (\alpha - \delta\xi)(x - 0.5), \\
\xi &= \min \left[ 0, \min_{x \in [0,1]} \zeta \int_0^x Z^2(s)h(s)ds \right],
\end{aligned} \tag{9}$$

where  $\zeta$  is given by the user and  $\xi$  is also considered for mean-centering constraints, similarly to  $\alpha$ . Note that there are two additional parameters  $\omega$  and  $\psi$  in  $h(x)$ .  $\omega$  is a unique zero of  $h(x)$  and  $\psi$  is a slope of  $h(x)$ , controlling its steepness at  $\omega$ . Truncated normal prior distributions are assigned to these parameters as

$$\psi \sim \mathcal{N}(m_{0,\psi}, v_{0,\psi}^2)I(\psi > 0), \quad \omega \sim \mathcal{N}(m_{0,\omega}, v_{0,\omega}^2)I(0 < \omega < 1).$$

For the shape restricted models, the full conditional distributions for the parameters  $\boldsymbol{\theta}$ ,  $\sigma^2$ ,  $\omega$  and  $\psi$  do not have closed-form expressions. For the MCMC sampling steps, we therefore use adaptive Metropolis algorithms for  $\boldsymbol{\theta}$ , slice sampling algorithms for  $\sigma^2$  and additional hyperparameters for  $\boldsymbol{\theta}$ , and numerical integration using the trapezoidal rule to obtain the basis functions,  $\varphi^a(\cdot)$  and  $\varphi^b(\cdot)$ . Additional details can be found in [Lenk and Choi \(2017\)](#). Further, as shown in the representations of U-shaped (8) and S-shaped functions (9), all of the parameters other than two additional parameters,  $\omega$  and  $\psi$  for  $h(x)$ , and the integral  $Z^2h$  in computing  $f(x)$  are the same as in the BSAR models with monotonicity and convexity. Using a finite approximation of  $Z_J$ , the same MCMC sampling steps can be used with additional adaptive Metropolis steps for  $\omega$  and  $\psi$  using truncated normal distributions and numerical integrations to compute  $Z_J^2h$  over a fine grid. Further details are also given in the supplementary materials in the online appendix of [Lenk and Choi \(2017\)](#).

### 2.3. The BSAQ model

We next consider a partial linear additive model for quantile regression. A Bayesian analysis of a quantile regression assumes that the error term has an asymmetric Laplace distribution (e.g., see [Yu and Moyeed 2001](#) and [Sriram, Ramamoorthi, and Ghosh 2013](#)), with density

$$\mathcal{AL}_p(\epsilon; \sigma^2) = \frac{p(1-p)}{\sigma^2} \exp \left\{ -\frac{|\epsilon| + (2p-1)\epsilon}{2\sigma^2} \right\}, \tag{10}$$

where  $\sigma^2$  is a positive scale parameter and  $p$  is a fixed value in  $(0, 1)$ . Based on the asymmetric Laplace distribution in (10), we have the following partial linear additive quantile regression model, using the aforementioned spectral analysis approach:

$$\begin{aligned}
y_i &= \mathbf{w}_i^\top \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{i,k}) + \eta_1 u_i + \eta_2 \sigma u_i \epsilon_i, \quad i = 1, \dots, n, \\
u_i &\sim \text{Exp}(\sigma^{-2}), \quad i = 1, \dots, n,
\end{aligned}$$



where  $\eta_1 = (1-2p)/\{p(1-p)\}$ ,  $\eta_2^2 = 2/\{p(1-p)\}$ , and  $\pi(u)$  is a density function of an exponential distribution with mean  $\sigma^2$ ,  $\pi(u) = \sigma^{-2}e^{-u/\sigma^2}$ . As in the BSAR model in Section 2.1, we use the Bayesian spectral representation for the unknown additive regression functions  $f_k(\cdot)$ ,  $k = 1, \dots, K$ . We refer to this model as the Bayesian spectral analysis quantile regression (BSAQ) model.

Note that the joint posterior distribution for the BSAQ is proportional to

$$p(\boldsymbol{\beta}, \boldsymbol{\theta}_J, \mathbf{u}, \tau^2, \gamma \mid \text{Data}) \propto \prod_{i=1}^n \mathcal{N}(y_i; \mathbf{w}_i^\top \boldsymbol{\beta} + f_J(x_i) + \eta_1 u_i, \eta_2^2 \sigma^2 u_i) \mathcal{N}(\boldsymbol{\beta}; \mathbf{m}_{0,\beta}, \sigma^2 \mathbf{V}_{0,\beta}) \\ \times \prod_{j=1}^J N(\theta_j; 0, \sigma^2 \tau^2 \exp[-j\gamma]) \mathcal{IG}\left(\tau^2; \frac{r_{0,\tau}}{2}, \frac{s_{0,\tau}}{2}\right) \mathcal{Exp}(\gamma; \omega_0) \times \prod_{i=1}^n \mathcal{Exp}(u_i; \sigma^{-2}),$$

where we use a location-scale mixture representation of exponential and normal distributions for  $\mathcal{AL}_p$ , which is commonly used to facilitate the posterior sampling algorithm. Thus, the MCMC sampling steps are the same as those described in Sections 2.1 and 2.2 for the BSAR model with or without shape restrictions, with the exception of an additional sampling step that updates latent mixing parameters from generalized inverse Gaussian distributions (see, e.g., Kozumi and Kobayashi 2011, Alhamzawi and Yu 2013, and Jo *et al.* 2016 for detailed sampling procedures with posterior distributions). That is, we update the latent variables  $\mathbf{u} = (u_1, \dots, u_n)$  from

$$[u_i \mid \text{others}] \sim \mathcal{GIG}\left(\frac{1}{2}, \tilde{\delta}, \tilde{\xi}_i\right), \quad \tilde{\delta} = \frac{2 + \eta_1^2/\eta_2^2}{\sigma^2}, \quad \tilde{\xi}_i = \frac{(y_i - \mathbf{w}_i^\top \boldsymbol{\beta} - \sum_{j=1}^J \theta_j \varphi_j(x_i))^2}{\eta_2^2 \sigma^2}.$$

The shape restrictions in (5)–(9) are also incorporated into the BSAQ model, and the posterior sampling steps are identical to those of BSAR with shape restrictions with the previous additional sampling step for the latent mixing parameters  $\mathbf{u}$ .

Furthermore, we consider error terms  $\{\epsilon_i\}$  as a random sample from the Dirichlet process scale mixture of an asymmetric Laplace distribution for flexible nonparametric modeling for the error distribution (see, e.g., Jo *et al.* 2016 and Kottas and Krnjajić 2009). Specifically, we model the error distribution as follows:

$$\epsilon_i \stackrel{\text{iid}}{\sim} \int \mathcal{AL}_p(\epsilon; \sigma^2) dG(\sigma^2), \quad G \sim DP(M, G_0), \quad M \sim \mathcal{Ga}(a_M, b_M), \quad G_0 = \mathcal{IG}\left(\sigma^{-2}; \frac{r_{0,\sigma}}{2}, \frac{s_{0,\sigma}}{2}\right),$$

for which the “no-gaps” algorithm of MacEachern and Müller (1998) is also used.

## 2.4. The GBSAR model

Generalized linear models (GLMs), introduced by Nelder and Wedderburn (1972), provide a unified framework for the regression analysis of normal and non-normal response variables, assuming that the mean of the response variable is linearly associated with the predictors by a known smooth function, called a *link function*. A GAM (Hastie and Tibshirani 1990) combines the properties of GLMs with those of additive non-/semiparametric models by relaxing the linearity assumption on the actual relationship between response variable and predictors. Here, we consider the GAM approach using a spectral analysis regression, referred to as the generalized linear BSAR (GBSAR) model. Specifically, we deal with semiparametric

generalized additive models for discrete responses, including binary, Poisson and negative binomial data.

*GBSAR: Binary response regression*

Consider a binary response variable  $y_i \in \{0, 1\}$ ,  $i = 1, \dots, n$  for a collection of  $n$  subjects associated with  $p$  linear effect predictors  $\mathbf{w}_i = (w_{i1}, \dots, w_{ip})^\top$  and  $K$  nonlinear effect predictors  $\mathbf{x}_i = (x_{i,1}, \dots, x_{i,K})^\top$ . The Bayesian generalized partial linear model for binary response is given as follows:

$$y_i \stackrel{\text{iid}}{\sim} \mathcal{B}(1, \mu_i), \quad 1 \leq i \leq n, \quad (11)$$

$$g(\mu_i) = \mathbf{w}_i^\top \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{i,k}),$$

where  $g(\cdot)$  is a *link function*,  $\boldsymbol{\beta}$  is a  $p$ -dimensional vector of coefficients for linear effects, and  $f_k(\cdot)$  is an unknown nonlinear function of the scalar  $x_{i,k} \in [0, 1]$ , which is modeled using spectral analysis models of Gaussian processes. Typically, the *link function*  $g(\cdot)$  in the binary regression is chosen as the cumulative distribution function (CDF) of some continuous distributions defined on the real line, such as, for example, the standard normal distribution and the logistic distribution. In the former case, the binary regression model is referred to as the *probit model*, and the latter is called the *logit model* (see, e.g., McCullagh and Nelder 1989, Hastie and Tibshirani 1990 and Dey, Ghosh, and Mallick 2000).

**GBSAR with the probit regression.** For the probit link,  $g^{-1}(\mu) = \Phi(\mu)$ , where  $\Phi(\mu)$  denotes the CDF of the standard normal distribution, Albert and Chib (1993) proposed a simple data augmentation approach that renders the full conditional distributions of the model parameters equivalent to those under the Bayesian normal regression model. Thus, the GBSAR with the probit link represents the model in (11) using auxiliary variables, as follows:

$$y_i = I(z_i > 0),$$

$$z_i = \mathbf{w}_i^\top \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{i,k}) + \epsilon_i, \quad (12)$$

$$\epsilon_i \stackrel{\text{iid}}{\sim} \mathcal{N}(0, 1).$$

The main advantage of Albert and Chib (1993)'s approach in (12) is that it enables the use of the Gibbs sampling algorithm to simulate posterior samples of the unknown parameters in the GBSAR with the probit regression model. That is, assuming  $K = 1$  for simplicity, the joint posterior of all unknown parameters is proportional to

$$\prod_{i=1}^n \{I(z_i \geq 0)I(y_i = 1) + I(z_i < 0)I(y_i = 0)\} \times \prod_{i=1}^n \exp\left(-\frac{1}{2}(z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(x_i))^2\right)$$

$$\times \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{m}_\beta)^\top \mathbf{V}_\beta^{-1}(\boldsymbol{\beta} - \mathbf{m}_\beta)\right) \times \prod_{j=1}^J \left\{ (\tau^2 \exp(-j\gamma))^{-\frac{1}{2}} \exp\left(-\frac{\theta_j^2}{2\tau^2 \exp(-j\gamma)}\right) \right\}$$

$$\times \exp(-w_0\gamma) \times (\tau^2)^{-\frac{r_0}{2}-1} \exp\left(-\frac{s_0}{2\tau^2}\right),$$

and, thus, the MCMC sampling steps are the same as those in BSAR with the Gibbs sampling and slice sampling methods, except for an additional step to update the latent variable  $z_i$  from a form of the truncated normal posterior distribution,

$$z_i | \text{all others} \sim \begin{cases} \mathcal{N}(\mathbf{w}_i^\top \boldsymbol{\beta} + f(x_i), 1) I(z_i \geq 0) & \text{if } y_i = 1, \\ \mathcal{N}(\mathbf{w}_i^\top \boldsymbol{\beta} + f(x_i), 1) I(z_i \leq 0) & \text{if } y_i = 0. \end{cases}$$

**GBSAR with the logistic regression.** As an alternative to the probit link, the logit link,  $g^{-1}(\mu) = 1/(1 + e^{-\mu})$ , is also widely used. From the missing-data mechanism of [Albert and Chib \(1993\)](#), the GBSAR model (11) with the logit link is defined by

$$\begin{aligned} y_i &= I(z_i > 0), \\ z_i &= \mathbf{w}_i^\top \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{i,k}) + \epsilon_i, \\ \epsilon_i &\stackrel{\text{iid}}{\sim} \mathcal{L}o(0, 1), \end{aligned} \tag{13}$$

where  $\mathcal{L}o(0, 1)$  denotes a standard logistic distribution ([Devroye 1986](#)) with a density function given by  $\mathcal{L}o(\epsilon; 0, 1) = \frac{e^\epsilon}{(1 + e^\epsilon)^2}$ . The binary regression model with a logit link in (13), called the *logistic regression model*, is typically preferred to the probit regression model (12) for most statistical applications because of the easy interpretation of the regression coefficients. In the implementation of GBSAR logistic regression, we exploit the hierarchical representation of the logistic distribution and the data-augmentation approach for the posterior sampling proposed by [Holmes and Held \(2006\)](#). Specifically, [Holmes and Held \(2006\)](#) represented the logistic distribution as a scale mixture of a normal distribution ([Andrew and Mallows 1974](#)):

$$\begin{aligned} \epsilon_i | \lambda_i &\stackrel{\text{iid}}{\sim} \mathcal{N}(0, \lambda_i), \\ \lambda_i &= (2\psi_i)^2, \quad \psi_i \sim \mathcal{K}\mathcal{S}, \end{aligned}$$

where  $\mathcal{K}\mathcal{S}$  denotes the Kolmogorov-Smirnov distribution (see, e.g., [Devroye 1986](#); [Holmes and Held 2006](#)) given by  $\mathcal{K}\mathcal{S}(\psi) = \sum_{n=-\infty}^{\infty} (-1)^n e^{-2n^2\psi^2}$ ,  $\psi \geq 0$ . Note that a further set of auxiliary variables  $\lambda_i$ ,  $i = 1, \dots, n$ , facilitates the posterior sampling algorithm of the logistic regression model in (13).

Thus, the MCMC sampling procedures are the same as those of probit regression using the algorithm of [Albert and Chib \(1993\)](#) except for the additional steps for updating  $z_i$  and  $\psi_i$  from the logistic distribution as a scale mixture of the normal distribution.

- Update  $z_i$ ,  $i = 1, \dots, n$ :

$$z_i | \text{all others} \propto \begin{cases} \mathcal{L}o(z_i; \mathbf{w}_i^\top \boldsymbol{\beta} + f(x_i), 1) I(z_i > 0) & \text{if } y_i = 1, \\ \mathcal{L}o(z_i; \mathbf{w}_i^\top \boldsymbol{\beta} + f(x_i), 1) I(z_i \leq 0) & \text{if } y_i = 0, \end{cases}$$

- To update  $\psi_i$ ,  $i = 1, \dots, n$ , we first re-parameterize  $\lambda_i = (2\psi_i)^2$  and then update  $\lambda_i$  using a rejection sampling algorithm with a proposal generated from the generalized inverse Gaussian distribution,  $\mathcal{G}\mathcal{I}\mathcal{G}(0.5, 1, R_i^2)$ , where  $R_i^2 = [z_i - \mathbf{w}_i^\top \boldsymbol{\beta} - f(x_i)]^2$ .

Alternatively, because the full conditional distributions of the unknown parameters do not have standard forms, the Metropolis-Hastings algorithm, which is typically used in the Bayesian logistic regression model, can be directly applied to posterior sampling without using the missing-data mechanism of [Albert and Chib \(1993\)](#) and [Holmes and Held \(2006\)](#). Specifically, we generate posterior samples from the joint posterior distribution based on the adaptive Metropolis algorithms (e.g., [Haario, Saksman, and Tamminen 2001](#) and [Roberts and Rosenthal 2009](#)) using normal distributions with adaptively updated variances as proposal distributions. The shape restrictions in Section 2.2 are also incorporated into the GBSAR for binary regression, and the same posterior sampling algorithms are applied as in BSAR with shape restrictions.

### *GBSAR: Count response regression*

When the response is an unbounded count,  $y = 0, 1, 2, \dots$ , we can use count response regression models. In this case, two regression models for counts are commonly considered: the Poisson regression and the negative binomial regression. The Poisson regression model assumes that the response variables  $\mathbf{y} = (y_1, \dots, y_n)$  are independently Poisson distributed with the rate parameters  $\lambda_i$ ,  $i = 1, \dots, n$  and that the logarithm of  $\lambda_i$  can be modeled by a regression function of predictors. The negative binomial regression model extends the Poisson regression model by introducing an additional parameter that is known to easily handle the over-dispersion problem (see, e.g., [Hilbe 2011](#)).

**GBSAR with Poisson regression.** We work with the following GBSAR for the Poisson additive model:

$$\begin{aligned} y_i | \lambda_i &\stackrel{\text{iid}}{\sim} \mathcal{P}(\lambda_i), \quad i = 1, \dots, n, \\ \log(\lambda_i) &= \mathbf{w}_i^\top \boldsymbol{\beta} + \sum_{k=1}^K f_k(x_{i,k}), \end{aligned} \tag{14}$$

where  $\boldsymbol{\beta}$  is a  $p$ -dimensional vector of the coefficients for the linear effects,  $\mathbf{w}_i$  is the corresponding design vector, and the  $\{f_k(\cdot)\}$  are unknown nonlinear functions of the scalars  $x_{i,k} \in [0, 1]$ , modeled using spectral analysis models of Gaussian processes, as before. Note that because the full conditional distributions of the unknown parameters in (14) do not have standard forms, we also use the adaptive Metropolis algorithms.

That is, we generate posterior samples from the joint posterior distribution based on the adaptive Metropolis algorithms using normal distributions with adaptively updated variances as proposal distributions. Specifically, we illustrate the process for generating candidates for  $\boldsymbol{\beta}$ :

- Generate the candidates of  $\boldsymbol{\beta}_t$  with a proposal distribution given at iteration  $t$  by

$$Q_t(\boldsymbol{\beta}, \cdot) = \begin{cases} \mathcal{N}(\boldsymbol{\beta}_{t-1}, (0.1)^2 I_d/d) & \text{for } t \leq 2d, \\ 0.95 \mathcal{N}(\boldsymbol{\beta}_{t-1}, (2.38)^2 \Sigma_t/d) + 0.05 \mathcal{N}(\boldsymbol{\beta}_{t-1}, (0.1)^2 I_d/d) & \text{for } t > 2d, \end{cases}$$

where  $d$  is the dimension of  $\boldsymbol{\beta}$ , and  $\Sigma_t$  is the empirical covariance estimate of  $\boldsymbol{\beta}$ , sequentially computed as

$$\Sigma_t = \frac{t-1}{t} \Sigma_{t-1} + \frac{1}{t} \left\{ t \bar{\boldsymbol{\beta}}_{t-1} \bar{\boldsymbol{\beta}}_{t-1}^\top - (t+1) \bar{\boldsymbol{\beta}}_t \bar{\boldsymbol{\beta}}_t^\top + \boldsymbol{\beta}_t \boldsymbol{\beta}_t^\top \right\},$$

where  $\bar{\beta}_{t+1} = t^{-1} \sum_{i=0}^t \beta_i$ .

- Then, the candidates are accepted with the product of the prior distributions in Sections 2.1 and 2.2 for BSAR with or without shape restrictions and the Poisson likelihood from (14).

Note that the remaining parameters with spectral coefficients and other hyperparameters are updated as before.

**GBSAR with negative binomial regression.** The Poisson regression model can be of limited use in several disciplines because empirical count data sets typically exhibit overdispersion, usually caused by dependencies, variation between responses or violations in the distributional assumptions. The negative binomial distribution can better deal with such overdispersed count data and, thus, the negative binomial regression model is often preferred to the Poisson regression model (see, e.g., Hilbe 2014 and Fahrmeir, Kneib, Lang, and Marx 2013).

The Bayesian negative binomial additive model with GBSAR is given as follows:

$$y_i \mid \beta, \theta, \kappa \stackrel{\text{ind}}{\sim} \mathcal{NB} \left[ \exp \left\{ \mathbf{w}_i^\top \beta + \sum_{k=1}^K f_k(x_{i,k}) \right\}, \kappa \right], \quad i = 1, \dots, n, \quad (15)$$

where  $\kappa > 0$  is a dispersion parameter, assigned a gamma distribution with shape parameter  $r_{0,\kappa}$  and the rate parameter  $s_{0,\kappa}$  as a prior. The negative binomial model (15) can also be expressed as a continuous mixture of Poisson distributions, where the mixing distribution of the Poisson rate is a gamma distribution (see, e.g., Koop, Poirier, and Tobias 2007 and Luts and Wand 2015):

$$y_i \mid \lambda_i \stackrel{\text{ind}}{\sim} \mathcal{P}(\lambda_i), \quad i = 1, \dots, n, \quad (16)$$

$$\lambda_i \mid \beta, \theta, \kappa \stackrel{\text{ind}}{\sim} \mathcal{Ga} \left( \kappa, \kappa \exp \left\{ -\mathbf{w}_i^\top \beta - \sum_{k=1}^K f_k(x_{i,k}) \right\} \right).$$

Here, the parameters  $\kappa$  and  $\beta$  are updated using the adaptive Metropolis algorithm, and the posterior sampling procedures of the remaining unknown parameters are the same as before. Note that the negative binomial distribution was originally derived as a limiting case of the Poisson-gamma mixture distribution (Greenwood and Yule 1920), and that, in the current version of **bsamGP**, both the (15) and (16) approaches are available for implementing the GBSAR negative binomial regression. The shape restrictions in Section 2.2 are also incorporated into the GBSAR for count regression, and the same posterior sampling algorithms are applied as in BSAR with shape restrictions, using the joint posterior distribution proportional to the product of either the Poisson or negative binomial likelihood and the prior distributions described in Sections 2.1 and 2.2.

## 2.5. The BSAD model

In addition to regression models, we include a Bayesian semiparametric density estimation procedure using a spectral representation of Gaussian processes as in Lenk (2003). The semiparametric density model consists of a parametric component, specified by an exponential

family, and a nonparametric component, specified by a logistic Gaussian process (see, e.g., Leonard 1978 and Lenk 1988, 1991). The logistic Gaussian process, defined as the logistic transformation of a Gaussian process, provides a flexible model for estimating unknown densities using the covariance structure (see, e.g., Lenk 1991; Tokdar and Ghosh 2007; Riihimäki and Vehtari 2014).

Specifically, consider the situation in which the observations  $y_1, \dots, y_n$  constitute a random sample from an absolutely continuous probability density  $f$  with respect to a known, dominating measure  $G$  on the support  $\mathcal{Y}$ . A semiparametric density estimation model using a logistic Gaussian process prior (Lenk 2003), which we refer to as the Bayesian spectral analysis density estimation (BSAD) hereafter, is given as follows:

$$f(y|\boldsymbol{\beta}, Z) = \frac{\exp[\mathbf{w}(y)^\top \boldsymbol{\beta} + Z(y)]}{\int_{\mathcal{Y}} \exp[\mathbf{w}(x)^\top \boldsymbol{\beta} + Z(x)] dG(x)}, \quad y \in \mathcal{Y}, \quad (17)$$

where  $\mathbf{w}(y) = [w_1(y), \dots, w_p(y)]^\top$  is a  $p$ -vector of non-constant functions specifying an exponential family,  $\boldsymbol{\beta}$  is a  $p$ -vector of unknown coefficients, and  $Z$  is a zero-mean, second-order Gaussian process using the spectral representation  $Z(y) = \sum_{j=1}^{\infty} \theta_j \varphi_j(y)$ ,  $\varphi_j(y) = \sqrt{2} \cos(\pi j G(y))$ , as before. A normal prior for  $\{\theta_j\}$  similar to that of BSAR in (4) is considered

$$\theta_j \sim \mathcal{N}\left(0, \tau^2 \exp[-\gamma c_j]\right) \quad \text{for } \tau^2 > 0 \quad \text{and } \gamma > 0.$$

However, two possible parameterizations of  $c_j$  are considered, namely  $c_j = j$  for the **geometric** smoother and  $c_j = \log(j+1)$  for the **algebraic** smoother (Lenk 1999, 2003). Other unknown parameters have the same prior specifications as in the BSAR with a normal distribution for  $\boldsymbol{\beta}$  and in the T smoother for  $\tau^2$  and  $\gamma$ . In order to generate posterior samples in BSAD, given the logistic transformation in (17), the MCMC sampling procedures of BSAR are employed to estimate  $\boldsymbol{\beta}$ , spectral coefficients, and additional hyperparameters by transforming the problem from density estimation into the semiparametric regression based on a discrete version of the likelihood. Additional details can be found in Lenk (2003).

### 3. Basic implementation of bsamGP

The **bsamGP** package for R provides tools that implement the Bayesian spectral analysis models described in the previous section. In this section, we demonstrate the general usage of the main functions and methods of **bsamGP** using two synthetic data sets: one with a normal semiparametric regression for the BSAR and the other with a semiparametric density estimation for the BSAD. To assess the adequacy of the semiparametric model against the parametric model, we also provide the log marginal likelihoods for both cases. More importantly, we demonstrate how to specify shape restrictions on the semiparametric regression models, the BSAR, BSAQ, and GBSAR, with various options in **bsamGP** using benchmark data sets.

#### 3.1. The BSAR model with synthetic data

The main function in **bsamGP** for the analysis of normal regression data is a function `bsar()` that fits a semiparametric regression model using the BSAR approach described in Section 2.1.



To illustrate the use of the `bsar` function for the BSAR model, we generate synthetic data with  $n = 100$  observations from

$$y_i = 2w_i + f(x_i) + \epsilon_i, \quad \epsilon_i \sim \mathcal{N}(0, 1^2), \quad (18)$$

where  $w_i$  follows a normal distribution with mean 0 and standard deviation 0.5 and  $f(x) = 5 - 10x + 8 \exp[-100(x - 0.3)^2] - 8 \exp[-100(x - 0.7)^2]$  is the “double-normal” function used in [Lenk \(1999\)](#). Here, the predictor values  $x_i$  are generated from a uniform distribution,  $x_i \stackrel{\text{iid}}{\sim} \text{Unif}(0, 1)$ . The data are simulated using the following code:

```
R> set.seed(1)
R> n <- 100
R> f <- function(x) {
+   5 - 10 * x + 8 * exp(- 100 * (x - 0.3) ^ 2) -
+     8 * exp(- 100 * (x - 0.7) ^ 2)
+ }
R> w <- rnorm(n, sd = 0.5)
R> x <- runif(n)
R> y <- 2 * w + f(x) + rnorm(n, sd = 1)
```

To fit the data with the BSAR, we first need to specify the number of basis functions and the hyperparameters as follows:

```
R> nbasis <- 50
R> prior <- list(beta_m0 = numeric(2), beta_v0 = diag(100, 2), w0 = 2,
+   tau2_m0 = 1, tau2_v0 = 100, sigma2_m0 = 1, sigma2_v0 = 1000)
```

In total, 60,000 draws from the MCMC sampling method implemented in the `bsar` function were completed. A transition (burn-in) period of 10,000 samples was considered, and the chain was sub-sampled at every tenth iteration to obtain a final posterior sample size of 5,000. The following code illustrates the MCMC specification:

```
R> mcmc <- list(nblow = 10000, nskip = 10, smcmc = 5000, ndisp = 5000)
```

Then, the BSAR model is fitted using the following commands:

```
R> fout <- bsar(y ~ w + fs(x), nbasis = nbasis, mcmc = mcmc, prior = prior,
+   shape = "Free", marginal.likelihood = TRUE, spm.adequacy = TRUE)
```

**bsamGP** facilitates an intuitive formula interface to describe the semiparametric additive model where `fs()` indicates that nonparametric smoothing is applied to a variable `x` with spectral analysis regression in BSAM. Note that no shape constraint in the regression function is considered with option `shape = "Free"` in the specification of the BSAR model. Furthermore, we use the logarithm of the marginal likelihoods and the Bayes factor to test the adequacy of the semiparametric model with the options `marginal.likelihood = TRUE` and `spm.adequacy = TRUE`.

Note that the **coda** ([Plummer, Best, Cowles, and Vines 2006](#)) package can be used to perform convergence diagnostics based on output objects in `fout` and produce trace and density plots with the `plot` method in **coda**. For illustration, we display these plots of  $\beta$  and  $\sigma^2$  in [Figure 1](#) using the following commands.

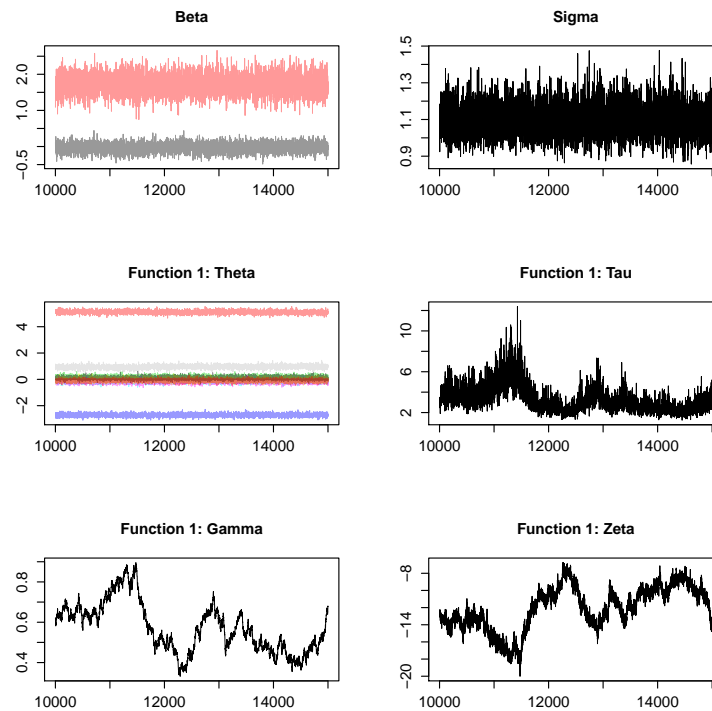


Figure 1: Trace plots for the parameters of the BSAR model.

```
R> library("coda")
R> post <- as.mcmc(data.frame(beta = fout$mcmc.draws$beta,
+   sigma2 = fout$mcmc.draws$sigma2))
R> plot(post)
```

Alternatively, the generic `plot` method for `fout` can be used to produce trace plots for each parameter using the following command.

```
R> plot(fout)
```

The posterior summary is produced from the object `fout` of the `bsar` function using either the `print` or the `summary` command. The code using the `summary` command and some of the results for the posterior summary are given as follows:

```
R> summary(fout)
```

```
Number of Cosine basis functions    = 50
Number of observations              = 100
Number of covariates (no intercept) = 1

MCMC transition draws              = 10000
MCMC draws saved for estimation    = 5000
Save every nskip draws            = 10
```

```
MCMC draws total           = 60000
```

```
R-Square                   = 0.9711
```

```
Log Integrated Likelihood
```

```
LIL Gelfand & Dey         = -231.8385
```

```
LIL Newton & Raftery (biased) = -157.5351
```

```
LIL Parametric            = -335.6345
```

```
H0: Parametric versus H1: Semiparametric
```

```
Log Bayes Factor (BF[01]) = 103.796
```

```
beta
```

	PostM	PostStd	PostM/STD
(Intercept)	-0.03426094	0.1240964	-0.2760832
w	1.72447915	0.2589314	6.6599846

```
sigma
```

```
PostM sigma = 1.095848
```

```
PostS sigma = 0.08495285
```

```
...
```

From the above summary results, we can test the adequacy of the semiparametric model against that of the parametric model using the option `spm.adequacy = TRUE` as mentioned previously. This shows that the Bayes factor clearly favors the semiparametric model over the parametric model for the simulated data set. The marginal likelihoods are computed with the option `marginal.likelihood = TRUE` using the methods of [Gelfand and Dey \(1994\)](#) and [Newton and Raftery \(1994\)](#).

The left panel of Figure 2 displays a scatter plot of the simulated data along with the true regression function, the estimated regression function, and point-wise 95% equal-tail credible intervals. The posterior mean estimate is very close to the true mean function (i.e., the “double-normal” function). The following code obtains the estimated mean function under the BSAR model and the true function and is used to draw the plot displayed in the left panel of Figure 2.

```
R> fit <- fitted(fout, HPD = FALSE)
R> plot(fit, ggplot2 = FALSE)
R> lines(fit$xgrid, f(fit$xgrid), lwd = 3, lty = 2)
```

**bsamGP** also provides the `predict` method to easily assess an out-of-sample posterior prediction. Specifically, the `predict` function for objects created by **bsamGP** is provided to allow the drawing of samples from the posterior predictive distributions and, thus, to obtain the unobserved response predicted by the model after the MCMC sampling is completed. The following code illustrates the use of the `predict` method in an out-of-sample prediction with 100 test data points. The `type = "response"` option *by default* in `predict` provides the posterior predictive estimates of the future (unobserved) observations given the current observations by averaging the conditional distribution of the future observations over all possible parameters weighted by their posterior distributions, whereas `type = "mean"` returns

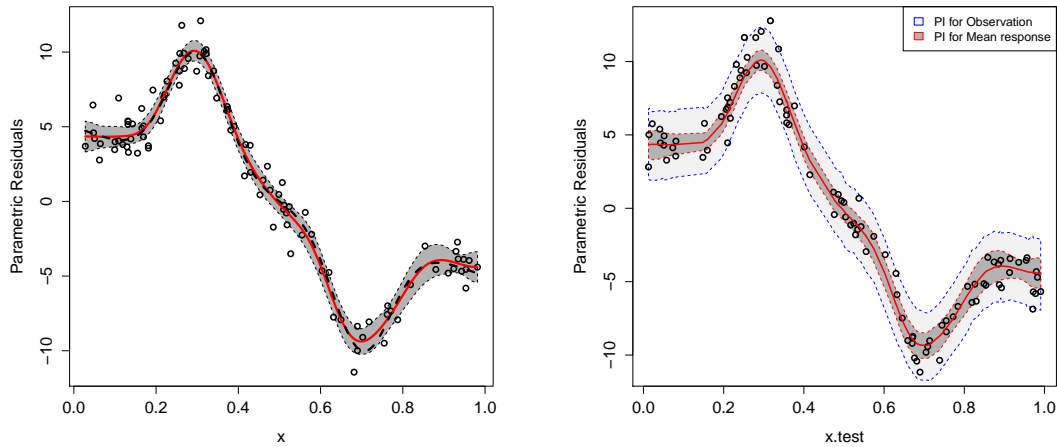


Figure 2: Left: Plot of the true and estimated mean regression curves with the true function (black dashed line), posterior mean estimates (red solid line), and point-wise 95% equal-tail credible intervals (gray region). Right: Plot of the posterior predictive mean regression curve (red solid line) with point-wise 95% equal-tail predictive intervals (PI for mean response, gray region) for the mean regression curve and those (PI for observation, light gray) for out-of-sample test data by the predict method.

posterior predictive regression curves. The right panel of Figure 2 portrays the posterior predictive mean of regression curves with point-wise 95% equal-tail intervals for posterior predictive estimates of future observations, specifically, out-of-sample test data, as well as those for posterior predictive regression curves superimposed with the test data. Note that the latter for the regression curves, that is, the noise-free observations, is narrower than the former for the predicted value of  $\tilde{y}$ , the future observation with noise.

```
R> set.seed(2)
R> n.test <- 100
R> w.test <- rnorm(n.test, sd = 0.5)
R> x.test <- runif(n.test)
R> fit.pred.tobs <- predict(fout, newp = w.test, newnp = x.test,
+   type = "response")
R> fit.pred.tmean <- predict(fout, newp = w.test, newnp = x.test,
+   type = "mean")
```

Note that similar plots to those in Figure 2 can be drawn by the `plot` method using the `ggplot2` (Wickham 2009) package with the option `ggplot2 = TRUE` option *by default*.

Alternatively, the simulated data set in (18) can be fitted based on the assumption of an unknown error distribution using the function `bsardpm` with a Dirichlet process mixture model. By default, the function `bsardpm` utilizes a location-scale mixture, as described in Section 2.1, and the option of a scale mixture can be specified as `location = FALSE` in `bsardpm`. To complete the model specification of `bsardpm`, we additionally specify hyperparameters on  $\kappa$  and  $M$  as default values. Then, the BSAR model with a Dirichlet process mixture is fitted using the following R code, and the abridged output of `fout.dpm` is given as follows:

```
R> prior <- append(prior, list(kappa_r0 = 1, kappa_s0 = 100, tmass_a = 2,
+   tmass_b = 4))
R> mcmc <- list(nblow = 20000, nskip = 10, smcmc = 5000)
R> fout.dpm <- bsardpm(y ~ w + fs(x), nbasis = nbasis, mcmc = mcmc,
+   prior = prior, shape = "Free")
R> summary(fout.dpm)

...
Log Pseudo Marginal Likelihood (LPML)
LPML Mukhopadhyay & Gelfand                = -158.8072

R-Square                                     = 0.9716

Number of Clusters
nclass
      1      2      3      4
0.8808 0.1128 0.0062 0.0002

Total mass
      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.003818 0.117289 0.201338 0.243443 0.328480 1.230444

beta
      PostM PostStd PostM/STD
w1 1.718656 0.252336 6.810983
...
```

With the `plot` command on the fitted object, **bsamGP** provides a density plot of residuals for diagnostics as well as a fitted mean curve with point-wise 95% credible intervals, as shown in Figure 3.

```
R> fit.dpm <- fitted(fout.dpm, HPD = TRUE)
R> plot(fit.dpm)
```

### 3.2. The BSAR model with shape restrictions: Benchmark data sets

In this subsection, we illustrate how to fit the BSAR, BSAQ, and GBSAR models with shape restrictions using benchmark data sets based on the shape-constraint approaches described in Sections 2.2–2.4. As in the BSAR illustration in Section 3.1, the major function in the **bsamGP** package is `bsar`. The option of no shape constraint is specified as `shape = "Free"`, and the shape-constrained regression models are fitted by specifying the `shape` argument to impose various types of shape restrictions, as follows:

```
bsar(formula, xmin, xmax, nbasis, nint, mcmc = list(), prior = list(),
      shape = c("Free", "Increasing", "Decreasing", "IncreasingConvex",
                "DecreasingConcave", "IncreasingConcave", "DecreasingConvex",
                "IncreasingS", "DecreasingS", "IncreasingRotatedS",
                "DecreasingRotatedS", "InvertedU", "Ushape"),
      marginal.likelihood = TRUE, spm.adequacy = FALSE)
```

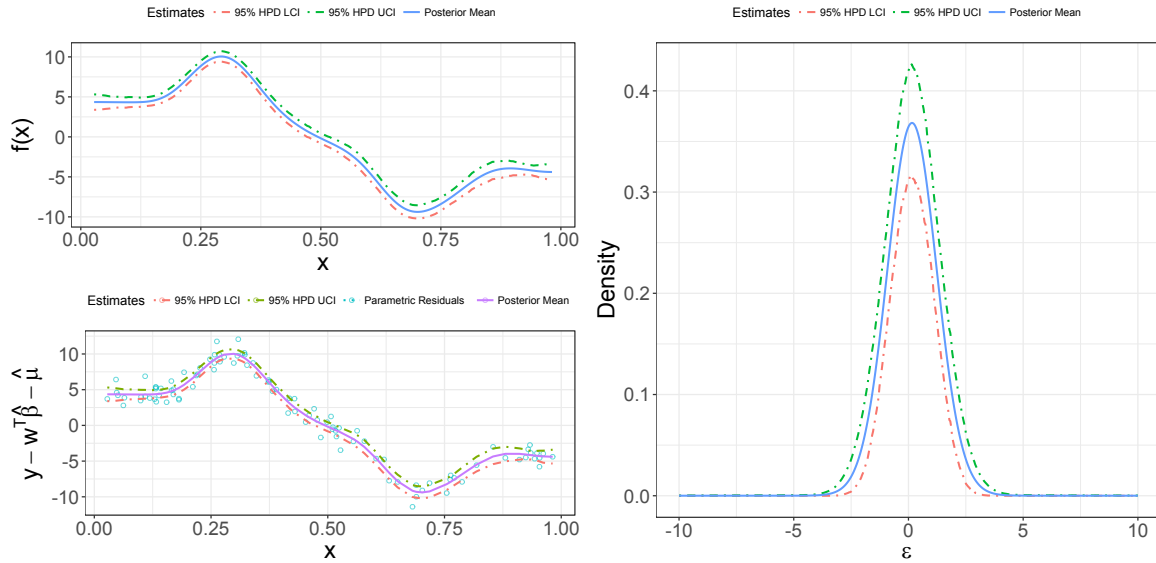


Figure 3: Left panel: The fitted regression curves produced by the `plot` method for ‘`bsamdpm`’ objects. Right panel: The density plot of residuals for diagnostics produced by the `plot` method for ‘`bsamdpm`’ objects.

More detailed explanations and demonstrations of the **bsamGP** package for shape-constrained regressions are presented using benchmark data sets for BSAR, BSAQ, and GBSAR models in the remainder of this subsection.

#### *Electricity demand data: BSAR with shape restriction*

To illustrate the usage of the argument `shape` for the BSAR with shape restrictions, we first consider a data set on demand for electricity, which consists of 288 quarterly observations in Ontario for the period from 1971 to 1994, taken from [Yatchew \(2003\)](https://www.economics.utoronto.ca/yatchew/) and available from <https://www.economics.utoronto.ca/yatchew/>. This data set has been analyzed using a cubic spline model (Engle, Granger, Rice, and Weiss 1986), a partial linear model (Yatchew 2003), and shape-restricted regression models (Lenk and Choi 2017), among others. The R package **bsamGP** includes the data set `Elec.demand`, which contains six variables: electricity demand (`enerm`), gross domestic product (`gdp`), price of electricity (`pelec`), price of natural gas (`pgas`), and the number of heating (`hddqm`) and cooling (`cddqm`) degree days relative to a reference temperature. We use `Elec.demand` to show how the results in Lenk and Choi (2017) can be reproduced.

```
R> data("Elec.demand", package = "bsamGP")
R> attach(Elec.demand)
R> y <- log(enerm / gdp)
R> w <- log(pelec / pgas)
R> x <- cddqm - hddqm
R> xmin <- min(x)
R> xmax <- max(x)
R> xmid <- (xmax + xmin) / 2
```



```
R> xrange <- xmax - xmin
```

Following the approach of [Lenk and Choi \(2017\)](#), we fit a model to `Elec.demand` using the `bsar` function in `bsamGP` with shape restrictions. We use the logarithm of the ratio of electricity demand to gross domestic product as the response variable and the log price ratio of electricity to natural gas and temperature, which is the number of heating and cooling degree days relative to 68°F, as the linear and nonlinear fixed effects, respectively. For the fixed effect of temperature, we enforce shape restrictions on its nonlinear regression function.

To fit the BSAR model with a shape restriction, we use the same hyperparameters as those in the BSAR model without shape restrictions in [Section 3.1](#) with the exception of the additional hyperparameters used in the nonparametric component with shape restrictions. For the hyperparameters of the nonparametric component, we fix  $v_{\theta_0}^2 = 100^2$ ,  $(r_{0,\tau}/s_{0,\tau}) = 1$ ,  $(2r_{0,\tau}/s_{0,\tau}^2) = 100$ ,  $\omega_0 = 2$ ,  $m_{0,\alpha} = 0$ , and  $v_{0,\alpha}^2 = 50^2$ . For the variance parameter, we choose  $(r_{0,\sigma}/s_{0,\sigma}) = 1$  and  $(2r_{0,\sigma}/s_{0,\sigma}^2) = 1000$ . Finally, we use  $m_{0,\psi} = 1000$ ,  $v_{0,\psi}^2 = 10000$ ,  $m_{0,\omega} = Q_2$ , and  $v_{0,\omega}^2 = R/8$  for hyperparameters of the inflection point  $\omega$  and the slope  $\psi$ . Here,  $Q_2$  is the midpoint and  $R$  is the range of  $x$  ([Lenk and Choi 2017](#)). Further, to identify satisfactory variances of the adaptive Metropolis algorithm, we initialize a Markov chain with 10,000 samples and then obtain 5,000 samples of MCMC simulation, which are thinned from the original MCMC sample by a factor of 10% after a burn-in period of 10,000 samples ([Lenk and Choi 2017](#)).

```
R> prior <- list(theta0_m0 = 0, theta0_s0 = 100, tau2_m0 = 1, tau2_v0 = 100,
+   w0 = 2, beta_m0 = numeric(2), beta_v0 = diag(100, 2), sigma2_m0 = 1,
+   sigma2_v0 = 1000, alpha_m0 = 3, alpha_s0 = 50, iflagpsi = 1,
+   psifixed = 1000, omega_m0 = xmid, omega_s0 = xrange / 8)
R> mcmc <- list(nblow0 = 10000, nblow = 10000, nskip = 10, smcmc = 5000)
```

The following R code illustrates how the BSAR models are fitted with a shape restriction using the `bsar` function and how the posterior inferences are subsequently extracted from these models. We estimate the *monotone decreasing* regression function of temperature, specified in the argument `shape`, and use 50 basis functions for the spectral representation.

```
R> fout1 <- bsar(y ~ w + fs(x), nbasis = 50, mcmc = mcmc, prior = prior,
+   shape = "Decreasing", marginal.likelihood = TRUE, verbose = TRUE)
R> fit1 <- fitted(fout1)
R> plot(fit1)
R> summary(fout1)
```

Note that we specify the `verbose` option to be active, `verbose = TRUE`, so that the `bsar` function displays its running procedure with progress bars, acceptance rates in adaptation, burn-in periods, and sampling phases. The output of `fout1` given by the `summary` method provides the log-integrated likelihood (LIL) and the fit statistics that are summarized in [Table 1](#) as well as the posterior estimates of all unknown parameters.

[Figure 4](#) shows the parametric residuals  $\hat{y}_i - \hat{\beta}^\top \mathbf{w}_i$  and the posterior mean estimates  $\hat{f}$  versus temperature (number of cooling or heating degree days) for the BSAR with a monotone decreasing restriction, obtained from the `plot` command, as in the previous R code. The plots

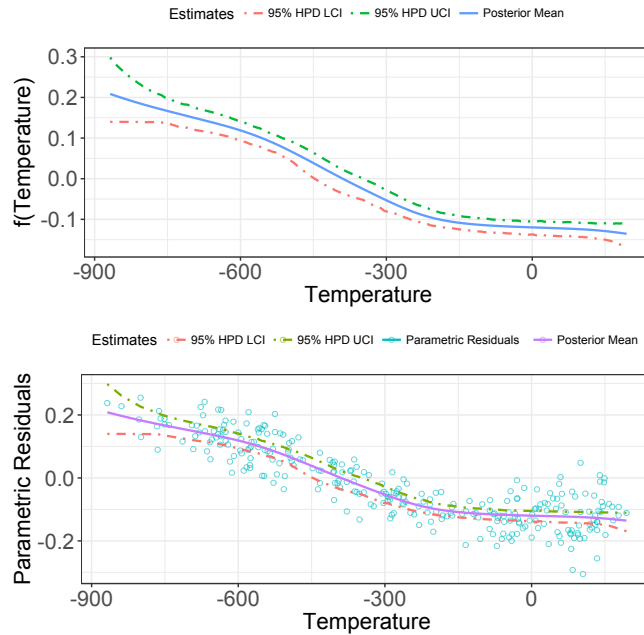


Figure 4: Estimated monotone decreasing regression curves for electric demand data: posterior mean estimates (blue and purple solid lines), point-wise 95% HPD intervals (in green and red dash-dotted lines), and parametric residuals (blue dots).

	Decreasing	DecreasingConvex	DecreasingS
LIL	237.1036	164.1054	63.8932
R-Square	0.8063	0.7921	0.8084
Error Standard Deviation	0.0999	0.1000	0.0981
Intercept	-1.5803	-1.5786	-1.5785
Log Price Ratio	-0.0749	-0.0710	-0.0746

Table 1: Posterior summaries for electricity demand data.

also include the point-wise 95% highest posterior density (HPD) intervals for the estimated model.

In addition to the restriction of monotonicity, using the same data set, we consider two other types of shape restrictions, namely convexity and an S-shape, by specifying the argument `shape`, as required. The following R code illustrates how the BSAR models are fitted with these shape restrictions.

```
R> fout2 <- bsar(y ~ w + fs(x), nbasis = 50, mcmc = mcmc, prior = prior,
+   shape = "DecreasingConvex", marginal.likelihood = TRUE)
R> fout3 <- bsar(y ~ w + fs(x), nbasis = 50, mcmc = mcmc, prior = prior,
+   shape = "DecreasingS", marginal.likelihood = TRUE)
```

We summarize the fit statistics and estimated parameters of all three models in Table 1. The monotone decreasing model has a marginally better LIL than those of the model with convexity and the S-shaped model, as discussed in [Lenk and Choi \(2017\)](#). All of the models

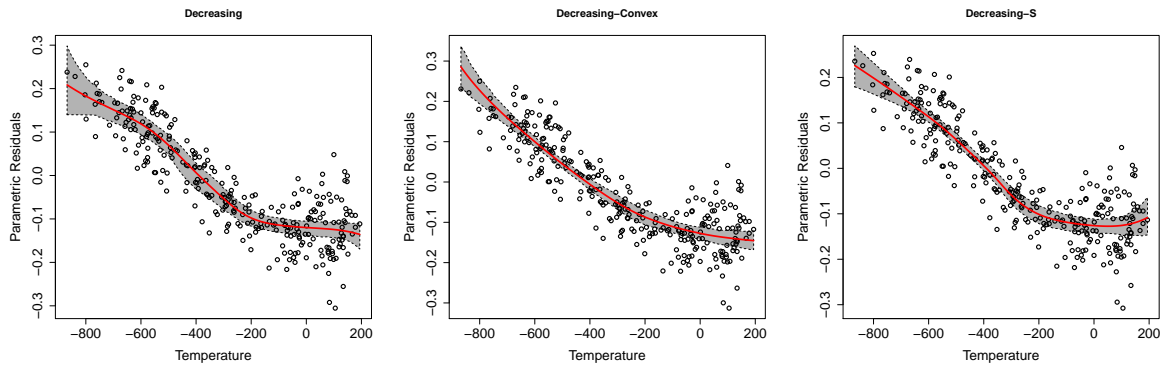


Figure 5: Estimated mean regression curves of all models for electric demand data: posterior mean estimates (red solid line), point-wise 95% HPD intervals (in the gray area), and parametric residuals (black dots).

indicate that demand decreases as the price of electricity increases relative to the price of natural gas, as shown in Figure 5.

#### *Micronutrient blood plasma data: BSAQ with shape restrictions*

As a real data example of the BSAQ model with shape restrictions, we consider data on  $n = 314$  healthy patients from a study of micronutrient levels in blood plasma (Nierenberg, Stukel, Baron, Dain, and Greenberg 1989), available from [http://lib.stat.cmu.edu/datasets/Plasma\\_Retinol](http://lib.stat.cmu.edu/datasets/Plasma_Retinol). This data set was originally used to investigate the relationship between personal characteristics and plasma concentrations of micronutrients such as retinol or beta-carotene. Recently, Meyer *et al.* (2011) analyzed the same data set for the effects of age, body mass index (BMI), and smoking status on the log of blood plasma beta-carotene using semiparametric additive normal regression models. They considered a shape-constrained regression model to fit the data, with a decreasing function of BMI and an increasing function of age, based on shape-restricted splines.

The R package **bsamGP** includes the data set `plasma`, which contains 14 variables: 12 personal characteristics (age, sex, smoking status, etc.) and two variables (beta-carotene and retinol) for the plasma concentrations of micronutrients. For a more detailed description of the data, see Nierenberg *et al.* (1989). Following the work of Meyer *et al.* (2011), we focus on the relationship between the log of blood plasma beta-carotene (`betaplasm`) and smoking status (current or not), `age`, and body mass index (`bmi`), which is defined as  $\text{weight}/\text{height}^2$  in the units  $kg/m^2$ . Here, we are interested in quantile regression functions rather than the mean regression of Meyer *et al.* (2011), and, thus, we apply the BSAQ by imposing the same shape restrictions as Meyer *et al.* (2011) for illustration purposes. Specifically, in order to evaluate the effects of age, BMI, and smoking status on the blood plasma concentration of beta-carotene, we consider the `bsaq` function, estimating the model given by

$$\log(\text{betaplasm}) = \beta_0 + \beta \text{smoke} + f_1(\text{bmi}) + f_2(\text{age}) + \epsilon, \quad \epsilon \sim \mathcal{AL}_p(\epsilon; \sigma^2), \quad (19)$$

where `smoke` denotes a design matrix containing a dummy variable for smoking status (“current smoker” or “not”), and the functions  $f_1$  and  $f_2$  are assumed to be decreasing and increasing, respectively.

To fit the BSAQ model (19), we use the hyperparameters  $m_{0,\beta} = 0$ ,  $V_{0,\beta} = 100I$ ,  $v_{\theta_0}^2 = 100^2$ ,  $(r_{0,\tau}/s_{0,\tau}) = 1$ ,  $(2r_{0,\tau}/s_{0,\tau}^2) = 100$ ,  $\omega_0 = 2$ ,  $(r_{0,\sigma}/s_{0,\sigma}) = 1$ , and  $(2r_{0,\sigma}/s_{0,\sigma}^2) = 1000$  and an MCMC simulation of length 60,000 with a burn-in period of 10,000 samples, and final posterior samples with a size of 5,000 at every tenth iteration, both of which are the same as before in `bsar`. The following R code illustrates how to fit the additive BSAQ model with shape restrictions for the `plasma` data with the specification of prior distributions and MCMC parameters in the `bsaq` function.

```
R> data("plasma", package = "bsamGP")
R> attach(plasma)
R> prior <- list(theta0_m0 = 0, theta0_s0 = 100, tau2_m0 = 1, tau2_v0 = 100,
+   w0 = 2, beta_m0 = numeric(2), beta_v0 = diag(100, 2), sigma2_m0 = 1,
+   sigma2_v0 = 1000)
R> mcmc <- list(nblow0 = 10000, nblow = 10000, nskip = 10, smcmc = 5000)
R> foutBSAQ <- bsaq(log(betaplasma) ~ I(smoke == "Current_Smoker") +
+   fs(bmi) + fs(age), p = 0.5, nbasis = 50, mcmc = mcmc, prior = prior,
+   shape = c("Decreasing", "Increasing"))
R> fitBSAQ <- fitted(foutBSAQ)
R> plot(fitBSAQ)
```

Note that the predictors `bmi` and `age` are used for two nonparametric components and that `bsaq` automatically fits two-component additive nonparametric median regression functions using the argument `p = 0.5` with the shape restriction using the argument `shape = c("Decreasing", "Increasing")`. That is,  $f_1(\text{bmi})$  is a decreasing function of `bmi`, and  $f_2(\text{age})$  is an increasing function of `age`, as shown in Figure 6. The left panel of Figure 6 shows the relationship between `log(betaplasma)` and `bmi`, indicating an overall decreasing trend, whereas the right panel shows that the relationship between `log(betaplasma)` and `age` is increasing.

Other quantile regression functions can be considered by specifying the argument `p` with different quantile levels of interest in the function `bsaq`. As an illustration, we provide the quantile regression functions of `bmi` and `age` with shape restrictions using BSAQ models at  $p = \{0.1, 0.5, 0.9\}$  (see Figure 7). In each panel, the black line denotes the quantile regression function for current non-smokers, and the red line indicates the function for current smokers. All of the quantile regression function estimates for `bmi` and `log(betaplasma)` show overall negative trends, whereas the estimates for `age` and `log(betaplasma)` demonstrate increasing relationships, as enforced by the argument `shape` in `bsaq`. In both relationships, the level of `log(betaplasma)` for current non-smokers is higher than that for current smokers, which indicates that the level of `betaplasma` is influenced by the smoking status.

Alternatively, the data set `plasma` can be fitted based on the assumption of an unknown error distribution using the function `bsaqdpm` with a Dirichlet process scale mixture model for BSAQ, as described in Section 2.3. The following R code illustrates how to fit the additive BSAQ model with shape restrictions using `bsaqdpm` for `plasma` data. For an object returned by `bsaqdpm`, the `plot` method provides estimated quantile curves as well as an estimated density plot of residuals for diagnostics, as shown in Figure 8.

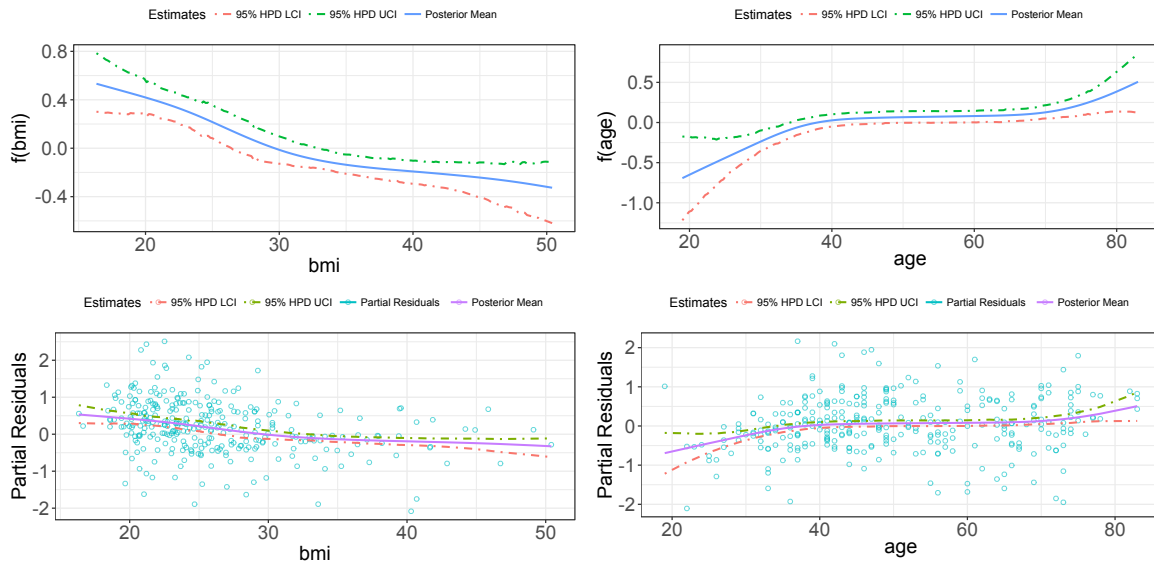


Figure 6: Posterior mean estimates for the effect of bmi and age on  $\log(\text{betaplasma})$  levels. The left panel shows the relationship between  $\log(\text{betaplasma})$  and bmi, and the right panel displays the relationship between  $\log(\text{betaplasma})$  and age.

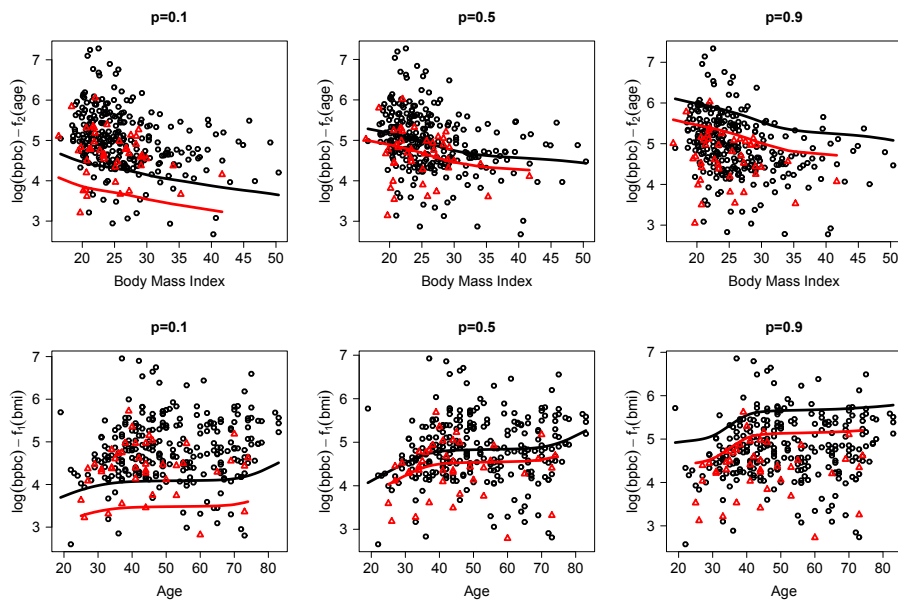


Figure 7: Posterior mean estimates for the quantile regression function of bmi and age on  $\log(\text{betaplasma})$  levels. The top panels show the relationship between  $\log(\text{betaplasma})$  and bmi, and the bottom panels display the relationship between  $\log(\text{betaplasma})$  and age.

```
R> prior <- append(prior, list(kappa_r0 = 1, kappa_s0 = 100, tmass_a = 2,
+   tmass_b = 4))
R> mcmc <- list(nblow = 20000, nskip = 10, smcmc = 5000)
```

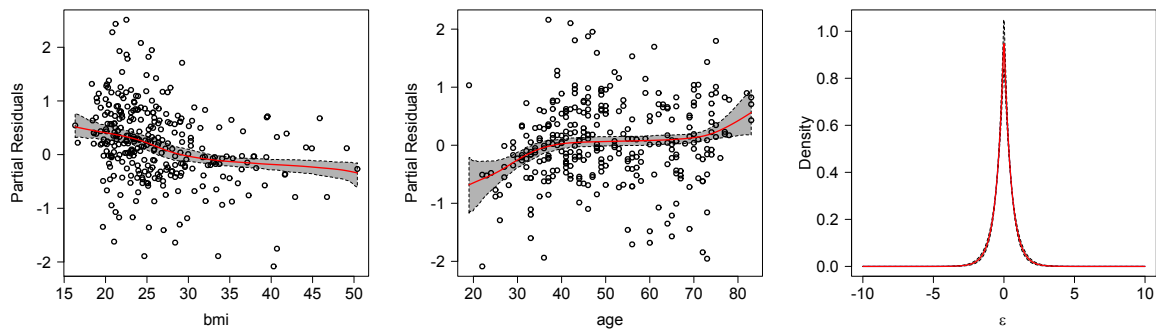


Figure 8: The left and middle panels display the posterior mean estimates (red line) and 95% HPD interval (gray shaded area) for the effect of `bmi` and `age` on `log(betaplasma)` levels with scatter plots of their parametric residuals. The right panel shows an estimated density function of these residuals.

```
R> foutBSAQ.dpm <- bsaqdpm(log(betaplasma) ~ I(smoke == "Current_Smoker") +
+   fs(bmi) + fs(age), p = 0.5, nbasis = 50, mcmc = mcmc, prior = prior,
+   shape = c("Decreasing", "Increasing"))
R> fitBSAQ.dpm <- fitted(foutBSAQ.dpm)
```

#### *Wage-union membership data: GBSAR binary regression with shape restrictions*

In order to illustrate the GBSAR model for binary responses with shape restrictions, we consider the data on wages and union membership of [Berndt \(1991\)](#), as discussed in [Ruppert, Wand, and Carroll \(2003\)](#) and [Crainiceanu, Ruppert, and Wand \(2005\)](#), among others. The data consists of a random sample of 534 US workers on 11 variables from the 1985 Current Population Survey (CPS), available from [http://lib.stat.cmu.edu/datasets/CPS\\_85\\_Wages](http://lib.stat.cmu.edu/datasets/CPS_85_Wages) and also from the R package **SemiPar** ([Wand 2018](#)).

The R package **bsamGP** includes the data set `wage.union`, which has one dependent variable (an indicator of union membership, `union`), and ten predictor variables, including six dummy variables for southern residence (`south`), `race`, occupational status (`occupation`), `sector`, `sex`, and marital status (`married`), and four continuous variables for `experience`, `wages`, `age`, and `education`. [Ruppert et al. \(2003\)](#) and [Crainiceanu et al. \(2005\)](#) analyzed the wage-union membership data using a semiparametric additive logistic regression model with splines because the standard linear, quadratic and cubic regressions were not adequate for identifying the features of union membership probability. They used `wages`, `age`, `education`, `race`, `gender`, and `south` as predictor variables and modeled the first three predictors (`wages`, `age`, and `education`) with penalized splines. They found that the effect of `wages` appears to be non-monotonic and that the probability of union membership tends to increase with `age` and decrease with `education`. Here, we re-analyze the same data set with the generalized semiparametric additive logistic regression, using the `gbsar` function of **bsamGP**. We assume that `age` and `education` are monotonically related to the probability of union membership. The R code illustrates how to fit the GBSAR model with a logit link for the binary data using the arguments `family = "bernoulli"`, `link = "logit"` and the default prior distributions in `gbsar`.



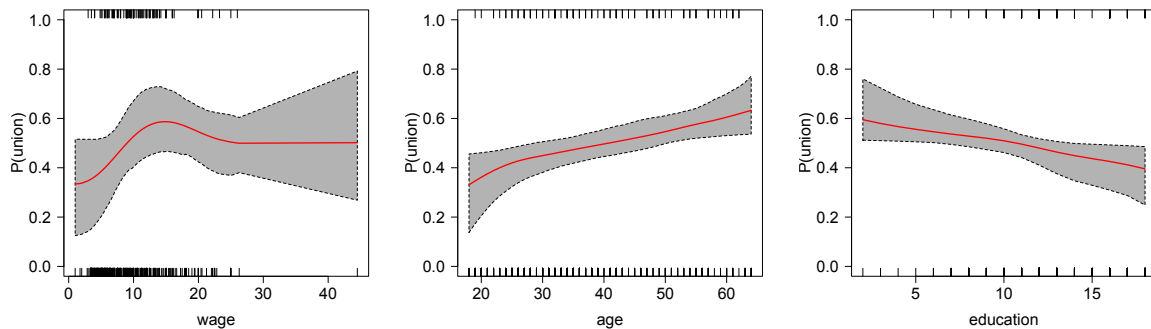


Figure 9: Posterior mean estimates and point-wise 95% credible intervals for the effect of `wage`, `age`, and `education` on the estimated probability of union membership using a logit link.

```
R> data("wage.union", package = "bsamGP")
R> attach(wage.union)
R> race <- ifelse(race == 3, 1, 0)
R> mcmc <- list(nblow0 = 10000, nblow = 10000, nskip = 10, smcmc = 5000)
R> foutLogit <- gbsar(union ~ race + sex + south + fs(wage) + fs(age) +
+   fs(education), family = "bernoulli", link = "logit", mcmc = mcmc,
+   nbasis = 50, shape = c("Free", "Decreasing", "Increasing"))
R> fitLogit <- fitted(foutLogit, HPD = FALSE)
```

Note that the `gbsar` function for the logistic regression uses adaptive Metropolis algorithms (Haario *et al.* 2001; Atchadé and Rosenthal 2005; Roberts and Rosenthal 2009) to generate posterior samples *by default*. To draw posterior samples from the algorithm of Holmes and Held (2006), the user should utilize the `gbsar` function with the option `algorithm = "KS"`.

The analysis results using `bsamGP` are shown in Figure 9. The plots show the effects of the three predictors (`wage`, `age` and `education`) modeled nonparametrically. The Y-axes are the estimated probabilities of union membership. Each plot displays the posterior mean estimates and 95% point-wise credible intervals. The non-monotonic behavior of the probability of union membership is captured similarly to that found in Ruppert *et al.* (2003) and Crainiceanu *et al.* (2005). In addition, the monotonic (increasing with `age` and decreasing with `education`) relationship between the probability of union membership and `age` and `education` is captured reasonably well, as shown in Figure 9. In particular, the probability of union membership nearly linearly increases with `age` and decreases with `education`.

Note that the same data can also be fitted with a probit regression using the `gbsar` function by replacing the argument `link = "logit"` with `link = "probit"`, as in the R code below.

```
R> foutPbt <- gbsar(union ~ race + sex + south + fs(wage) + fs(education) +
+   fs(age), nbasis = 50, family = "bernoulli", link = "probit",
+   shape = c("Free", "Decreasing", "Increasing"))
```

*Mortality-temperature data: GBSAR count response regression with shape restrictions*

The health effects of exposure to extreme temperatures have been frequently investigated, and many epidemiological studies (e.g., Curriero, Heiner, Samet, Zeger, Strug, and Patz 2002; Anderson and Bell 2009; Gasparrini *et al.* 2015) have confirmed that when temperatures are extremely high or low, there is a significant association with increased mortality. Furthermore, the mortality rate has an exposure-response curve with respect to temperature with a specific shape. Consequently, we illustrate the shape-restricted semiparametric count response regression models for Poisson and negative binomial data using the arguments `family` and `link` in `gbsar` with a benchmark data set on the mortality-temperature relationship.

The data set is based on real applications from Gasparrini *et al.* (2015) for assessing the effect of mean temperature on mortality. In particular, the data include regional daily death occurrences in England and Wales in the United Kingdom (UK), which are publicly available from the Office for National Statistics, and daily mean temperature estimated from minimum and maximum temperatures, which are available from the British Atmospheric Data Centre. These data are described in detail in related epidemiological studies and recent publications (see, e.g., Armstrong *et al.* 2011; Gasparrini, Armstrong, Kovats, and Wilkinson 2012; Gasparrini *et al.* 2015). Specifically, the data set is part of that collected in the UK and used for a multi-country observation study in Gasparrini *et al.* (2015) based on a first-stage time-series quasi-Poisson regression model with B-splines. Details on the data set are given in the supplementary appendix of Gasparrini *et al.* (2015), and the data set can be obtained from the supplementary web page at [https://github.com/gasparrini/2015\\_gasparrini\\_Lancet\\_Rcodedata](https://github.com/gasparrini/2015_gasparrini_Lancet_Rcodedata).

In order to illustrate `gbsar`, we use 5,113 daily observations from London for the period from 1993 to 2006, included as the data set `London.Mortality` in **bsamGP**. The data set `London.Mortality` consists of 5,113 observations on the number of deaths (`death`), mean daily temperature (`tmean` in °C), and other related variables, such as the relative humidity, maximum/minimum temperatures, and others (`rh`, `tmax`, `tmin`, `date`, `dewp`). Here, for simplicity, we seek only the effect of temperature on the mortality rate in order to demonstrate the `gbsar` function. We use a Poisson semiparametric regression model with a count response variable `death` and a continuous predictor `tmean`, and compare the fitted results, with or without appropriate shape restrictions. In particular, the U-shaped mortality-temperature relationship (e.g., Muggeo 2008; Anderson and Bell 2009; Muggeo 2010) can be identified and tested using the argument `shape` in the `gbsar` function.

The following R code illustrates the use of `gbsar()` in the **bsamGP** package for a Poisson semiparametric regression with shape restrictions. We specify the Poisson family with the log link function, with and without shape restrictions. We set the number of basis functions to 50 and use the default hyperparameters, as before. Additional hyperparameters for  $\omega$  and  $\psi$  in the U-shaped restriction are specified *by default* as  $m_{0,\psi} = 100$ ,  $v_{0,\psi}^2 = 100^2$ ,  $m_{0,\omega} = Q_2$ , and  $v_{0,\omega}^2 = R/8$ , where  $Q_2$  is the midpoint and  $R$  is the range of `tmean`, respectively. Note that we set the adaptation iterations to 10,000 (only applicable for shape restricted cases) and obtain 1,000 MCMC samples at every ten steps after 10,000 burn-in iterations.

```
R> data("London.Mortality", package = "bsamGP")
R> mcmc <- list(nblow0 = 10000, blow = 10000, nskip = 10, smcmc = 1000,
+             maxmodmet = 10)
R> fout.free <- with(London.Mortality, gbsar(death ~ fs(tmean), mcmc = mcmc,
```

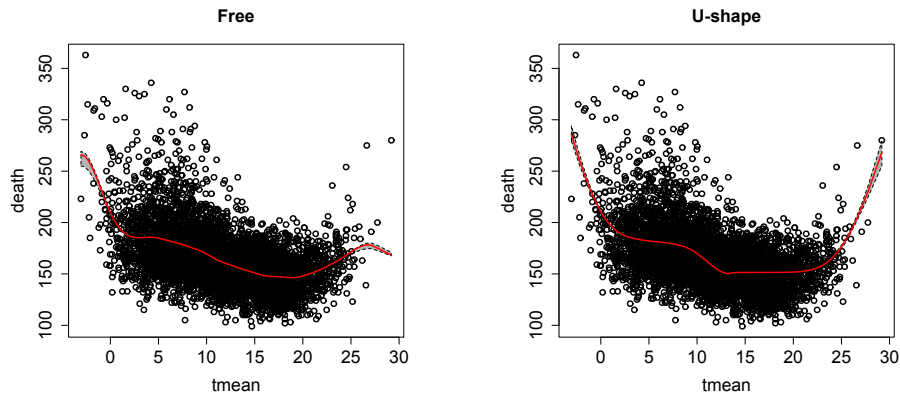


Figure 10: Posterior estimates obtained from the `gbsar` function for the London mortality data. The black dots represent daily death counts for given temperatures, the red line represents the estimated mean of daily death counts, and the gray shaded area represents the 95% HPD fitted from the GBSAR model.

```
+ family = "poisson", link = "log", nbasis = 50, shape = "Free"))
R> fit.free <- fitted(fout.free)
R> fout.U <- with(London.Mortality, gbsar(death ~ fs(tmean), mcmc = mcmc,
+ family = "poisson", link = "log", nbasis = 50, shape = "Ushape"))
R> fit.U <- fitted(fout.U)
```

Figure 10 displays the fitted mean of daily death occurrences superimposed with the number of deaths obtained from the GBSAR Poisson regression with and without a shape restriction. As shown in Figure 10, both unrestricted and U-shaped exposure-response curves identify the nonlinear exposure-response curves in the association between temperature with increased mortality, indicating that high and low temperatures have an adverse effect on mortality.

The adequacy of the shape restrictions can be tested by computing the marginal likelihoods. The marginal likelihoods (LIL) given in the abridged output of `fout.free` and `fout.U` below show that the unrestricted Poisson semiparametric regression has a larger LIL and would be favored over that with a U-shaped restriction by the observed data set, in terms of model selection. However, note that the U-shaped model is not always correctly identified by the marginal likelihood and that when the two models are consistent with the true function, such as the unrestricted model *versus* the U-shaped model with additional parameters, as in the current exposure-response curve, the marginal likelihood may prefer a simpler model, as discussed in [Lenk and Choi \(2017\)](#). Further, the mortality rate from the shape-restricted model clearly shows the U-shaped exposure-response curve estimates and confirms the empirical findings on the shape of the curve in previous mortality-temperature studies (e.g., [Curriero et al. 2002](#); [Muggeo 2008](#); [Anderson and Bell 2009](#)). In addition, an estimate of  $\omega$ , the abscissa of the minimum of the U-shaped curve, is provided as approximately 13.1°C, the temperature that minimizes mortality, as shown in the abridged output of `fout.U`.

```
R> summary(fout.free)
...
Log Integrated Likelihood
```

```
LIL Gelfand & Dey = -26911.14
LIL Newton & Raftery (biased) = -26724.28
...
```

```
R> summary(fout.U)
```

```
...
Log Integrated Likelihood
LIL Gelfand & Dey = -26965.14
LIL Newton & Raftery (biased) = -26803.36
...
omega is inflection point of squish function
Posterior mean omega = 13.12218
Posterior stdev omega = 0.0179773
...
```

In addition, we check the adequacy of the Poisson regression *versus* the negative binomial regression with a U-shaped restriction for the data set by computing the marginal likelihoods for model selection. To fit the negative binomial regression, we specify the arguments in `gbsar` as `family = "negative.binomial"` and `link = "log"`, as illustrated in the following R code and abridged output:

```
R> fout.U.nb <- gbsar(death ~ fs(tmean), family = "negative.binomial",
+   link = "log", nbasis = 50, shape = "Ushape", mcmc = mcmc,
+   prior = prior)
R> fit.U.nb = fitted(fout.U.nb)
R> summary(fout.U.nb)
```

```
...
Log Integrated Likelihood
LIL Gelfand & Dey = -23668.52
LIL Newton & Raftery (biased) = -23434.95
...
omega is inflection point of squish function
Posterior mean omega = 13.95253
Posterior stdev omega = 0.08368391
...
```

Table 2 summarizes the computed marginal likelihoods of the two candidate models (U-shaped Poisson, and U-shaped negative binomial). Based on the results in Table 2, the U-shaped negative binomial semiparametric regression model provides the better fit for the data set than the U-shaped Poisson model.

Note that the data set was originally used in statistical analyses based on the over-dispersed Poisson distribution (Gasparri *et al.* 2012) and the time-series quasi-Poisson model (Gasparri *et al.* 2015) associated with different time lags. However, the current analysis considers only the mean temperature as a predictor of the mortality rate in order to illustrate the

	U-shape Poisson	U-shape negative binomial
LIL (Gelfand & Dey)	-26965.14	- <b>23668.52</b>
LIL (Newton & Raftery)	-26803.36	- <b>23434.95</b>

Table 2: LIL of London mortality data.

`gbsar` function of **bsamGP**. Thus, the negative binomial distribution could be favored over the Poisson distribution for the observed data, which is probably because over-dispersion is better explained by the negative binomial distribution.

Here, we use a *default* prior specification in the negative binomial regression with an additional hyperparameter, that is, the dispersion parameter  $\kappa$  with the shape parameter  $r_{0,\kappa}$  and the rate parameter  $s_{0,\kappa}$ , to be weakly informative by setting a prior mean of 1 and a prior variance of 100. Note that the negative binomial distribution can be represented as a Poisson-gamma mixture and that **bsamGP** also implements the mixture approach to the negative binomial regression by specifying `family = "poisson.gamma"`.

### 3.3. Density estimation with the BSAD model

In addition to regression modeling, the BSAM provides a semiparametric density estimation procedure for the BSAD model, as described in Section 2.4. Here, we explain how to use the **bsamGP** package for the BSAD model as the final illustration of **bsamGP** using synthetic data. The data used in this analysis are based on a random sample of size  $n = 500$  from a semiparametric density  $f(y) \propto \exp\{h_1(y)\beta_1 + h_2(y)\beta_2 + z(y)\}$ . Here, the parametric component has a truncated gamma distribution with  $h_1(y) = \log(y)$ ,  $\beta_1 = 3$ ,  $h_2(y) = y$ ,  $\beta_2 = -1$ , defined on  $(0, 10)$ , and the nonparametric component has a sigmoid function  $z(y) = \tanh(y - 5)$ . In order to generate samples from the semiparametric model, we utilize rejection sampling with a uniform envelope on  $(0, 10)$ . The following R code illustrates the process of generating the synthetic data from the semiparametric model.

```
R> set.seed(1)
R> nSample <- 500
R> a <- 0
R> b <- 10
R> zFunc <- function(x, center) tanh(x - center)
R> paraFunc <- function(x, a) drop(3 * log(x - a) - x)
R> semiFunc <- function(x, a, b) exp(paraFunc(x, a) + zFunc(x, (b - a) / 2))
R> InvAccRate <- optimize(function(x) semiFunc(x, a, b) / dunif(x, a, b),
+   maximum = TRUE, interval = c(a, b))$objective
R> u <- runif(nSample * InvAccRate)
R> y <- runif(nSample * InvAccRate, a, b)
R> x <- y[u < semiFunc(y, a, b) / (InvAccRate * dunif(y, a, b))][1:nSample]
```

To fit the data, we use the geometric smoother prior and generate 1,000 posterior samples with 20,000 burn-in iterations and a thinning of five from the MCMC algorithm of the `bsad` function. In estimating the semiparametric model, we set the maximum number of basis functions to 98 and the number of grid points to 101. The R command is given as follows:

```
R> mcmc <- list(nblow = 20000, nskip = 5, smcmc = 1000)
```

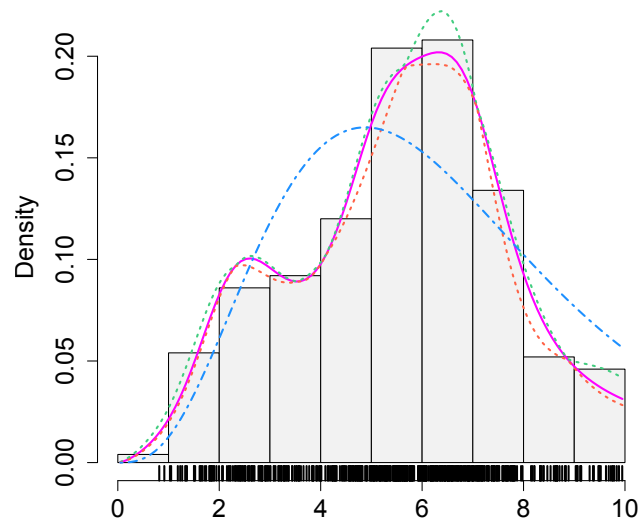


Figure 11: Posterior estimates of the `bsad` function for the simulated data generated from the semiparametric model, overlaid with a histogram of the data. The dashed blue line denotes the parametric model and the solid magenta line represents the posterior mean of the semiparametric model with dotted green and orange lines representing the 95% equal-tail credible intervals.

```
R> fout <- bsad(x, xmin = a, xmax = b, nint = 101, MaxNCos = 98,
+   mcmc = mcmc, smoother = "geometric", parametric = "gamma")
```

The estimated density curve based on the BSAD model can be drawn with the `plot` method as displayed in Figure 11.

```
R> fit <- fitted(fout, HPD = FALSE)
R> plot(fit)
```

The posterior summary is also produced for the object returned by the `bsad` function using the `summary` command. The code and partial output of the posterior probability of each model and the natural logarithm of the marginal likelihoods are given as follows:

```
R> summary(fout)

...
Posterior Probabilities of Parametric vs Semi-parametric with kappa
  Para SemiPara
    0      1
Ln Marginal Distribution of the Data
  Para      Semi      SemiMK
-1062.375 -1037.290 -1040.604
...
```

As shown in Figure 11 and by the marginal likelihood in the posterior summary, the semiparametric model is more appropriate for fitting the simulated data than the parametric model is and, thus, provides a reasonable estimate of the probability density function.



## 4. Conclusion

This article has presented a user-friendly new R package called **bsamGP** for fitting semiparametric models using Gaussian process priors. The **bsamGP** package provides easy access to fitting Bayesian semiparametric models for regression and density estimation. The **bsamGP** package utilizes spectral representations in order to simplify the posterior computation of Gaussian processes through the linearization of the covariance structure. **bsamGP** contains popular parametric linear regression models, semiparametric partial linear regression models, semiparametric density estimation models, semiparametric quantile regression models, and generalized semiparametric models. All semiparametric regression models in **bsamGP** allow for additive models and nonlinear regression functions with, for example, monotonic, monotonic convex or concave, S-shaped, and U-shaped constraints. **bsamGP** uses compiled Fortran 90 code to maximize computational efficiency. Thus, the implementation is fast and reliable, making it viable for various Bayesian semiparametric models with shape restrictions to be applied more widely in practice.

Our package can be expanded in several directions. The current version allows for normal prior distributions for the coefficients of the linear regression and two smoothing prior distributions (T-smoother and Lasso smoother) for the nonparametric function. Thus, it could be expanded to allow spike-and-slab prior distributions for variable selection in the linear regression as well as for the nonparametric function with shape restrictions, which improves the model fit when the true unknown function is on the boundary of the constraints for the shape restrictions, as discussed in Jo *et al.* (2016) and Lenk and Choi (2017). For longitudinal data or clustered data, semiparametric additive mixed effects models can be considered in the BSAM framework by incorporating random effects and measurement errors in the covariates (e.g., Choi, Jo, Park, and Lenk 2017), and such models can be extended to and implemented by the **bsamGP** package. Furthermore, a companion software application is being developed as either a Python-based (van Rossum *et al.* 2011) graphical user interface (GUI) or an interactive web application using **shiny** (Chang, Cheng, Allaire, Xie, and McPherson 2019) in the R package **bsamGP**. These applications will have the same functionality as that of **bsamGP** but will allow practitioners to use **bsamGP** easily by means of a GUI instead of a command-line user interface in R.

## Acknowledgments

Research of Seongil Jo was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2017R1D1A3B03035235). Research of Taeryon Choi was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2016R1D1A1B03932178).

## References

- Albert JH, Chib S (1993). “Bayesian Analysis of Binary and Polychotomous Response Data.” *Journal of the American Statistical Association*, **88**(422), 669–679. doi:10.1080/01621459.1993.10476321.

- Alhamzawi R, Yu K (2013). “Conjugate Priors and Variable Selection for Bayesian Quantile Regression.” *Computational Statistics & Data Analysis*, **64**, 209–219. doi:10.1016/j.csda.2012.01.014.
- Anderson BG, Bell ML (2009). “Weather-Related Mortality: How Heat, Cold, and Heat Waves Affect Mortality in the United States.” *Epidemiology*, **20**(2), 205–213. doi:10.1097/ede.0b013e318190ee08.
- Andrew D, Mallows C (1974). “Scale Mixtures of Normal Distributions.” *Journal of the Royal Statistical Society B*, **36**(1), 99–102. doi:10.1111/j.2517-6161.1974.tb00989.x.
- Armstrong BG, Chalabi Z, Fenn B, Hajat S, Kovats S, Milojevic A, Wilkinson P (2011). “Association of Mortality with High Temperatures in a Temperate Climate: England and Wales.” *Journal of Epidemiology & Community Health*, **65**(4), 340–345. doi:10.1136/jech.2009.093161.
- Atchadé YF, Rosenthal JS (2005). “On Adaptive Markov Chain Monte Carlo Algorithms.” *Bernoulli*, **11**(5), 815–828. doi:10.3150/bj/1130077595.
- Banerjee A, Dunson DB, Tokdar ST (2013). “Efficient Gaussian Process Regression for Large Datasets.” *Biometrika*, **100**(1), 75–89. doi:10.1093/biomet/ass068.
- Banerjee S, Gelfand AE, Finley AO, Sang H (2008). “Gaussian Predictive Process Models for Large Spatial Data Sets.” *Journal of Royal Statistical Society B*, **70**(4), 825–848. doi:10.1111/j.1467-9868.2008.00663.x.
- Berndt E (1991). *The Practice of Econometrics: Classical and Contemporary*. Addison-Wesley.
- Bornkamp B, Ickstadt K (2009). “Bayesian Nonparametric Estimation of Continuous Monotone Functions with Applications to Dose-Response Analysis.” *Biometrics*, **65**(1), 198–205. doi:10.1111/j.1541-0420.2008.01060.x.
- Bürkner PC (2017). “**brms**: An R Package for Bayesian Multilevel Models Using Stan.” *Journal of Statistical Software*, **80**(1), 1–28. doi:10.18637/jss.v080.i01.
- Carpenter B, Gelman A, Hoffman M, Lee D, Goodrich B, Betancourt M, Brubaker MA, Guo J, Li P, Riddell A (2016). “Stan: A Probabilistic Programming Language.” *Journal of Statistical Software*, **76**(1), 1–37. doi:10.18637/jss.v076.i01.
- Chang W, Cheng J, Allaire JJ, Xie Y, McPherson J (2019). **shiny**: Web Application Framework for R. R package version 1.3.2, URL <https://CRAN.R-project.org/package=shiny>.
- Choi T, Jo S, Park B, Lenk PJ (2017). “Bayesian Hierarchical Mixed Effects Models for Meta Analysis of Cadmium Toxicity Studies.” Unpublished manuscript.
- Congdon P (2006). *Bayesian Statistical Modelling*. Wiley Series in Probability and Statistics, 2nd edition. John Wiley & Sons.
- Congdon P (2014). *Applied Bayesian Modelling*. Wiley Series in Probability and Statistics, 2nd edition. John Wiley & Sons.

- Crainiceanu CM, Ruppert D, Wand MP (2005). “Bayesian Analysis for Penalized Spline Regression Using **WinBUGS**.” *Journal of Statistical Software*, **14**(14), 1–24. doi:10.18637/jss.v014.i14.
- Curriero FC, Heiner KS, Samet JM, Zeger SL, Strug L, Patz JA (2002). “Temperature and Mortality in 11 Cities of the Eastern United States.” *American Journal of Epidemiology*, **155**(1), 80–87. doi:10.1093/aje/155.1.80.
- Curtis SM, Ghosh SK (2011). “A Variable Selection Approach to Monotonic Regression with Bernstein Polynomials.” *Journal of Applied Statistics*, **38**(5), 961–976. doi:10.1080/02664761003692423.
- Damien P, Wakefield JC, Walker S (1999). “Gibbs Sampling for Bayesian Nonconjugate and Hierarchical Models Using Auxiliary Variables.” *Journal of the Royal Statistical Society B*, **61**(2), 331–344. doi:10.1111/1467-9868.00179.
- Devroye L (1986). *Nonuniform Random Variate Generation*. Springer-Verlag.
- Dey D, Ghosh SK, Mallick BK (2000). *Generalized Linear Models: A Bayesian Perspective*. Marcel Dekker. doi:10.1201/9781482293456.
- Duan JA, Guindani M, Gelfand AE (2007). “Generalized Spatial Dirichlet Process Models.” *Biometrika*, **94**(4), 809–825. doi:10.1093/biomet/asm071.
- Engle RF, Granger CWJ, Rice J, Weiss A (1986). “Semiparametric Estimates of the Relation Between Weather and Electricity Sales.” *Journal of the American Statistical Association*, **81**(394), 310–320. doi:10.2307/2289218.
- Fahrmeir L, Kneib T, Lang S, Marx B (2013). *Regression: Models, Methods and Applications*. Springer-Verlag.
- Ferguson TS (1973). “A Bayesian Analysis of Some Nonparametric Problems.” *The Annals of Statistics*, **1**(2), 209–230. doi:10.1214/aos/1176342360.
- Gasparrini A, Armstrong B, Kovats S, Wilkinson P (2012). “The Effect of High Temperatures on Cause-Specific Mortality in England and Wales.” *Occupational and Environmental Medicine*, **69**(1), 56–61. doi:10.1136/oem.2010.059782.
- Gasparrini A, Guo Y, Hashizume M, Lavigne E, Zanobetti A, Schwartz J, Tobias A, Tong S, Rocklöv J, Forsberg B, *et al.* (2015). “Mortality Risk Attributable to High and Low Ambient Temperature: A Multicountry Observational Study.” *The Lancet*, **386**(9991), 369–375. doi:10.1016/s0140-6736(14)62114-0.
- Gelfand AE, Dey DK (1994). “Bayesian Model Choice: Asymptotics and Exact Calculations.” *Journal of the Royal Statistical Society B*, **56**(3), 501–514. doi:10.1111/j.2517-6161.1994.tb01996.x.
- Gelfand AE, Kottas A, MacEachern SN (2005). “Bayesian Nonparametric Spatial Modeling with Dirichlet Process Mixing.” *Journal of the American Statistical Association*, **100**(471), 1021–1035. doi:10.1198/016214504000002078.

- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2013). *Bayesian Data Analysis*. 3rd edition. Chapman & Hall/CRC. doi:10.1201/b16018.
- Golchi S, Bingham DR, Chipman H, Campbell DA (2015). “Monotone Emulation of Computer Experiments.” *SIAM/ASA Journal on Uncertainty Quantification*, **3**(1), 370–392. doi:10.1137/140976741.
- Gramacy R (2016). “**laGP**: Large-Scale Spatial Modeling via Local Approximate Gaussian Processes in R.” *Journal of Statistical Software*, **72**(1), 1–46. doi:10.18637/jss.v072.i01.
- Gramacy RB (2007). “**tgp**: An R Package for Bayesian Nonstationary, Semiparametric Non-linear Regression and Design by Treed Gaussian Process Models.” *Journal of Statistical Software*, **19**(9), 1–46. doi:10.18637/jss.v019.i09.
- Greenwood M, Yule U (1920). “An Inquiry into the Nature of Frequency Distributions Representative of Multiple Happenings with Particular Reference to the Occurrence of Multiple Attacks of Disease or of Repeated Accidents.” *Journal of the Royal Statistical Society*, **83**(2), 255–270. doi:10.2307/2341080.
- Grenander U (1981). *Abstract Inference*. John Wiley & Sons.
- Haario H, Saksman E, Tamminen J (2001). “An Adaptive Metropolis Algorithm.” *Bernoulli*, **7**(2), 223–242. doi:10.2307/3318737.
- Hastie TJ, Tibshirani RJ (1990). *Generalized Additive Models*. Chapman & Hall, London.
- Hilbe JM (2011). *Negative Binomial Regression*. Cambridge University Press.
- Hilbe JM (2014). *Modeling Count Data*. Cambridge University Press.
- Holmes CC, Held L (2006). “Bayesian Auxiliary Variable Models for Binary and Multinomial Regression.” *Bayesian Analysis*, **1**(1), 145–168. doi:10.1214/06-ba105.
- Jara A, Hanson T, Quintana F, Müller P, Rosner G (2011). “**DPpackage**: Bayesian Semi-And Nonparametric Modeling in R.” *Journal of Statistical Software*, **40**(5), 1–30. doi:10.18637/jss.v040.i05.
- Jo S, Choi T, Park B, Lenk PJ (2019). **bsamGP: Bayesian Spectral Analysis Models Using Gaussian Process Priors**. R package version 1.2.2, URL <https://CRAN.R-project.org/package=bsamGP>.
- Jo S, Roh T, Choi T (2016). “Bayesian Spectral Analysis Models for Quantile Regression with Dirichlet Process Mixtures.” *Journal of Nonparametric Statistics*, **28**(1), 177–206. doi:10.1080/10485252.2015.1124877.
- Kaufman CG, Sain SR (2010). “Bayesian Functional ANOVA Modeling Using Gaussian Process Prior Distributions.” *Bayesian Analysis*, **5**(1), 123–149. doi:10.1214/10-ba505.
- Koop G, Poirier DJ, Tobias JL (2007). *Bayesian Econometric Methods*. Cambridge University Press.

- Kottas A, Krnjajić M (2009). “Bayesian Semiparametric Modelling in Quantile Regression.” *Scandinavian Journal of Statistics*, **36**(2), 297–319. doi:10.1111/j.1467-9469.2008.00626.x.
- Kozumi H, Kobayashi G (2011). “Gibbs Sampling Methods for Bayesian Quantile Regression.” *Journal of Statistical Computation and Simulation*, **81**(11), 1565–1578. doi:10.1080/00949655.2010.496117.
- Kreider DL, Kuller RG, Ostberg DR, Perkins FW (1966). *An Introduction to Linear Analysis*. Addison-Wesley.
- Lenk P (1988). “The Logistic Normal Distribution for Bayesian Nonparametric Predictive Densities.” *Journal of the American Statistical Association*, **83**(402), 509–516. doi:10.1080/01621459.1988.10478625.
- Lenk P (1991). “Towards a Practicable Bayesian Nonparametric Density Estimator.” *Biometrika*, **78**(3), 531–543. doi:10.2307/2337022.
- Lenk P (1999). “Bayesian Inference for Semiparametric Regression Using a Fourier Representation.” *Journal of the Royal Statistical Society B*, **61**(4), 863–879. doi:10.1111/1467-9868.00207.
- Lenk P (2003). “Bayesian Semiparametric Density Estimation and Model Verification Using a Logistic-Gaussian Process.” *Journal of Computational and Graphical Statistics*, **12**(3), 548–565. doi:10.1198/1061860032021.
- Lenk P, Choi T (2017). “Bayesian Analysis of Shape-Restricted Functions Using Gaussian Process Priors.” *Statistica Sinica*, **27**(1), 43–69. doi:10.5705/ss.202015.0096.
- Leonard T (1978). “Density Estimation, Stochastic Processes, and Prior Information.” *Journal of Royal Statistical Society B*, **40**(2), 113–146. doi:10.1111/j.2517-6161.1978.tb01655.x.
- Lin L, Dunson DB (2014). “Bayesian Monotone Regression Using Gaussian Process Projection.” *Biometrika*, **101**(2), 303–317. doi:10.1093/biomet/ast063.
- Lunn D, Spiegelhalter D, Thomas A, Best N (2009). “The BUGS Project: Evolution, Critique and Future Directions.” *Statistics in Medicine*, **28**(25), 3049–3067. doi:10.1002/sim.3680.
- Luts J, Wand MP (2015). “Variational Inference for Count Response Semiparametric Regression.” *Bayesian Analysis*, **10**(4), 991–1023. doi:10.1214/14-ba932.
- MacEachern SN, Müller P (1998). “Estimating Mixture of Dirichlet Process Models.” *Journal of Computational and Graphical Statistics*, **7**(2), 223–338. doi:10.2307/1390815.
- Marley J, Wand M (2010). “Non-Standard Semiparametric Regression via **BRugs**.” *Journal of Statistical Software*, **37**(5), 1–30. doi:10.18637/jss.v037.i05.
- McCullagh P, Nelder JA (1989). *Generalized Linear Models*. 2nd edition. Chapman & Hall.
- Meyer MC, Hackstadt AJ, Hoeting JA (2011). “Bayesian Estimation and Inference for Generalised Partial Linear Models Using Shape-Restricted Splines.” *Journal of Nonparametric Statistics*, **23**(4), 867–884. doi:10.1080/10485252.2011.597852.

- Muggeo VM (2010). “Analyzing Temperature Effects on Mortality within the R Environment: The Constrained Segmented Distributed Lag Parameterization.” *Journal of Statistical Software*, **32**(12), 1–17. doi:[10.18637/jss.v032.i12](https://doi.org/10.18637/jss.v032.i12).
- Muggeo VMR (2008). “Modeling Temperature Effects on Mortality: Multiple Segmented Relationships with Common Break Points.” *Biostatistics*, **9**(4), 613–620. doi:[10.1093/biostatistics/kxm057](https://doi.org/10.1093/biostatistics/kxm057).
- Murphy KP (2012). *Machine Learning: A Probabilistic Perspective*. MIT press.
- Neal RM (2003). “Slice Sampling.” *The Annals of Statistics*, **31**(3), 705–767. doi:[10.1214/aos/1056562461](https://doi.org/10.1214/aos/1056562461).
- Nelder JA, Wedderburn RWM (1972). “Generalized Linear Models.” *Journal of the Royal Statistical Society A*, **135**(3), 370–384. doi:[10.2307/2344614](https://doi.org/10.2307/2344614).
- Newton MA, Raftery AE (1994). “Approximate Bayesian Inference with the Weighted Likelihood Bootstrap.” *Journal of the Royal Statistical Society B*, **56**(1), 3–48. doi:[10.1111/j.2517-6161.1994.tb01956.x](https://doi.org/10.1111/j.2517-6161.1994.tb01956.x).
- Nierenberg D, Stukel T, Baron J, Dain B, Greenberg E (1989). “Determinants of Plasma Levels of Beta-Carotene and Retinol.” *American Journal of Epidemiology*, **130**(3), 511–521. doi:[10.1093/oxfordjournals.aje.a115365](https://doi.org/10.1093/oxfordjournals.aje.a115365).
- O’Hagan A (1978). “Curve Fitting and Optimal Design for Prediction.” *Journal of the Royal Statistical Society B*, **40**(1), 1–42. doi:[10.1111/j.2517-6161.1978.tb01643.x](https://doi.org/10.1111/j.2517-6161.1978.tb01643.x).
- Paciorek C (2007). “Bayesian Smoothing with Gaussian Processes Using Fourier Basis Functions in the **spectralGP** Package.” *Journal of Statistical Software*, **19**(2), 1–38. doi:[10.18637/jss.v019.i02](https://doi.org/10.18637/jss.v019.i02).
- Paciorek CJ, Lipshitz B, Zhuo W, Prabhat, Kaufman CG, Thomas RC (2015). “Parallelizing Gaussian Process Calculations in R.” *Journal of Statistical Software*, **63**(10), 1–23. doi:[10.18637/jss.v063.i10](https://doi.org/10.18637/jss.v063.i10).
- Plummer M (2003). “**JAGS**: A Program for Analysis of Bayesian Graphical Models Using Gibbs Sampling.” In K Hornik, F Leisch, A Zeileis (eds.), *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*. Technische Universität Wien, Vienna, Austria. URL <https://www.R-project.org/conferences/DSC-2003/Proceedings/Plummer.pdf>.
- Plummer M, Best N, Cowles K, Vines K (2006). “**coda**: Convergence Diagnosis and Output Analysis for MCMC.” *R News*, **6**(1), 7–11. URL <https://CRAN.R-project.org/doc/Rnews/>.
- Rasmussen CE, Nickisch H (2010). “Gaussian Processes for Machine Learning (**GPML**) Toolbox.” *Journal of Machine Learning Research*, **11**, 3011–3015.
- Rasmussen CE, Williams CK (2006). *Gaussian Processes for Machine Learning*. MIT Press. doi:[10.7551/mitpress/3206.001.0001](https://doi.org/10.7551/mitpress/3206.001.0001).



- R Core Team (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Riihimäki J, Vehtari A (2014). “Laplace Approximation for Logistic Gaussian Process Density Estimation and Regression.” *Bayesian Analysis*, **9**(2), 425–448. doi:10.1214/14-ba872.
- Roberts GO, Rosenthal JS (2009). “Examples of Adaptive MCMC.” *Journal of Computational and Graphical Statistics*, **18**(2), 349–367. doi:10.1198/jcgs.2009.06134.
- Ruppert D, Wand MP, Carroll RJ (2003). *Semiparametric Regression*. Cambridge University Press.
- Shi JQ, Murray-Smith R, Titterton DM (2005). “Hierarchical Gaussian Process Mixtures for Regression.” *Statistics and Computing*, **15**(1), 31–41. doi:10.1007/s11222-005-4787-7.
- Sriram K, Ramamoorthi RV, Ghosh P (2013). “Posterior Consistency of Bayesian Quantile Regression Based on the Misspecified Asymmetric Laplace Density.” *Bayesian Analysis*, **8**(2), 479–504. doi:10.1214/13-ba817.
- Stan Development Team (2017). *Stan Modeling Language User’s Guide and Reference Manual, Version 2.16.0*. URL <http://mc-stan.org/>.
- The MathWorks Inc (2019). *MATLAB – The Language of Technical Computing, Version R2019a*. Natick. URL <http://www.mathworks.com/products/matlab/>.
- Tokdar ST (2007). “Towards a Faster Implementation of Density Estimation with Logistic Gaussian Process Priors.” *Journal of Computational and Graphical Statistics*, **16**(3), 633–655. doi:10.1198/106186007x210206.
- Tokdar ST, Ghosh JK (2007). “Posterior Consistency of Logistic Gaussian Process Priors in Density Estimation.” *Journal of Statistical Planning and Inference*, **137**(1), 34–42. doi:10.1016/j.jspi.2005.09.005.
- van Rossum G, et al. (2011). *Python Programming Language*. URL <https://www.python.org/>.
- Vanhatalo J, Riihimäki J, Hartikainen J, Jylänki P, Tolvanen V, Vehtari A (2013). “GPstuff: Bayesian Modeling with Gaussian Processes.” *Journal of Machine Learning Research*, **14**, 1175–1179.
- Wand M (2018). **SemiPar**: *Semiparametric Regression*. R package version 1.0-4.2, URL <https://CRAN.R-project.org/package=SemiPar>.
- Wang B, Shi JQ (2014). “Generalized Gaussian Process Regression Model for Non-Gaussian Functional Data.” *Journal of the American Statistical Association*, **109**(507), 1123–1133. doi:10.1080/01621459.2014.889021.
- Wang X, Berger JO (2016). “Estimating Shape Constrained Functions Using Gaussian Processes.” *SIAM/ASA Journal on Uncertainty Quantification*, **4**(1), 1–25. doi:10.1137/140955033.



- Wickham H (2009). **ggplot2**: *Elegant Graphics for Data Analysis*. Springer-Verlag. doi:  
[10.1007/978-3-319-24277-4](https://doi.org/10.1007/978-3-319-24277-4).
- Wikle CK (2002). “Spatial Modelling of Count Data: A Case Study in Modelling Breeding Bird Survey Data on Large Spatial Domains.” In *Spatial Cluster Modelling*, pp. 199–209. Chapman & Hall/CRC, Boca Raton.
- Wood S (2016). “Just Another Gibbs Additive Modeler: Interfacing JAGS and **mgcv**.” *Journal of Statistical Software*, **75**(7), 1–15. doi:[10.18637/jss.v075.i07](https://doi.org/10.18637/jss.v075.i07).
- Wood SN (2017). *Generalized Additive Models: An Introduction with R*. 2nd edition. Chapman & Hall/CRC.
- Yang J, Zhu H, Choi T, Cox DD (2016). “Smoothing and Mean-Covariance Estimation of Function Data with a Bayesian Hierarchical Model.” *Bayesian Analysis*, **11**(3), 649–670. doi:[10.1214/15-ba967](https://doi.org/10.1214/15-ba967).
- Yatchew A (2003). *Semiparametric Regression for the Applied Econometrician*. Cambridge University Press.
- Yu K, Moyeed RA (2001). “Bayesian Quantile Regression.” *Statistics & Probability Letters*, **54**(4), 437–447. doi:[10.1016/s0167-7152\(01\)00124-9](https://doi.org/10.1016/s0167-7152(01)00124-9).
- Zhu H, Vannucci M, Cox DD (2010). “A Bayesian Hierarchical Model for Classification with Selection of Functional Predictors.” *Biometrics*, **66**(2), 463–473. doi:[10.1111/j.1541-0420.2009.01283.x](https://doi.org/10.1111/j.1541-0420.2009.01283.x).

### **Affiliation:**

Seongil Jo  
Department of Statistics (Institute of Applied Statistics)  
Chonbuk National University, Republic of Korea  
E-mail: [statjs@jbnu.ac.kr](mailto:statjs@jbnu.ac.kr)

Taeryon Choi (*corresponding author*)  
Department of Statistics  
Korea University, Republic of Korea  
E-mail: [trchoi@korea.ac.kr](mailto:trchoi@korea.ac.kr)  
URL: <http://statlab2.korea.ac.kr/>

Beomjo Park  
Department of Statistics  
Korea University, Republic of Korea

Peter Lenk  
Stephen M. Ross School of Business  
University of Michigan, United States of America