Locally Adaptive Bayesian Smoothing using Shrinkage Priors

（縮小事前分布を用いた局所適合的なベイズ平滑化）

学位取得年月　2024 年 9 月

鬼塚 貴広

Locally Adaptive Bayesian Smoothing using Shrinkage Priors

Takahiro Onizuka
Graduate School of Advanced Science
and Engineering, Mathematics Program
September 2024

Abstract

This thesis deals with locally adaptive Bayesian smoothing methods to estimate the non-stationary trend of quantile and boundary. The underlying idea of the proposals is based on trend filtering which is one of the smoothing methods in nonparametric statistics. All proposed models are constructed by assuming asymmetric Laplace or truncated normal distribution as a working likelihood, and are adopted shrinkage priors such as horseshoe prior to achieve locally adaptive shrinkage. We first summarize technical background of selected topics in Chapter 2. The Bayesian quantile trend filtering methods for time series and spatial data are established in Chapters 3 and 4 respectively. In Chapter 5, we propose the Bayesian boundary trend filtering by introducing the approximation of truncated normal likelihood. Finally, we conclude this thesis and discuss several future works in Chapter 6.

Acknowledgements

I would like to express my deepest gratitude to Dr. Shintaro Hashimoto of Hiroshima University. His guidance, encouragement, valuable comments, and helpful suggestions led to this thesis. I would also like to thank Dr. Shonosuke Sugasawa of Keio University, Dr. Yasuyuki Hamura of Kyoto University, and Mr. Fumiya Iwashige of Hiroshima University for supporting my research. I wish to thank Professor Hirofumi Wakaki and Professor Hirokazu Yanagihara of Hiroshima University for their encouragement and helpful advice. My research was supported in part by JST, the establishment of university fellowships towards the creation of science technology innovation, Grant Number JPMJFS2129. Finally, I thank my parents and other relatives.

Takahiro Onizuka

Contents

Chapter 1

Introduction

This thesis deals with the Bayesian smoothing methods for different statistics such as quantiles (Chapter 3 and 4) and boundaries (Chapter 5) of time series or spatial data. Smoothing or trend estimation is an important issue in investigating characteristics of data, and such methods have been applied in various scientific fields such as astronomical spectroscopy (e.g. Politsch et al., 2020), biometrics (e.g. Faulkner et al., 2020), bioinformatics (e.g. Eilers and De Menezes, 2005), economics (e.g. Yamada, 2022; Daouia et al., 2016) and environmetrics (e.g. Brantley et al., 2020; Tibshirani et al., 2011) among others. To focus on the smoothing of data, we consider the sequence model $y_i = \theta_i + \varepsilon_i \quad (i = 1, \ldots, n)$, where y_i is an observation, θ_i is a true trend and ε_i is a noise. Let n be a sample size, and $y = (y_1, \ldots, y_n)^\top$ and $\theta = (\theta_1, \ldots, \theta_n)^\top$. For estimating underlying trends θ, the ℓ_1 trend filtering (Kim et al., 2009; Tibshirani, 2014) is one of the popular methods that can flexibly capture local abrupt changes in trends, compared with spline methods, and it is formulated by the minimizer of the following penalized problem:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}} \ell(y - \theta) + \lambda \| D_n^{(k+1)} \theta \|_1, \tag{1.1}$$

where $\ell(\cdot)$ is a quadratic loss, $D_n^{(k+1)}$ is a $(n-k-1) \times n$ difference operator matrix of order $k+1$, and $\lambda > 0$ is a tuning constant. The ℓ_1 trend filtering is known to be a special case of the generalized lasso proposed by Tibshirani and Taylor (2011). Furthermore, fast and efficient optimization algorithms for trend filtering have also been proposed (e.g. Ramdas and Tibshirani, 2016). Due to such advantages in terms of flexibility and computation, extensions of the original trend filtering to spatial data (Wang et al., 2015) and functional data (Wakayama and Sugasawa, 2023, 2024) have been considered. However, the majority of existing studies focus on estimating mean

trends with a homogeneous variance structure, and these methods may not work well in the presence of outliers or data with heterogeneous variance. Additionally, we are often interested in estimating quantiles rather than means. Recently, by replacing the loss function $\ell(\cdot)$ with the check loss function, Brantley et al. (2020) proposed a quantile version of trend filtering (QTF). Theoretical properties of quantile trend filtering have been shown by Madrid Padilla and Chatterjee (2022).

The main difficulty in applying the optimization-based trend filtering as considered in Brantley et al. (2020) is that uncertainty quantification for trend estimation is not straightforward. Moreover, the frequentist formulation includes tuning parameters in the regularization, but the data-dependent selection of the tuning parameter is not obvious, especially with quantile smoothing.

A reasonable alternative is to employ a Bayesian formulation for trend filtering by introducing priors. In general, the Bayesian formulation for trend filtering is based on the model:

$$y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim f(\cdot), \quad D_n^{(k+1)}\theta \sim \pi(\cdot), \quad (i = 1, \ldots, n), \tag{1.2}$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent errors following the working likelihood f, and π correspond to the prior density functions. A simple Bayesian counterpart that corresponds to penalized square loss is a combination of the Gaussian likelihood on f and Laplace prior distribution on π (e.g. Roualdes, 2015). The resulting posterior mode is the same as that of the solution of the problem (1.1). To achieve locally adaptive smoothing, Faulkner and Minin (2018) proposed a more flexible Bayesian trend filtering via global-local shrinkage priors. Kowal et al. (2019) also considered a Bayesian formulation based on a dynamic shrinkage process under Gaussian likelihood to estimate the mean trend. To estimate quantile trends, assuming a more flexible probability distribution called extended asymmetric Laplace distribution in (1.2), Barata et al. (2022) also proposed the extended dynamic quantile linear model. The Bayesian approach has some advantages: 1) capable of full probabilistic uncertainty quantification through posterior distribution, 2) flexible shrinkage by using global-local shrinkage priors (Carvalho et al., 2010), and 3) estimating regularization parameters from Markov chain Monte Carlo method.

The remainder of the thesis is organized as follows.

- Chapter 2: Technical background.

We summarized the common fundamentals throughout this article. Under the linear regression model, we introduce mean or quantile estimation and Bayesian inference including a brief introduction to global-local shrinkage priors. Mathematical formulation of trend filtering is also described.

- Chapter 3: Calibrated variational Bayesian inference for quantile smoothing.

  In this chapter, we provide fast Bayesian quantile trend filtering for time series data. To induce locally adaptive Bayesian inference on trends, we introduce general shrinkage priors, and to quickly compute the posterior distribution, we develop calibrated mean-field variational Bayes approximations, which guarantee that the frequentist coverage of credible intervals obtained from the approximated posterior is a specified nominal level. Simulation and empirical studies show that the proposed algorithm is computationally much more efficient than the Gibbs sampler and tends to provide stable inference results, especially for high/low quantiles. This chapter is based on Onizuka et al. (2024a).

- Chapter 4: Locally adaptive Bayesian spatial quantile smoothing.

  Spatial trend estimation under potential heterogeneity is an important problem in extracting spatial characteristics and hazards such as criminal activity. In this chapter, we propose a Bayesian quantile trend filtering method to estimate the non-stationary trend of quantiles on graphs and apply it to crime data in Tokyo between 2013 and 2017. By modeling multiple observation cases, we can estimate the potential heterogeneity of spatial crime trends over multiple years in the application. To induce locally adaptive Bayesian inference on trends, we introduce general shrinkage priors for graph differences. Introducing so-called shadow priors with multivariate distribution for local scale parameters and mixture representation of the asymmetric Laplace distribution, we provide a simple Gibbs sampling algorithm to generate posterior samples. The numerical performance of the proposed method is demonstrated through simulation studies. This chapter is based on Onizuka et al. (2024b).

- Chapter 5: Locally adaptive Bayesian boundary smoothing.

  Estimating boundary curves has many applications such as economics, climate science, and medicine. Bayesian trend filtering has been developed as

one of the locally adaptive smoothing methods to estimate the non-stationary trend of data. This study develops a Bayesian trend filtering for estimating the boundary trend. To this end, the truncated multivariate normal working likelihood and global-local shrinkage priors based on the scale mixtures of the normal distribution are introduced. In particular, well-known horseshoe prior for difference leads to locally adaptive shrinkage estimation for boundary trend. However, the full conditional distributions of the Gibbs sampler involve high-dimensional truncated multivariate normal distribution. To overcome the difficulty of sampling, an approximation of truncated multivariate normal distribution is employed. Using the approximation, the proposed models lead to an efficient Gibbs sampling algorithm via the Pólya-Gamma data augmentation. The proposed method is also extended by considering a nearly isotonic constraint. The performance of the proposed method is illustrated through some numerical experiments and real data examples. This chapter is based on Onizuka et al. (2024).

Some concluding remarks are given in Chapter 6. In particular, the results and contribution of the thesis are summarized.

Chapter 2

Technical background of selected topics

2.1 Mean and quantile estimation in frequentist and Bayesian perspective

Let y_i be a response variable and x_i be a covariate variable. In this section, we consider the following regression problem:

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,$$

where ε_i $(i = 1, \ldots, n)$ is an independent and identically distributed error term. For simplicity, we assume linear regression model $f(x_i) = x_i^\top \beta$, where $\beta \in \mathbb{R}^k$ is an unknown parameter. Moreover, it is re-expressed by

$$y = X\beta + \varepsilon, \tag{2.1}$$

where $y = (y_1, \ldots, y_n)^\top$, $X \in \mathbb{R}^{n \times k}$ is a covariate matrix, and $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\top$.

2.1.1 Frequentist methods

First, if we assume $E[\varepsilon_i] = 0$ $(i = 1, \ldots, n)$ in (2.1), then $E[y_i] = f(x_i)$, which f is mean of data y_i. For estimating $f(x_i) = x_i^\top \beta$, the ordinary least squares estimate $\hat{\beta}$ is defined by

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^k}{\operatorname{argmin}} \, \|y - X\beta\|_2^2,$$

where $\|y - X\beta\|_2^2 = \sum_{i=1}^n (y_i - x_i^\top \beta)^2$ is the quadratic loss function.

Next, for conditional cumulative distribution function $F(y_i \mid x_i)$ of y_i, the conditional pth quantile is defined by

$$Q_{y_i|x_i}(p) = \inf_{y_i \in \mathbb{R}} \{y_i \mid F(y_i \mid x_i) > p\},$$

where $p \in (0, 1)$ is a quantile level. When $f(x_i) = x_i^\top \beta$ is pth quantile of y_i, which means $f(x_i) = Q_{y_i|x_i}(p)$, quantile regression proposed by Koenker and Bassett Jr (1978) to estimate pth quantile of y under the regression model (2.1) is given by solving the following optimization problem:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^k}{\operatorname{argmin}} \rho_p(y - X\beta).$$

where $\rho_p(\cdot)$ is the check loss function defined by

$$\rho_p(r) = \sum_{i=1}^{n} r_i\{p - 1(r_i < 0)\}. \tag{2.2}$$

2.1.2 Bayesian inference

As a natural Bayesian model to estimate the mean of y under (2.1), it is often assumed that $\varepsilon_i \sim N(0, \sigma^2)$ $(i = 1, \ldots, n)$, where σ^2 is a scale parameter. Then, the conditional distribution of y is represented by

$$p(y \mid \beta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{-\frac{1}{2\sigma^2} \|y - X\beta\|_2^2\right\}.$$

The term in the exponential function is similar to the quadratic loss in the frequentist method. To obtain the posterior distribution, we assume the prior distribution $\pi(\beta)$ and $\pi(\sigma^2)$ such as Gaussian prior and inverse-gamma prior for β and σ^2, respectively.

In Bayesian quantile regression, the asymmetric Laplace distribution $\mathrm{AL}(p, \sigma^2)$ is often assumed for the error term ε_i (Yu and Moyeed, 2001). The asymmetric Laplace distribution has the following probability density function

$$f_{\mathrm{AL}(p)}(x) = \frac{p(1-p)}{\sigma^2} \exp\left\{-\rho_p\left(\frac{x}{\sigma^2}\right)\right\}, \tag{2.3}$$

where p is a fixed constant which characterizes the quantile level, σ^2 is a scale parameter, and $\rho_p(\cdot)$ is a check loss function defined by (2.2). As well as the Gaussian assumption for mean estimation, the term in the exponential function corresponds to the scaled check loss in the frequentist method. The prior distributions for β and

σ^2 are also assumed.

2.2 Bayesian shrinkage

In the regression model (2.1), the shrinkage estimation of β under sparsity is usually considered. As a famous frequentist method, the lasso estimator is given by the following penalized minimization:

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^k}{\operatorname{argmin}} \left\{ \|y - X\beta\|_2^2 + \lambda \|\beta\|_1 \right\},$$

where λ is a tuning parameter and $\|\cdot\|_1$ is a L_1 norm. From this formulation, the Bayesian lasso (Park and Casella, 2008) is proposed by

$$p(y \mid \beta, \sigma^2) = \frac{1}{\sqrt{(2\pi\sigma^2)^n}} \exp\left\{ -\frac{1}{2\sigma^2} \|y - X\beta\|_2^2 \right\},$$
$$\pi(\beta \mid \sigma^2, \lambda^*) = \prod_{j=1}^{k} \frac{\lambda^*}{2\sigma} \exp\left(-\frac{\lambda^* |\beta_j|}{\sigma} \right),$$

where $\pi(\beta \mid \sigma^2, \lambda^*)$ is the Laplace distribution with scale parameter σ. When σ and λ^* are fixed, the MAP (maximum a posteriori) estimator is equal to the frequentist lasso estimator $\hat{\beta}$ with tuning parameter $\lambda^* = 2\sigma\lambda$. The Laplace prior leads to the shrinkage toward zero due to the density mass at zero. For efficient sampling, the scale mixture representation of the Laplace prior $\pi(\beta \mid \sigma^2, \lambda^*)$ is also given as follows:

$$\pi(\beta_j \mid \sigma^2, \lambda^*, \tau_j) \sim N(0, \sigma^2\tau_j^2), \quad \pi(\tau_j^2 \mid \lambda^*) \sim \mathrm{Exp}(\lambda^2/2) \quad (j = 1, \ldots, k).$$

As the other way to induce shrinkage toward zero, the spike-and-slab prior which has the point mass at zero and heavy-tail is considered. Recently, as a normal scale mixture distribution, the global-local shrinkage priors to induce shrinkage toward zero are also considered:

$$\beta_j \mid \lambda_j, \tau \sim N(0, \lambda_j^2\tau^2), \quad \lambda_j \sim \pi(\cdot), \quad \tau \sim \pi(\cdot),$$

where λ_j is a local parameter that leads to individual (coordinate-wise) shrinkage and τ^2 is a global parameter that leads to common shrinkage. One of the global-local shrinkage priors is horseshoe prior proposed by Carvalho et al. (2010), which

assumes

$$\lambda_j \sim C^+(0,1), \quad \tau \sim C^+(0,1).$$

Global-local shrinkage priors for non-Gaussian data have also been developed in recent years (e.g. Datta and Dunson, 2016; Hamura et al., 2022b, 2024).

2.3 Trend filtering

Let $y_i = \theta_i + \varepsilon_i \quad (i = 1, \ldots, n)$ be a sequence model, where y_i is an observation, θ_i is a true trend and ε_i is a noise. The estimate of ℓ_1 trend filtering (Kim et al., 2009) is given by solving the optimization problem

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \, \|y - \theta\|_2^2 + \lambda \|D_n^{(k+1)} \theta\|_1, \tag{2.4}$$

where $y = (y_1, \ldots, y_n)^\top$, $\theta = (\theta_1, \ldots, \theta_n)^\top$, $\lambda > 0$ is a tuning constant, and $D_n^{(k+1)}$ is a $(n-k-1) \times n$ difference operator matrix of order $k+1$, which is defined by

$$D_n^{(1)} = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1) \times n}, \quad D_n^{(k+1)} = D_{n-k}^{(1)} D_n^{(k)}. \tag{2.5}$$

Depending on the different order k, we can express various smoothing such as piecewise constant, linear, quadratic, and so forth (see e.g. Tibshirani, 2014). The trend filtering is also called the fused lasso for $k = 0$.

Brantley et al. (2020) proposed quantile trend filtering, defined as the optimization problem

$$\hat{\theta}_p = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \, \rho_p(y - \theta) + \lambda \|D_n^{(k+1)} \theta\|_1, \tag{2.6}$$

where $\hat{\theta}_p$ is pth quantile trend, $\lambda > 0$ is a tuning constant and $\rho_p(\cdot)$ is a check loss function given in (2.2). To solve the problem (2.6), Brantley et al. (2020) proposed a parallelizable alternating direction method of multipliers (ADMM) algorithm, and also proposed the selection of smoothing parameters λ using a modified criterion based on the extended Bayesian information criterion.

Chapter 3

Fast and locally adaptive quantile smoothing using calibrated variational approximations

3.1 Introduction

In this study, we provided a Bayesian quantile trend filtering method. To construct a fast and efficient algorithm, we employed the asymmetric Laplace distribution as a working likelihood (Yu and Moyeed, 2001; Sriram et al., 2013) and two types of shrinkage priors; Laplace (Park and Casella, 2008) and horseshoe (Carvalho et al., 2010) priors. From the data augmentation strategies of asymmetric Laplace distribution by Kozumi and Kobayashi (2011) and the shrinkage priors, we construct an efficient Gibbs sampling algorithm and a mean-field variational Bayes (MFVB) algorithm. The variational Bayes method has the following several characteristics: 1) it gives the quick calculation of point estimates, 2) the MFVB algorithm tends to provide narrower credible intervals than that of Gibbs sampling (e.g. Blei et al., 2017). Moreover, in the general Bayesian framework under the assumption of the true sampling distribution or the working likelihood, the (possibly) misspecified model such as asymmetric Laplace likelihood may produce invalid credible intervals. To overcome such problems, we proposed a new simulation-based calibration algorithm for the credible intervals of the variational posterior distribution. It is expected to give valid credible intervals for the MFVB posterior distribution fast. We showed the performance of the proposed methods through simulation studies and real data examples.

The remainder of the chapter is structured as follows: In Section 3.2, we formu-

late a Bayesian quantile trend filtering method and provide Gibbs sampling and variational Bayes algorithms. In Section 3.3, we describe the main proposal of this study, a new calibration algorithm for approximating posterior distribution with variational Bayes approximation. In Section 3.4, we illustrate simulation studies to compare the performance of the proposed methods. In Section 3.5, we apply the proposed methods to real data examples. Concluding remarks are presented in Section 3.6. Additional information on the proposed algorithms and numerical experiments are provided in the Appendix 3.7. R code implementing the proposed methods is available at the GitHub repository (`https://github.com/Takahiro-Onizuka/BQTF-VB`).

3.2 Bayesian quantile trend filtering

3.2.1 Bayesian formulation and shrinkage priors for differences

To conduct Bayesian inference of the quantile trend corresponding to the frequentist model (2.6), we often use the following model:

$$y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{AL}(p, \sigma^2), \quad i = 1, \ldots, n, \tag{3.1}$$

where θ_i and σ^2 are unknown parameters, p is a fixed quantile level, and $\mathrm{AL}(p, \sigma^2)$ denotes the asymmetric Laplace distribution with the density function (2.3). Although the model assumes that a single variable is observed at each point i, multiple observations per grid point often appear in practice. Therefore, following Heng et al. (2023), this study considered the following model which accounts for multiple observations per grid point:

$$y_{ij} = \theta_p(x_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathrm{AL}(p, \sigma^2), \quad i = 1, \ldots, n, \quad j = 1, \ldots, N_i, \tag{3.2}$$

where $\theta_p(x)$ is a p-th quantile in the location x, n is the number of locations data are observed, and N_i is the amount of data for each location x_i. It is a natural generalization of the sequence model (3.1). Note that the model (3.2) is a nonparametric quantile regression with a single covariate, and the proposed approach can easily be generalized for additive regression with multiple covariates. For simplicity in notation, we denoted $\theta_p(x_i) = \theta_i$ for the remainder of the chapter.

The model (3.2) for the likelihood corresponds to the loss function in frequentist

method (2.6), and then shrinkage priors on differences are introduced to induce the smoothness of the trends which corresponds to the penalty term in the frequentist method. For Bayesian model, we define the $(k+1)$th order difference operator D as

$$D = \begin{pmatrix} I_{k+1} & O \\ D_n^{(k+1)} \end{pmatrix}, \tag{3.3}$$

where I_{k+1} is $(k+1) \times (k+1)$ identity matrix, O is zero matrix and $D_n^{(k+1)}$ is $(n-k-1) \times n$ standard difference matrix given in (2.5). We consider flexible shrinkage priors on $D\theta$, and the priors are represented by

$$D\theta \mid \tau^2, \sigma^2, w \sim N_n(0, \sigma^2 W) \;\text{ with }\; W = \mathrm{diag}(w_1^2, \ldots, w_{k+1}^2, \tau^2 w_{k+2}^2, \ldots, \tau^2 w_n^2),$$

where $w = (w_1, \ldots, w_n)$ represents local shrinkage parameters for each element in $D\theta$ and τ^2 is a global shrinkage parameter. Since the D is non-singular matrix, the prior of θ can be rewritten as

$$\theta \mid \tau^2, \sigma^2, w \sim N_n(0, \sigma^2 (D^\top W^{-1} D)^{-1}). \tag{3.4}$$

Note that since $(D\theta)_i = \theta_i \sim N(0, w_i^2)$ for $i = 1, \ldots, k+1$, w_i $(i = 1, \ldots, k+1)$ is not related to shrinkage of difference. For this reason, we assumed the conjugate inverse gamma distribution $\mathrm{IG}(a_{w_i}, b_{w_i})$ for w_i^2. As mentioned in Section 2.2, introducing the specific priors for w_i $(i = k+2, \ldots, n)$ and τ, various degrees of shrinkage can be expressed. In particular, for $i = k+2, \ldots, n$, we considered two types of distribution; $w_i \sim \mathrm{Exp}(1/2)$ and $w_i \sim C^+(0,1)$, which were motivated from Laplace or Bayesian lasso prior (Park and Casella, 2008) and horseshoe prior (Carvalho et al., 2010), respectively. Regarding the other parameters, we assigned $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$ and $\tau \sim C^+(0, C_\tau)$. The default choice of hyperparameters is $a_\sigma = b_\sigma = 0.1$ and $C_\tau = 1$.

Remark 3.2.1. Tibshirani (2014) extend the difference matrix (2.5) for the situation where data is observed at an irregular grid, and we also apply the adjusted difference operator to the proposed method. We assume that the locations of data $x = (x_1, \ldots, x_n)$ have the ordering $x_1 < x_2 < \cdots < x_n$ and $x_{j+1} - x_j$ is not constant. Note that this issue is related to nonparametric quantile regression. When the locations $x \in \mathbb{R}^n$ are irregular and strictly increasing, an adjusted difference operator

for $k \geq 1$ is defined by

$$D_n^{(x,k+1)} = D_{n-k}^{(x,1)} \mathrm{diag}\left(\frac{k}{x_{k+1} - x_1}, \ldots, \frac{k}{x_n - x_{n-k}}\right) D_n^{(x,k)}$$

where $D_n^{(x,1)} = D_n^{(1)}$. It is a natural generalization of (2.5) because $D_n^{(x,k+1)} = D_n^{(k+1)}$ for $x_1 = 1, x_2 = 2, \ldots, x_n = n$ (see also Heng et al., 2023). Then, the matrix D for our model is also given by

$$D = \begin{pmatrix} I_{k+1} & O \\ D_n^{(x,k+1)} \end{pmatrix},$$

where O is a zero matrix.

3.2.2 Gibbs sampler

We first derived a Gibbs sampler which is one of the Markov chain Monte Carlo (MCMC) methods. For the simple computation of the posterior distribution, the stochastic representation of the asymmetric Laplace distribution (Kozumi and Kobayashi, 2011) is utilized. For $\varepsilon_{ij} \sim \mathrm{AL}(p, \sigma^2)$, we have the following stochastic expression:

$$\varepsilon_{ij} = \psi z_{ij} + \sqrt{\sigma^2 z_{ij} t^2} u_{ij}, \quad \psi = \frac{1-2p}{p(1-p)}, \quad t^2 = \frac{2}{p(1-p)},$$

where $u_{ij} \sim N(0,1)$ and $z_{ij} \mid \sigma^2 \sim \mathrm{Exp}(1/\sigma^2)$ for $i = 1, \ldots, n$ and $j = 1, \ldots, N_i$. From the above expression, the conditional likelihood function of y_{ij} is given by

$$p(y_{ij} \mid \theta_i, z_{ij}, \sigma^2) = (2\pi t^2 \sigma^2)^{-1/2} z_{ij}^{-1/2} \exp\left\{-\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 \sigma^2 z_{ij}}\right\}. \tag{3.5}$$

Under the prior (3.4), the full conditional distributions of θ and z_i are given by

$$\begin{aligned}
&\theta \mid y, z, \sigma^2, \gamma^2 \sim N_n\left(A^{-1}B, \sigma^2 A^{-1}\right), \\
&z_{ij} \mid y_{ij}, \theta_i, \sigma^2 \sim \mathrm{GIG}\left(\frac{1}{2}, \frac{(y_{ij} - \theta_i)^2}{t^2\sigma^2}, \frac{\psi^2}{t^2\sigma^2} + \frac{2}{\sigma^2}\right), \quad i = 1, \ldots, n, \ j = 1, \ldots, N_i,
\end{aligned}$$

respectively, where

$$\begin{aligned}
A &= D^\top W^{-1} D + \frac{1}{t^2} \mathrm{diag}\left(\sum_{j=1}^{N_1} z_{1j}^{-1}, \ldots, \sum_{j=1}^{N_n} z_{nj}^{-1}\right), \\
B &= \left(\frac{1}{t^2}\sum_{j=1}^{N_1}\left(\frac{y_{1j}}{z_{1j}} - \psi\right), \ldots, \frac{1}{t^2}\sum_{j=1}^{N_n}\left(\frac{y_{nj}}{z_{nj}} - \psi\right)\right)^\top
\end{aligned}$$

and $\mathrm{GIG}(a, b, c)$ denotes the generalized inverse Gaussian distribution. The full conditional distribution of the scale parameter σ^2 is given by

$$\sigma^2 \mid y, \theta, z, w, \tau^2 \sim \mathrm{IG}\left(\frac{n+3N}{2} + a_\sigma, \alpha_{\sigma^2}\right),$$

where $N = \sum_{i=1}^{n} N_i$ is the number of observed values and

$$\alpha_{\sigma^2} = \sum_{i=1}^{n}\sum_{j=1}^{N_i} \frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 z_{ij}} + \frac{1}{2}\theta^\top D^\top W^{-1} D\theta + \sum_{i=1}^{n}\sum_{j=1}^{N_i} z_{ij} + b_\sigma.$$

By using the augmentation of the half-Cauchy distribution (Makalic and Schmidt, 2015), the full conditional distribution of the global shrinkage parameter τ^2 is given by

$$\tau^2 \mid \theta, w, \sigma^2, \xi \sim \mathrm{IG}\left(\frac{n-k}{2}, \frac{1}{2\sigma^2}\sum_{i=k+1}^{n}\frac{\eta_i^2}{w_i^2} + \frac{1}{\xi}\right), \quad \xi \mid \tau^2 \sim \mathrm{IG}\left(\frac{1}{2}, \frac{1}{\tau^2} + 1\right),$$

where ξ is an augmented parameter for τ^2. For $i = 1, \ldots, k+1$, we assumed the prior $\mathrm{IG}(a_{w_i}, b_{w_i})$ for w_i, and then the full conditional distribution of w_i is given by

$$w_i^2 \mid \theta, \sigma^2 \sim \mathrm{IG}\left(\frac{1}{2} + a_{w_i}, \frac{\eta_i^2}{2\sigma^2} + b_{w_i}\right), \quad i = 1, \ldots, k+1.$$

The full conditional distribution of local shrinkage parameter w_i $(i = k+2, \ldots, n)$ depends on the choice of the prior, either Laplace or horseshoe prior.

- **(Laplace-type prior)** For Laplace-type prior, we set $\tau^2 = 1$ and assume $w_i \mid \gamma^2 \sim \mathrm{Exp}(\gamma^2/2)$ for $i = k+2, \ldots, n$. Then we have $(D\theta)_i \sim \mathrm{Lap}(\gamma)$. Noting that $\gamma \sim C^+(0,1)$, sampling from the standard half-Cauchy prior is equivalent to $\gamma^2 \mid \nu \sim \mathrm{IG}(1/2, 1/\nu)$ and $\nu \sim \mathrm{IG}(1/2, 1)$ using the augmentation technique (Makalic and Schmidt, 2015). Hence, the full conditional distributions of w_i, γ^2 and ν are, respectively, given by

$$w_i^2 \mid \theta, \sigma^2, \gamma^2 \sim \mathrm{GIG}\left(\frac{1}{2}, \frac{\eta_i^2}{\sigma^2}, \gamma^2\right),$$
$$\gamma^2 \mid w, \nu \sim \mathrm{GIG}\left(n - k - \frac{3}{2}, \frac{2}{\nu}, \sum_{i=k+2}^{n} w_i^2\right), \quad \nu \mid \gamma^2 \sim \mathrm{IG}\left(\frac{1}{2}, \frac{1}{\gamma^2} + 1\right).$$

- **(Horseshoe-type prior)** For Horseshoe-type prior, we assume $w_i \sim C^+(0,1)$ for $i = k+2, \ldots, n$. By using the representation $w_i^2 \mid \nu_i \sim \mathrm{IG}(1/2, 1/\nu_i)$ and

$\nu_i \sim \mathrm{IG}(1/2, 1)$, the full conditional distributions of w_i and ν_i are, respectively, given by

$$w_i^2 \mid \theta, \sigma^2, \tau^2, \nu_i \sim \mathrm{IG}\left(1, \frac{1}{\nu_i} + \frac{\eta_i^2}{2\sigma^2\tau^2}\right), \quad \nu_i \mid w_i \sim \mathrm{IG}\left(\frac{1}{2}, \frac{1}{w_i^2} + 1\right).$$

3.2.3 Variational Bayes approximation

When the sample size is large, the MCMC algorithm presented in Section 3.2.2 can be computationally intensive due to the large number of parameters that need to be sampled. To reduce the computational resources required, the variational Bayes approximation (e.g. Blei et al., 2017; Tran et al., 2021) of the joint posterior is often considered, which is the method to approximate an intractable posterior distribution by using a simpler probability distribution. Since the variational Bayes method does not require sampling from the posterior distribution like MCMC, and it searches for an optimal variational posterior by using the optimization method. In particular, we employed the mean-field variational Bayes (MFVB) approximation algorithms that require the forms of full conditional distributions as given in Section 3.2.2.

The variational distribution $q^*(\theta) \in \mathcal{Q}$ is defined by the minimizer of the Kullback-Leibler divergence from $q(\theta)$ to the true posterior distribution $p(\theta|y)$

$$q^* = \arg\min_{q \in \mathcal{Q}} \mathrm{KL}(q \| p(\cdot \mid y)) = \arg\min_{q \in \mathcal{Q}} \int q(\theta) \log \frac{q(\theta)}{p(\theta \mid y)} d\theta. \tag{3.6}$$

If θ is decomposed as $\theta = (\theta_1, \ldots, \theta_K)$ and parameters $\theta_1, \ldots, \theta_K$ are mutually independent, each variational posterior can be updated as

$$q(\theta_k) \propto \exp(\mathrm{E}_{\theta_{-k}}[\log p(y, \theta)]) = \exp\left(\int q(\theta_k) \log p(y, \theta) d\theta_{-k}\right), \quad k = 1, \ldots, K,$$

where θ_{-k} denotes the parameters other than θ_k and $\mathrm{E}_{\theta_{-k}}[\cdot]$ denotes the expectation under the probability density given parameters except for θ_k. Such a form of approximation is known as the MFVB approximation. If the full conditional distribution of θ_k has a familiar form, the above formula is easy to compute. According to the Gibbs sampling algorithm in Section 3.2.2, we used the following form of variational posteriors:

$$q(\theta, z, \sigma^2, \tau^2, \xi) = q(\theta) q(z) q(\sigma^2) q(\tau^2) q(\xi),$$

where

$$
\begin{aligned}
&q(\theta) \sim N_n(A^{-1}B, (E_{1/\sigma^2}A)^{-1}), \quad q(z_{ij}) \sim \mathrm{GIG}\left(\frac{1}{2}, \alpha_{z_{ij}}, \beta_{z_{ij}}\right), \\
&q(\sigma^2) \sim \mathrm{IG}\left(\frac{n+3N}{2} + a_\sigma, \alpha_{\sigma^2}\right), \quad q(\tau^2) \sim \mathrm{IG}\left(\frac{n-k}{2}, \alpha_{\tau^2}\right), \\
&q(\xi) \sim \mathrm{IG}\left(\frac{1}{2}, E_{1/\tau^2} + 1\right).
\end{aligned} \tag{3.7}
$$

For $i = 1, \ldots, k+1$, we assume the prior $\mathrm{IG}(a_{w_i}, b_{w_i})$ for w_i, and then the variational distribution of w_i is given by

$$
q(w_i^2) \sim \mathrm{IG}\left(\frac{1}{2} + a_{w_i}, \frac{1}{2}E_{\eta_i^2}E_{1/\sigma^2} + b_{w_i}\right).
$$

The variational distributions of the other parameters depended on the specific choice of the distributional form of $\pi(w_i)$ $(i = k+2, \ldots, n)$, which are provided as follows:

- **(Laplace-type prior)** The variational distributions for w_i^2 $(i = k+2, \ldots, n)$, γ^2 and ν are given by

$$
\begin{aligned}
&q(w_i^2) \sim \mathrm{GIG}\left(\frac{1}{2}, \alpha_{w_i^2}, E_{\gamma^2}\right), \\
&q(\gamma^2) \sim \mathrm{GIG}\left(n-k-\frac{3}{2}, 2E_{1/\nu}, \sum_{i=k+2}^{n} E_{w_i^2}\right), \quad q(\nu) \sim \mathrm{IG}\left(\frac{1}{2}, E_{1/\gamma^2} + 1\right),
\end{aligned}
$$

- **(Horseshoe-type prior)** The variational distributions for w_i^2 and ν_i $(i = k+2, \ldots, n)$ are given by

$$
q(w_i^2) \sim \mathrm{IG}(1, \alpha_{w_i^2}), \quad q(\nu_i) \sim \mathrm{IG}\left(\frac{1}{2}, E_{1/w_i^2} + 1\right),
$$

To obtain the variational parameters in each distribution, we update the parameters by using the coordinate ascent algorithm (e.g. Blei et al., 2017). The two proposed variational algorithms based on the above variational distributions are given in Algorithm 1 and 2. Note that we set $\epsilon = 10^{-4}$ as the convergence criterion in the simulation study, e_i is a unit vector that the ith component is 1, $d_i^\top$ is the ith row of difference matrix D, and $K_c(\cdot)$ is the modified Bessel function of the second kind with order c in Algorithms 1 and 2.

Algorithm 1 — Variational Bayes approximation under Laplace prior.

Initialize: $E_{z_{ij}}, E_{1/z_{ij}}, E_{1/w_i}, E_{1/\sigma^2}, E_{\gamma^2}, E_{1/\nu} > 0$ $(j = 1, \ldots, N_i, i = 1, \ldots, n)$. Set $E_{1/\tau^2} = 1, E_{1/\xi} = 0$ under Laplace prior.

1. Cycle the following:

(i)
$$A \leftarrow \frac{1}{t^2}\mathrm{diag}\left(\sum_{j=1}^{N_1} E_{1/z_{1j}}, \ldots, \sum_{j=1}^{N_n} E_{1/z_{nj}}\right) + D^\top \hat{W}^{-1} D,$$

$$B \leftarrow \frac{1}{t^2}\left(C - \psi 1_n\right), \quad C \leftarrow \left(\sum_{j=1}^{N_1} y_{1j} E_{1/z_{1j}}, \ldots, \sum_{j=1}^{N_n} y_{nj} E_{1/z_{nj}}\right)^\top,$$

$$\hat{W}^{-1} \leftarrow \mathrm{diag}(E_{1/w_1^2}, \ldots, E_{1/w_{k+1}^2}, E_{1/\tau^2}E_{1/w_{k+2}^2}, \ldots, E_{1/\tau^2}E_{1/w_n^2}),$$

$$E_{\theta_i} \leftarrow (A^{-1}B)_i, \quad E_{\theta_i^2} \leftarrow e_i^\top (E_{\sigma^2}^{-1} A^{-1} + A^{-1} B B^\top A^{-1}) e_i,$$

$$E_{\eta_i^2} \leftarrow d_i^\top (E_{\sigma^2}^{-1} A^{-1} + A^{-1} B B^\top A^{-1}) d_i \quad (i = 1, \ldots, n),$$

$$\alpha_{\sigma^2} \leftarrow \frac{1}{2t^2}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(y_{ij}^2 E_{1/z_{ij}} - 2\psi y_{ij} + \psi^2 E_{z_{ij}} - 2(E_{1/z_{ij}} y_{ij} - \psi)E_{\theta_i} + E_{\theta_i^2}E_{1/z_{ij}}\right)$$

$$+ \frac{1}{2}\sum_{i=1}^{k+1} E_{\eta_i^2}E_{1/w_i^2} + \frac{1}{2}\sum_{i=k+2}^{n} E_{\eta_i^2}E_{1/w_i^2}E_{1/\tau^2} + \sum_{i=1}^{n}\sum_{j=1}^{N_i} E_{z_{ij}} + b_\sigma,$$

$$E_{\sigma^2} \leftarrow \frac{2\alpha_{\sigma^2}}{n + 3N + 2a_\sigma - 2}, \quad E_{1/\sigma^2} \leftarrow \frac{n + 3N + 2a_\sigma}{2\alpha_{\sigma^2}},$$

$$\alpha_{z_{ij}} \leftarrow \frac{1}{t^2}(y_{ij}^2 - 2y_{ij}E_{\theta_i} + E_{\theta_i^2})E_{1/\sigma^2}, \quad \beta_{z_{ij}} \leftarrow \left(\frac{\psi^2}{t^2} + 2\right)E_{1/\sigma^2},$$

$$E_{z_{ij}} \leftarrow \frac{\sqrt{a_{z_{ij}}}K_{3/2}(\sqrt{a_{z_{ij}}b_{z_{ij}}})}{\sqrt{b_{z_{ij}}}K_{1/2}(\sqrt{a_{z_{ij}}b_{z_{ij}}})},$$

$$E_{1/z_{ij}} \leftarrow \frac{\sqrt{b_{z_{ij}}}K_{3/2}(\sqrt{a_{z_{ij}}b_{z_{ij}}})}{\sqrt{a_{z_{ij}}}K_{1/2}(\sqrt{a_{z_{ij}}b_{z_{ij}}})} - \frac{1}{a_{z_{ij}}} \quad (j = 1, \ldots, N_i, i = 1, \ldots, n),$$

$$E_{1/w_i^2} \leftarrow (1 + 2a_{w_i})/(E_{\eta_i^2}E_{1/\sigma^2} + 2b_{w_i}) \quad (i = 1, \ldots, k+1)$$

(ii)
$$\alpha_{w_i^2} \leftarrow E_{1/\sigma^2}E_{\eta_i^2}, \quad E_{w_i^2} \leftarrow \frac{\sqrt{\alpha_{w_i^2}}K_{3/2}(\sqrt{\alpha_{w_i^2}E_{\gamma^2}})}{\sqrt{E_{\gamma^2}}K_{1/2}(\sqrt{\alpha_{w_i^2}E_{\gamma^2}})},$$

$$E_{1/w_i^2} \leftarrow \frac{\sqrt{E_{\gamma^2}}K_{3/2}(\sqrt{a_{w_i^2}E_{\gamma^2}})}{\sqrt{a_{w_i^2}}K_{1/2}(\sqrt{a_{w_i^2}E_{\gamma^2}})} - \frac{1}{\alpha_{w_i^2}} \quad (i = k+2, \ldots, n),$$

$$E_{\gamma^2} \leftarrow \frac{\sqrt{2E_{1/\nu}}K_{n-k-1/2}(\sqrt{2E_{1/\nu}\sum_{i=k+2}^{n}E_{w_i^2}})}{\sqrt{\sum_{i=k+2}^{n}E_{w_i^2}}K_{n-k-3/2}(\sqrt{2E_{1/\nu}\sum_{i=k+2}^{n}E_{w_i^2}})},$$

$$E_{1/\gamma^2} \leftarrow \frac{\sqrt{\sum_{i=k+2}^{n}E_{w_i^2}}K_{n-k-1/2}(\sqrt{2E_{1/\nu}\sum_{i=k+2}^{n}E_{w_i^2}})}{\sqrt{2E_{1/\nu}}K_{n-k-3/2}(\sqrt{2E_{1/\nu}\sum_{i=k+2}^{n}E_{w_i^2}})},$$

$$E_{1/\nu} \leftarrow \frac{1}{2(E_{1/\gamma^2} + 1)}.$$

2. For iteration ℓ in step 1 and convergence criterion $\epsilon > 0$, if $|E_{\theta_i}^{(\ell)} - E_{\theta_i}^{(\ell-1)}| < \epsilon$, stop the algorithm.

Algorithm 2 — Variational Bayes approximation under horseshoe prior.

Initialize: $E_{z_{ij}}, E_{1/z_{ij}}, E_{1/w_i}, E_{1/\tau^2}, E_{1/\sigma^2}, E_{1/\xi^2}, E_{1/\nu_i} > 0$ $(j = 1, \ldots, N_i, i = 1, \ldots, n)$.

1. Cycle the following:

   (i) Same update as (i) in Algorithm 1.

   (ii) $$\alpha_{w_i^2} \leftarrow E_{1/\nu_i} + \frac{1}{2} E_{1/\sigma^2} E_{1/\tau^2} E_{\eta_i^2},$$
   $$E_{1/w_i^2} \leftarrow \frac{1}{a_{w_i^2}}, \quad E_{1/\nu_i^2} \leftarrow \frac{1}{2(E_{1/w_i^2} + 1)} \quad (i = k+2, \ldots, n),$$
   $$\alpha_{\tau^2} \leftarrow \frac{1}{2} \sum_{i=k+2}^{n} E_{\eta_i^2} E_{1/w_i^2} E_{1/\sigma^2} + E_{1/\xi}, \quad E_{1/\tau^2} \leftarrow \frac{n-k}{2\alpha_{\tau^2}},$$
   $$E_{1/\xi} \leftarrow \frac{1}{2(E_{1/\tau^2} + 1)}.$$

2. For some iteration ℓ in step 1 and convergence criterion $\epsilon > 0$, if $|E_{\theta_i}^{(\ell)} - E_{\theta_i}^{(\ell-1)}| < \epsilon$, stop the algorithm.

3.3 Calibrated variational Bayes approximation

The main proposal of this study is described below. When we use the mean field variational Bayes method, the posterior credible intervals are calculated based on the quantile of the variational posterior. In the proposed model, the variational distribution of the parameter of interest θ_i is represented by the normal distribution $N(\mu_i, \Sigma_{ii})$, where the mean μ_i and variance Σ_{ii} are defined in Section 3.2.3. Although the variational approximation provides the point estimate quickly, the corresponding credible interval tends to be narrow in general (e.g. Wand et al., 2011; Blei et al., 2017). Additionally, it is well-known that the credible interval can be affected by model misspecification, as addressed by Sriram et al. (2013) and Sriram (2015) in the Bayesian linear quantile regression. Hence, if the asymmetric Laplace working likelihood in the proposed model is misspecified, the proposed model would not have been able to provide valid credible intervals even if we use the MCMC algorithm.

As presented in the previous subsection, the conditional prior and likelihood of θ were given by (3.4) and (3.5), respectively. Here we add a common (non-random)

scale parameter λ, and then replace (3.4) and (3.5) with

$$p(y_{ij} \mid \theta_i, z_{ij}, \sigma^2) = (2\pi t^2 \lambda \sigma^2)^{-1/2} z_{ij}^{-1/2} \exp\left\{ -\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 \lambda \sigma^2 z_{ij}} \right\},$$
$$p(\theta \mid \sigma^2, \tau, w) = (2\pi\lambda\sigma^2)^{-n/2} |D^\top W^{-1} D|^{1/2} \exp\left(-\frac{1}{2\lambda\sigma^2} \theta^\top D^\top W^{-1} D\theta \right),$$

respectively. Based on these representations, the variational posterior of θ is given by

$$q(\theta) \sim N_n(\mu, \lambda\Sigma).$$

The constant λ in the likelihood and conditional prior controls the scale of the variational posterior. Indeed, it is natural that the scale of the posteriors was determined by the scale of the likelihood and prior. If the scale parameter λ is given locally for each θ_i (i.e. λ_i), then the variational posterior of θ_i is also given by $q(\theta_i) \sim N_n(\mu_i, \lambda_i \Sigma_{ii})$ for each i. We used the formulation to calibrate credible intervals after the point estimation. The proposed calibration algorithm is given in Algorithm 3.

Algorithm 3 is similar to the calibration method for general Bayes credible regions proposed by Syring and Martin (2019), but the proposed algorithm drastically differs from the existing calibration method in that it computes variational Bayes posteriors for B times while the calibration method by Syring and Martin (2019) runs MCMC algorithms for B times. Thus, the proposed algorithm is computationally much faster than the existing calibration algorithm. Further, compared with the Gibbs sampler presented in Section 3.2.2, steps 3 and 4 in Algorithm 3 can be parallelized so that a significant reduction of computational costs can be attained with the proposed method.

After we obtain the optimal value of λ using Algorithm 3, we use $q(\theta) \sim N_n(\mu^*, \lambda\Sigma^*)$ as the calibrated variational posterior distribution. We then construct the calibrated credible interval of θ_i $(i = 1, \ldots, n)$ by calculating the quantile of the marginal distribution of $N_n(\mu^*, \lambda\Sigma^*)$. Here we used the residual bootstrap method (e.g. Efron, 1982) to obtain bootstrap samples in Algorithm 3. Since this algorithm is based on such semiparametric bootstrap sampling, robust uncertainty quantification can be expected even if the asymmetric Laplace assumption is violated.

Remark 3.3.1. In Algorithm 3, we employed the residual bootstrap using 50%

Algorithm 3 — Calibration of variational posterior.

For calibration of variational posterior at the quantile level p, we set the monotonically increasing sequence $1 = \lambda_1 < \lambda_2 < \dots$ and run the following four steps:

1. Estimate the variational posterior for p-th quantile trend filtering $q(\theta) \approx N_n(\mu^*, \Sigma^*)$ using the observed data $y = (y_1, \dots, y_n)$.

2. Run the variational algorithm and estimate the variational posterior for 0.5-th quantile $q(\theta) \approx N_n(\mu_{50\%}, \Sigma_{50\%})$ using the observed data y.

3. Generate B bootstrap samples $y^{(1)}, \dots, y^{(B)}$ based on the residuals $y - \mu_{50\%}$, and calculate the variational posteriors as $N_n(\mu^{(1)}, \Sigma^{(1)}), \dots, N_n(\mu^{(B)}, \Sigma^{(B)})$ using bootstrap samples $y^{(1)}, \dots, y^{(B)}$.

4. Regarding $\{\mu^{(1)}, \dots, \mu^{(B)}\}$ as B posterior samples, for $i = 1, \dots, n$, evaluate the empirical coverage probability

$$\hat{c}_{\alpha,i}(\lambda_\ell) = \frac{1}{B} \sum_{b=1}^{B} 1\{\mu_i^{(b)} \in C_\alpha(\mu_i^*, \lambda_\ell \Sigma_{ii}^*)\},$$

   where $C_{\alpha,i}(\mu_i^*, \lambda_\ell \Sigma_{ii}^*)$ is $100(1-\alpha)\%$ credible intervals under $N(\mu_i^*, \lambda_\ell \Sigma_{ii}^*)$ ($\ell = 1, 2, \dots$). Then, selecte the optimal value $\hat{\lambda}$ so that

$$\hat{\lambda}_i = \mathrm{argmin}_\lambda \{\hat{c}_{\alpha,i}(\lambda) - (1-\alpha)\},$$

   and $q(\theta_i) \sim N_n(\mu_i^*, \hat{\lambda}_i \Sigma_{ii}^*)$ is the calibrated variational posterior distribution.

---

quantile trend estimate as a fitted value when we estimated any quantile level. At first glance, it might seem like it is better to use bootstrap sampling based on residue $y - \mu_p$, where μ_p is p-th quantile trend estimate. However, since we aim to re-sample from the empirical distribution of the original dataset y, the use of a 50% quantile trend estimated as a fitted value in residual bootstrap is reasonable in practice. This is the critical point of Algorithm 1.

To show the theoretical results of Algorithm 3 is not easy as well as the algorithm by Syring and Martin (2019) because it needs to evaluate the approximation errors of both bootstrap and variational approximations. We confirm the proposed algorithm through numerical experiments in the next section.

3.4 Simulation studies

We illustrate the performance of the proposed method through simulation studies.

3.4.1 Simulation setting

To evaluate the performance of the proposed methods, we considered the following data-generating processes (see also Faulkner and Minin, 2018; Brantley et al., 2020):

$$y_i = f(x_i) + \varepsilon(x_i), \quad i = 1, \ldots, 100,$$

where $f(x)$ is a true function and $\varepsilon(x)$ is a noise function. First, we assumed the following two true functions:

- **Piecewise constant (PC)**

$$\begin{aligned} f(x) = 2.5 \cdot I(1 \leq x \leq 20) + I(21 \leq x \leq 40) \\ + 3.5 \cdot I(41 \leq x \leq 60) + 1.5 \cdot I(61 \leq x \leq 100) \end{aligned}$$

- **Varying smoothness (VS)**

$$f(x) = 2 + \sin(4x - 2) + 2\exp(-30(4x - 2)^2).$$

Since the scenario (PC) has three change points at $x = 21$, 41, and 61, we aim to assess the ability to capture a jumping structure. The second scenario (VS) is smooth and has a rapid change near $x = 50$, which is reasonable to confirm the shrinkage effect of the proposed methods and the adaptation of localized change. As noise functions, we considered the following three scenarios that represented the heterogeneous variance and various types of model misspecification.

(I) Gaussian noise: $\varepsilon(x) \sim N(0, (1 + x^2)^2/16)$.

(II) Beta noise: $\varepsilon(x) \sim \text{Beta}(1, 11 - 10x)$.

(III) Mixed normal noise: $\varepsilon(x) \sim xN(-0.2, 0.5) + (1 - x)N(0.2, 0.5)$.

For each scenario, simulated true quantile trends are summarized in Figure 3.5 of the Appendix 3.7. True quantile trends were computed from the quantiles of pointwise noise distributions. We next introduce the details of simulations. We used the two MCMC methods (denoted by MCMC-HS and MCMC-Lap), two non-calibrated variational Bayes methods (denoted by VB-HS and VB-Lap), and two calibrated variational Bayes methods (denoted by CVB-HS and CVB-Lap), where HS and Lap are the horseshoe and Laplace priors, respectively. Note that we implemented

CVB without parallelization although the bootstrap calibration steps in CVB can be parallelized. To compare with the frequentist method, we used the quantile trend filtering based on the ADMM algorithm proposed by Brantley et al. (2020), where the penalty parameter of Brantley's method was determined by the extended Bayesian information criteria. The method can be implemented using their `R` package in `https://github.com/halleybrantley/detrendr`. For the order of trend filtering, we considered $k = 0$ for (PC) and $k = 1$ for (VS). Note that $k = 0, 1$ express the piecewise constant and the piecewise linear, respectively. We generated 7,500 posterior samples by using the Gibbs sampler presented in Section 3.2.2, and then only every 10th scan was saved (thinning). As criteria to measure the performance, we adopted the mean squared error (MSE), mean absolute deviation (MAD), mean credible interval width (MCIW), and coverage probability (CP) which are defined by

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\theta_i - \hat{\theta}_i)^2, \quad \text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|\theta_i - \hat{\theta}_i|,$$
$$\text{MCIW} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{97.5,i} - \hat{\theta}_{2.5,i}, \quad \text{CP} = \frac{1}{n}\sum_{i=1}^{n} I(\hat{\theta}_{2.5,i} \leq \theta_i^* \leq \hat{\theta}_{97.5,i}),$$

respectively, where $\hat{\theta}_{100(1-\alpha),i}$ represent the $100(1-\alpha)\%$ posterior quantiles of θ_i and θ_i^* are true quantiles of y at location x_i. Additional simulation results under a different true function are provided in the Appendix 3.7.

3.4.2 Simulation results

We show the simulation results for each scenario. Note that the point estimates of the variational Bayes method were the same as those of the calibrated variational method because the difference between them was only the variance of the variational posterior distribution. Hence, we omitted the results of the CVB-HS and CVB-Lap in Tables 3.1 and 3.3. The frequentist quantile trend filtering by Brantley et al. (2020) is denoted by "ADMM".

Piecewise constant. We summarized the numerical results of the point estimate and uncertainty quantification in Tables 3.1 and 3.2, respectively. From Table 3.1, we observed that the point estimates of the MCMC-HS method performed the best in all cases, and the frequentist ADMM method performed the worst in terms of MSE and MAD. For uncertainty quantification, it was shown that the MCMC methods

have reasonable coverage probabilities for center quantiles such as 0.25, 0.5, and 0.75 except for the case of beta distributed noise, while the MCMC methods for extremal quantiles such as 0.05 and 0.95 appear to be far away from the nominal coverage rate 0.95. The MCIW of the VB-HS and VB-Lap methods tended to be shorter than that of the MCMC therefore, the corresponding coverage probabilities were extremely underestimated. However, the CVB-HS and CVB-Lap methods could quantify the uncertainty in almost all cases including extremal quantiles. We also show one-shot simulation results under the Gaussian noise in Figure 3.1. As shown in the figure, the credible intervals of CVB-HS are similar to those of MCMC-HS for 0.25, 0.50, and 0.75 quantiles. Furthermore, the calibrated credible intervals by CVB-HS are wider than those of the MCMC for extremal quantiles.

Table 3.1: Average values of MSE and MAD based on 100 replications for piecewise constant with $k = 0$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.046**	**0.013**	**0.009**	**0.013**	**0.046**	**0.168**	**0.083**	**0.069**	**0.083**	**0.168**
VB-HS	0.094	*0.033*	*0.026*	*0.034*	*0.053*	*0.198*	*0.133*	*0.112*	*0.132*	*0.182*
MCMC-Lap	0.105	0.050	0.040	0.050	0.106	0.263	0.173	0.154	0.172	0.266
VB-Lap	*0.091*	0.042	0.033	0.038	0.059	0.213	0.155	0.139	0.149	0.191
ADMM	0.384	0.040	0.045	0.102	0.304	0.357	0.147	0.168	0.190	0.343
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.004*	**0.003**	**0.004**	**0.007**	*0.020*	*0.026*	**0.027**	**0.040**	**0.056**	*0.109*
VB-HS	**0.001**	*0.006*	*0.009*	*0.014*	*0.020*	**0.014**	*0.039*	*0.058*	*0.082*	0.112
MCMC-Lap	0.013	0.015	0.015	0.019	0.038	0.078	0.081	0.086	0.102	0.156
VB-Lap	0.059	0.009	0.011	*0.014*	**0.019**	0.083	0.060	0.072	0.086	**0.106**
ADMM	0.793	**0.003**	0.011	0.101	0.442	0.437	0.041	0.085	0.164	0.411
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.089**	**0.036**	**0.029**	**0.039**	**0.102**	**0.235**	**0.136**	**0.123**	**0.141**	**0.250**
VB-HS	0.247	0.080	*0.067*	0.085	*0.125*	0.316	*0.209*	*0.190*	0.216	0.275
MCMC-Lap	0.191	0.094	0.078	0.095	0.204	0.365	0.240	0.217	0.242	0.379
VB-Lap	*0.137*	*0.079*	0.069	*0.079*	0.126	*0.282*	0.213	0.201	*0.215*	*0.273*
ADMM	0.315	0.183	0.123	0.170	0.236	0.386	0.271	0.266	0.284	0.351

Varying smoothness. The results for the (VS) scenario are reported in Tables 3.3 and 3.4. From Table 3.3, the MCMC-HS method also performed well relative to the other methods in terms of point estimation, while the variational Bayes methods under horseshoe prior also provided comparable point estimates. Different from the (PC) scenario, the MCMC methods had slightly worse coverage probabilities. In particular, the MCMC-HS and MCMC-Lap under the mixed normal noise, which

Table 3.2: Average values of MCIW and CP based on 100 replications for piecewise constant with $k = 0$. The CP values above 90% are represented in bold.

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.549	0.447	0.412	0.437	0.543	0.757	**0.950**	**0.970**	**0.943**	0.761
CVB-HS	1.034	0.620	0.471	0.596	1.015	**0.929**	**0.924**	**0.904**	**0.920**	**0.922**
VB-HS	0.117	0.205	0.221	0.205	0.117	0.189	0.477	0.611	0.462	0.187
MCMC-Lap	0.775	0.764	0.785	0.759	0.762	0.806	**0.923**	**0.960**	**0.926**	0.800
CVB-Lap	1.039	0.822	0.606	0.792	0.990	**0.926**	**0.952**	**0.910**	**0.953**	**0.941**
VB-Lap	0.404	0.562	0.596	0.556	0.408	0.568	0.852	**0.907**	0.871	0.607
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.113	0.137	0.186	0.245	0.295	**0.975**	**0.959**	**0.940**	**0.916**	0.676
CVB-HS	0.282	0.202	0.207	0.360	0.640	**0.992**	**0.938**	0.864	0.894	**0.926**
VB-HS	0.034	0.069	0.094	0.099	0.063	0.805	0.682	0.573	0.405	0.167
MCMC-Lap	0.377	0.326	0.326	0.309	0.390	**0.952**	**0.927**	**0.902**	0.807	0.756
CVB-Lap	0.281	0.305	0.307	0.385	0.407	**0.908**	**0.929**	0.899	0.889	0.835
VB-Lap	0.214	0.274	0.307	0.296	0.264	0.855	0.916	0.899	0.827	0.676
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.871	0.730	0.696	0.730	0.863	0.797	**0.943**	**0.972**	**0.947**	0.784
CVB-HS	1.536	0.926	0.780	0.942	1.577	**0.929**	**0.931**	**0.919**	**0.928**	**0.955**
VB-HS	0.177	0.318	0.342	0.316	0.177	0.212	0.453	0.518	0.434	0.206
MCMC-Lap	1.129	1.135	1.135	1.125	1.097	0.830	**0.940**	**0.963**	**0.942**	0.795
CVB-Lap	1.573	1.155	0.891	1.159	1.572	**0.947**	**0.960**	**0.920**	**0.962**	**0.954**
VB-Lap	0.548	0.778	0.818	0.770	0.556	0.575	0.852	0.897	0.851	0.595

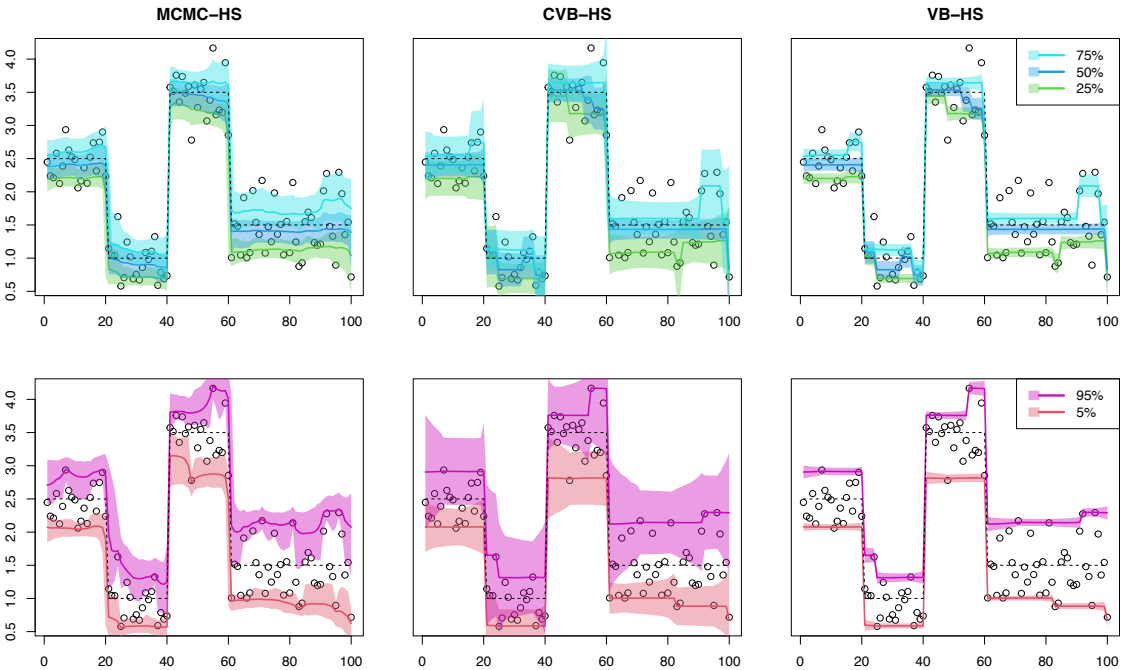

Figure 3.1: One-shot simulation results under piecewise constant and Gauss noise. The order of trend filtering is $k = 0$ for all methods.

is a relatively high degree of misspecification, appeared to be far from the nominal coverage rate of 0.95. Although the MCIW of the variational Bayes methods without calibration also tended to be shorter than that of MCMC, the calibrated variational Bayes dramatically improved the coverage even under the mixed normal case. We also show one-shot simulation results under the Gaussian noise in Figure 3.2.

Finally, we assessed the efficiency of posterior computation. To this end, we calculated the raw computing time and effective sample size per unit time. The latter is defined as the effective sample size divided by the computation time in seconds. Note that the effective sample size for the variational Bayes methods (VB and CVB) was 7,500 since i.i.d. samples could be drawn from variational posterior distributions. The values averaged over 100 replications of simulating datasets are presented in Table 3.5. The results show that the proposed algorithm provides posterior samples much more efficiently than the MCMC algorithm. Such computationally efficient property of the proposed method is a benefit of a novel combination of variational approximation and posterior calibration.

Table 3.3: Average values of MSE and MAD based on 100 replications for varying smoothness with $k = 1$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.068*	*0.034*	**0.017**	**0.019**	**0.041**	**0.179**	**0.119**	**0.097**	**0.104**	**0.156**
VB-HS	0.133	**0.026**	*0.020*	*0.025*	*0.056*	0.229	**0.119**	*0.105*	*0.117*	*0.180*
MCMC-Lap	**0.064**	0.039	0.026	0.028	0.057	*0.194*	*0.138*	0.122	0.128	0.188
VB-Lap	0.097	0.052	0.027	0.028	0.058	0.209	0.143	0.121	0.128	0.190
ADMM	0.177	0.075	0.031	0.035	0.097	0.237	0.163	0.123	0.137	0.230
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.004**	**0.004**	**0.004**	**0.007**	**0.014**	**0.042**	**0.037**	**0.046**	**0.058**	**0.092**
VB-HS	0.101	*0.005*	*0.006*	*0.009*	*0.020*	0.107	0.044	*0.053*	*0.069*	*0.106*
MCMC-Lap	*0.006*	*0.005*	*0.006*	*0.009*	0.022	*0.054*	0.046	0.057	0.072	0.115
VB-Lap	0.022	*0.005*	*0.006*	0.010	0.021	0.067	*0.043*	0.056	0.072	0.113
ADMM	0.204	0.057	0.011	0.018	0.094	0.171	0.100	0.068	0.088	0.199
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.147*	0.130	0.077	**0.055**	**0.090**	**0.253**	0.214	0.181	**0.172**	**0.230**
VB-HS	0.167	**0.062**	**0.047**	*0.056*	0.114	0.279	**0.181**	**0.162**	*0.178*	*0.258*
MCMC-Lap	**0.108**	*0.085*	*0.061*	0.060	*0.112*	*0.255*	*0.198*	*0.180*	0.188	0.265
VB-Lap	0.147	0.106	0.072	0.066	0.114	0.274	0.206	0.186	0.194	0.263
ADMM	0.195	0.088	0.057	0.066	0.138	0.284	0.202	0.174	0.192	0.287

Table 3.4: Average values of MCIW and CP based on 100 replications for varying smoothness with $k = 1$. The CP values above 90% are represented in bold.

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.498	0.476	0.447	0.444	0.446	0.730	0.890	**0.930**	**0.903**	0.728
CVB-HS	1.010	0.592	0.496	0.572	1.097	**0.911**	**0.935**	**0.936**	**0.937**	**0.956**
VB-HS	0.124	0.196	0.209	0.195	0.118	0.206	0.505	0.584	0.518	0.215
MCMC-Lap	0.545	0.571	0.562	0.553	0.524	0.760	0.897	**0.929**	**0.914**	0.756
CVB-Lap	1.335	0.890	0.602	0.821	1.240	**0.961**	**0.960**	**0.937**	**0.977**	**0.968**
VB-Lap	0.207	0.326	0.369	0.348	0.227	0.346	0.705	0.788	0.726	0.364
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.162	0.178	0.213	0.241	0.260	**0.931**	**0.941**	**0.925**	0.898	0.713
CVB-HS	0.445	0.259	0.209	0.347	0.724	**0.927**	**0.950**	0.884	**0.935**	**0.969**
VB-HS	0.051	0.085	0.103	0.104	0.065	0.502	0.656	0.606	0.487	0.219
MCMC-Lap	0.251	0.248	0.282	0.306	0.304	0.953	0.952	0.942	0.903	0.736
CVB-Lap	0.522	0.362	0.268	0.458	0.673	**0.968**	**0.976**	**0.923**	**0.969**	**0.954**
VB-Lap	0.108	0.164	0.199	0.201	0.136	0.723	0.874	0.854	0.756	0.368
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.600	0.636	0.654	0.655	0.666	0.692	0.838	0.874	0.876	0.743
CVB-HS	1.271	0.854	0.747	0.850	1.540	**0.909**	**0.926**	**0.927**	**0.934**	**0.961**
VB-HS	0.172	0.286	0.303	0.284	0.178	0.211	0.490	0.558	0.496	0.226
MCMC-Lap	0.727	0.755	0.765	0.768	0.755	0.753	0.885	**0.907**	0.891	0.749
CVB-Lap	1.780	1.214	0.891	1.249	1.972	**0.963**	**0.942**	**0.924**	**0.969**	**0.979**
VB-Lap	0.279	0.400	0.458	0.455	0.324	0.330	0.647	0.711	0.677	0.394

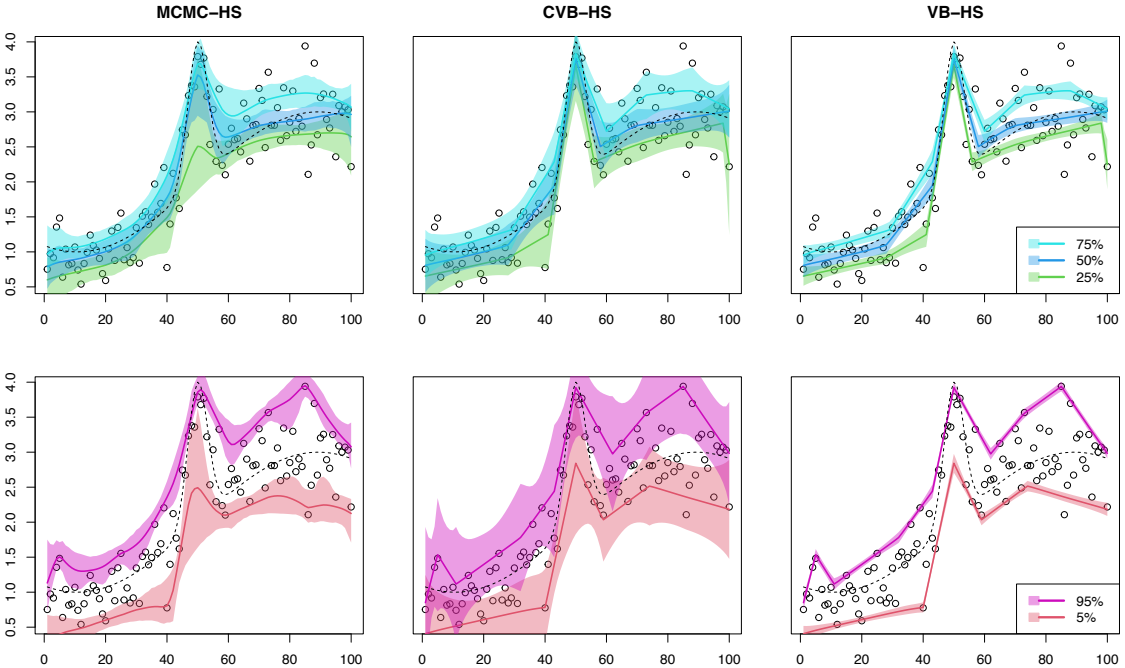

Figure 3.2: One-shot simulation results under varying smoothness and Gauss noise. The order of trend filtering is $k = 1$ for all methods.

Table 3.5: Average values of raw computing time and effective sample size per unit time based on 100 replications for all scenarios.

(PC) Piecewise constant										
	Computation time (second)					ESS (per second)				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	33	32	32	32	33	13	39	45	40	13
CVB-HS	13	9	11	9	12	603	869	699	867	613
MCMC-Lap	37	36	36	36	37	12	81	109	82	13
CVB-Lap	12	9	9	7	11	662	848	862	1120	687
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	32	32	32	33	13	61	59	41	13
CVB-HS	8	7	9	8	10	1006	1089	874	991	792
MCMC-Lap	37	37	37	37	37	11	79	126	74	12
CVB-Lap	6	5	8	4	6	1306	1654	894	1791	1287
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	31	31	32	32	13	35	40	35	13
CVB-HS	16	11	12	10	15	493	724	606	731	497
MCMC-Lap	36	36	36	36	36	13	78	98	79	14
CVB-Lap	15	12	9	9	13	513	634	807	883	586
(VS) Varying smoothness										
	Computation time (second)					ESS (per second)				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	32	32	32	32	13	31	37	35	14
CVB-HS	12	8	11	8	13	620	897	679	923	575
MCMC-Lap	36	35	35	35	36	13	53	72	65	13
CVB-Lap	16	13	11	11	15	471	588	684	713	519
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	33	33	33	33	33	12	48	45	38	14
CVB-HS	9	7	10	7	11	861	1118	780	1075	699
MCMC-Lap	36	36	36	36	36	11	78	105	69	12
CVB-Lap	11	6	9	6	9	740	1202	813	1181	879
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	32	32	32	32	14	25	30	30	15
CVB-HS	14	10	13	10	16	555	756	593	779	481
MCMC-Lap	36	36	36	36	36	14	36	49	51	14
CVB-Lap	22	20	17	19	23	359	392	463	412	335

3.5 Real data analysis

We apply the proposed method to two real data examples.

3.5.1 Nile data

We first applied the proposed methods to the famous Nile river data (Cobb, 1978; Balke, 1993). The data contains measurements of the annual flow of the river Nile from 1871 to 1970, and we found an apparent change point near 1898. We considered $k = 0$ and compared the three methods, that is MCMC-HS, CVB-HS, and ADMM. We generated 60,000 posterior samples after discarding the first 10,000 posterior samples as burn-in, and then only every 10th scan was saved. For the Bayesian methods, we adopted $(a, b) = (1, 3)$ as hyperparameters in the inverse gamma prior to σ^2. The resulting estimates of quantiles and the corresponding 95% credible intervals are shown in Figure 3.3. In terms of point estimation, the horseshoe prior appears to capture the piecewise constant structures well, and the point estimates of CVB-HS and ADMM are comparable for all quantiles. For uncertainty quantification, the lengths of credible intervals of the MCMC-HS and CVB-HS are comparable for 25%, 50%, and 75% quantiles, while the CVB-HS method has wider credible intervals than those of the MCMC method, especially for extremal quantiles such as 5% and 95% (see also Table 3.6). This is consistent with the simulation results in Section 3.4.2. In Table 3.7, we provided the effective sample size per unit time of the proposed algorithm and MCMC, which showed significant improvement of computational efficiency by the proposed method. Hence, we could conclude that the proposed algorithm performs better than the MCMC for this application, in terms of both qualities of inference and computational efficiency.

Table 3.6: Average lengths of credible intervals for real data examples.

	Nile data (Section 3.5.1)					Munich rent data (Section 3.5.2)				
	0.05	0.25	0.5	0.75	0.95	0.1	0.3	0.5	0.7	0.9
MCMC-HS	0.10	0.10	0.10	0.11	0.14	0.98	0.88	0.93	0.90	0.95
CVB-HS	0.23	0.12	0.10	0.10	0.21	2.33	1.37	1.15	1.30	1.88

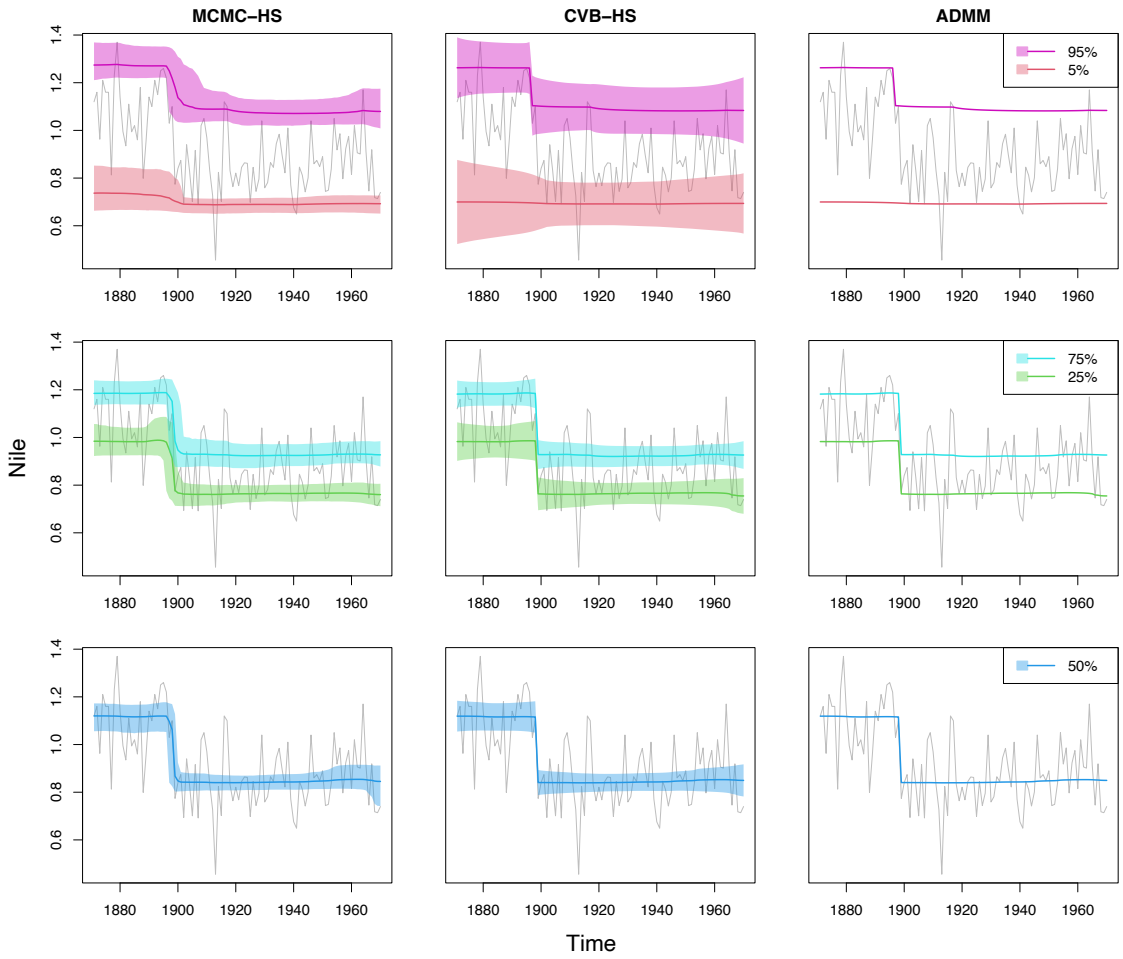

Figure 3.3: Point estimates and 95% credible intervals for Nile data.

Table 3.7: Effective sample size per unit time for real data examples.

	Nile data (Section 3.5.1)					Munich rent data (Section 3.5.2)				
	0.05	0.25	0.5	0.75	0.95	0.1	0.3	0.5	0.7	0.9
MCMC-HS	4	7	7	6	5	4	7	7	6	5
CVB-HS	47619	52632	47619	55556	48780	4029	6233	1943	6124	2949

3.5.2 Munich rent data

The proposed methods can also be applied to multiple observations with an irregular grid. We used Munich rent data (`https://github.com/jrfaulkner/spmrf`) which includes the value of rent per square meter and floor space in Munich, Germany (see also Rue and Held, 2005; Faulkner and Minin, 2018; Heng et al., 2023). The data has multiple observations per location and an irregular grid. Let the response $y = (y_1, \ldots, y_n)$ be the rent and the location $x = (x_1, \ldots, x_n)$ be the floor size. At the location x_i, the response y_i has multiple observations per location, that is, $y_i = (y_{i1}, \ldots, y_{iN_i})^\top \in \mathbb{R}^{N_i}$. Furthermore, the difference $x_{j+1} - x_j$ is not always constant, therefore the floor spaces are unequally spaced. This is a different situation from the example in Section 3.5.1. The data contains $N = \sum_{i=1}^{n} N_i = 2,035$ observations and the floor space (or location) has 134 distinct values. We applied the third-order adjusted difference operator defined in Remark 3.2.1 to the proposed Bayesian quantile trend filtering methods (i.e. MCMC-HS and CVB-HS with $k = 2$). Since Brantley's quantile trend filtering method (Brantley et al., 2020) cannot be applied to the data with multiple observations per location, we applied the quantile smoothing spline method by Nychka et al. (2017) as a frequentist competitor. The method could be implemented by using `qsreg` function in `R` package `fields`. The details of the method are provided in Nychka et al. (1995) and Oh et al. (2004), and the smoothing parameter was chosen by using cross-validation. For these methods, we analyzed the five quantile levels such as 10%, 30%, 50%, 70% and 90%. For the Bayesian methods, we generated 60,000 posterior samples after discarding the first 10,000 posterior samples as burn-in, and then only every 10th scan was saved.

The results of the point estimate and credible interval are shown in Figure 3.4. The frequentist smoothing spline method is denoted by "Spline" in Figure 3.4. The CVB-HS and Spline methods gave comparable baseline estimates, while the MCMC-HS method provided slightly smoother point estimates than the other two methods, especially for the large floor size. These decreasing trends mean that the houses with small floor sizes have a greater effect on their rent. Such a trend was also observed in the Bayesian mean trend filtering by Faulkner and Minin (2018). Compared with the MCMC-HS, the CVB-HS method has a wider length of 95% credible intervals for large values of floor size. The phenomenon appears to be reasonable because the data are less in such regions. Additionally, two Bayesian methods provided almost the same results for the center quantile levels such as 30%, 50%, and 70%, while the

credible intervals of CVB-HS were wider than those of the MCMC-HS, especially for extremal quantile levels such as 90% and 10% (see also Table 3.6). This indicates that the MCMC-HS method possibly underestimated extremal quantile regions. We again computed the effective sample size per unit time of the proposed algorithm and the MCMC, and the results are given in Table 3.7. From the results, we concluded that the proposed algorithm performs better than the MCMC for this application, in terms, not only of the quality of inference but also of computational efficiency.

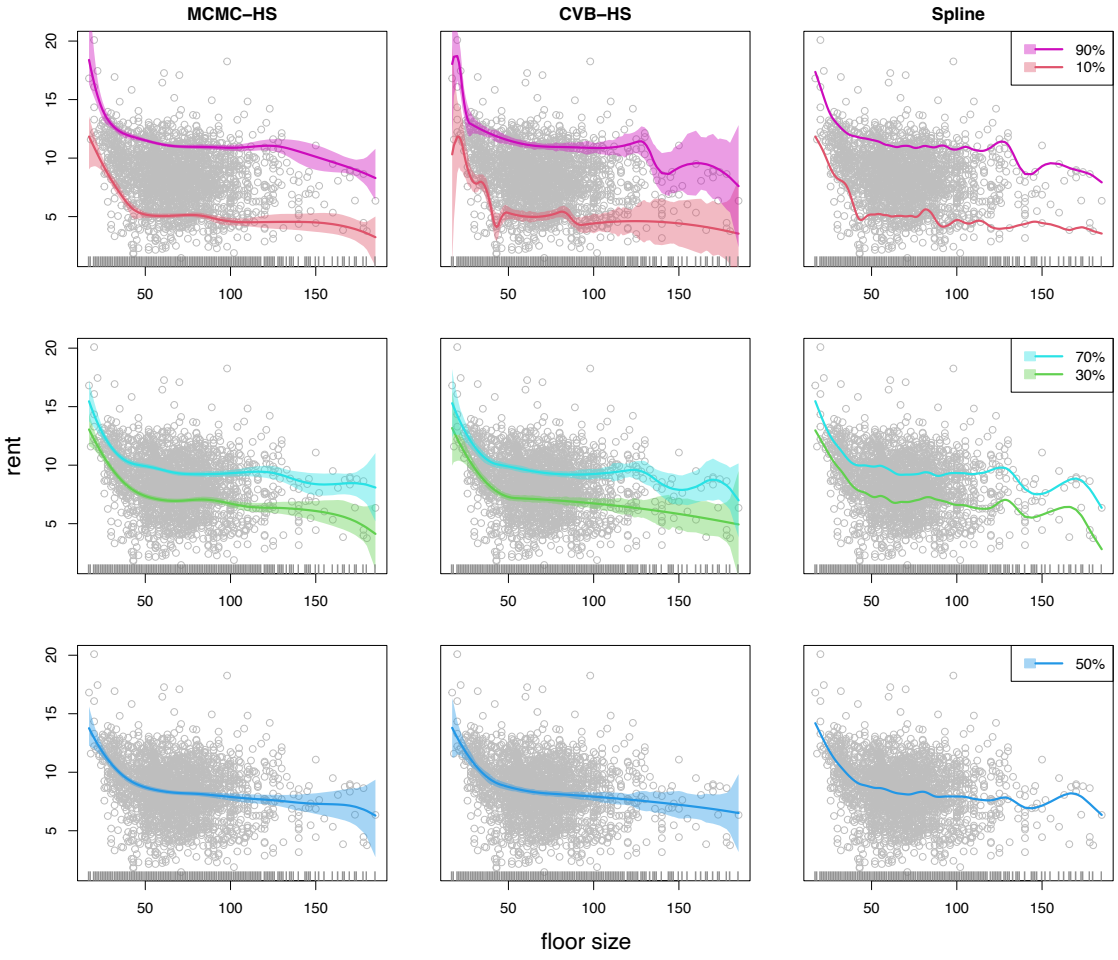

Figure 3.4: Point estimates and 95% credible intervals for Munich rent data.

3.6 Concluding remarks

This study proposed a quick and accurate calibration algorithm for credible intervals using a mean-field variational Bayes method. The proposed CVB method can reasonably calibrate credible intervals with possible model misspecifications. In nu-

merical experiments, it was shown that the proposed method worked especially well in the inference for high/low quantile levels. We also showed that the computational efficiency of the proposed CVB methods is higher than the MCMC versions in terms of the efficient sample size and computation time. If computation time is not a concern, then MCMC-based methods may be capable of providing accurate point and interval estimates. However, the estimation of low/high quantiles tends to be unstable, and if the model is misspecified, the estimation of other quantile points will also be unstable. The method of Syring and Martin (2019) could also be used in such cases, but the proposed CVB method is capable of parallel computation and is thus much more computationally efficient. Finally, as the drawback of the variational posterior approximations, the proposed CVB method may not accurately reflect prior beliefs about parameters. It was observed that the CVB-HS and Spline had remarkably similar results in terms of the trajectories of the trends in the Munich rent example. We believe this is due to the variational approximation of the posterior distribution. However, the proposed method still has the advantage of providing point estimation results comparable to those of the optimization method and of allowing the quick and accurate evaluation of uncertainty under finite samples.

Furthermore, it may be more suitable to use a skewed distribution as a variational distribution of the quantile trends which can provide asymmetric credible intervals. However, when a different variational distribution was adopted, the mean-filed approximation algorithm used in this chapter was no longer applicable, therefore a detailed investigation extends the scope of this study.

3.7 Appendix

This section provides algorithm details and additional information on simulation study and real data examples.

3.7.1 Full conditional distributions in Gibbs sampler

We provide the details of the full conditional distributions.

- The full conditional distribution of θ.

$$p(\theta \mid \tau^2, w, z, \sigma^2, y) \propto \prod_{i=1}^{n} \prod_{j=1}^{N_i} \exp\left(-\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2\sigma^2 z_{ij}}\right) \exp\left(-\frac{1}{2\sigma^2}\theta^\top D^\top W^{-1} D\theta\right)$$

$$\propto \exp\left\{-\frac{1}{2\sigma^2}(\theta - A^{-1}B)^\top A(\theta - A^{-1}B)\right\},$$

which is $N_n(A^{-1}B, \sigma^2 A^{-1})$, where

$$a_i = \sum_{j=1}^{N_i} \frac{1}{z_{ij}}, \quad b_i = \sum_{j=1}^{N_i} \frac{(y_{ij} - \psi z_{ij})}{z_{ij}}, \quad c = \left(\sum_{j=1}^{N_1} y_{1j} - \psi z_{1j}, \ldots, \sum_{j=1}^{N_n} y_{nj} - \psi z_{nj}\right)^\top$$

$$A = D^\top W^{-1} D + \frac{1}{t^2}\text{diag}\left(\sum_{j=1}^{N_1} z_{1j}^{-1}, \ldots, \sum_{j=1}^{N_n} z_{nj}^{-1}\right),$$

$$B = \left(\frac{1}{t^2}\sum_{j=1}^{N_1}\left(\frac{y_{1j}}{z_{1j}} - \psi\right), \ldots, \frac{1}{t^2}\sum_{j=1}^{N_n}\left(\frac{y_{nj}}{z_{nj}} - \psi\right)\right)^\top.$$

- The full conditional distributions of z_{ij} for $i = 1, \ldots, n$ and $j = 1, \ldots, N_i$.

$$p(z_{ij} \mid \theta_i, y, \sigma^2) \propto (z_{ij})^{-1/2} \exp\left(-\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2\sigma^2 z_{ij}}\right) \exp\left(-\frac{z_{ij}}{\sigma^2}\right)$$
$$\propto (z_{ij})^{-1+1/2} \exp\left\{-\frac{1}{2}\left(\frac{(y_{ij} - \theta_i)^2}{t^2\sigma^2}\frac{1}{z_{ij}} + \left(\frac{\psi^2}{t^2} + 2\right)\frac{1}{\sigma^2} z_{ij}\right)\right\},$$

which is $\text{GIG}\left(1/2, (y_{ij} - \theta_i)^2/(t^2\sigma^2), (\psi^2/t^2 + 2)\, z_{ij}/\sigma^2\right)$.

- The full conditional distribution of σ^2.

$$p(\sigma^2 \mid \theta, \tau^2, w, y, z)$$
$$\propto \prod_{i=1}^{n}\prod_{j=1}^{N_i}(\sigma^2)^{-1/2} \exp\left(-\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2\sigma^2 z_{ij}}\right) \times (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\theta^\top D^\top W^{-1} D\theta\right)$$
$$\times \prod_{i=1}^{n}\prod_{j=1}^{N_i}(\sigma^2)^{-1} \exp\left(-\frac{z_{ij}}{\sigma^2}\right) \times (\sigma^2)^{-1-a_\sigma} \exp\left(-\frac{b_\sigma}{\sigma^2}\right)$$
$$\propto (\sigma^2)^{-1-(3N+n)/2+a_\sigma}$$
$$\times \exp\left\{-\frac{1}{\sigma^2}\left(\sum_{i=1}^{n}\sum_{j=1}^{N_i}\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 z_{ij}} + \frac{1}{2}\theta^\top D^\top W^{-1} D\theta + \sum_{i=1}^{n}\sum_{j=1}^{N_i} z_{ij} + b_\sigma\right)\right\},$$

which is $\text{IG}\left((3N + n)/2 + a_\sigma, \alpha_{\sigma^2}\right)$, where

$$\alpha_{\sigma^2} = \sum_{i=1}^{n}\sum_{j=1}^{N_i}\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 z_{ij}} + \frac{1}{2}\theta^\top D^\top W^{-1} D\theta + \sum_{i=1}^{n}\sum_{j=1}^{N_i} z_{ij} + b_\sigma.$$

- The full conditional distribution of τ^2.

$$p(\tau^2 \mid \theta, \sigma^2, w) \propto |W|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\theta^\top D^\top W^{-1} D\theta\right) (\tau^2)^{-1-1/2} \exp\left(-\frac{1}{\tau^2\xi}\right)$$
$$\propto (\tau^2)^{-1-(n-k)/2} \exp\left\{-\frac{1}{\tau^2}\left(\sum_{i=k+2}^{n} \frac{\eta_i^2}{2\sigma^2 w_i^2} + \frac{1}{\xi}\right)\right\},$$

  which is $\mathrm{IG}\left((n-k)/2, \sum_{i=k+2}^{n} \eta_i^2/(2\sigma^2 w_i^2) + 1/\xi\right)$.

- The full conditional distribution of ξ.

$$p(\xi \mid \tau^2) \propto \exp\left(-\frac{1}{\tau^2\xi}\right) \xi^{-1-1/2} \exp\left(-\frac{1}{\xi}\right)$$
$$\propto \xi^{-1-1/2} \exp\left\{-\frac{1}{\xi}\left(\frac{1}{\tau^2} + 1\right)\right\},$$

  which is $\mathrm{IG}\left(1/2, 1/\tau^2 + 1\right)$.

- The full conditional distributions of w_i^2 for $i = 1, \ldots, k+1$.

$$p(w_i^2 \mid \theta_i, \sigma^2) \propto (w_i^2)^{-1/2} \exp\left(-\frac{\eta_i^2}{2w_i^2\sigma^2}\right) (w_i^2)^{-1-a_{w_i}} \exp\left(-\frac{b_{w_i}}{w_i^2}\right)$$
$$\propto (w_i^2)^{-1-(1/2+a_{w_i})} \exp\left(-\left(\frac{\eta_i^2}{2\sigma^2} + b_{w_i}\right)\frac{1}{w_i}\right),$$

  which is $\mathrm{IG}\left(1/2 + a_{w_i}, \eta_i^2/2\sigma^2 + b_{w_i}\right)$.

The full conditional distributions of w_i for $i = k+2, \ldots, n$ and their argumentation parameters are derived separately for Laplace-type and horseshoe-type priors.

Laplace-type

For $i = k+2, \ldots, n$, we assume $w_i^2 \sim \mathrm{Exp}(\gamma^2/2)$.

- The full conditional distributions of w_i for $i = k+2, \ldots, n$.

$$p(w_i^2 \mid \theta_i, \gamma^2) \propto |W|^{-1/2} \exp\left(-\frac{1}{2}\sum_{i=1}^{n} \frac{\eta_i^2}{\sigma^2 w_i^2}\right) \exp\left(-\frac{\gamma^2}{2} w_i^2\right)$$
$$\propto (w_i^2)^{-1+1/2} \exp\left(-\frac{1}{2}\left(\frac{\eta_i^2}{\sigma^2}\frac{1}{w_i} - \gamma^2 w_i^2\right)\right),$$

  which is $\mathrm{GIG}\left(1/2, \eta_i^2/\sigma^2, \gamma^2\right)$.

- The full conditional distribution of γ^2

$$p(\gamma^2 \mid w, \nu^2) \propto \prod_{i=k+2}^{n} \left(\frac{\gamma^2}{2}\right) \exp\left(-\frac{\gamma^2}{2} w_i^2\right) (\gamma^2)^{-1-1/2} \exp\left(-\frac{1}{\gamma^2 \nu}\right)$$
$$\propto (\gamma^2)^{-1+(n-k-1-1/2)} \exp\left(-\frac{1}{2}\left(\gamma^2 \sum_{i=k+2}^{n} w_i^2 + \frac{2}{\nu}\frac{1}{\gamma^2}\right)\right),$$

  which is GIG $\left(n-k-3/2, 2/\nu, \sum_{i=k+2}^{n} w_i^2\right)$.

- The full conditional distribution of ν.

$$p(\nu \mid \gamma^2) \propto \exp\left(-\frac{1}{\gamma^2 \nu}\right) \nu^{-1-1/2} \exp\left(-\frac{1}{\nu}\right)$$
$$\propto \nu^{-1-1/2} \exp\left(-\left(\frac{1}{\gamma^2}+1\right)\frac{1}{\nu}\right),$$

  which is IG $(1/2, 1/\gamma^2 + 1)$.

Horseshoe-type

For $i = k+2, \ldots, n$, we assume $w_i \sim C^+(0,1)$.

- The full conditional distributions of w_i for $i = k+2, \ldots, n$.

$$p(w_i^2 \mid \theta_i) \propto (w_i^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\theta^\top D^\top W^{-1} D\theta\right) (w_i^2)^{-1-1/2} \exp\left(-\frac{1}{w_i^2 \nu_i}\right)$$
$$\propto (w_i^2)^{-1-1} \exp\left(-\left(\frac{\eta_i^2}{2\tau^2\sigma^2} + \frac{1}{\nu_i}\right)\frac{1}{w_i^2}\right),$$

  which is IG $(1, \eta_i^2/(2\tau^2\sigma^2) + 1/\nu_i)$.

- The full conditional distributions of ν_i for $i = k+2, \ldots, n$.

$$p(\nu_i \mid w_i^2) \propto \exp\left(-\frac{1}{w_i^2 \nu_i}\right) \nu^{-1-1/2} \exp\left(-\frac{1}{\nu_i}\right)$$
$$\propto \nu_i^{-1-1/2} \exp\left(-\left(\frac{1}{w_i^2}+1\right)\frac{1}{\nu_i}\right),$$

  which is IG $(1/2, 1/w_i^2 + 1)$.

3.7.2 Variational distributions

We summarize the derivations of variational distributions.

- The variational distribution of θ.

$$
\begin{aligned}
q(\theta) &\propto \exp\left(\mathrm{E}_{z,\tau^2,w,\sigma^2}\left[-\frac{1}{2t^2\sigma^2}\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(-\frac{(y_{ij}-\theta_i-\psi z_{ij})^2}{2t^2\sigma^2 z_{ij}}\right)+\left(-\frac{1}{2\sigma^2}\theta^\top D^\top W^{-1}D\theta\right)\right]\right)\\
&\propto \exp\left(-\frac{1}{2}(\theta-A^{-1}B)^\top(\mathrm{E}_{1/\sigma^2}A)(\theta-A^{-1}B)\right),
\end{aligned}
$$

  which is $N_n(A^{-1}B,(\mathrm{E}_{1/\sigma^2}A)^{-1})$, where

$$
\begin{aligned}
A &= D^\top\hat{W}^{-1}D+\frac{1}{t^2}\mathrm{diag}\left(\sum_{j=1}^{N_1}E_{1/z_{1j}},\ldots,\sum_{j=1}^{N_n}E_{1/z_{nj}}\right),\\
B &= \left(\frac{1}{t^2}\sum_{j=1}^{N_1}\left(y_{1j}E_{1/z_{1j}}-\psi\right),\ldots,\frac{1}{t^2}\sum_{j=1}^{N_n}\left(y_{nj}E_{1/z_{nj}}-\psi\right)\right)^\top\\
\hat{W}^{-1} &= \mathrm{diag}(E_{1/w_1^2},\ldots,E_{1/w_{k+1}^2},E_{1/\tau^2}E_{1/w_{k+2}^2},\ldots,E_{1/\tau^2}E_{1/w_n^2}),\\
E_{1/w_i^2} &= \mathrm{E}_{w_i^2}[1/w_i^2],\quad E_{1/\tau^2}=\mathrm{E}_{\tau^2}[1/\tau^2],\\
E_{1/\sigma^2} &= \mathrm{E}_{\sigma^2}[1/\sigma^2],\quad E_{1/z_{ij}}=\mathrm{E}_{z_{ij}}[1/z_{ij}]
\end{aligned}
$$

- The variational distributions of z_{ij} for $i=1,\ldots,n$ and $j=1,\ldots,N_j$.

$$
\begin{aligned}
q(z_{ij}) &\propto \exp\left(\mathrm{E}_{\theta,\tau^2,w,\sigma^2}\left[\log(z_{ij})^{-1/2}+\left(-\frac{(y_{ij}-\theta_i-\psi z_{ij})^2}{2t^2\sigma^2 z_{ij}}\right)+\left(-\frac{z_{ij}}{\sigma^2}\right)\right]\right)\\
&\propto (z_{ij})^{-1/2}\exp\left(-\frac{1}{2}\left\{\mathrm{E}_{\sigma^2}\left[\frac{1}{\sigma^2}\right]\mathrm{E}_\theta\left[\frac{(y_{ij}-\theta_i)^2}{t^2}\right]\frac{1}{z_{ij}}+\left(\frac{\psi^2}{t^2}-2\right)\mathrm{E}_{\sigma^2}\left[\frac{1}{\sigma^2}\right]z_{ij}\right\}\right),
\end{aligned}
$$

  which is GIG $\left(1/2,\alpha_{z_{ij}},\beta_{z_{ij}}\right)$, where

$$
\begin{aligned}
\alpha_{z_{ij}} &= \frac{1}{t^2}E_{1/\sigma^2}(y_{ij}^2-2y_{ij}E_\theta+E_{\theta^2}),\quad \beta_{z_{ij}}=\left(\frac{\psi^2}{t^2}-2\right)E_{1/\sigma^2},\\
E_{\theta_i} &= \mathrm{E}_{\theta_i}[\theta_i]=(A^{-1}B)_i,\quad E_{\theta_i^2}=\mathrm{E}_{\theta_i}[\theta_i^2]=e_i^\top(E_{\sigma^2}^{-1}A^{-1}+A^{-1}BB^\top A^{-1})e_i,
\end{aligned}
$$

  where e_i is a unit vector that the ith component is 1.

- The variational distribution of σ.

$$
\begin{aligned}
q(\sigma^2) \propto \exp\Bigg(\mathrm{E}_{\theta,\tau^2,w,z}\Bigg[&\log(\sigma^2)^{-N/2}+\sum_{i=1}^{n}\sum_{j=1}^{N_i}\left(-\frac{(y_{ij}-\theta_i-\psi z_{ij})^2}{2t^2\sigma^2 z_{ij}}\right)\\
&+\log(\sigma^2)^{-n/2}-\left(\frac{1}{2\sigma^2}\theta^\top D^\top W^{-1}D\theta\right)+\log(\sigma^2)^{-N}\\
&-\sum_{i=1}^{n}\sum_{j=1}^{N}\frac{z_{ij}}{\sigma^2}+\log(\sigma^2)^{-1-a_\sigma}-\frac{b_\sigma}{\sigma^2}\Bigg]\Bigg)
\end{aligned}
$$

$$\propto (\sigma^2)^{-1-(n+3N)/2-a_\sigma} \exp\left(-\frac{\alpha_{\sigma^2}}{\sigma^2}\right),$$

which is $\mathrm{IG}\left((n+3N)/2+a_\sigma, \alpha_{\sigma^2}\right)$, where

$$\alpha_{\sigma^2} = \sum_{i=1}^{n}\sum_{j=1}^{N_i} \frac{1}{2t^2}\left\{\left(y_{ij}^2 - 2y_{ij}E_{\theta_i} + E_{\theta_i^2}\right)E_{1/z_{ij}} - 2\psi\left(y_{ij} - E_{\theta_i}\right) + \psi^2 E_{z_{ij}}\right\}$$
$$+ \frac{1}{2}\left(\sum_{i=1}^{k+1} E_{\eta_i^2}E_{w_i} + \sum_{i=k+2}^{n} E_{\eta_i^2}E_{1/\tau^2}E_{1/w_i^2}\right) + \sum_{i=1}^{n}\sum_{j=1}^{N} E_{z_{ij}} + b_\sigma,$$
$$E_{\eta_i^2} = d_i^\top (E_{\sigma^2}^{-1}A^{-1} + A^{-1}BB^\top A^{-1})d_i.$$

- The variational distribution of τ^2.

$$q(\tau^2) \propto \exp\left(\mathrm{E}_{\theta,\sigma^2,w,\xi}\left[\log|W|^{1/2} - \frac{1}{2\sigma^2}\theta^\top D^\top W^{-1} D\theta + \log(\tau^2)^{-1-1/2} - \frac{1}{\tau^2\xi}\right]\right)$$
$$\propto (\tau^2)^{-1-(n-k)/2} \exp\left\{-\left(\frac{1}{2}E_{1/\sigma^2}\sum_{i=k+2}^{n} E_{\eta_i^2}E_{1/w_i^2} + E_{1/\xi}\right)\frac{1}{\tau^2}\right\},$$

which is $\mathrm{IG}\left((n-k)/2, E_{1/\sigma^2}/2\sum_{i=k+2}^{n} E_{\eta_i^2}E_{1/w_i^2} + E_{1/\xi}\right)$, where $E_{1/\xi} = \mathrm{E}_\xi[1/\xi]$

- The variational distribution of ξ.

$$q(\xi) \propto \xi^{-1-1/2} \exp\left(-\mathrm{E}_{\tau^2}\left[\frac{1}{\tau^2}\right]\frac{1}{\xi} - \frac{1}{\xi}\right)$$
$$\propto \xi^{-1-1/2} \exp\left\{-\left(E_{1/\tau^2}+1\right)\frac{1}{\xi}\right\},$$

which is $\mathrm{IG}\left(1/2, E_{1/\tau^2}+1\right)$.

- The variational distributions of w_i^2 for $i = 1, \ldots, k+1$.

$$q(w_i^2) \propto \exp\left(\log(w_i^2)^{-1/2} - \mathrm{E}_{\theta,\sigma^2}\left[\frac{\eta_i^2}{2w_i^2\sigma^2}\right]\right)(w_i^2)^{-1-a_{w_i}} \exp\left(-\frac{b_{w_i}}{w_i^2}\right)$$
$$\propto (w_i^2)^{-1-(1/2+a_{w_i})} \exp\left(-\left(\frac{1}{2}E_{\eta_i^2}E_{1/\sigma^2} + b_{w_i}\right)\frac{1}{w_i}\right),$$

which is $\mathrm{IG}\left(1/2 + a_{w_i}, E_{\eta_i^2}E_{1/\sigma^2}/2 + b_{w_i}\right)$.

The variational distributions of w_i for $i = k+2, \ldots, n$ and their argumentation parameters are derived separately for Laplace-type and horseshoe-type priors.

Laplace-type

In Laplace-type prior, we set $\tau^2 = 1$. For $i = k+2, \ldots, n$, we assume $w_i^2 \sim \text{Exp}(\gamma^2/2)$.

- The variational distributions of w_i for $i = k+2, \ldots, n$.

$$\begin{aligned} q(w_i^2) &\propto \exp\left(\log(w_i^2)^{-1/2} - \mathrm{E}_{\theta,\sigma^2}\left[\frac{\eta_i^2}{2w_i^2\sigma^2}\right] - \mathrm{E}_{\gamma^2}\left[\frac{\gamma^2}{2}w_i^2\right]\right) \\ &\propto (w_i^2)^{-1+1/2}\exp\left(-\frac{1}{2}\left(E_{\eta_i^2}E_{1/\sigma^2}\frac{1}{w_i} + E_{\gamma^2}w_i^2\right)\right), \end{aligned}$$

  which is GIG $\left(1/2, E_{\eta_i^2}E_{1/\sigma^2}, E_{\gamma^2}\right)$.

- The variational distributions of γ^2.

$$\begin{aligned} q(\gamma^2) &\propto \exp\left(\log(\gamma^2)^{n-k-1} - \frac{\gamma^2}{2}\sum_{i=k+2}^{n}\mathrm{E}_w[w_i^2] + \log(\gamma^2)^{-1-1/2} - \frac{1}{\gamma^2}\mathrm{E}_\nu\left[\frac{1}{\nu}\right]\right) \\ &\propto (\gamma^2)^{-1+(n-k-1-1/2)}\exp\left(-\frac{1}{2}\left(\gamma^2\sum_{i=k+2}^{n}E_{w_i^2} + 2E_{1/\nu}\frac{1}{\gamma^2}\right)\right), \end{aligned}$$

  which is GIG $\left(n-k-3/2, 2E_{1/\nu}, \sum_{i=k+2}^{n}E_{w_i^2}\right)$, where $E_{1/\nu} = \mathrm{E}_\nu[1/\nu]$.

- The variational distributions of ν.

$$\begin{aligned} q(\nu) &\propto \exp\left(-\mathrm{E}_{\gamma^2}\left[\frac{1}{\gamma^2}\right]\frac{1}{\nu}\right)\nu^{-1-1/2}\exp\left(-\frac{1}{\nu}\right) \\ &\propto \nu^{-1-1/2}\exp\left(-\left(E_{1/\gamma^2}+1\right)\frac{1}{\nu}\right), \end{aligned}$$

  which is IG $\left(1/2, E_{1/\gamma^2}+1\right)$, where $E_{1/\gamma^2} = \mathrm{E}_{\gamma^2}[1/\gamma^2]$.

Horseshoe-type

For $i = k+2, \ldots, n$, we assume $w_i \sim C^+(0,1)$.

- The variational distributions of w_i $(i = k+2, \ldots, n)$.

$$\begin{aligned} q(w_i^2) &\propto \exp\left(\log(w_i^2)^{-1/2} - \frac{1}{2}\mathrm{E}_{\theta,\tau^2,\sigma^2}\left[\frac{1}{\sigma^2}\theta^\top D^\top W^{-1}D\theta\right] + \log(w_i^2)^{-1-1/2} - \mathrm{E}_{\nu_i}\left[\frac{1}{\nu_i}\right]\frac{1}{w_i^2}\right) \\ &\propto (w_i^2)^{-1-1}\exp\left(-\left(\frac{1}{2}E_{\eta_i^2}E_{1/\tau^2}E_{1/\sigma^2} + E_{1/\nu_i}\right)\frac{1}{w_i^2}\right), \end{aligned}$$

  which is IG $\left(1, E_{\eta_i^2}E_{1/\tau^2}E_{1/\sigma^2}/2 + E_{1/\nu_i}\right)$, where $E_{1/\nu_i} = \mathrm{E}_{\nu_i}[1/\nu_i]$.

- The variational distributions of ν_i $(i = k+2, \ldots, n)$.

$$q(\nu_i) \propto \exp\left(-\mathrm{E}_{w_i^2}\left[\frac{1}{w_i^2}\right]\frac{1}{\nu_i}\right)\nu_i^{-1-1/2}\exp\left(-\frac{1}{\nu_i}\right)$$
$$\propto \nu_i^{-1-1/2}\exp\left(-\left(E_{1/w_i^2}+1\right)\frac{1}{\nu_i}\right),$$

  which is IG $\left(1/2, E_{1/w_i^2}+1\right)$.

3.7.3 Additional information on simulation studies

Smooth Gaussian process

In Section 4 of the main manuscript, we provided simulation results under (PC) piecewise constant and (VS) varying smoothness. In this subsection, we provide simulation results under the smooth Gaussian process (denoted by (GP)) as a true data-generating scenario. The plots of simulated true quantiles are shown in Figure 3.5. The additional description and results under (GP) are as follows.

(GP) Smooth Gaussian process

$$f \sim \mathrm{GP}(\mu, \Sigma), \quad \Sigma_{i,j} = \sigma_f^2 \exp\{-(t_j - t_i)^2/(2\rho^2)\}.$$

The scenario (GP) generates observations from the Gaussian process with squared exponential covariance function (see also Faulkner and Minin, 2018). We set $\mu = 2$, $\sigma_f^2 = 1$ and $\rho = 10$. The function f was generated with the same random number seed for all scenarios (Gaussian, Beta, and Mixed normal noise functions). The aim of the scenario (GP) is to test the ability of the proposed methods to handle a smoothly varying function with no local change. To this end, we consider $k = 2$ in this scenario.

The results are represented in Table 3.8. The raw computing time and effective sample size per unit time are also reported in Table 3.9. Note that the simulation is conducted without parallel computation as well as those of Section 4 in the main text.

Multiple observations

We provide simulation results under multiple observations per location. The setting of simulation is as follows. Data was simulated by the equation (5) in the

Table 3.8: Average values of MSE, MAD, MCIW and CP based on 100 replications for Gaussian process with $k = 2$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.084	*0.020*	**0.016**	**0.019**	**0.040**	*0.187*	*0.109*	**0.099**	**0.107**	**0.153**
VB-HS	0.174	0.024	0.019	0.023	0.048	0.243	0.118	0.107	0.117	0.168
MCMC-Lap	0.093	0.021	*0.017*	*0.020*	*0.043*	0.195	*0.110*	0.100	*0.108*	*0.159*
VB-Lap	**0.048**	**0.019**	*0.017*	*0.020*	0.048	**0.170**	**0.107**	*0.100*	0.110	0.171
ADMM	*0.050*	0.024	0.023	0.026	0.049	0.174	0.120	0.115	0.123	0.172
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.004*	**0.003**	**0.005**	**0.007**	**0.014**	0.041	*0.037*	**0.048**	*0.062*	**0.092**
VB-HS	0.009	*0.004*	*0.006*	*0.008*	0.017	*0.037*	0.041	0.054	0.067	0.101
MCMC-Lap	**0.003**	**0.003**	**0.005**	**0.007**	*0.015*	0.039	**0.036**	**0.048**	**0.061**	*0.096*
VB-Lap	**0.003**	**0.003**	**0.005**	**0.007**	0.018	**0.030**	**0.036**	*0.049*	0.063	0.104
ADMM	*0.004*	*0.004*	*0.006*	0.009	0.017	0.040	0.039	0.055	0.069	0.103
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.169	0.064	*0.037*	*0.041*	**0.089**	0.278	*0.177*	*0.150*	**0.158**	**0.234**
VB-HS	0.205	0.087	0.043	0.047	0.106	0.309	0.196	0.159	0.169	0.254
MCMC-Lap	0.171	0.067	*0.037*	**0.040**	*0.094*	0.282	0.178	**0.149**	**0.158**	*0.242*
VB-Lap	**0.100**	**0.043**	**0.035**	*0.041*	0.103	**0.250**	**0.164**	**0.149**	*0.160*	0.252
ADMM	*0.106*	*0.056*	0.046	0.053	0.111	*0.259*	0.187	0.170	0.183	0.261

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.510	0.478	0.462	0.458	0.414	0.738	**0.906**	**0.935**	**0.908**	0.723
CVB-HS	1.267	0.616	0.486	0.599	1.161	**0.950**	**0.943**	**0.921**	**0.941**	**0.980**
VB-HS	0.145	0.217	0.232	0.216	0.124	0.221	0.546	0.610	0.538	0.235
MCMC-Lap	0.520	0.494	0.478	0.470	0.422	0.748	**0.913**	**0.944**	**0.916**	0.720
CVB-Lap	1.534	0.864	0.589	0.792	1.226	**0.994**	**0.996**	**0.972**	**0.991**	**0.982**
VB-Lap	0.185	0.291	0.309	0.289	0.182	0.333	0.721	0.777	0.713	0.330
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.161	0.179	0.214	0.244	0.233	**0.948**	**0.946**	**0.923**	0.891	0.693
CVB-HS	0.490	0.272	0.214	0.350	0.725	**0.991**	**0.962**	0.885	**0.933**	**0.972**
VB-HS	0.045	0.087	0.109	0.111	0.066	0.474	0.674	0.621	0.507	0.215
MCMC-Lap	0.161	0.182	0.221	0.252	0.239	**0.956**	**0.950**	**0.936**	**0.902**	0.692
CVB-Lap	0.608	0.377	0.271	0.459	0.737	**0.993**	**0.990**	**0.951**	**0.981**	**0.977**
VB-Lap	0.069	0.119	0.149	0.154	0.102	0.770	0.855	0.808	0.697	0.307
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.674	0.740	0.706	0.680	0.628	0.693	**0.905**	**0.937**	**0.901**	0.728
CVB-HS	1.652	0.921	0.746	0.847	1.650	**0.966**	**0.939**	**0.929**	**0.934**	**0.976**
VB-HS	0.203	0.329	0.348	0.317	0.193	0.218	0.515	0.609	0.559	0.236
MCMC-Lap	0.693	0.758	0.725	0.702	0.646	0.710	**0.913**	**0.946**	**0.916**	0.724
CVB-Lap	1.909	1.145	0.878	1.134	1.783	**0.993**	**0.995**	**0.981**	**0.993**	**0.985**
VB-Lap	0.274	0.425	0.452	0.421	0.275	0.324	0.693	0.769	0.697	0.333

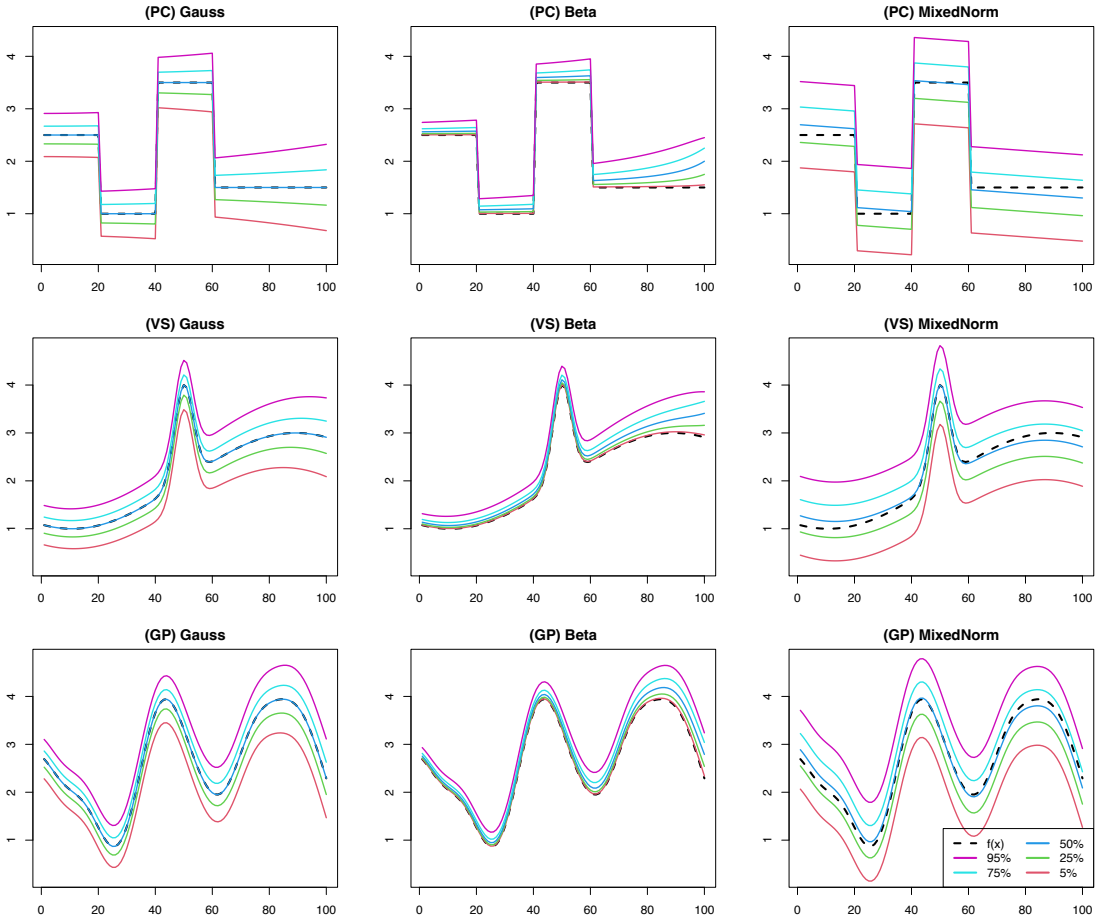

Figure 3.5: Simulated true quantile trends.

Table 3.9: Average values of raw computing time and effective sample size per unit time based on 100 replications for GP scenarios.

	Computation time (second)					ESS (per second)				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	32	32	32	32	14	38	44	41	15
CVB-HS	13	8	11	8	13	575	943	692	980	576
MCMC-Lap	35	35	35	35	35	13	56	69	59	14
CVB-Lap	20	17	17	12	16	378	442	431	627	475
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	32	32	32	32	13	48	45	42	14
CVB-HS	9	7	8	7	11	859	1144	907	1113	714
MCMC-Lap	35	35	35	35	35	11	73	88	59	13
CVB-Lap	30	8	12	8	10	257	915	639	997	728
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	32	32	32	32	32	15	36	42	39	15
CVB-HS	15	10	14	9	16	503	732	526	849	487
MCMC-Lap	35	35	35	35	35	14	51	59	53	14
CVB-Lap	29	32	25	20	24	259	241	304	375	320

main manuscript, where the number of data at each location N_i $(i = 1, \ldots, n)$ was simulated from a multinomial distribution with outcomes 2, 3, and 4 with probabilities 1/3. We used the same noise distributions as those of Section 4 in the main manuscript (Gauss, Beta, and Mixed normal). The number of locations was set as $n = 100$, while the total sample size N was 200 to 400 because of the randomness of N_i. As true data-generating functions, (PC), (VS) and (GP) were adopted. The convergence criterion of variational Bayes methods was set as 10^{-3}, which is slightly different from that of the main simulation. We compared the proposed methods with the quantile smoothing spline method (denoted by Spline for short) by Nychka et al. (2017) as a frequentist method, which was also applied to Munich rent data in Section 5.2 of the main manuscript. The reason is that Brantley's quantile trend filtering cannnot apply to multiple observations per location.

The results of the simulations are summarized in Tables 3.10, 3.11 and 3.12. From these results, the MCMC-HS method was the best and the VB-HS method was at least better than the Spline method for almost all scenarios in terms of point estimation. Compared with the results of a single observation in the main

manuscript, the MSE and MAD were smaller for each method, and it seems that a larger sample size induces more accurate point estimation. Although the CVB method tended to be over-coverage, it gave wider credible intervals than those of the MCMC method as seen in the MCIW, and conservatively improved coverage probability as well as single observation.

More locations

In Section 4 of the main manuscript, we set the number of locations to $n = 100$. We considered the same simulation scenarios as those of Section 4, but we set $n = 200$ instead of $n = 100$. We note that if we change the number of locations, the (GP) function presented in Subsection 3.7.3 changes due to random sampling. To avoid the problem, we considered a linear completion between the original data points. Furthermore, we only compared the following four methods: MCMC-HS, CVB-HS, VB-HS, and ADMM.

The results are shown in Table 3.13, 3.14 and 3.15. The MSE and MAD for all methods were improved over the results of a smaller sample size in most cases. The point estimation of the MCMC-HS method performed well in many cases as well as the results of $n = 100$ observations in terms of point estimation. The coverage probabilities of the MCMC-HS method were also similar to those of $n = 100$ observations, which were under 0.90 for extreme quantiles such as 0.05 and 0.95. However, the CVB-HS method provided values over 0.90 except for 0.5 quantile level under (PC) and Beta.

3.7.4 Additional information on real data example

We confirm that the proposed algorithm works well by making a simple diagnosis of sampling efficiency. Here, we especially consider the sampling efficiency in the two real data analyses. We generated 60,000 posterior samples under MCMC in section 2.3 after discarding the first 10,000 posterior samples as burn-in, and then only every 10th scan was saved. For Nile data in Section 5.1 and Munich rent data in Section 5.2, sample paths and autocorrelations of the posterior samples in three selected points are provided in Figures 3.6 and 3.7, respectively. The mixing and autocorrelation of the MCMC algorithm seem satisfactory.

Table 3.10: Average values of MSE, MAD, MCIW and CP based on 100 replications for multiple observations (PC) with $k = 0$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0322**	**0.0052**	**0.0033**	**0.0049**	**0.0332**	**0.1350**	**0.0535**	**0.0436**	**0.0518**	**0.1387**
VB-HS	*0.0387*	*0.0182*	*0.0132*	*0.0180*	*0.0400*	*0.1516*	*0.0971*	*0.0812*	*0.0946*	*0.1564*
MCMC-Lap	0.0582	0.0205	0.0167	0.0208	0.0576	0.1917	0.1110	0.0996	0.1111	0.1921
VB-Lap	0.0589	0.0190	0.0154	0.0190	0.0587	0.1950	0.1060	0.0950	0.1053	0.1974
Spline	0.1154	0.0671	0.0548	0.0672	0.1141	0.2214	0.1586	0.1473	0.1582	0.2247
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.0011*	**0.0013**	**0.0022**	**0.0037**	**0.0155**	0.0180	**0.0177**	**0.0276**	**0.0412**	**0.0948**
VB-HS	**0.0008**	*0.0029*	*0.0048*	*0.0079*	*0.0162*	**0.0109**	*0.0270*	*0.0416*	*0.0606*	*0.0991*
MCMC-Lap	0.0046	0.0040	0.0056	0.0092	0.0247	0.0433	0.0398	0.0514	0.0702	0.1236
VB-Lap	0.0042	0.0040	0.0056	0.0093	0.0308	0.0351	0.0386	0.0511	0.0709	0.1420
Spline	0.1120	0.0580	0.0424	0.0659	0.1042	0.1240	0.0992	0.1021	0.1257	0.1736
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0656**	**0.0136**	**0.0094**	**0.0143**	**0.0669**	**0.2004**	**0.0889**	**0.0728**	**0.0885**	**0.2016**
VB-HS	*0.0822*	*0.0397*	*0.0281*	*0.0406*	*0.0827*	*0.2277*	*0.1507*	*0.1239*	*0.1519*	*0.2286*
MCMC-Lap	0.1124	0.0433	0.0355	0.0443	0.1129	0.2754	0.1642	0.1473	0.1650	0.2745
VB-Lap	0.1043	0.0381	0.0313	0.0397	0.1061	0.2649	0.1526	0.1372	0.1545	0.2658
Spline	0.1559	0.0866	0.0763	0.0871	0.1595	0.2915	0.2053	0.1919	0.2067	0.2958

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.391	0.271	0.246	0.272	0.394	0.723	**0.938**	**0.960**	**0.947**	0.720
CVB-HS	1.097	0.664	0.429	0.675	1.125	**0.983**	**0.991**	**0.967**	**0.994**	**0.985**
VB-HS	0.108	0.181	0.187	0.184	0.110	0.228	0.568	0.675	0.602	0.219
MCMC-Lap	0.476	0.533	0.532	0.537	0.476	0.734	**0.941**	**0.964**	**0.943**	0.732
CVB-Lap	1.296	0.678	0.497	0.690	1.323	**0.981**	**0.979**	**0.953**	**0.982**	**0.985**
VB-Lap	0.203	0.342	0.362	0.344	0.204	0.288	0.804	0.871	0.809	0.275
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.054	0.084	0.124	0.171	0.216	**0.972**	**0.951**	**0.932**	**0.902**	0.606
CVB-HS	0.454	0.266	0.194	0.425	0.748	**1.000**	**0.987**	**0.934**	**0.985**	**0.983**
VB-HS	0.026	0.065	0.086	0.097	0.056	0.845	0.795	0.702	0.543	0.184
MCMC-Lap	0.171	0.205	0.247	0.276	0.238	**0.942**	**0.950**	**0.939**	0.889	0.619
CVB-Lap	0.496	0.311	0.231	0.451	0.797	**0.992**	**0.976**	**0.912**	**0.970**	**0.964**
VB-Lap	0.070	0.136	0.172	0.177	0.090	0.737	0.863	0.836	0.710	0.161
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.628	0.439	0.404	0.439	0.615	0.750	**0.935**	**0.957**	**0.927**	0.742
CVB-HS	1.520	1.007	0.677	1.009	1.551	**0.985**	**0.996**	**0.978**	**0.993**	**0.985**
VB-HS	0.169	0.272	0.276	0.273	0.169	0.229	0.524	0.632	0.527	0.217
MCMC-Lap	0.734	0.792	0.778	0.790	0.727	0.746	0.944	0.962	0.941	0.741
CVB-Lap	1.789	0.950	0.702	0.958	1.809	**0.987**	**0.982**	**0.953**	**0.979**	**0.987**
VB-Lap	0.319	0.497	0.517	0.496	0.318	0.336	0.806	0.867	0.805	0.339

Table 3.11: Average values of MSE, MAD, MCIW and CP based on 100 replications for multiple observations (VS) with $k = 1$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0246**	**0.0079**	**0.0063**	**0.0073**	**0.0255**	**0.1168**	**0.0667**	**0.0597**	**0.0638**	**0.1192**
VB-HS	*0.0313*	0.0135	0.0104	0.0138	*0.0331*	*0.1325*	0.0846	*0.0749*	0.0846	*0.1370*
MCMC-Lap	0.0346	0.0117	0.0099	0.0113	0.0351	0.1438	0.0838	0.0777	0.0828	0.1458
VB-Lap	0.0320	*0.0115*	*0.0098*	*0.0112*	0.0334	0.1373	*0.0831*	0.0773	*0.0821*	0.1415
Spline	0.0400	0.0151	0.0126	0.0149	0.0423	0.1558	0.0955	0.0878	0.0949	0.1615
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.0016	**0.0017**	**0.0024**	**0.0036**	**0.0118**	0.0254	**0.0249**	**0.0329**	**0.0426**	**0.0817**
VB-HS	0.0017	0.0026	0.0042	0.0065	*0.0122*	*0.0209*	0.0289	0.0414	0.0561	*0.0842*
MCMC-Lap	*0.0015*	*0.0020*	*0.0032*	*0.0051*	0.0159	0.0269	0.0284	0.0389	0.0526	0.0977
VB-Lap	**0.0013**	*0.0020*	*0.0032*	*0.0051*	0.0142	**0.0199**	*0.0282*	*0.0392*	*0.0527*	0.0930
Spline	0.0016	0.0022	0.0036	0.0057	0.0144	0.0239	0.0310	0.0430	0.0567	0.0949
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0494**	**0.0191**	**0.0152**	**0.0178**	**0.0514**	**0.1702**	**0.1048**	**0.0914**	**0.1007**	**0.1758**
VB-HS	*0.0649*	0.0282	*0.0226*	0.0275	0.0677	*0.1979*	*0.1274*	*0.1119*	*0.1253*	*0.2030*
MCMC-Lap	0.0695	*0.0280*	0.0229	*0.0255*	0.0732	0.2092	0.1298	0.1174	0.1265	0.2153
VB-Lap	0.0651	0.0287	0.0228	0.0257	*0.0664*	0.2003	0.1304	0.1164	0.1259	0.2034
Spline	0.0925	0.0354	0.0291	0.0351	0.0917	0.2422	0.1501	0.1352	0.1493	0.2422

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.337	0.277	0.275	0.278	0.341	0.731	0.897	**0.921**	**0.913**	0.734
CVB-HS	1.134	0.592	0.403	0.617	1.172	**0.994**	**0.989**	**0.965**	**0.994**	**0.994**
VB-HS	0.104	0.154	0.163	0.159	0.105	0.260	0.556	0.645	0.584	0.254
MCMC-Lap	0.403	0.373	0.370	0.374	0.400	0.740	**0.918**	**0.944**	**0.925**	0.734
CVB-Lap	1.150	0.580	0.425	0.599	1.184	**0.993**	**0.989**	**0.967**	**0.993**	**0.993**
VB-Lap	0.170	0.228	0.240	0.230	0.172	0.387	0.735	0.787	0.743	0.378
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.084	0.110	0.136	0.162	0.200	**0.940**	**0.931**	**0.904**	0.864	0.648
CVB-HS	0.082	0.131	0.191	0.233	0.237	0.887	**0.927**	**0.931**	**0.900**	0.738
VB-HS	0.030	0.067	0.085	0.090	0.057	0.655	0.747	0.670	0.538	0.226
MCMC-Lap	0.117	0.146	0.182	0.213	0.225	**0.950**	**0.951**	**0.934**	0.891	0.648
CVB-Lap	0.514	0.280	0.213	0.394	0.781	**0.999**	**0.990**	**0.951**	**0.989**	**0.993**
VB-Lap	0.052	0.097	0.121	0.130	0.093	0.790	0.857	0.808	0.707	0.312
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.486	0.418	0.406	0.415	0.491	0.720	0.882	**0.911**	0.893	0.717
CVB-HS	0.719	0.647	0.592	0.646	0.733	0.844	**0.957**	**0.962**	**0.956**	0.852
VB-HS	0.159	0.226	0.232	0.223	0.158	0.247	0.527	0.613	0.536	0.230
MCMC-Lap	0.604	0.528	0.521	0.529	0.606	0.742	0.890	**0.919**	**0.906**	0.737
CVB-Lap	1.390	0.792	0.598	0.797	1.448	**0.990**	**0.976**	**0.952**	**0.984**	**0.991**
VB-Lap	0.252	0.312	0.327	0.317	0.256	0.383	0.671	0.741	0.687	0.380

Table 3.12: Average values of MSE, MAD, MCIW and CP based on 100 replications for multiple observations (GP) with $k = 2$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0188**	**0.0079**	**0.0065**	**0.0076**	**0.0204**	**0.1035**	**0.0693**	**0.0635**	**0.0676**	**0.1080**
VB-HS	0.0267	0.0109	0.0085	0.0102	0.0264	0.1189	0.0791	0.0718	0.0765	0.1231
MCMC-Lap	*0.0206*	*0.0083*	*0.0069*	*0.0080*	*0.0224*	*0.1094*	*0.0703*	*0.0646*	*0.0688*	*0.1136*
VB-Lap	0.0222	0.0089	0.0074	0.0086	0.0234	0.1122	0.0729	0.0669	0.0713	0.1168
Spline	0.0394	0.0140	0.0110	0.0135	0.0419	0.1547	0.0915	0.0817	0.0900	0.1605
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.0011	**0.0014**	**0.0023**	**0.0035**	**0.0084**	0.0219	*0.0234*	*0.0329*	*0.0434*	**0.0702**
VB-HS	*0.0010*	0.0021	0.0032	0.0050	0.0095	*0.0161*	0.0262	0.0375	0.0510	0.0755
MCMC-Lap	**0.0008**	**0.0014**	**0.0023**	**0.0035**	0.0094	0.0201	**0.0225**	**0.0321**	**0.0433**	*0.0751*
VB-Lap	**0.0008**	*0.0016*	*0.0025*	*0.0039*	*0.0093*	**0.0139**	0.0241	0.0339	0.0457	0.0752
Spline	*0.0010*	0.0018	0.0028	0.0048	0.0139	0.0171	0.0267	0.0370	0.0515	0.0933
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0476**	**0.0192**	**0.0155**	**0.0180**	**0.0427**	**0.1656**	**0.1084**	**0.0975**	**0.1064**	**0.1630**
VB-HS	0.0900	0.0228	0.0181	0.0223	0.0538	0.2018	0.1181	0.1060	0.1175	0.1810
MCMC-Lap	*0.0524*	*0.0197*	*0.0158*	*0.0187*	0.0480	*0.1747*	*0.1102*	*0.0989*	*0.1086*	0.1731
VB-Lap	0.0672	0.0206	0.0165	0.0199	*0.0478*	0.1838	0.1130	0.1013	0.1120	*0.1723*
Spline	0.0925	0.0332	0.0265	0.0331	0.0917	0.2421	0.1453	0.1287	0.1446	0.2422

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.291	0.281	0.279	0.281	0.295	0.727	0.895	**0.914**	0.899	0.715
CVB-HS	1.077	0.551	0.374	0.566	1.086	**0.995**	**0.989**	**0.961**	**0.995**	**0.996**
VB-HS	0.103	0.153	0.160	0.153	0.106	0.280	0.569	0.629	0.590	0.286
MCMC-Lap	0.305	0.294	0.294	0.294	0.308	0.732	**0.907**	**0.930**	**0.911**	0.723
CVB-Lap	1.065	0.562	0.430	0.585	1.102	**0.995**	**0.994**	**0.987**	**0.996**	**0.996**
VB-Lap	0.131	0.181	0.192	0.181	0.133	0.364	0.687	0.752	0.707	0.369
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.074	0.103	0.133	0.157	0.170	**0.944**	**0.926**	**0.907**	0.852	0.665
CVB-HS	0.524	0.252	0.185	0.375	0.743	**1.000**	**0.990**	**0.943**	**0.990**	**0.997**
VB-HS	0.029	0.063	0.079	0.085	0.058	0.664	0.745	0.652	0.527	0.249
MCMC-Lap	0.072	0.104	0.137	0.162	0.179	**0.950**	**0.940**	**0.921**	0.871	0.661
CVB-Lap	0.529	0.258	0.192	0.379	0.760	**1.000**	**0.992**	**0.959**	**0.994**	**0.997**
VB-Lap	0.036	0.072	0.092	0.099	0.075	0.806	0.824	0.762	0.638	0.311
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.449	0.422	0.416	0.420	0.428	0.707	0.867	**0.907**	0.883	0.696
CVB-HS	1.327	0.771	0.547	0.769	1.290	**0.989**	**0.988**	**0.958**	**0.987**	**0.995**
VB-HS	0.161	0.220	0.233	0.221	0.158	0.275	0.541	0.614	0.540	0.266
MCMC-Lap	0.478	0.441	0.436	0.439	0.461	0.712	0.881	**0.921**	0.896	0.703
CVB-Lap	1.346	0.810	0.656	0.818	1.308	**0.991**	**0.995**	**0.988**	**0.995**	**0.996**
VB-Lap	0.198	0.264	0.278	0.264	0.198	0.348	0.643	0.721	0.644	0.349

Table 3.13: Average values of MSE, MAD, MCIW and CP based on 100 replications for 200 locations (PC) with $k = 0$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.0397	**0.0060**	**0.0040**	**0.0062**	**0.0404**	0.1528	**0.0547**	**0.0459**	**0.0565**	**0.1542**
VB-HS	*0.0358*	*0.0176*	*0.0122*	*0.0172*	*0.0375*	*0.1479*	0.0934	0.0765	0.0921	0.1511
ADMM	**0.0320**	0.0238	0.0143	0.0202	0.0366	**0.1309**	0.1187	0.0912	0.1111	0.1335
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.0052	**0.0016**	**0.0024**	**0.0042**	0.0209	0.0214	**0.0200**	**0.0310**	**0.0447**	0.1112
VB-HS	**0.0005**	0.0026	*0.0043*	*0.0077*	**0.0144**	**0.0100**	*0.0265*	*0.0404*	*0.0601*	**0.0944**
ADMM	*0.0009*	*0.0021*	0.0045	0.0130	*0.0185*	*0.0130*	0.0344	0.0514	0.0927	*0.0975*
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0777**	**0.0201**	**0.0144**	**0.0181**	**0.0811**	*0.2202*	**0.0958**	**0.0824**	**0.0942**	*0.2234*
VB-HS	*0.0866*	*0.0427*	*0.0325*	*0.0440*	*0.0898*	0.2296	*0.1510*	*0.1274*	*0.1530*	0.2309
ADMM	0.0891	0.0638	0.0471	0.0590	0.0940	**0.2184**	0.1879	0.1596	0.1830	**0.2162**

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.474	0.286	0.251	0.280	0.477	0.746	**0.942**	**0.957**	**0.927**	0.746
CVB-HS	1.003	0.507	0.374	0.497	0.991	**0.973**	**0.967**	**0.943**	**0.962**	**0.968**
VB-HS	0.115	0.187	0.194	0.185	0.117	0.237	0.619	0.724	0.625	0.235
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.066	0.094	0.136	0.185	0.270	**0.976**	**0.948**	**0.920**	0.895	0.625
CVB-HS	0.359	0.229	0.161	0.327	0.662	**0.997**	**0.975**	0.892	**0.947**	**0.977**
VB-HS	0.032	0.071	0.090	0.098	0.062	0.871	0.822	0.720	0.566	0.207
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.758	0.463	0.425	0.455	0.758	0.778	**0.924**	**0.949**	**0.927**	0.778
CVB-HS	1.435	0.766	0.629	0.756	1.449	**0.979**	**0.968**	**0.949**	**0.962**	**0.975**
VB-HS	0.177	0.281	0.293	0.282	0.177	0.232	0.564	0.672	0.560	0.237

Table 3.14: Average values of MSE, MAD, MCIW and CP based on 100 replications for 200 locations (VS) with $k = 1$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0279**	**0.0101**	**0.0090**	**0.0104**	**0.0279**	**0.1264**	**0.0759**	**0.0701**	**0.0751**	**0.1240**
VB-HS	*0.0365*	*0.0147*	*0.0124*	*0.0162*	*0.0362*	*0.1449*	*0.0889*	*0.0810*	*0.0921*	*0.1439*
ADMM	0.0719	0.0151	0.0138	0.0166	0.0405	0.1659	0.0927	0.0880	0.0968	0.1488
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.0023*	**0.0019**	**0.0027**	**0.0040**	**0.0115**	*0.0300*	**0.0277**	**0.0365**	**0.0460**	**0.0812**
VB-HS	**0.0022**	0.0025	0.0038	0.0060	*0.0134*	**0.0244**	*0.0300*	*0.0419*	0.0559	*0.0884*
ADMM	0.0946	*0.0023*	*0.0037*	*0.0055*	0.0223	0.1058	0.0311	0.0427	*0.0552*	0.1010
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0622**	**0.0269**	**0.0204**	**0.0237**	**0.0576**	**0.1858**	**0.1207**	**0.1080**	**0.1165**	**0.1854**
VB-HS	*0.0750*	*0.0310*	*0.0241*	*0.0316*	0.0795	*0.2116*	*0.1337*	*0.1175*	*0.1337*	0.2195
ADMM	0.0950	0.0336	0.0289	0.0356	*0.0732*	0.2168	0.1416	0.1322	0.1456	*0.2064*

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.362	0.321	0.315	0.322	0.354	0.725	**0.903**	**0.923**	**0.904**	0.732
CVB-HS	1.060	0.522	0.418	0.506	1.045	**0.988**	**0.973**	**0.943**	**0.955**	**0.982**
VB-HS	0.120	0.187	0.192	0.186	0.121	0.264	0.625	0.692	0.618	0.280
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.098	0.125	0.151	0.175	0.212	0.927	**0.935**	**0.905**	0.874	0.683
CVB-HS	0.460	0.251	0.187	0.332	0.677	**0.998**	**0.982**	**0.920**	**0.962**	**0.986**
VB-HS	0.038	0.079	0.097	0.103	0.066	0.692	0.784	0.698	0.583	0.246
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.515	0.490	0.479	0.477	0.540	0.711	0.890	**0.918**	0.891	0.745
CVB-HS	1.315	0.721	0.637	0.690	1.383	**0.984**	**0.963**	**0.962**	**0.953**	**0.985**
VB-HS	0.182	0.268	0.281	0.267	0.182	0.272	0.589	0.676	0.593	0.263

Table 3.15: Average values of MSE, MAD, MCIW and CP based on 100 replications for 200 locations (GP) with $k = 2$. The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively. The CP values above 90% are represented in bold.

	MSE					MAD				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	**0.0329**	**0.0111**	**0.0094**	**0.0115**	**0.0277**	**0.1309**	**0.0812**	**0.0747**	**0.0815**	**0.1239**
VB-HS	0.1083	0.0131	*0.0107*	*0.0142*	*0.0349*	0.1854	*0.0877*	*0.0788*	*0.0885*	*0.1408*
ADMM	*0.0409*	*0.0199*	0.0180	0.0218	0.0397	*0.1531*	0.1077	0.1017	0.1108	0.1516
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.0016*	**0.0017**	**0.0028**	**0.0044**	**0.0100**	0.0271	**0.0264**	**0.0373**	**0.0491**	**0.0777**
VB-HS	**0.0015**	*0.0023*	*0.0034*	*0.0054*	*0.0120*	**0.0206**	*0.0284*	*0.0406*	*0.0537*	*0.0854*
ADMM	**0.0015**	0.0028	0.0047	0.0068	0.0133	0.0209	0.0330	0.0478	0.0613	0.0904
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	*0.1091*	**0.0337**	**0.0231**	**0.0265**	**0.0615**	**0.2281**	**0.1367**	**0.1177**	**0.1272**	**0.1936**
VB-HS	0.1447	*0.0356*	**0.0231**	*0.0276*	*0.0770*	0.2562	*0.1375*	*0.1180*	*0.1299*	*0.2163*
ADMM	**0.0911**	0.0471	*0.0423*	0.0480	0.0939	*0.2400*	0.1719	0.1618	0.1715	0.2394

	MCIW					CP				
(I) Gauss	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.368	0.358	0.353	0.355	0.338	0.723	**0.914**	**0.937**	**0.917**	0.734
CVB-HS	1.105	0.528	0.440	0.522	1.065	**0.964**	**0.978**	**0.966**	**0.972**	**0.987**
VB-HS	0.139	0.195	0.206	0.196	0.130	0.280	0.632	0.709	0.642	0.300
(II) Beta	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.096	0.126	0.160	0.188	0.190	**0.947**	**0.948**	**0.913**	0.875	0.670
CVB-HS	0.462	0.252	0.204	0.333	0.693	**0.999**	**0.986**	**0.945**	**0.968**	**0.990**
VB-HS	0.037	0.081	0.100	0.105	0.072	0.635	0.803	0.705	0.596	0.267
(III) Mixed normal	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
MCMC-HS	0.5628	0.5995	0.5890	0.5730	0.5322	0.6995	**0.9126**	**0.9469**	**0.9162**	0.7287
CVB-HS	1.393	0.753	0.654	0.720	1.404	**0.962**	**0.967**	**0.965**	**0.963**	**0.987**
VB-HS	0.198	0.283	0.300	0.281	0.194	0.260	0.595	0.690	0.615	0.278

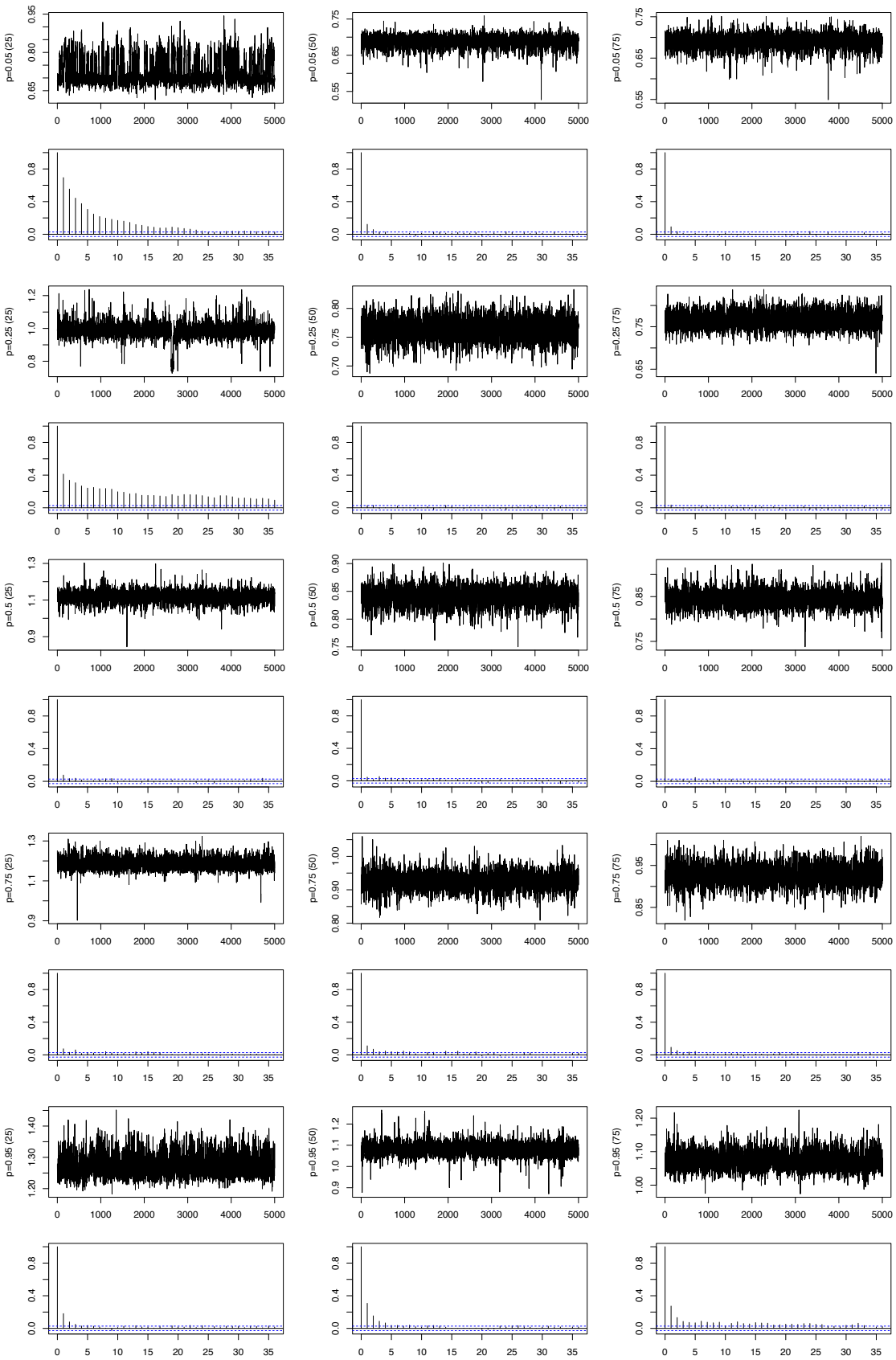

Figure 3.6: Trace plots and autocorrelations of posterior samples of $\theta_{25}, \theta_{50}, \theta_{75}$ (from left to right) for quantile levels $0.05, 0.25, 0.50, 0.75, 0.95$ (from top to bottom) in real data analysis of Nile data.

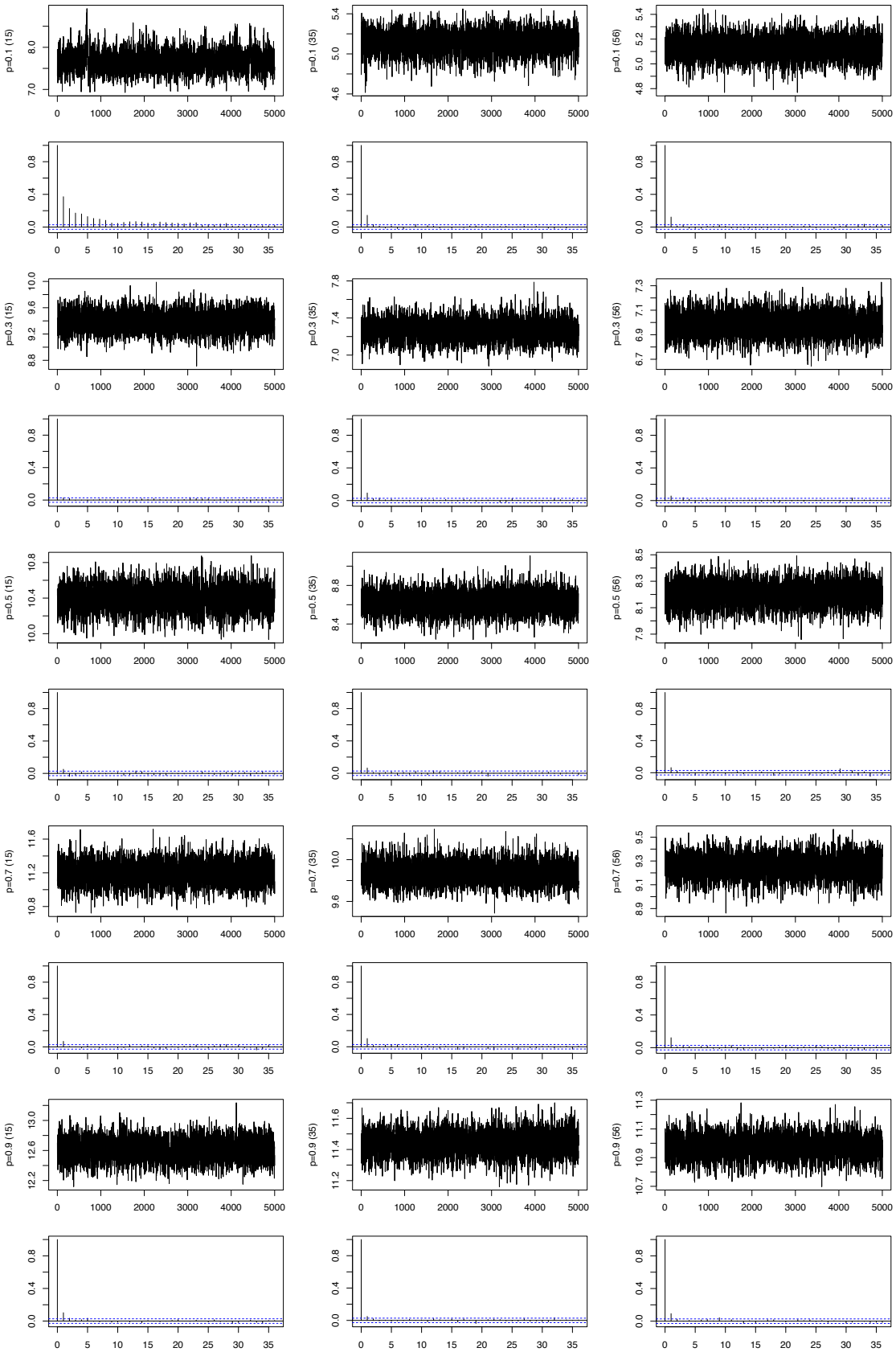

Figure 3.7: Trace plots and autocorrelations of posterior samples of $\theta_{15}, \theta_{35}, \theta_{56}$ (from left to right) for quantile levels $0.10, 0.30, 0.50, 0.70, 0.90$ (from top to bottom) in real data analysis of Munich rent data.

Chapter 4

Locally adaptive spatial quantile smoothing

4.1 Introduction

Estimating the spatial trend of the number of crimes is vital to ensure community safety and to respond quickly to incidents. For example, more police may be assigned to areas with a lot of crimes than to areas with few crimes. Tokyo metropolitan police department mentioned that crime predictions have some effects: 1) Efficient development of police officers, 2) Realization of improved public safety, 3) Improving police operations efficiency, and 4) Conducting effective patrols. Inference on the crime risk for each area is an important task for crime data analysis, and it has been revealed that crime can be controlled more effectively and efficiently by concentrating police enforcement efforts on high-risk spots and time (e.g. Braga, 2001). Since the number of crimes is often heterogeneous per region, the use of statistical models that take into account such heterogeneity is necessary. In Japan, University of Tsukuba Division of Policy and Planning Sciences Commons provides "GIS database of several police-recorded crimes at O-aza, chome in Tokyo, 2009–2017". The data contain the number of various crimes from 2009 to 2017 as well as spatial information and the area for each region. Recently, Hamura et al. (2021) and Yano et al. (2021) dealt with the data as zero-inflated count data and they proposed hierarchical Poisson models. It is known that crime data have spatial heterogeneity in the sense that most of the areas have little or no crime throughout multiple years, while others have a lot of crime yearly. Figure 4.1 shows the averaged values of violent crimes during 2013-2017 in Tokyo. The plot indicates that the distribution of violent crimes has spatial heterogeneity and there are several hotspots. Hamura et al. (2021) regarded

the hotspots as outliers and proposed a robust method for violent crimes in 2017. On the other hand, Yano et al. (2021) focused on the pickpocket (not violent crime) from 2012 to the first half of 2018 at 978 towns in eight wards, and considered a Bayesian prediction problem based on the Poisson distribution. Our goal in this study is to estimate the spatial high-risk trends of violent crime with uncertainty and to detect the potential risk throughout multiple years simultaneously. It is important to adaptively estimate trends without smoothing for potentially high-risk areas.

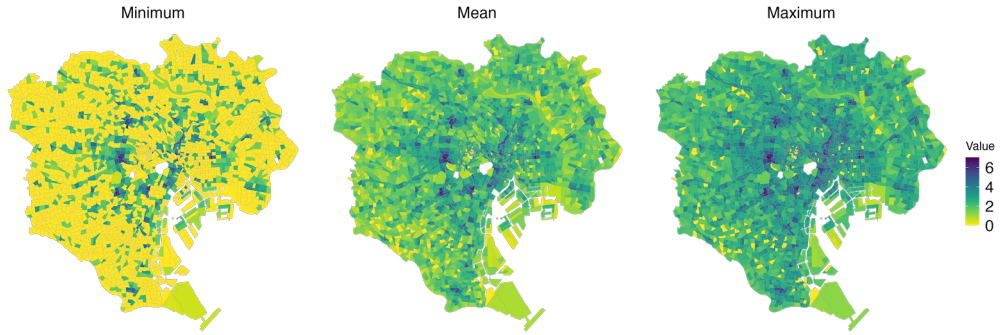

Figure 4.1: Spatial plot of $\log(1+Y)$ for crime density Y. From left to right, the minimum, mean, and maximum values for each area over five years from 2013 to 2017.

Spatial data with longitude and latitude information are considered point-level data. Statistical methods for such data have been developed. As a nonparametric Bayesian approach, Taddy (2010) considered the autoregressive mixture model and also provided an application of crime data analysis. On the other hand, data observed per region is known as areal data. Tokyo crime data considered in this Chapter is areal data, and it is constructed by the total number of crimes per region for a year. In other words, it does not make much sense to consider it as point-level data. For crime data as areal data, Balocchi and Jensen (2019) proposed a Bayesian linear regression model over time within a spatial correlation like conditional autoregressive formulation, and they applied their method to an analysis of violent crimes in Philadelphia. For the same data, Balocchi et al. (2023) also proposed the CAR-within-clusters model which assumes linear formulation and conditionally autoregression (CAR) model for each cluster, which deals with spatial discontinuity by introducing cluster and gives spatial continuity within a cluster. They recommend using the *crime density* defined as the number of crimes divided by the land area to deal with the difference in land size. The approach treats crime data as a

continuous value instead of count data. Following the study, we adopt the crime density in the Tokyo crime data, that is, we assume the continuous distribution on the distribution generating data in our modeling, not count data such as Poisson distribution.

In this study, we develop a quantile trend estimation for spatial data. The smoothing method has been studied in the context of function estimation to investigate the characteristics of the time series data. As mentioned in Section 2.3, one of the nonparametric methods to estimate underlying trends is the ℓ_1 trend filtering (Kim et al., 2009; Tibshirani, 2014). Wang et al. (2015) extend the original trend filtering to the spatial smoothing by assuming the graph structure. By accounting for covariates, Reich et al. (2011) proposed a Bayesian spatial quantile regression by introducing spatially varying basis-function coefficients. Castillo-Mateo et al. (2023) also considered spatial quantile autoregression for space-time dependence data, which is based on the Gaussian process model to capture spatial dependence over the grid cells.

There are some difficulties with these methods. The main difficulty in applying frequentist trend filtering is that uncertainty quantification is not straightforward. Moreover, the frequentist formulation includes tuning parameters that influence smoothness in the penalty term, but the data-dependent selection of the tuning parameter is not obvious, especially under quantile smoothing. While Bayesian methods are capable of mitigating these issues, the existing approach only focuses on time series data; thereby it cannot handle the smoothing of data on general graphs such as spatial data. Moreover, most of the studies focused on estimating a mean trend under a homogeneous variance structure, and these methods may not work well in data with heterogeneous variance. Nevertheless, quantile smoothing for spatial data has not been studied even from a frequentist perspective.

To overcome the issues, we extend the Bayesian quantile smoothing for time series data to Bayesian quantile trend filtering on general graphs including spatial neighboring structures, and also allow for multiple spatial data in which the number of samples for each location may be different. To this end, we employ the asymmetric Laplace distribution as a working likelihood (Yu and Moyeed, 2001), where the theoretical justification of using the likelihood is discussed in Sriram et al. (2013) and Sriram (2015). The novelty of the proposed approach is the construction of the prior distribution on the graph difference. In particular, we consider the horse-

shoe prior (Carvalho et al., 2010) as locally adaptive shrinkage priors for the graph differences. We introduce a novel hierarchical formulation for the prior, known as "shadow priors" that enhances the efficiency of posterior computation. Specifically, combining the data augmentation strategy by Kozumi and Kobayashi (2011), we develop a simple Gibbs sampling algorithm to generate posterior samples. We demonstrate the usefulness and wide applicability of proposed methods through extensive simulation studies and application to Tokyo crime data. We here present the advantage of the proposed trend filtering method compared with the existing Bayesian spatial methods: the simultaneous autoregressive (SAR) model and the Gaussian process (GP) model. In Figure 4.2, we show two examples of true quantile trends (adopted in simulation studies in Section 3.3), and their estimated results obtained by the proposed method (BQTF-HS) as well as SAR and GP models. It is observed that BQTF-HS tends to provide better estimation results than both SAR and GP models, successfully taking account of local changes and the smoothness of the true trend. Note that similar advantages of trend filtering were confirmed in the context of smoothing mean parameters (Tibshirani, 2014; Wang et al., 2015).

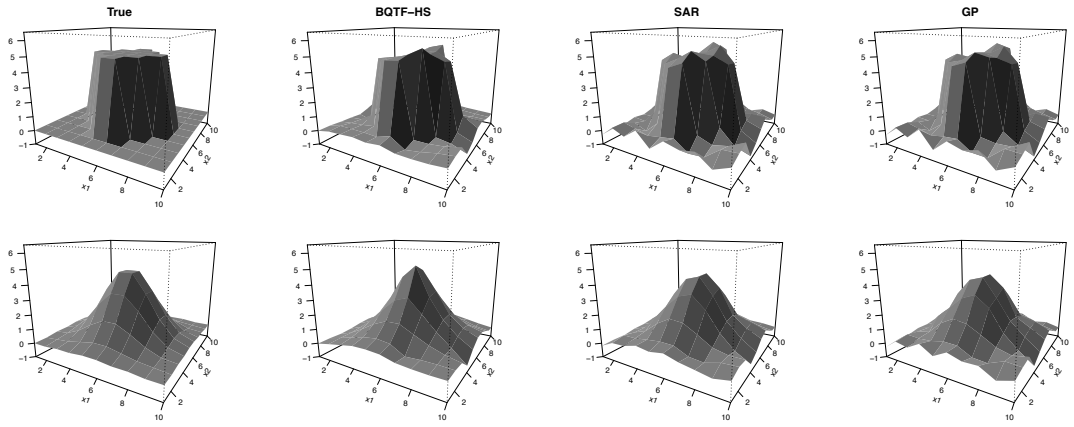

Figure 4.2: The examples of three methods for the 0.5-th quantile level. The left panels are two true signals. The estimates of the proposed methods under horseshoe and $k = 1$, SAR models, and GP models for two signals from the second from left to right.

This chapter is organized as follows: In Section 4.2, we propose a new Bayesian trend filtering method to estimate quantiles and construct an efficient posterior sampling algorithm based on Gibbs sampling. In Section 4.3, we illustrate some simulation studies to compare the performance of proposed methods. In Section 4.4, we apply the proposed methods to violent crime data in Tokyo. Additional

numerical results are provided in the Appendix 4.6. `R` code implementing the proposed methods is available in the GitHub repository (URL: `https://github.com/Takahiro-Onizuka/BSQS`).

4.2 Bayesian quantile trend filtering on graphs

4.2.1 Shrinkage priors on graph differences

We will consider the following model (see e.g. Onizuka et al., 2024a):

$$y_{ij} = \theta(x_i) + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \mathrm{AL}(p, \sigma^2), \quad i = 1, \ldots, n, \quad j = 1, \ldots, N_i, \tag{4.1}$$

where y_{ij} is a jth observation in the location x_i, $\theta(x_i) = \theta_i$ is a common quantile to $y_{i1}, \ldots, y_{iN_i}$ in the location x_i, N_i is the number of data per each location x_i. Here $\mathrm{AL}(p, \sigma^2)$ denotes the asymmetric Laplace distribution (2.3), where p is a fixed constant that characterizes the quantile level and σ^2 is a scale parameter. Note that the model (4.1) handles a situation with multiple observations per grid point, and $(y_{i1}, \ldots, y_{iN_i})$ are marginally correlated due to the common $\theta(x_i)$.

Suppose that spatial location $x = (x_1, \ldots, x_n)$ has a graph structure, and then $\theta_1, \ldots, \theta_n$ are on general graphs (including the standard trend filtering as a linear chain graph). The assumption is commonly used because the areal data has an adjacency relation and the simultaneous/conditional autoregressive models are also based on graph structure. Following Wang et al. (2015), let $G = (V, E)$ be an undirected graph with vertex set $V = \{1, \ldots, n\}$ and edge set E. We assume that $|V| = n$ and $|E| = m$. For $k = 0$, if $e_\ell = (i, j) \in V$, then $D^{(1)}$ has ℓ-th row

$$D^{(1)}_\ell = (0, \ldots, 0, \underbrace{1}_{i}, 0 \ldots, 0, \underbrace{-1}_{j}, 0, \ldots, 0), \tag{4.2}$$

where $1 \leq \ell \leq m$. For a graph G, the graph difference operator of order $k+1$ is denoted by $D^{(k+1)}$. When $k \geq 1$, graph difference operator $D^{(k+1)}$ is defined by

$$D^{(k+1)} = \begin{cases} (D^{(1)})^\top D^{(k)} & \text{for odd } k, \\ D^{(1)} D^{(k)} & \text{for even } k. \end{cases} \tag{4.3}$$

Here, we have $D^{(k+1)} \in \mathbb{R}^{n \times n}$ for odd k and $D^{(k+1)} \in \mathbb{R}^{m \times n}$ for even k. We note that the first-order graph difference operator $D^{(1)}$ is a natural generalization of the usual first-order difference operator in (2.5), and if we consider the linear chain

graph corresponding to time series data, then they coincide. The k controls the smoothness of the estimated trend. For example, $k = 0$ represents the assumption that the trend to estimate is piecewise constant like the upper left in Figure 4.2. $k \geq 1$ corresponds to the piecewise polynomial trend with degree k as an estimate for the unknown spatial trend. In other words, the estimate of θ_i has a relationship with its neighboring values like a polynomial function, which is similar to a local linear/polynomial regression. Empirically, we recommend $k = 1$ to capture changes and avoid over-fitting.

Let D be a $m \times n$ full-rank matrix representing a general difference operator on a graph, and we consider flexible shrinkage priors on $D\theta$. When m is smaller than n as in a linear chain graph, D can be transformed to $n \times n$ non-singular matrix (see also Onizuka et al., 2024a). We here assume that $m \geq n$ since the number of edges is typically larger than that of nodes. We consider the prior $D\theta \mid \tau^2, \sigma^2, w \sim N_n(0, \tau^2\sigma^2 W)$ with a diagonal covariance matrix $W = \mathrm{diag}(w_1^2, \ldots, w_m^2)$, where $w = (w_1, \ldots, w_m)$ represents local shrinkage parameters for each element in $D\theta$ and τ^2 is a global shrinkage parameter. When $m = n$, the prior can be rewritten as

$$\theta \mid \tau^2, \sigma^2, w \sim N_n(0, \sigma^2\tau^2(D^\top W^{-1} D)^{-1}).$$

Our idea is to use the above prior form even under $m > n$, noting that the covariance matrix $(D^\top W^{-1} D)^{-1}$ is still non-singular under $m > n$. The density function of the conditional prior of θ is given by

$$\pi(\theta \mid \tau^2, \sigma^2, w) = (2\pi\sigma^2\tau^2)^{-n/2} |D^\top W^{-1} D|^{1/2} \exp\left(-\frac{1}{2\sigma^2\tau^2}\theta^\top D^\top W^{-1} D\theta\right). \quad (4.4)$$

Now, we consider the prior for w. The standard approach is the use of an independent prior $\pi(w) = \prod_{i=1}^m \pi(w_i)$, and some familiar distribution is used for $\pi(w_i)$, for example, exponential prior or inverse gamma prior. However, the full conditional distribution of w is not a familiar form due to the term $|D^\top W^{-1} D|^{1/2}$ in the density (4.4). Therefore, it is not easy to construct an efficient Gibbs sampler. Alternatively, we consider the following joint prior:

$$\pi(w) \propto |D^\top W^{-1} D|^{-1/2} |W|^{-1/2} \prod_{i=1}^m \pi(w_i), \quad (4.5)$$

where $\pi(w_i)$ is a proper univariate distribution. For a square matrix D such that k is odd, the joint prior equals the product of the standard prior $\pi(w_i)$. As shown

in Subsection 4.2.2, the resulting full conditional distributions of w are familiar forms under well-known priors for local shrinkage parameters. As a result, we can construct a Gibbs sampler for the proposed method. Such priors given in (4.5) are known as "shadow priors", and are used to improve the mixing of Markov chain Monte Carlo (MCMC) algorithm (e.g. Liechty et al., 2009) or to construct tractable full conditional distributions (e.g. Liu et al., 2014; Xu and Ghosh, 2015). Note that these works demonstrate that the use of shadow prior has little effect on posterior inference.

As an univariate distribution $\pi(w_i)$ in (4.5), we consider two types of distributions, $w_i \sim \text{Exp}(1/2)$ and $w_i \sim C^+(0,1)$. These priors are motivated by the Bayesian lasso prior (Park and Casella, 2008) and horseshoe prior (Carvalho et al., 2010), respectively. Regarding the other parameters, we assign $\sigma^2 \sim \text{IG}(a_\sigma, b_\sigma)$ and $\tau \sim C^+(0, C_\tau)$, where a_σ, b_σ and C_τ are fixed hyper-parameters.

The proposed prior for θ belongs to a class of general priors, described as

$$\theta \mid \sigma^2, \tau^2, \rho \sim N(0, \sigma^2\tau^2 Q(\rho)). \tag{4.6}$$

Note that the simultaneous autoregressive (SAR) and Gaussian process (GP) prior are popular approaches for spatial smoothing and the priors can also be expressed as (4.6) with different matrix $Q(\rho)$ from that of the proposed prior. The two priors will be compared through simulation studies and more detailed explanations are provided in Section 4.3.

Note that the three conditional priors of θ include σ^2 in the scale although σ^2 is the scale parameter of the likelihood (see equations (4.4) and (4.6)). The formulation has been often used for the conditional normal prior (e.g. Polson and Scott, 2012) and induces the advantage that the scale of the prior is automatically adjusted when units of observations are changed.

4.2.2 Markov chain Monte Carlo algorithm

To develop an efficient posterior computation algorithm via Gibbs sampling, we employ the stochastic representation of the asymmetric Laplace distribution (Kozumi and Kobayashi, 2011). For $\varepsilon_{ij} \sim \text{AL}(p, \sigma^2)$, we have the following argumentation

$$\varepsilon_{ij} = \psi z_{ij} + \sqrt{\sigma^2 z_{ij} t^2} u_{ij}, \quad \psi = \frac{1-2p}{p(1-p)}, \quad t^2 = \frac{2}{p(1-p)},$$

where $u_{ij} \sim N(0,1)$ and $z_{ij} \mid \sigma^2 \sim \text{Exp}(1/\sigma^2)$ for $i = 1, \ldots, n$. From the above expression, the conditional likelihood function of y_{ij} is given by

$$p(y_{ij} \mid \theta_i, z_{ij}, \sigma^2) = (2\pi t^2 \sigma^2)^{-1/2} z_{ij}^{-1/2} \exp\left\{ -\frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 \sigma^2 z_{ij}} \right\}.$$

Then, under the conditionally Gaussian prior of θ in (4.4), the full conditional distributions of z_i and θ are given by

$$\begin{aligned}
&\theta \mid y, z, \sigma^2, \gamma^2 \sim N_n\left(A^{-1}B, \sigma^2 A^{-1}\right), \\
&z_{ij} \mid y_{ij}, \theta_i, \sigma^2 \sim \text{GIG}\left(\frac{1}{2}, \frac{(y_{ij} - \theta_i)^2}{t^2\sigma^2}, \frac{\psi^2}{t^2\sigma^2} + \frac{2}{\sigma^2}\right), \quad i = 1, \ldots, n, \ j = 1, \ldots, N_i,
\end{aligned}$$

where

$$\begin{aligned}
A &= \frac{1}{\tau^2} D^\top W^{-1} D + \frac{1}{t^2} \text{diag}\left(\sum_{j=1}^{N_1} z_{1j}^{-1}, \ldots, \sum_{j=1}^{N_n} z_{nj}^{-1} \right), \\
B &= \left(\sum_{j=1}^{N_1} \frac{y_{1j} - \psi z_{1j}}{t^2 z_{1j}}, \ldots, \sum_{j=1}^{N_n} \frac{y_{nj} - \psi z_{nj}}{t^2 z_{nj}} \right)^\top
\end{aligned}$$

and $\text{GIG}(a, b, c)$ denotes the generalized inverse Gaussian distribution. The full conditional distributions of the scale parameter of observations, σ^2, and global shrinkage parameter τ^2 are given by

$$\begin{aligned}
&\sigma^2 \mid y, \theta, z, w, \tau^2 \sim \text{IG}\left(\frac{n + 3N}{2} + a_\sigma, \beta_{\sigma^2}\right), \\
&\tau^2 \mid \theta, w, \sigma^2, \xi \sim \text{IG}\left(\frac{n+1}{2}, \frac{1}{2\sigma^2}\theta^\top D^\top W^{-1} D\theta + \frac{1}{\xi}\right), \qquad \xi \mid \tau^2 \sim \text{IG}\left(\frac{1}{2}, \frac{1}{\tau^2} + 1\right), \\
&\beta_{\sigma^2} = \sum_{i=1}^{n} \sum_{j=1}^{N_i} \frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 z_{ij}} + \frac{\theta^\top D^\top W^{-1} D\theta}{2\tau^2} + \sum_{i=1}^{n} \sum_{j=1}^{N_i} z_{ij} + b_\sigma,
\end{aligned}$$

where N is the number of total data and ξ is an augmented parameter for τ^2. The full conditional distributions of the other parameters depend on the specific choice of the distributional form of $\pi(w_i)$, which are summarized as follows.

- **(Laplace-type prior)** The full conditional distributions of θ, z_i, and σ^2 have already been mentioned. For the Laplace-type prior, we give $\tau^2 = 1$ and $w_i \mid \gamma^2 \sim \text{Exp}(\gamma^2/2)$. In this condition, we can model that $(D\theta)_i \sim \text{Lap}(\gamma)$. Because our condition is $\gamma \sim C^+(0, 1)$, by using the representation that if $\text{IG}(\gamma^2 \mid 1/2, 1/\nu)$ and $\text{IG}(\nu \mid 1/2, 1/a^2)$, then $\gamma \sim C^+(0, a)$, the full conditional

distributions of w_i, γ^2 and ν are given by

$$w_i^2 \mid \theta, \sigma^2, \gamma^2, \nu \sim \mathrm{GIG}\left(\frac{1}{2}, \frac{\eta_i^2}{\sigma^2}, \gamma^2\right),$$
$$\gamma^2 \mid w, \nu \sim \mathrm{GIG}\left(m - \frac{1}{2}, \frac{2}{\nu}, \sum_{i=1}^{m} w_i^2\right), \quad \nu \mid \gamma^2 \sim \mathrm{IG}\left(\frac{1}{2}, \frac{1}{\gamma^2} + 1\right),$$

where $\mathrm{GIG}(a, b, c)$ is the generalized inverse Gaussian distribution and $\eta_i = (D\theta)_i$.

- **(Horseshoe-type prior)** The full conditional distributions of θ, z_i, σ^2 and τ^2 have already been mentioned. For the Horseshoe-type prior, $w_i \sim C^+(0, 1)$. By using the representation that $w_i^2 \mid \nu_i \sim \mathrm{IG}(1/2, 1/\nu_i)$ and $\nu_i \sim (1/2, 1)$, the full conditional distributions of w_i and ν_i are given by

$$w_i^2 \mid \theta, \sigma^2, \gamma^2, \nu \sim \mathrm{IG}\left(1, \frac{1}{\nu_i} + \frac{\eta_i^2}{2\sigma^2\tau^2}\right), \quad \nu_i \mid w_i \sim \mathrm{IG}\left(\frac{1}{2}, \frac{1}{w_i^2} + 1\right),$$

where $\mathrm{IG}(a, b)$ is the inverse Gamma distribution and $\eta_i = (D\theta)_i$.

4.3 Simulation studies

We illustrate the performance of the proposed method through simulation studies.

4.3.1 Simulation setting

We show simulation studies for data on 2-D lattice graphs. We formulate the data-generating process as follows. Let $G = (V, E)$ be a 2-D lattice graph. We set $V = \{1, \ldots, 100\}$ and $|E| = 180$ for the graph G. The edges are defined by whether the lattice is adjacent or not. A more general graph structure is also considered in the Appendix 4.6. Noisy data were generated from the model $y_{ij} = f(x_i) + \varepsilon(x_i)$ ($i = 1, \ldots, 100$, $j = 1, \ldots, 5$), where $x_i = (x_{i1}, x_{i2})$ is a two-dimensional coordinate, and $f(x)$ and $\epsilon(x)$ are true and noise functions, respectively. Based on the model 3.2, we generated five data for each location i. The following two true functions were considered:

- Two block structure

$$f(x_i) = 5 \ \ (\text{center}), \quad \text{and} \quad f(x_i) = 0 \ \ (\text{other}),$$

- Exponential function

$$f(x_i) = 5\exp\left(-\frac{1}{2}(x_i-\mu)^\top\Sigma^{-1}(x_i-\mu)\right), \quad \mu = (5.5, 5.5), \quad \Sigma = 3I_2,$$

where $x_i = (x_{i1}, x_{i2})$, $x_{i1}, x_{i2} = 1, 2, \ldots, 10$ and I_n is $n \times n$ identity matrix. These functions are shown in Figure 4.3. As noise functions $\epsilon(x)$, we considered the following three structures:

(I) Homogeneous: $\epsilon(x_i) \sim N(0, 1)$.

(II) Block heterogeneous:

$$\epsilon(x_i) \sim \begin{cases} N(0, 0.5^2) & (1 \leq x_{i1} \leq 5,\ 1 \leq x_{i2} \leq 5) \\ N(0, 2^2) & (6 \leq x_{i1} \leq 10,\ 6 \leq x_{i2} \leq 10) \\ N(0, 1) & (\text{otherwise}) \end{cases}.$$

(III) Smooth heterogeneous:

$$\epsilon(x_i) \sim \begin{cases} N(0, 0.5^2) & (1 \leq x_{i1} \leq 4,\ 1 \leq x_{i2} \leq 4) \\ N(0, 1) & (x_{i1} = 5, 6,\ 1 \leq x_{i2} \leq 6 \text{ or } 1 \leq x_{i1} \leq 6,\ x_{i2} = 5, 6) \\ N(0, 1.5^2) & (x_{i1} = 7, 8,\ 1 \leq x_{i2} \leq 8 \text{ or } 1 \leq x_{i1} \leq 8,\ x_{i2} = 7, 8) \\ N(0, 2^2) & (\text{otherwise}) \end{cases}.$$

The two-block structure is a reasonable function to verify the ability to capture the jump without smoothing. In the exponential function, we examine the ability to estimate a continuous curve with noisy data. The noise (I) represents spatial homogeneity, while the noise may not be realistic in practical situations. In noise (II) and (III), the aim is to verify how well the proposed method can handle spatial heterogeneity. In particular, noise (III) has a stronger degree of spatial heterogeneity than noise (II). The visualizations of these noise distributions are given in the Appendix 4.6. Combining two true structures and three noise functions, we consider six scenarios. Scenarios (i), (ii), and (iii) are based on two block structure and noise functions (I), (II), and (III), respectively. Scenarios (iv), (v), and (vi) are based on exponential structure and noise functions (I), (II), and (III), respectively. Since the noise function is homogeneity, scenarios (i) and (iv) are easier, and scenario (iv) would especially be the easiest because of its smoothness and homogeneity. Scenario (iii) constructed by a two-block structure and smooth heterogeneous noise

would be the most difficult for two reasons: hard to capture the jump points and heavy heterogeneity.

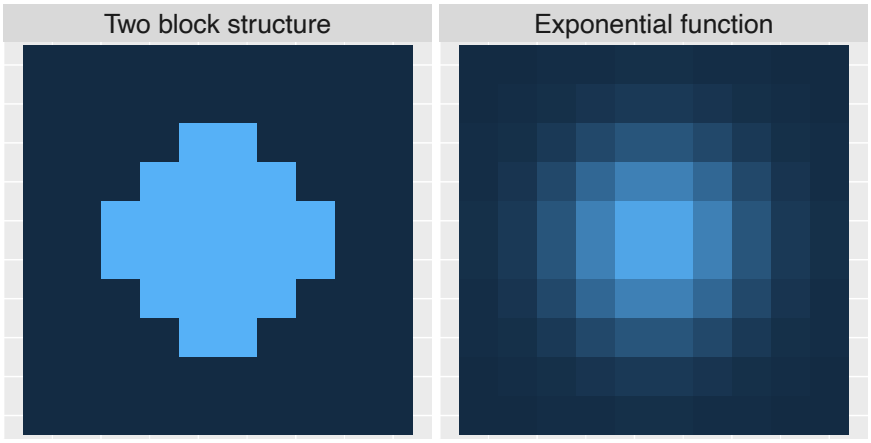

Figure 4.3: Two types of true function $f(x)$.

We used the two proposed methods (denoted by BQTF-HS and BQTF-Lap), where HS and Lap are the horseshoe and Laplace priors, respectively. Although there is no previous research about spatial quantile smoothing, to evaluate the performance of the proposed method, we compare the BQTF methods with the following three methods:

- SAR: Bayesian simultaneous autoregressive (SAR) quantile model, which is based on graph structure as well as BQTF. The SAR prior takes the form of (4.6) and is based on a graph structure with a contingency matrix Ω. The matrix $Q(\rho)$ is given by $Q(\rho)^{-1} = (I_n - \rho\Omega)^\top (I_n - \rho\Omega)$, and the parameter ρ controls the effect of the spatial correlation. The MCMC algorithm is summarized in the Appendix 4.6. Note that the parameter ρ was sampled by the random walk Metropolis-Hastings (MH) algorithm.

- GP: Bayesian quantile smoothing under Gaussian process prior, which the prior takes the form of (4.6) and is based on the location of a data observed point. The matrix $Q(\rho)$ is given by $Q_{ij}(\rho) := (Q(\rho))_{i,j} = \exp\left(-\|x_i - x_j\|/(2\rho)\right)$, where ρ also controls the effect of the spatial dependence between the location x_i and x_j. The MCMC algorithm is summarized in the Appendix 4.6. Note that the parameter ρ was sampled by the random walk MH algorithm.

- qgam: Additive quantile regression which is the frequentist method proposed

by Fasiolo et al. (2021). The method can be implemented by using their `qgam` R package. Let $x_i = (x_{i1}, x_{i2})$ be a two-dimensional coordinate and y_i be a observed data. Then the corresponding estimate $\hat{\theta}_i$ of θ_i is obtained by the sum of functions $\hat{\theta}_i = \hat{q}_1(x_{i1}) + \hat{q}_2(x_{i2})$, where $\hat{q}_1(\cdot)$ and $\hat{q}_2(\cdot)$ are nonparameteric estimates of quantile functions.

Note that the detailed posterior computation algorithms of SAR and GP methods are presented in Appendix 4.6. Since the SAR and GP models are based on the Gaussian type prior, we could not expect locally adaptive smoothing. If the data y_i is generated from a simple true function $f(x_i) = f_1(x_{i1}) + f_2(x_{i2})$, then the qgam method would give a pretty good smoothing, but it is unrealistic in a practical situation and the simulation setting is more complicated. The other methods such as spatial regression models are also compared and the results are summarized in the Appendix 4.6. For the Bayesian methods, we generated 7,500 posterior samples, and then only every 10th scan was saved, and the order of trend filtering was set as $k = 1$ (i.e. area-wise linear trend). We estimate five quantile levels: 0.1, 0.3, 0.5, 0.7, and 0.9. To evaluate the performance of estimates, we adopt the mean squared error (MSE), the mean absolute deviation (MAD), the mean credible interval width (MCIW), and the coverage probability (CP) which are defined by

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\theta_i^* - \hat{\theta}_i)^2, \quad \text{MAD} = \frac{1}{n}\sum_{i=1}^{n}|\theta_i^* - \hat{\theta}_i|,$$
$$\text{MCIW} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{97.5,i} - \hat{\theta}_{2.5,i}, \quad \text{CP} = \frac{1}{n}\sum_{i=1}^{n} I(\hat{\theta}_{2.5,i} \le \theta_i^* \le \hat{\theta}_{97.5,i}),$$

respectively, where $\hat{\theta}_{100(1-\alpha),i}$ represent the $100(1-\alpha)\%$ posterior quantiles of θ_i and θ_i^* are true quantiles at location x_i. These values were averaged over 100 replications of simulating datasets.

4.3.2 Simulation result

Simulation results are shown in Tables 4.1 and 4.2. Note that MCIW and CP are reported only for Bayesian methods. From Tables 4.1 and 4.2, the proposed BQTF method under horseshoe prior tends to provide a reasonable point estimate not only in the case of homogeneous but also for heterogeneous variances. When the true structure is the exponential function (such as scenarios (iv), (v), and (vi)), the proposed two methods provide comparable point estimates, and the additive quantile

regression has smaller MSE and MAD than that for (i), (ii) and (iii) scenarios. However, it is observed that the additive quantile regression does not work well for any scenario compared with the proposed methods. The MAD of the proposed BQTF-HS is smaller than that of the SAR and GP models for all cases, while the SAR model is sometimes the best for exponential function in terms of MSE because of the smooth trend structure. In comparison between the SAR and GP models, the SAR model is better than the GP model in terms of MSE and MAD. For uncertainty quantification, while BQTF methods have reasonable coverage probabilities for the 50% quantile trend, the coverage probabilities of 95% credible intervals for extremal quantiles such as 0.1 and 0.9 seem to be far away from the nominal coverage rate of 0.95. Note that the mean credible interval width (MCIW) is the order of HS, Lap, SAR, and GP.

4.4 Application to crime trend analysis in Tokyo

We apply the proposed methods to spatial data analysis. We used the "GIS database of number of police-recorded crime at O-aza, chome in Tokyo, 2009–2017", which was provided by University of Tsukuba Division of Policy and Planning Sciences Commons. The "chome" represents a specific area, block or street within a city or town. For example, "3-chome, Shinjuku" would refer to the third block within Shinjuku town in Tokyo. The data contains the number of crimes in Tokyo, and we focus on the violent crime data in particular. We used the number of violent crimes from some 23 wards in Tokyo for five years (from 2013 to 2017) whose number of locations is $n = 3,125$ and the sample size is $N = 3125 \times 5 = 15,625$. The number of edges is 8,996. The edges are constructed based on the 5 nearest neighbor searches. Namely, when x_j is in 5 nearest neighbors of x_i, then we connect x_i and x_j even if they are not adjacent to each other on the map. Since the data also involve information on the area (km^2) of each region, we define $Y = (Y_1, \ldots, Y_{3125})$ as the values of the number of violent crimes divided by the area for each region, which are called *crime density* as we mentioned in Section 4.1. Using the value of Y may be reasonable because the larger the area, the greater the number of crimes in general. Balocchi et al. (2023) also used the crime density normalized by the area. Moreover, we use the value on the log scale as data $y = \log(1 + Y)$. Such a transformation is popular in the literature (see also Balocchi and Jensen, 2019).

Table 4.1: Average values of MSE, MAD, MCIW, and CP based on 100 replications for scenarios (i), (ii), and (iii) (two-block structure). The minimum values of MSE and MAD are represented in bold.

	Scenario (i)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.264**	**0.158**	**0.138**	**0.152**	**0.248**	**0.377**	**0.283**	**0.266**	**0.280**	**0.371**
Lap	0.337	0.212	0.193	0.211	0.339	0.460	0.359	0.344	0.360	0.463
SAR	0.345	0.224	0.206	0.224	0.345	0.469	0.377	0.362	0.378	0.470
GP	0.347	0.217	0.199	0.217	0.354	0.470	0.372	0.356	0.373	0.476
qgam	3.985	2.461	1.877	2.269	3.062	1.256	1.266	1.195	1.266	1.425
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.322	1.287	1.262	1.261	1.290	0.822	0.920	0.934	0.921	0.827
Lap	1.400	1.554	1.552	1.541	1.395	0.799	0.910	0.921	0.906	0.801
SAR	1.539	1.706	1.706	1.693	1.534	0.826	0.923	0.935	0.920	0.829
GP	1.525	1.712	1.717	1.703	1.510	0.826	0.928	0.942	0.927	0.826
	Scenario (ii)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.464**	**0.272**	**0.244**	**0.256**	**0.451**	**0.454**	**0.337**	**0.317**	**0.328**	**0.447**
Lap	0.542	0.311	0.283	0.306	0.541	0.524	0.400	0.381	0.399	0.526
SAR	0.537	0.316	0.289	0.312	0.522	0.531	0.415	0.395	0.414	0.529
GP	0.545	0.310	0.282	0.309	0.549	0.533	0.409	0.389	0.410	0.538
qgam	3.973	2.419	1.891	2.219	3.208	1.315	1.241	1.197	1.256	1.459
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.440	1.409	1.369	1.378	1.388	0.814	0.911	0.920	0.909	0.814
Lap	1.543	1.689	1.675	1.678	1.528	0.801	0.907	0.919	0.907	0.799
SAR	1.694	1.854	1.839	1.838	1.688	0.824	0.921	0.932	0.920	0.825
GP	1.681	1.864	1.858	1.852	1.662	0.823	0.924	0.937	0.924	0.823
	Scenario (iii)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.651**	**0.395**	**0.346**	**0.375**	**0.602**	**0.581**	**0.438**	**0.408**	**0.426**	**0.559**
Lap	0.740	0.431	0.389	0.424	0.740	0.643	0.493	0.469	0.489	0.642
SAR	0.736	0.427	0.387	0.422	0.721	0.648	0.501	0.479	0.499	0.644
GP	0.760	0.422	0.378	0.416	0.768	0.656	0.497	0.472	0.494	0.661
qgam	3.792	2.295	1.908	2.174	2.984	1.366	1.258	1.204	1.255	1.422
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.814	1.798	1.767	1.764	1.783	0.795	0.896	0.912	0.899	0.806
Lap	1.932	2.076	2.063	2.061	1.911	0.790	0.901	0.916	0.899	0.794
SAR	2.092	2.254	2.240	2.232	2.073	0.820	0.920	0.930	0.916	0.825
GP	2.071	2.276	2.273	2.270	2.048	0.815	0.925	0.938	0.925	0.821

Table 4.2: Average values of MSE, MAD, MCIW, and CP based on 100 replications for scenarios (iv), (v), and (vi) (exponential function). The minimum values of MSE and MAD are represented in bold.

	Scenario (iv)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.135**	**0.082**	0.075	0.082	**0.135**	**0.280**	**0.218**	**0.207**	**0.214**	**0.274**
Lap	0.184	**0.082**	**0.072**	**0.081**	0.186	0.341	0.224	0.212	0.226	0.342
SAR	0.195	0.087	0.076	0.085	0.184	0.357	0.234	0.220	0.233	0.347
GP	0.251	0.119	0.107	0.125	0.273	0.405	0.275	0.262	0.285	0.423
qgam	0.577	0.450	0.386	0.418	0.510	0.522	0.503	0.504	0.522	0.551
	MCIW					CP				
HS	0.994	0.982	0.983	0.986	1.021	0.816	0.907	0.921	0.915	0.846
Lap	1.160	1.148	1.136	1.143	1.169	0.857	0.952	0.964	0.953	0.863
SAR	1.275	1.259	1.244	1.243	1.263	0.887	0.967	0.975	0.970	0.893
GP	1.333	1.424	1.433	1.445	1.348	0.878	0.961	0.972	0.961	0.872
	Scenario (v)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.284**	0.125	0.109	0.122	0.282	**0.355**	**0.246**	0.228	**0.242**	**0.349**
Lap	0.313	0.116	0.095	0.115	0.312	0.397	0.248	**0.225**	0.246	0.396
SAR	0.289	**0.112**	**0.093**	**0.110**	**0.272**	0.398	0.251	0.229	0.247	0.385
GP	0.385	0.152	0.131	0.166	0.418	0.457	0.293	0.272	0.304	0.475
qgam	0.686	0.472	0.398	0.432	0.619	0.588	0.515	0.508	0.528	0.601
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.141	1.075	1.067	1.082	1.154	0.816	0.907	0.926	0.915	0.832
Lap	1.306	1.228	1.208	1.225	1.302	0.849	0.947	0.959	0.949	0.853
SAR	1.434	1.339	1.308	1.321	1.405	0.876	0.963	0.971	0.961	0.883
GP	1.514	1.533	1.525	1.557	1.517	0.858	0.959	0.969	0.955	0.853
	Scenario (vi)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.348**	**0.169**	0.142	**0.151**	**0.311**	**0.423**	**0.305**	**0.278**	**0.287**	**0.400**
Lap	0.435	0.170	0.140	0.157	0.408	0.487	0.307	0.280	0.296	0.470
SAR	0.424	**0.169**	**0.138**	**0.151**	0.362	0.495	0.313	0.283	0.295	0.456
GP	0.579	0.224	0.182	0.221	0.587	0.578	0.361	0.327	0.359	0.584
qgam	0.916	0.598	0.492	0.505	0.655	0.718	0.591	0.565	0.579	0.660
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.392	1.293	1.272	1.273	1.413	0.798	0.890	0.911	0.901	0.833
Lap	1.634	1.518	1.477	1.484	1.613	0.839	0.938	0.953	0.943	0.853
SAR	1.794	1.664	1.606	1.593	1.720	0.872	0.959	0.969	0.960	0.888
GP	1.887	1.886	1.853	1.894	1.872	0.849	0.957	0.971	0.960	0.860

We regard five years of data as multiple observation data per location. The data is shown in Figure 4.4, and the plot indicates that spatial trends have not changed over the years. Additionally, the histograms of y for each year and all years are also shown in Figure 4.5, which represent that the distribution of observed data is the same for all years. Although there are some hotspots for each year in Figure 4.4, some high-risk areas have overlapped throughout the five years. They tend to be particularly common in downtown areas, and such areas can be seen as ones with potentially high risk. In this section, our goal is to estimate spatial quantile trends and detect potential hotspots. In particular, since we are interested in the median and high-risk cases of criminal activity, we estimate 50% and 90% quantile trends. We adopt the proposed Bayesian quantile trend filtering under horseshoe prior (BQTF-HS) and compare the performance with two methods: the SAR model and the additive quantile regression (qgam) using latitude and longitude as covariates. The GP model has a high computation cost because the covariance matrix is not a sparse matrix, unlike the BQTF and SAR models. Therefore, although the GP models can be applied to the example, we only consider the above two methods as competitors. For the Bayesian methods, we generated 50,000 posterior samples, and then the first 10,000 samples were discarded and only every 40th scan was saved. The order of trend filtering is $k = 1$. The estimated quantile trends are shown in Figure 4.6. Note that if the estimate has a negative value, then it is plotted as zero. The proposed BQTF method seems to capture the zero-inflated data throughout five years. For the 90% quantile trend, the BQTF method provides the adaptive smoothing that detects not only high-risk spots but also low-risk spots and gives smoothing a high quantile trend. The estimate of the SAR model is not smoother than the BQTF methods and is similar to raw data shown in Figure 4.4. On the other hand, the additive quantile regression (qgam) method results in over-shrinkage and clearly can not achieve a locally adaptive smoothing. In other words, the qgam method cannot detect hotspots, and the areas that seem to be not hotspots also have a green or blue color. Therefore, we can conclude that the proposed BQTF method gives a more reasonable estimate of potential quantile trend than the SAR model and qgam, which are not smooth or producing over-shrinkage results. The six hotspots detected by the proposed are shown in Figure 4.6, which are the main stations (Shinjuku, Ikebukuro, Shibuya, Shinbashi, Tokyo, and Akihabara) in the Yamanote line, and the areas are filled by blue, which have high values. Such areas seem to be outliers in

Hamura et al. (2021), and the same result is observed. Moreover, the lower left area filled by yellow is considered lower-risk in terms of 50% trend, and the spatial effect analyzed in Hamura et al. (2021) also has small values in these areas. However, as seen in the 90% trend, it seems that the risk in these areas is not low potentially. The potential risk is not clear from the 50% quantile trend or the other method. Hence, the BQTF method provides a locally adaptive smoothing for high quantile trends and captures latent heterogeneity by treating five years of data as multiple observations.

Figure 4.4: Spatial plot of $\log(1 + Y)$ for crime density Y based on raw data from 2013 to 2017.

4.5 Concluding remarks

In this study, we proposed a Bayesian quantile trend filtering (BQTF) method on graphs under continuous shrinkage priors, which enables us to estimate quantile trends for spatial data. We also provide a simple Gibbs sampler by introducing a kind of shadow prior. Through simulation studies, it is shown that the BQTF estimates under the horseshoe prior provide locally adaptive smoothing in the sense

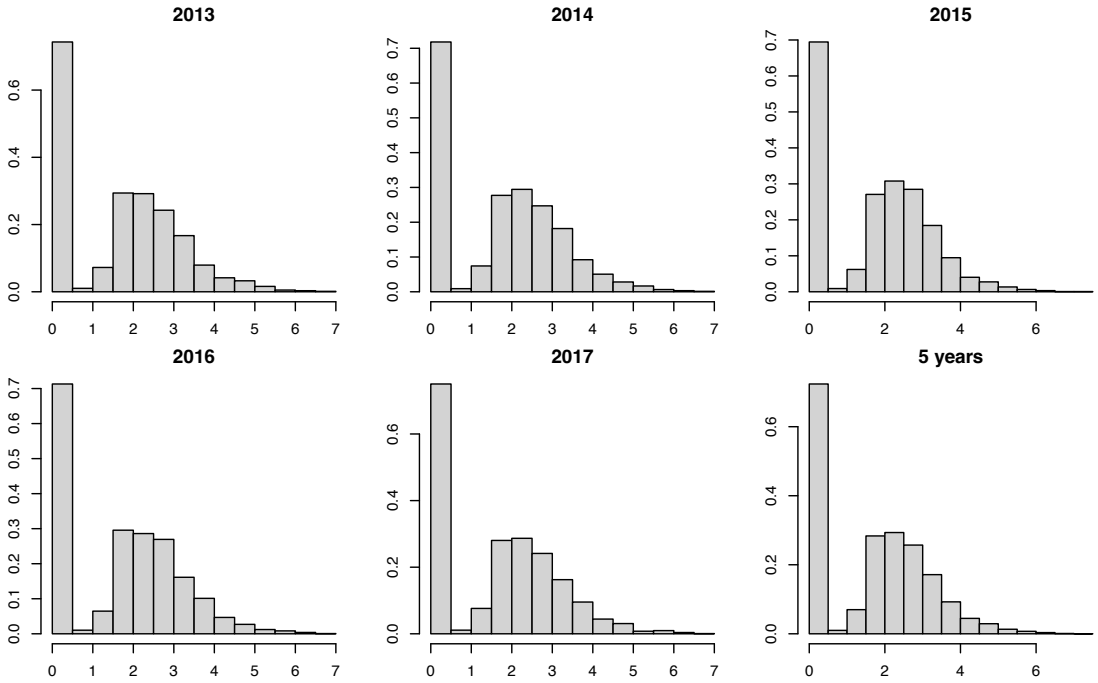

Figure 4.5: The histogram of the crime densities for each year and five-year data.

of capturing the change of quantile trends and estimating the smooth quantile trends. The application of the violent crime data in Tokyo gives interesting results in that the proposed method provides locally adaptive quantile smoothing for all quantiles and detects hotspots focusing on a low quantile level.

In terms of application, locally adaptive smoothing of point-referenced spatial data is an important problem in future work, while we consider the smoothing of areal data. The trend estimation of point-level data has also been studied (see also Lum and Gelfand, 2012). Since the observation points of these data are different between years, the proposed methods cannot be used as is to detect potential hotspots throughout multiple years.

4.6 Appendix

This section provides additional information for the simulation study, the MCMC algorithm of the other methods, and the additional analysis of Tokyo Crime data.

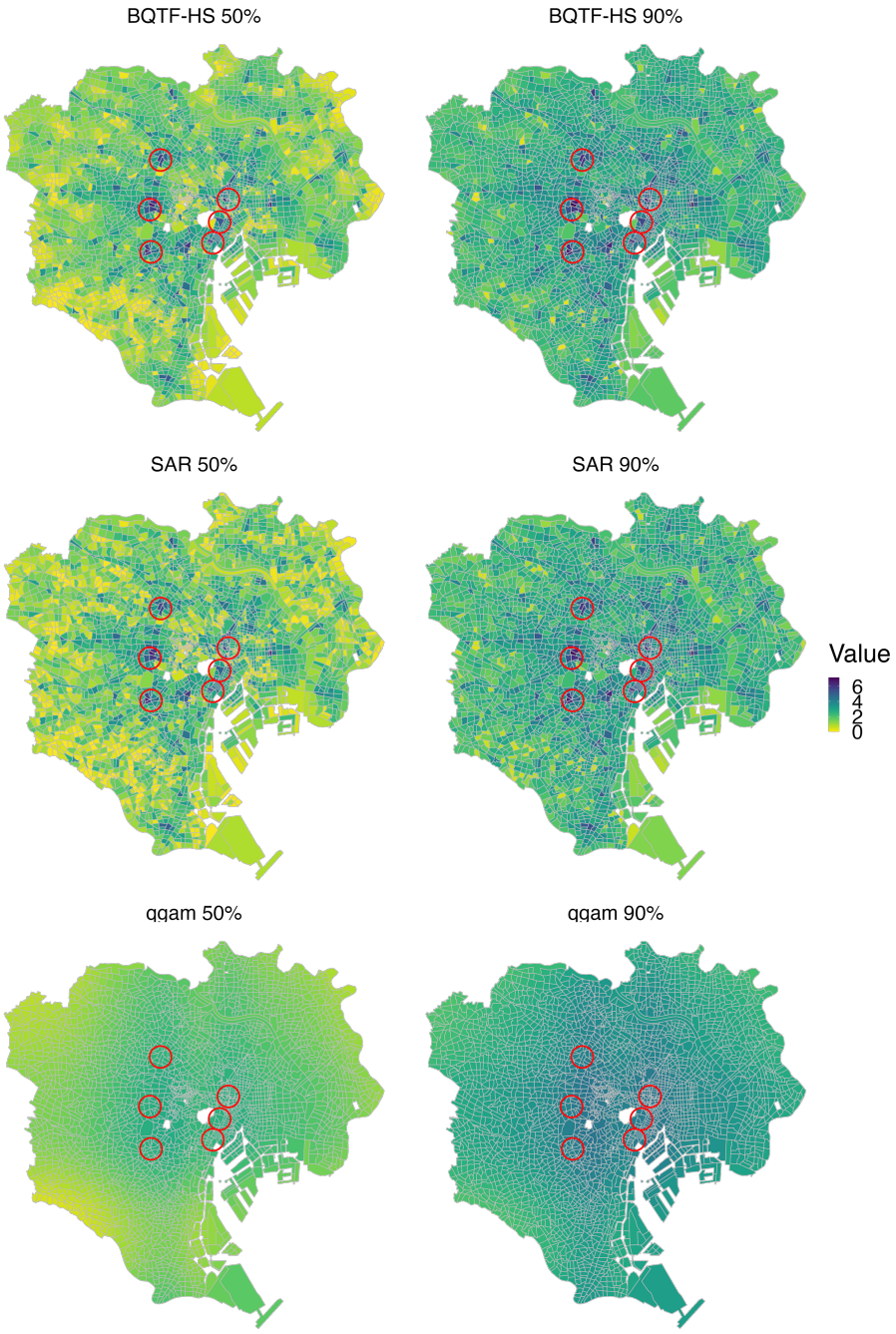

Figure 4.6: Estimated trends via BQTF-HS, SAR, and qgam from top to bottom for two quantile levels: 50% (left) and 90% (right). The six red points are the main stations (Shinjuku, Ikebukuro, Shibuya, Shinbashi, Tokyo, and Akihabara) in Tokyo.

4.6.1 Additional information for simulation study

The MCMC algorithms of SAR and GP models

In this subsection, we summarized the MCMC algorithms of SAR and GP models compared with the proposed methods in simulation studies. Since the prior of θ is $\theta \sim N_n(0, \sigma^2\tau^2 Q(\rho))$ for both methods, the algorithm of SAR and GP prior is equal. Because $\theta \sim N_n(0, \sigma^2\tau^2 Q(\rho))$ is assumed instead of $\theta \mid \sigma^2, \tau^2, w \sim N_n(0, \sigma^2\tau^2(D^\top W^{-1} D)^{-1})$ in the proposed methods, the algorithm is directly given as follows:

- Sample θ from

$$\begin{aligned}
&\theta \mid y, \sigma^2, z, \tau^2, \rho \sim N(A^{-1}B, \sigma^2 A^{-1}),\\
&A = \frac{1}{\tau^2}Q^{-1} + \frac{1}{t^2}\mathrm{diag}\left(\sum_{j=1}^{N_1} z_{1j}^{-1}, \ldots, \sum_{j=1}^{N_n} z_{nj}^{-1}\right),\\
&B = \left(\sum_{j=1}^{N_1} \frac{y_{1j} - \psi z_{1j}}{t^2 z_{1j}}, \ldots, \sum_{j=1}^{N_n} \frac{y_{nj} - \psi z_{nj}}{t^2 z_{nj}}\right)^\top, \quad Q = Q(\rho).
\end{aligned}$$

- Sample σ^2 from

$$\begin{aligned}
&\sigma^2 \mid \theta, y, z \sim \mathrm{IG}\left(a_\sigma + \frac{n+3N}{2}, \beta_\sigma\right),\\
&\beta_\sigma = b_\sigma + \sum_{i=1}^{n}\sum_{j=1}^{N_i} \frac{(y_{ij} - \theta_i - \psi z_{ij})^2}{2t^2 z_{ij}} + \sum_{i=1}^{n}\sum_{j=1}^{N_i} z_{ij} + \frac{1}{\tau^2}\theta^\top Q^{-1}\theta.
\end{aligned}$$

- If $\tau^2 \sim \mathrm{IG}(a_\tau, b_\tau)$ is assumed for the prior of τ^2, sample τ^2 from

$$\tau^2 \mid \theta, \sigma^2 \sim \mathrm{IG}\left(a_\tau + n/2, b_\tau + \frac{1}{2\sigma^2}\theta^\top Q^{-1}\theta\right).$$

- The parameter ρ is sampled with the random walk MH.

True five quantiles under the three noise distribution

Figure 4.7 summarizes the plot of true five quantile trends under three noise distribution ϵ in the simulation study (see also Section 3). We considered the three scenarios: that is (I) homogeneous, (II) block heterogeneous, and (III) smooth heterogeneous.

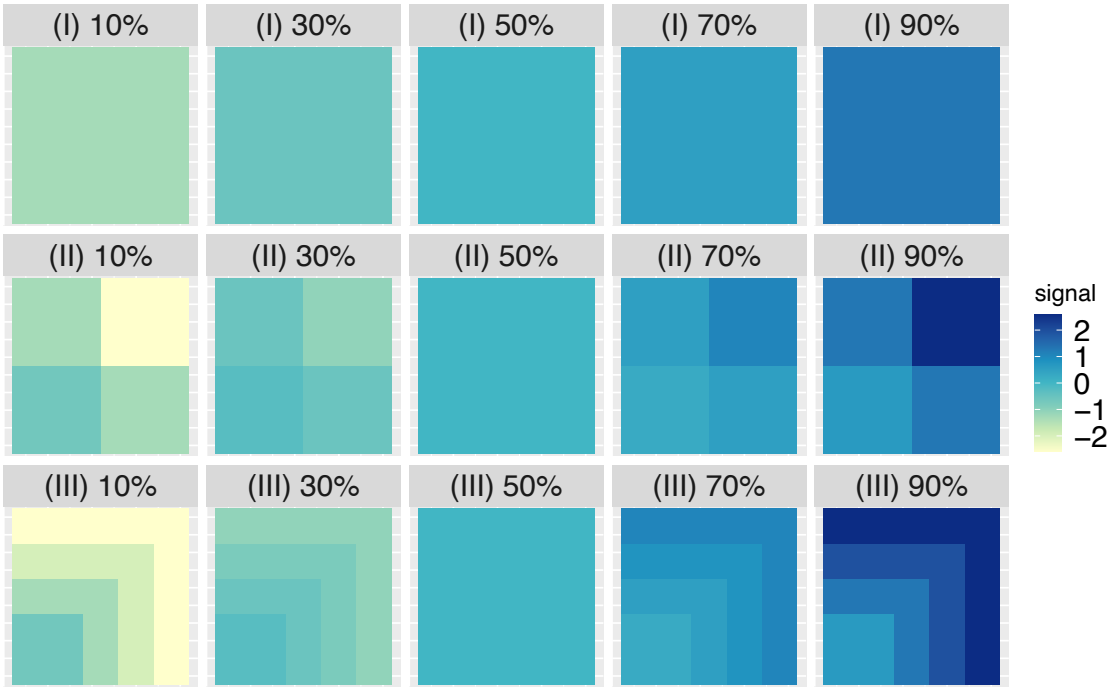

Figure 4.7: The true p-th quantiles ($p = 0.1, 0.3, 0.5, 0.7, 0.9$) for three noise distributions from top to bottom: (I), (II), (III).

Computation time and efficiency

We provide the raw computing time and the sampling efficiency of the MCMC algorithm in the simulation study. We calculated the effective sample size per unit time, defined as the effective sample size divided by the computation time in seconds. The results are reported in Tables 4.3 and 4.4. Although the effective sample size (ESS) per computation time under Laplace prior is larger than the other methods, the computation times are similar. It is also observed that there are no differences between scenarios.

Table 4.3: Average values of effective sample size per unit time and raw computing time based on 100 replications for scenarios (i), (ii), and (iii).

	Scenario (i)									
	ESS (per second)					Compuation rime (second)				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	11	16	19	17	11	66	66	66	66	66
Lap	20	34	38	35	20	71	70	70	71	71
SAR	11	12	12	12	11	64	64	64	64	64
GP	11	12	12	12	11	66	66	66	66	66
	Scenario (ii)									
	ESS (per second)					Compuation rime (second)				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	11	17	20	18	11	66	66	66	66	66
Lap	20	36	40	36	20	71	70	70	70	71
SAR	11	12	12	12	11	64	64	64	64	64
GP	11	12	12	12	11	66	66	66	66	66
	Scenario (iii)									
	ESS (per second)					Compuation rime (second)				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	10	14	16	15	10	66	66	66	66	66
Lap	21	35	39	36	21	70	70	70	70	70
SAR	11	12	12	12	11	64	64	64	64	64
GP	11	12	12	12	11	66	66	66	66	66

Comparison with the other methods

In addition to the methods presented in Section 3, we compared the following methods.

- HS ($k = 0$): the proposed method under horseshoe prior and 1st order difference operator.

Table 4.4: Average values of effective sample size and raw computing time based on 100 replications for scenarios (iv), (v), and (vi).

	Scenario (iv)									
	ESS (per second)					Compuation rime (second)				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	10	11	11	11	10	65	65	65	65	65
Lap	21	35	37	35	21	70	69	69	69	70
SAR	11	12	12	12	11	64	64	64	64	64
GP	11	11	11	11	11	67	67	67	67	67
	Scenario (v)									
	ESS (per second)					Compuation rime (second)				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	10	11	11	11	9	65	65	65	65	65
Lap	22	36	39	37	21	70	69	69	69	70
SAR	11	12	12	12	11	64	64	64	64	64
GP	11	11	11	11	11	67	67	67	67	67
	Scenario (vi)									
	ESS (per second)					Compuation rime (second)				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	9	11	11	11	9	66	66	66	66	66
Lap	22	36	39	36	22	70	70	70	70	70
SAR	11	12	12	12	11	64	64	64	64	64
GP	11	11	11	11	11	67	67	67	67	67

- Lap ($k = 0$): the proposed method under Laplace prior and 1st order difference operator.

- spreg: the classical spatial regression with two covariates (two-dimensional coordinate) and spatially correlated error terms, which is implemented by using the package `spatialreg` in `R`. Since we can not estimate a quantile using the method, we only compare it with the 50% quantile trend (true signal).

The results are reported in Tables 4.5 and 4.6. We also showed the HS and Lap methods under $k = 1$ for comparison. For the lower quantile level, because the BQTF methods under $k = 0$ lead to strong shrinkage, the point estimates of $k = 0$ are worse than those of $k = 0$ in the 0.1-th quantile level, especially under Laplace prior. The estimates under $k = 0$ are better than those of $k = 0$ for the other quantile levels. As seen in the main manuscript, the results of $k = 1$ under the two-block structure are better than the existing methods, and then we adopt $k = 1$ for real data analysis. Compared with spreg in Table 4.6, it is observed that the method does not work as well as the qgam method presented in the main manuscript.

2-D random graph

We consider a more general graph based on the 2-D lattice graph in the main simulation study. We set a new 2-D graph with an additional edge drawn on the diagonal, in which the number of vertexes and edges are $|V| = 100$ and $|E| = 342$, and then the twenty hundred edges are selected from the edge set randomly. The true structures of the 2-D lattice graph and the 2-D random graph are shown in Figure 4.8. On the graph structure, the simulation studies based on two true signals (two-block structure and exponential function) and three noise distributions ((I), (II), and (III)) are set as the additional simulation study. The results are shown in Table 4.7 and 4.8. Note that the results of the compared GP and qgam methods are the same as the 2-D lattice graph because they are only based on the location of the area, not the graph structure. For the two-block structure, the MSE and MAD of the BQTF-HS are the smallest in almost all cases. For the exponential function ((iv), (v), and (vi)), the SAR and the GP models are sometimes better than the proposed methods in center quantiles such as 0.3, 0.5, and 0.7 due to smooth trend. Note that the estimates under the 2-D random graph are worse than those under the 2-D lattice graph because the graph is not straightforward, unlike the GP model.

Table 4.5: Average values of MSE, MAD, MCIW, and CP based on 100 replications for scenarios (i), (ii), and (iii) (two-block structure).

	Scenario (i)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS ($k=0$)	0.545	0.052	0.046	0.053	0.106	0.329	0.174	0.165	0.176	0.247
HS ($k=1$)	0.264	0.158	0.138	0.152	0.248	0.377	0.283	0.266	0.280	0.371
Lap ($k=0$)	5.649	0.128	0.115	0.125	0.221	1.283	0.277	0.264	0.276	0.373
Lap ($k=1$)	0.337	0.212	0.193	0.211	0.339	0.460	0.359	0.344	0.360	0.463
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS ($k=0$)	0.957	1.011	1.015	1.012	1.027	0.857	0.975	0.983	0.978	0.914
HS ($k=1$)	1.322	1.287	1.262	1.261	1.290	0.822	0.920	0.934	0.921	0.827
Lap ($k=0$)	0.228	1.363	1.355	1.353	1.318	0.247	0.947	0.955	0.945	0.876
Lap ($k=1$)	1.400	1.554	1.552	1.541	1.395	0.799	0.910	0.921	0.906	0.801
	Scenario (ii)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS ($k=0$)	0.319	0.117	0.088	0.119	0.328	0.368	0.223	0.190	0.224	0.369
HS ($k=1$)	0.464	0.272	0.244	0.256	0.451	0.454	0.337	0.317	0.328	0.447
Lap ($k=0$)	5.946	0.189	0.168	0.188	0.572	1.546	0.311	0.290	0.308	0.483
Lap ($k=1$)	0.542	0.311	0.283	0.306	0.541	0.524	0.400	0.381	0.399	0.526
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS ($k=0$)	1.186	1.139	1.116	1.140	1.176	0.842	0.950	0.976	0.949	0.847
HS ($k=1$)	1.440	1.409	1.369	1.378	1.388	0.814	0.911	0.920	0.909	0.814
Lap ($k=0$)	0.294	1.474	1.448	1.460	1.446	0.261	0.939	0.949	0.931	0.841
Lap ($k=1$)	1.543	1.689	1.675	1.678	1.528	0.801	0.907	0.919	0.907	0.799
	Scenario (iii)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS ($k=0$)	0.347	0.129	0.101	0.129	0.349	0.422	0.257	0.227	0.255	0.416
HS ($k=1$)	0.651	0.395	0.346	0.375	0.602	0.581	0.438	0.408	0.426	0.559
Lap ($k=0$)	6.774	0.265	0.230	0.249	1.232	1.780	0.386	0.362	0.376	0.708
Lap ($k=1$)	0.740	0.431	0.389	0.424	0.740	0.643	0.493	0.469	0.489	0.642
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS ($k=0$)	1.478	1.398	1.373	1.401	1.477	0.848	0.960	0.974	0.962	0.845
HS ($k=1$)	1.814	1.798	1.767	1.764	1.783	0.795	0.896	0.912	0.899	0.806
Lap ($k=0$)	0.382	1.767	1.726	1.737	1.670	0.059	0.929	0.940	0.927	0.784
Lap ($k=1$)	1.932	2.076	2.063	2.061	1.911	0.790	0.901	0.916	0.899	0.794

Table 4.6: Average values of MSE, MAD, MCIW, and CP based on 100 replications for all scenarios.

	MSE					
	(i)	(ii)	(iii)	(iv)	(v)	(vi)
HS	0.138	0.244	0.346	0.075	0.109	0.142
Lap	0.193	0.283	0.389	0.072	0.095	0.140
spreg	1.744	4.209	5.886	1.791	2.855	4.523
	MAD					
HS	0.266	0.317	0.408	0.207	0.228	0.278
Lap	0.344	0.381	0.469	0.212	0.225	0.280
spreg	0.926	1.783	2.140	1.239	1.455	1.859

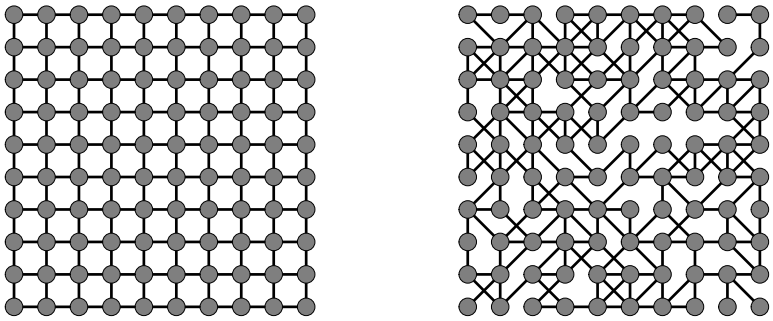

Figure 4.8: The true structures of the 2-D lattice graph (left) and the 2-D random graph (right).

Table 4.7: Average values of MSE, MAD, MCIW, and CP based on 100 replications for scenarios (i), (ii), and (iii) (two-block structure). The minimum values of MSE and MAD are represented in bold.

	Scenario (i)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.300**	**0.199**	**0.178**	**0.191**	**0.290**	**0.415**	**0.331**	**0.314**	**0.327**	**0.409**
Lap	0.346	0.214	0.196	0.213	0.348	0.468	0.365	0.351	0.366	0.469
SAR	0.344	0.214	0.196	0.214	0.345	0.468	0.370	0.354	0.369	0.470
GP	0.347	0.217	0.199	0.217	0.354	0.470	0.372	0.356	0.373	0.476
qgam	3.985	2.461	1.877	2.269	3.062	1.256	1.266	1.195	1.266	1.425
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.383	1.409	1.393	1.377	1.326	0.800	0.899	0.917	0.901	0.807
Lap	1.403	1.577	1.576	1.561	1.389	0.791	0.907	0.920	0.908	0.794
SAR	1.524	1.700	1.704	1.689	1.514	0.830	0.927	0.942	0.927	0.831
GP	1.525	1.712	1.717	1.703	1.510	0.826	0.928	0.942	0.927	0.826
	Scenario (ii)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.502**	**0.295**	**0.264**	**0.288**	**0.488**	**0.494**	**0.372**	**0.350**	**0.369**	**0.484**
Lap	0.553	0.309	0.283	0.311	0.558	0.533	0.403	0.385	0.404	0.534
SAR	0.540	0.300	0.275	0.307	0.534	0.532	0.403	0.385	0.408	0.532
GP	0.545	0.310	0.282	0.309	0.549	0.533	0.409	0.389	0.410	0.538
qgam	3.973	2.419	1.891	2.219	3.208	1.315	1.241	1.197	1.256	1.459
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.540	1.544	1.503	1.497	1.453	0.800	0.900	0.914	0.898	0.797
Lap	1.556	1.726	1.710	1.699	1.525	0.797	0.910	0.920	0.907	0.796
SAR	1.686	1.859	1.848	1.839	1.673	0.828	0.928	0.938	0.927	0.828
GP	1.681	1.864	1.858	1.852	1.662	0.823	0.924	0.937	0.924	0.823
	Scenario (iii)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.698**	0.423	**0.368**	**0.391**	**0.629**	**0.613**	**0.473**	**0.444**	**0.460**	**0.585**
Lap	0.765	0.434	0.387	0.422	0.755	0.659	0.498	0.473	0.494	0.652
SAR	0.747	**0.418**	0.373	0.411	0.731	0.653	0.494	0.469	0.490	0.648
GP	0.760	0.422	0.378	0.416	0.768	0.656	0.497	0.472	0.494	0.661
qgam	3.792	2.295	1.908	2.174	2.984	1.366	1.258	1.204	1.255	1.422
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.886	1.910	1.868	1.848	1.801	0.782	0.888	0.898	0.884	0.790
Lap	1.942	2.117	2.100	2.087	1.896	0.787	0.902	0.917	0.904	0.790
SAR	2.074	2.271	2.259	2.242	2.051	0.819	0.927	0.939	0.925	0.828
GP	2.071	2.276	2.273	2.270	2.048	0.815	0.925	0.938	0.925	0.821

Table 4.8: Average values of MSE, MAD, MCIW, and CP based on 100 replications for scenarios (iv), (v), and (vi) (exponential function). The minimum values of MSE and MAD are represented in bold.

	Scenario (iv)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.183**	0.125	0.114	0.119	**0.177**	**0.332**	0.273	0.263	**0.267**	**0.321**
Lap	0.219	**0.117**	**0.106**	**0.115**	0.222	0.371	**0.269**	**0.257**	0.268	0.370
SAR	0.237	0.123	0.109	0.118	0.226	0.391	0.278	0.263	0.274	0.382
GP	0.251	0.119	0.107	0.125	0.273	0.405	0.275	0.262	0.285	0.423
qgam	0.577	0.450	0.386	0.418	0.510	0.522	0.503	0.504	0.522	0.551
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.092	1.097	1.095	1.098	1.107	0.788	0.873	0.882	0.876	0.815
Lap	1.216	1.235	1.229	1.232	1.213	0.832	0.924	0.939	0.928	0.842
SAR	1.342	1.381	1.368	1.366	1.330	0.869	0.951	0.960	0.952	0.878
GP	1.333	1.424	1.433	1.445	1.348	0.878	0.961	0.972	0.961	0.872
	Scenario (v)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.357**	0.177	0.152	0.174	0.351	**0.419**	0.303	0.281	0.298	**0.405**
Lap	0.384	0.169	0.139	0.161	0.373	0.446	0.300	**0.272**	**0.292**	0.431
SAR	0.376	0.165	0.135	**0.155**	**0.338**	0.453	0.304	0.274	0.294	0.428
GP	0.385	**0.152**	**0.131**	0.166	0.418	0.457	**0.293**	**0.272**	0.304	0.475
qgam	0.686	0.472	0.398	0.432	0.619	0.588	0.515	0.508	0.528	0.601
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.249	1.192	1.189	1.203	1.242	0.777	0.878	0.898	0.880	0.798
Lap	1.392	1.344	1.315	1.318	1.350	0.824	0.923	0.941	0.927	0.829
SAR	1.541	1.495	1.458	1.454	1.491	0.856	0.945	0.959	0.949	0.866
GP	1.514	1.533	1.525	1.557	1.517	0.858	0.959	0.969	0.955	0.853
	Scenario (vi)									
	MSE					MAD				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	**0.450**	0.242	0.201	**0.202**	**0.374**	**0.492**	0.371	0.345	**0.345**	**0.447**
Lap	0.511	0.238	0.199	0.212	0.457	0.531	0.367	0.340	0.349	0.499
SAR	0.524	0.236	0.195	0.207	0.436	0.551	0.369	0.338	0.348	0.504
GP	0.579	**0.224**	**0.182**	0.221	0.587	0.578	**0.361**	**0.327**	0.359	0.584
qgam	0.916	0.598	0.492	0.505	0.655	0.718	0.591	0.565	0.579	0.660
	MCIW					CP				
	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
HS	1.502	1.441	1.411	1.409	1.497	0.767	0.855	0.876	0.872	0.813
Lap	1.696	1.630	1.585	1.578	1.646	0.812	0.910	0.929	0.923	0.842
SAR	1.884	1.842	1.776	1.748	1.804	0.852	0.947	0.959	0.949	0.876
GP	1.887	1.886	1.853	1.894	1.872	0.849	0.957	0.971	0.960	0.860

4.6.2 Additional information for Tokyo crime data analysis

In Section 4 of the main manuscript, the edges were constructed based on the 5 nearest neighbor searches. We compare the results with those of 3 and 7 nearest neighbors. The result is reported in Figure 4.9. From the figure, we can observe that the results for each number of nearest neighbors do not change very much. The number of edges for each graph is 5598, 8,996, and 12,398. It seems that the graph structure did not affect the smoothness in the example.

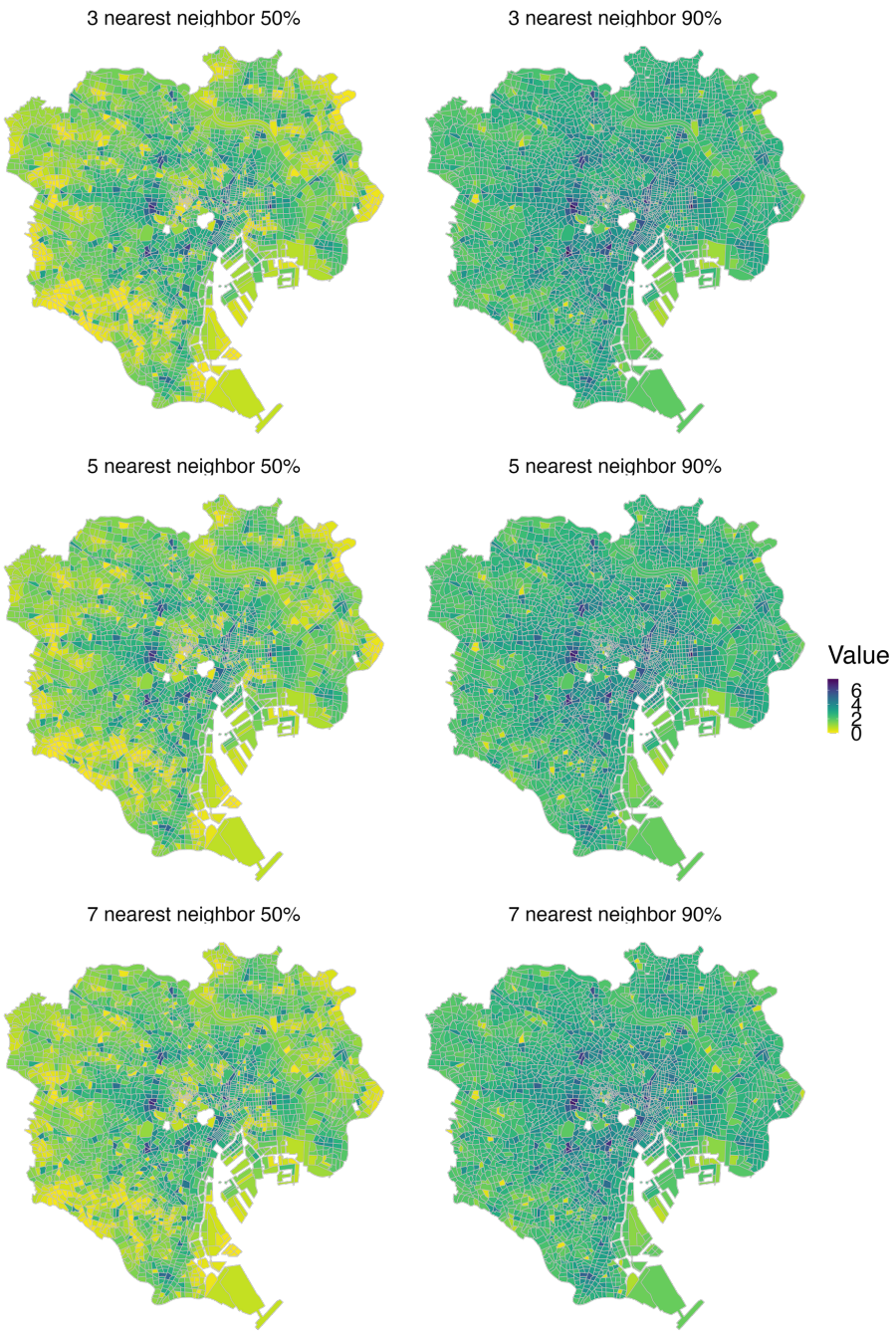

Figure 4.9: The three estimates under 3, 5, and 7 nearest neighbors from top to bottom.

Chapter 5

Locally adaptive Bayesian boundary smoothing

5.1 Introduction

Consider the nonparametric regression model

$$y_i = f(x_i) + \varepsilon_i, \quad i = 1, \ldots, n, \tag{5.1}$$

where $y_1, \ldots, y_n \in \mathbb{R}$ are outcomes, $x_1, \ldots, x_n \in \mathbb{R}$ are input points, f is the underlying function to be estimated, and $\varepsilon_1, \ldots, \varepsilon_n$ are independent errors which are not assumed to be centered, but to have one-sided support such as $(-\infty, 0]$. The regression function f describes the support frontier. For such models, some of the regularity conditions on the statistical model may be violated. For example, Smith (1994) considered the linear regression model with a class of one-sided error distributions including the exponential and Weibull distributions. In this case, the support of the distribution of the response variable depends on unknown parameters. He also showed that the asymptotic distribution of the estimator of the regression coefficient vector is non-normal. For this reason, the model (5.1) is generally called a "non-regular model". In the context of nonparametric regression, some authors proposed the methods and showed their theoretical properties for the model defined by (5.1) (see e.g. Deprins et al., 1984; Hall et al., 1998; Hall and Simar, 2002; Daouia et al., 2016; Reiß and Selk, 2017; Reiß and Schmidt-Hieber, 2020a,b; Selk et al., 2022). In particular, Daouia et al. (2017) provided an excellent `R` package to implement some nonparametric boundary estimation methods including shape restriction and robust estimation. There are many applications of the model (5.1), for instance,

in the microeconomic theory where the support boundary is considered as the set of the most efficient businesses or industries, in climate science where the trend of maximum values of temperature is important to come up with an environmental policy. Furthermore, these considerations often lead to the assumption of monotonicity/nearly monotonicity. In nonparametric regression or nonparametric curve fitting problems, shape constraints such as monotonicity and convexity are also useful when we have some prior information on the shape of data (see e.g. Robertson et al., 1988). For boundary curve estimation, Daouia et al. (2016) proposed spline smoothing methods under monotone/concave constraint, and provided an efficient optimization algorithm. We note that boundary regression models relate to nonparametric quantile regression with high or low quantile levels. The method is also called extremal quantile regression, and it might be interpreted as an exploratory tool rather than as a method for final boundary analysis. In this study, we are interested in the boundary regression curve, not the extremal quantile regression curve.

In this study, we develop a locally adaptive boundary smoothing method using Bayesian trend filtering. To this end, we use the truncated multivariate (tMVN) distribution as a working likelihood and shrinkage priors for differences. Using the scale mixtures of normal priors, we can easily derive Gibbs sampling algorithms, while one of the full conditional distributions is the tMVN distribution. It is well-known that sampling from the tMVN distribution is quite challenging even if the dimension of the parameter is moderately large (e.g. one hundred). To overcome this difficulty, we employ an approximation of the indicator function in the tMVN distribution by using the sigmoid function with a tuning scale parameter. The idea comes from the paper by Souris et al. (2018) on the Bayesian shape constraint regression (see also Ray et al., 2020). While they used the approximation for prior distribution, we adopt one to the likelihood function. Using such approximation, we provide an efficient Gibbs sampling algorithm using the Pólya-Gamma data augmentation (Polson et al., 2013). The shrinkage priors for differences are similar to Chapter 4. The Laplace prior corresponds to L_1-penalty in the original trend filtering and the horseshoe prior is known as a more flexible shrinkage prior in sparse Bayesian estimation by introducing global and local shrinkage parameters. We also extend the proposed model to the case of some shape constraints. In particular, we propose a nearly isotonic regression method for estimating support boundaries. The

nearly isotonic regression which was proposed by Tibshirani et al. (2011) is a kind of generalization of original isotonic regression because the method allows violations of monotonicity at some change points. In other words, the nearly isotonic regression is a penalized version of isotonic regression, and it has robustness against structural misspecification for the assumption of monotonicity. We illustrate the performance of the proposed method through some numerical experiments including real data examples and also provide the sensitivity analysis for selecting a tuning parameter of the sigmoid function.

The rest of this chapter is organized as follows. In Section 5.2, we formulate the proposed method and prior specification. An efficient Gibbs sampling algorithm using the approximated likelihood and data augmentation is also introduced. Some simulation studies and real data applications are given in Sections 5.3 and 5.4, respectively. R code implementing the proposed methods is available at the GitHub repository (URL: https://github.com/Takahiro-Onizuka/BBTF.git).

5.2 Bayesian boundary trend filtering

5.2.1 The idea of boundary trend filtering

We restate the original ℓ_1 trend filtering to estimate nonparametric regression model $y_i = f(x_i) + \varepsilon_i$ for $i = 1, \ldots, n$:

$$\min_{\theta \in \mathbb{R}^n} \sum_{i=1}^{n} (\theta_i - y_i)^2 + \lambda \| D_n^{(k+1)} \theta \|_1,$$

where $y = (y_1, \ldots, y_n)^\top$, $\theta = (\theta_1, \ldots, \theta_n)^\top = (f(x_1), \ldots, f(x_n))^\top$, $\lambda > 0$ is a tuning parameter which controls the smoothness of the trend, and $D_n^{(k+1)}$ is a $(n-k-1) \times n$ difference operator matrix of order $k+1$ defined in (2.5). Brantley et al. (2020) also propose the quantile trend filtering assuming check loss instead of quadratic loss, and these Beyaesian formulations are also considered (see e.g. Roualdes, 2015; Faulkner and Minin, 2018; Heng et al., 2023; Onizuka et al., 2024a). Although existing methods provide a reasonable estimate for mean or quantile trend, we can not apply such methods to nonparametric regression models with one-sided error as (5.1).

To this end, we consider the following optimization problem to estimate the

boundary trend:

$$\min_{\theta \geq y} \sum_{i=1}^{n} (\theta_i - y_i)^2 + \lambda \| D_n^{(k+1)} \theta \|_1, \tag{5.2}$$

where the relation $\theta \geq y$ between two vectors $y = (y_1, \ldots, y_n)^\top$ and $\theta = (\theta_1, \ldots, \theta_n)^\top$ implies $\theta_i \geq y_i$ for all i. Note that the constraint $\theta \geq y$ leads to the estimation of the upper boundary trend. When we estimate the lower boundary trend, we may consider the constraint $\theta \leq y$ instead of $\theta \geq y$. To the best of our knowledge, such boundary trend filtering has not been proposed in terms of trend filtering. As mentioned in Section 5.1, the Bayesian approach has several attractive properties for sparse estimation including trend filtering. For these reasons, we focus on the Bayesian boundary trend filtering in this study. We also see the advantages such as locally adaptive smoothing and uncertainty quantification through real data analysis in Section 5.4.2. We note that the purpose of trend filtering is to estimate "the value of the underlying function at the data points" and not to estimate the function itself. In other words, the trend filtering estimate $\hat{\theta}$ is not a function and it has a different purpose than the spline methods from which the function is obtained as an estimate.

5.2.2 Bayesian boundary trend filtering

Combining the model (1.2) and (5.2), we formulate the Bayesian boundary trend filtering. Without loss of generality, we only consider the estimation of the upper boundary trend. For each y_i $(i = 1, \ldots, n)$, we consider the following model:

$$y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim \mathrm{HN}(0, \sigma^2) \tag{5.3}$$

where $\mathrm{HN}(\mu, \sigma^2)$ represents the (upper truncated) half-normal distribution with location parameter μ and scale parameter σ^2. The conditional probability density function of y_i given θ_i in (5.3) is

$$p(y_i \mid \theta_i, \sigma^2) = \sqrt{\frac{2}{\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2} (y_i - \theta_i)^2 \right) 1_{\{y_i \leq \theta_i\}}(y_i), \quad i = 1, \ldots, n,$$

where $1_A(x)$ is an indicator function defined by $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise. The corresponding working likelihood of (θ, σ^2) is given by

$$L(\theta, \sigma^2 \mid y) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \theta_i)^2 \right) \prod_{i=1}^{n} 1_{\{y_i \leq \theta_i\}}(\theta_i), \tag{5.4}$$

which is that of the truncated multivariate normal distribution restricted to the region $\mathcal{C} = \{\theta \in \mathbb{R}^n \mid y_i \leq \theta_i,\ i = 1, \ldots, n\}$.

Next, we introduce shrinkage priors on differences $D\theta$, where D is the $k+1$ th order difference operator defined in (3.3) of Chapter 3. The shrinkage priors we consider here are the horseshoe, Laplace, and normal priors (see also Faulkner and Minin, 2018; Onizuka et al., 2024a). The horseshoe and Laplace priors can be represented by the scale mixtures of normal distribution:

$$D\theta \mid \tau^2, \sigma^2, u \sim N_n(0, \sigma^2 U), \tag{5.5}$$

where $U = \mathrm{diag}(u_1^2, \ldots, u_{k+1}^2, \tau^2 u_{k+2}^2, \ldots, \tau^2 u_n^2)$ with shrinkage parameters u_i and τ. Here, τ and u_i are called global and local parameters respectively, and this formulation enables locally adaptive smoothing. In (5.5), we employ a prior distribution on $D\theta$ which depends on error variance σ^2. The advantage of such a conditional prior is that the scale of the prior is automatically adjusted when we change units of observations. Such a formulation of the prior is widely used when we assume the normal prior (e.g. Polson and Scott, 2012). Since the matrix D is non-singular, the prior of θ can be rewritten as

$$\theta \mid \tau^2, \sigma^2, u \sim N_n(0, \sigma^2 (D^\top U^{-1} D)^{-1}). \tag{5.6}$$

We note that prior distributions for τ and u_i depend on the shrinkage priors. The shrinkage priors we consider in the chapter are expressed as the marginal priors of the following hierarchical priors.

- Horseshoe prior:

$$\begin{aligned} \theta \mid \tau^2, \sigma^2, u &\sim N_n(0, \sigma^2 (D^\top U^{-1} D)^{-1}), \\ \tau &\sim C^+(0, 1), \\ u_i^2 &\sim \mathrm{IG}(a_{u_i}, b_{u_i}) \quad (i = 1, \ldots, k+1), \\ u_i &\sim C^+(0, 1) \quad (i = k+2, \ldots, n), \end{aligned}$$

  where $U = \mathrm{diag}(u_1^2, \ldots, u_{k+1}^2, \tau^2 u_{k+2}^2, \ldots, \tau^2 u_n^2)$.

- Laplace prior:

$$\theta \mid \sigma^2, u \sim N_n(0, \sigma^2 (D^\top U^{-1} D)^{-1}),$$

$$u_i^2 \sim \mathrm{IG}(a_{u_i}, b_{u_i}) \quad (i = 1, \ldots, k+1),$$
$$u_i \sim \mathrm{Exp}(\gamma^2/2) \quad (i = k+2, \ldots, n), \quad \gamma^2 \sim \mathrm{IG}(a_\gamma, b_\gamma),$$

where $U = \mathrm{diag}(u_1^2, \ldots, u_{k+1}^2, u_{k+2}^2, \ldots, u_n^2)$.

- Normal prior:

$$\theta \mid \sigma^2, \tau^2, u_1, \ldots, u_{k+1} \sim N_n(0, \sigma^2(D^\top U^{-1} D)^{-1}),$$
$$u_i^2 \sim \mathrm{IG}(a_{u_i}, b_{u_i}) \quad (i = 1, \ldots, k+1), \quad \tau^2 \sim \mathrm{IG}(a_\tau, b_\tau),$$

where $U = \mathrm{diag}(u_1^2, \ldots, u_{k+1}^2, \tau^2, \ldots, \tau^2)$.

Here, $\mathrm{IG}(a, b)$ and $C^+(a, b)$ are the inverse-gamma distribution with shape a and rate b, and the half-Cauchy distribution with location a and scale b, respectively. We also assume the conjugate proper prior for σ^2 such as $\sigma^2 \sim \mathrm{IG}(a_\sigma, b_\sigma)$ for some hyper-parameters $a_\sigma > 0$ and $b_\sigma > 0$.

Such formulations of likelihood and prior distributions lead to tractable full conditional distribution for θ so that we can easily construct an efficient Gibbs sampler. From (5.4) and (5.6), the resulting full conditional distribution of θ is the truncated multivariate normal (tMVN) distribution. Efficient sampling algorithms for the tMVN distribution have been developed in recent years. For example, Pakman and Paninski (2014) proposed an exact Hamiltonian Markov chain algorithm, and Botev (2017) proposed accept-reject algorithms that create an exact sample from the tMVN distribution. However, it is known that sampling from the high-dimensional tMVN distribution is quite challenging even if the dimension of the parameter is moderately large (e.g. $n = 100$). To overcome such a sampling problem, we introduce the following approximation of the indicator function in the likelihood function (5.4). The idea comes from the paper by Souris et al. (2018). They proposed an approximation of the tMVN distribution via logistic sigmoid function $\sigma_\eta(\xi_i)$:

$$1(\xi_i \geq 0) \approx \sigma_\eta(\xi_i) = \frac{e^{\eta \xi_i}}{1 + e^{\eta \xi_i}}, \quad i = 1, \ldots, n. \tag{5.7}$$

Then, the approximate truncated multivariate normal likelihood $L_\eta(\theta \mid y)$ is represented by

$$L_\eta(\theta \mid y) \propto e^{-\frac{1}{2\sigma^2}(y-\theta)^\top (y-\theta)} \prod_{i=1}^{n} \left(\frac{e^{\eta(\theta_i - y_i)}}{1 + e^{\eta(\theta_i - y_i)}} \right). \tag{5.8}$$

The approximate distribution (5.8) is also called the *soft* truncated multivariate normal distribution by Souris et al. (2018). We can easily show that the approximate truncated multivariate normal likelihood $L_\eta(\theta \mid y)$ converges to the truncated multivariate normal likelihood $L(\theta \mid y)$ as $\eta \to \infty$ in the sense of L_1 convergence (see also Souris et al., 2018). Hence, the constant η controls the accuracy of the approximation, and we recommend a large value for η.

5.2.3 Shape constraints

There are many phenomena for which monotonic or concave constraints are appropriate such as the dose-response curve in medicine and the demand curve in economics. The most popular shape-constraint regression method is the isotonic regression. Although the isotonic regression is useful in application, the monotone assumption may be violated at a few points in practice. For example, the global warming in climate change (Tibshirani et al., 2011) and geological observations in seismology (Minami, 2020) indicate violations of the monotonicity. As a penalized isotonic regression, Tibshirani et al. (2011) proposed a nearly isotonic (NI) regression, and the corresponding estimate is defined by

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^n}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \theta_i)^2 + \lambda \sum_{i=1}^{n-1} (\theta_i - \theta_{i+1})_+,$$

where $(x)_+ = \max(x, 0)$ and $\lambda > 0$ is a tuning parameter that controls the violation of the monotone constraint. Ramdas and Tibshirani (2016) also applied nearly isotonic constraint to trend filtering. We now consider the Bayesian boundary trend filtering under the nearly isotonic constraint. First, the optimization problem is formulated as

$$\min_{\theta \geq y} \sum_{i=1}^{n} (y_i - \theta_i)^2 + \lambda_1 \| D_n^{(k+1)} \theta \|_1 + \lambda_2 \sum_{i=1}^{n-1} (\theta_i - \theta_{i+1})_+, \tag{5.9}$$

where $\lambda_1, \lambda_2 > 0$ are tuning parameters. If the constraint $\theta \geq y$ is removed, the optimization problem is nothing but the original nearly isotonic trend filtering proposed by Ramdas and Tibshirani (2016). The third term in (5.9) plays a role of nearly isotonic constraint. In other words, we impose a penalty when the monotonicity is violated. In the Bayesian context, the solution of the model (5.9) is equivalent to

the posterior mode under the prior which is proportional to

$$\exp\left\{-\lambda_1\|D^{(k+1)}\theta\|_1\right\}\exp\left\{-\lambda_2\sum_{i=1}^{n-1}(\theta_i-\theta_{i+1})_+\right\}. \tag{5.10}$$

We can easily deal with the first term in (5.10) like (5.6). For the second term in (5.10), it is useful to employ the variance-mean mixture representation (see e.g. Polson and Scott, 2013)

$$a^{-1}\exp\{-2c^{-1}(ax)_+\} = \int_0^\infty \phi_1(x \mid -av, cv)dv,$$

where $\phi_n(\cdot \mid a, b)$ represents n-dimensional Gaussian density with mean vector a and covariance matrix b. Using this identity, we introduce the prior distribution for the nearly isotonic constraint as

$$\exp\left\{-\frac{1}{\rho^2\sigma^2}\sum_{i=1}^{n-1}(\theta_i-\theta_{i+1})_+\right\} = \prod_{i=1}^{n-1}\int_0^\infty \phi_1(\theta_i-\theta_{i+1} \mid -v_i, 2\rho^2\sigma^2 v_i)dv_i, \tag{5.11}$$

and then the conditional prior of θ for nearly isotonic constraint is rewritten as

$$p(\theta \mid \rho^2, \sigma^2, v) = \prod_{i=1}^{n-1}\phi_1(\theta_i-\theta_{i+1} \mid -v_i, 2\rho^2\sigma^2 v_i) = \phi_{n-1}(P\theta \mid -v, V), \tag{5.12}$$

where $n-1$ is the length of the vector $P\theta$, $P = D_n^{(1)}$ and $V = \mathrm{diag}(2\rho^2\sigma^2 v_1, \ldots, 2\rho^2\sigma^2 v_{n-1})$ with a scale parameter ρ^2. Note that the scale parameter ρ^2 plays a role of tuning parameter λ_2 in (5.9), and we estimate ρ^2 from data assuming the prior distribution on ρ^2. In our numerical studies, we use a prior $\rho^2 \sim \mathrm{IG}(a_\rho, b_\rho)$. Introducing latent variables $v_1, \ldots, v_{n-1}$, the prior is written by conditional Gaussian distribution, and then the conditional prior of the shape-restricted trend filtering like (5.10) also becomes Gaussian distribution. While we only consider the monotonically increasing condition, if we assume a nearly decreasing or convex, then we may use $P = -D_n^{(1)} \in \mathbb{R}^{(n-1)\times n}$ or $P = D_n^{(2)} \in \mathbb{R}^{(n-2)\times n}$ respectively.

Remark 5.2.1 (Posterior propriety)**.** The prior defined by the left-hand side of (5.11) is improper. If we assume proper priors on the remaining parameters, then we can show that the joint posterior distribution is proper because the prior $p(\theta \mid \rho^2, \sigma^2) = \exp\{-1/(\rho^2\sigma^2)\sum_{i=1}^{n-1}(\theta_i-\theta_{i+1})_+\}$ is bounded by 1 for any θ. Hence, the

integral

$$\int p(y \mid \theta, \sigma^2) p(\theta \mid \sigma^2, \tau^2, u) p(\theta \mid \rho^2, \sigma^2) d\theta. \tag{5.13}$$

is bounded by the integral of the product of proper density functions.

5.2.4 Markov chain Monte Carlo algorithm

In this subsection, we construct an efficient posterior computation algorithm via the Markov chain Monte Carlo method. In the proposed model, we can construct a Gibbs sampler. First of all, we consider the sampling of θ from the posterior distribution. From (5.6), (5.8) and (5.12), the full conditional distribution of θ is given by the following form:

$$\phi_n(\theta \mid \mu_\theta, \Sigma_\theta) \prod_{i=1}^{n} \left(\frac{e^{\eta(\theta_i - y_i)}}{1 + e^{\eta(\theta_i - y_i)}} \right),$$

where μ_θ and Σ_θ are some mean vector and covariance matrix. To simplify the sampling, we put $\xi_i = \theta_i - y_i$ and $\xi = (\xi_1, \ldots, \xi_n)^\top$. Then the full conditional distribution of ξ is given by

$$\phi_n(\xi \mid A^{-1}b, A^{-1}) \prod_{i=1}^{n} \left(\frac{e^{\eta\xi_i}}{1 + e^{\eta\xi_i}} \right), \tag{5.14}$$

where the matrix A and vector b depend on the type of shrinkage priors. For sampling of ξ, the Polya-Gamma data augmentation proposed by Polson et al. (2013) can be applied, and then we can sample θ from the following three steps:

(i) Sample latent variables $\omega_i \mid \xi_i \sim \mathrm{PG}(1, \eta\xi_i)$ for $i = 1, \ldots, n$.

(ii) Sample $\xi \mid \omega \sim N_n(\mu_\omega, \Sigma_\omega)$, with

$$\Sigma_\omega = \left(\eta^2 \Omega + A\right)^{-1}, \quad \mu_\omega = \Sigma_\omega \left(\eta\kappa + b\right),$$

where $\omega_1, \ldots, \omega_n$ are latent variables, $\kappa = (1/2, \ldots, 1/2)^\top$, and $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n)$.

(iii) Set $\theta = \xi + y$.

In step 1, $\mathrm{PG}(b, c)$ is the Pólya-Gamma distribution with parameter $b > 0$ and $c \in \mathbb{R}$ (see Definition 1 in Polson et al. (2013)), and sampling from the distribution can be implemented by using `R` package `pgdraw`, for example.

Although we consider three shrinkage priors (horseshoe, Laplace, and normal priors), we only show the full conditional distributions for the horseshoe type prior. Since Gibbs sampling algorithms under the Laplace and normal type priors can also be derived in the same manner, we here omit them.

Under the horseshoe prior, using the mixture of inverse-gamma representation of the half-Cauchy distribution, all full conditional distributions are standard probability distributions (Makalic and Schmidt, 2015). By introducing latent variables ψ and ν_i for $i = k+2, \ldots, n$, it holds that

$$\begin{aligned}
&\tau \sim C^+(0,1) \iff \tau^2 \mid \psi \sim \mathrm{IG}(1/2, 1/\psi), \quad \psi \sim \mathrm{IG}(1/2, 1), \\
&u_i \sim C^+(0,1) \iff u_i^2 \mid \nu_i \sim \mathrm{IG}(1/2, 1/\nu_i), \quad \nu_i \sim \mathrm{IG}(1/2, 1), \quad i = k+2, \ldots, n.
\end{aligned}$$

Then we have the following Markov chain Monte Carlo algorithm under the horseshoe prior. We note that the full conditional distributions of θ and σ^2 depend on whether we assume the shape constraint or not. Furthermore, when we consider the shape constraint, we need to sample additional parameters ρ^2 and v_i for $i = 1, \ldots, n-1$ from the posterior.

Gibbs sampling algorithm under horseshoe prior

1. Sampling of θ

   - Draw $\omega_i \sim \mathrm{PG}(1, \eta e_i^\top \xi) = \mathrm{PG}(1, \eta \xi_i)$, independently for $i = 1, \ldots, n$.
   - Draw $\xi \sim N_n(\mu_\omega, \Sigma_\omega)$, with

   $$\begin{aligned}
   &\Sigma_\omega = \left(\eta^2 \Omega + A\right)^{-1}, \quad \mu_\omega = \Sigma_\omega \left(\eta \kappa + b\right), \\
   &\kappa = (1/2, \ldots, 1/2)^\top, \quad \Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n),
   \end{aligned}$$

   where A and b are as follows.

   - (Unconstraint)

   $$A = (I_n + D^\top U^{-1} D)/\sigma^2, \quad b = -D^\top U^{-1} D y.$$

   - (Nearly isotonic constraint)

   $$A = (I_n + D^\top U^{-1} D + PVP)/\sigma^2, \quad b = -(D^\top U^{-1} D y + P^\top V (Py + v)),$$

   where $P = D^{(1)}$ and $V = \mathrm{diag}(1/(2\rho^2 v_1), \ldots, 1/(2\rho^2 v_{n-1}))$.

- Set $\theta = \xi + y$

2. Sampling of σ^2

   - Draw $\sigma^2 \sim \mathrm{IG}\,(\alpha_\sigma, \beta_\sigma)$, where α_σ and β_σ are as follows.

     - (Unconstraint)

$$\alpha_\sigma = n + a_\sigma, \quad \beta_\sigma = \sum_{i=1}^{n}(y_i - \theta_i)^2/2 + \theta^\top D^\top U^{-1} D\theta/2 + b_\sigma.$$

     - (Nearly isotonic constraint)

$$\alpha_\sigma = (3n-1)/2 + a_\sigma,$$
$$\beta_\sigma = \sum_{i=1}^{n}(y_i - \theta_i)^2/2 + \theta^\top D^\top U^{-1} D\theta/2 + \sum_{i=1}^{n-1}(P\theta)_i^2/(2\rho^2 v_i) + b_\sigma.$$

3. Sampling of τ

   - Draw $\psi \sim \mathrm{IG}\,(1/2, 1/\tau^2 + 1)$.
   - Draw $\tau^2 \sim \mathrm{IG}\left((n-k)/2, \sum_{i=k+2}^{n}(D\theta)_i^2/(2\sigma^2) + 1/\psi\right)$.

4. Sampling of u

   - Draw $u_i^2 \sim \mathrm{IG}\,(1 + a_{u_i}, (D\theta)_i^2/(2\sigma^2) + b_\sigma)$, independently for $i = 1, \ldots, k+1$.
   - Draw $\nu_i^2 \sim \mathrm{IG}\,(1/2, 1/u_i + 1)$, independently for $i = k+2, \ldots, n$.
   - Draw $u_i^2 \sim \mathrm{IG}\,(1, (D\theta)_i^2/(2\sigma^2\tau^2) + 1/\nu_i)$, independently for $i = k+2, \ldots, n$.

5. Sampling of ρ^2 **(only when we assume the nearly isotonic constraint)**

   - Draw $v_i \sim \mathrm{GIG}\,(1/2, (P\theta)_i^2/(4\rho^2\sigma^2), 1/2\rho^2\sigma^2)$, independently for $i = 1, \ldots, n-1$.
   - Draw $\rho^2 \sim \mathrm{IG}\left((n-1)/2 + a_\rho, \sum_{i=1}^{n-1}(P\theta)_i^2/(2\sigma^2 v_i) + b_\rho\right)$.

In practical use, we specify hyper-parameters a_σ, b_σ, a_ρ, b_ρ, a_{u_i}, b_{u_i} for $i = 1, \ldots, k+1$. In our numerical studies, we set hyper-parameters $a_\sigma = b_\sigma = 0.1$, $a_\rho = b_\rho = 1$, and $a_{u_i} = b_{u_i} = 1$ for $i = 1, \ldots, k+1$ as default values.

5.3 Numerical experiments

We illustrate the performance of the proposed method through simulation studies. In Subsection 5.3.1, we deal with monotone functions as true regression functions. The aim is to compare the proposed methods with frequentist methods and to see the difference between unconstrained methods and nearly isotonic constraint methods. In Subsection 5.3.2, we will discuss the robustness of the shape constraint methods (nearly isotonic or monotone constraint) for structural misspecification such that the true boundary trend is not completely monotone.

5.3.1 Simulation (I): Estimation of monotone boundary

We generate the data from the model $y_i = f(x_i) + \varepsilon_i$ $(i = 1, \ldots, 100)$, where $f(x)$ and ε are a true function and a noise distribution, respectively. We assume the following two true functions:

(i) Square root (Sqrt): $f(x) = \sqrt{x}/2$

(ii) Piecewise constant (PC):

$$f(x) = 0.5 \cdot 1_{[1,20]}(x) + 1_{[21,40]}(x) + 2.5 \cdot I_{[41,60]}(x) + 3.5 \cdot 1_{[61,100]}(x)$$

We employ the (upper truncated) half-normal distribution $\mathrm{HN}(0, \sqrt{\pi/2}/\sigma)$ with location parameter 0 and variance parameter σ^2 to generate the noise ε_i. We consider the four scenarios: (a) $\sigma = 0.5$, (b) $\sigma = 1$, (c) $\sigma = 2$ and (d) mixtures of the half-normal $0.8 \times \mathrm{HN}(0, \sqrt{\pi/2}) + 0.2 \times \mathrm{HN}(0, \sqrt{\pi/2}/3)$. Hereafter, we often denote scenarios like (i-a) for example. We adopt the proposed methods: Bayesian boundary trend filtering under the horseshoe, Laplace, and normal priors (denoted by HS, Lap, and Nor) and nearly isotonic (NI) constrained Bayesian boundary trend filtering for each prior (denoted by HSNI, LapNI, and NorNI for short). As competitors, we consider the following frequentist methods:

- QS, CS, QSI, and CSI: Quadratic and Cubic spline methods (with/without isotonic constraint) for estimating boundary curve proposed by Daouia et al. (2016). The knot of the spline is selected by the Bayesian information criterion (BIC). R-code is provided by `R` package `npbr`.

- QTF: Quantile trend filtering method proposed by Brantley et al. (2020). The method solves the optimization problem using the alternating direction

method of multipliers (ADMM) algorithm, where the penalty parameter is determined by the extended Bayesian information criterion (eBIC). We use the quantile level 0.99 to estimate the extremal quantile trend. The method can be implemented by using their `R` package `detrendr`.

- FDH and LFDH: Classical nonparametric methods called the free disposal hull defined by Deprins et al. (1984) and its linearized version (see e.g. Hall and Park, 2002). The method can also be implemented by `R` package `npbr`.

For the proposed Bayesian methods, we generated 10500 posterior samples, then we removed the first 500 samples, and only every 5th scan was saved. For trend filtering methods (including QTF), we set the order of k as $k = 1$ (piecewise linear) and $k = 0$ (piecewise constant) for scenarios (i) and (ii) respectively. For the proposed method, we set $\eta = 500$.

To evaluate the performance of estimates, we adopted the root mean squared error (RMSE), the average length of the credible interval (AL), and the coverage probability (CP). These criteria are defined by $\text{RMSE} = \{n^{-1}\sum_{i=1}^{n}(f(x_i) - \hat{\theta}_i)^2\}^{1/2}$, $\text{AL} = n^{-1}\sum_{i=1}^{n}(\hat{\theta}_{97.5,i} - \hat{\theta}_{2.5,i})$, and $\text{CP} = n^{-1}\sum_{i=1}^{n} 1_{[\hat{\theta}_{2.5,i}, \hat{\theta}_{97.5,i}]}(f(x_i))$, respectively, where $\hat{\theta}_{100(1-\alpha),i}$ represent the $100(1-\alpha)\%$ posterior quantiles of θ_i. These values were averaged over 100 replications of simulating datasets. We only reported RMSE for frequentist competitors.

First, we show one-shot simulation results for some methods in Figure 5.1 when (b) $\sigma = 1$. We can observe that five methods (HS, HSNI, QS, QSI, and QTF) give reasonable estimates under scenario (i). For scenario (ii), the QS and QSI methods provide over-shrinkage estimates and the FDH and QTF methods can not capture some change points, while the proposed methods under the horseshoe prior give reasonable estimates for piecewise constant structure. A remarkable point is that the proposed methods under the horseshoe prior illustrate good performance for both scenarios.

We also report RMSE, AL, and CP averaged over 100 Monte Carlo replications in Tables 5.1 and 5.2. From Table 5.1, the results indicate that the proposed HS and HSNI methods provide more accurate point estimates than other Bayesian and frequentist methods except for scenarios (i-c) and (i-d). Focusing on scenario (i-d), we can observe that spline methods provide relatively smaller RMSE than those of the proposed Bayesian methods, and the proposed shape-constrained methods

significantly improve the RMSE of unconstraint methods. From Table 5.2, the coverage probabilities under the HS and HSNI methods are larger than the nominal level of 0.95 except for a few cases, whereas the average length of intervals of the HS and HSNI methods tend to be smaller than other methods under scenario (ii). Although the proposed methods under the Laplace prior also have reasonable coverage probabilities, the average lengths of intervals tend to be wider than those of the horseshoe prior.

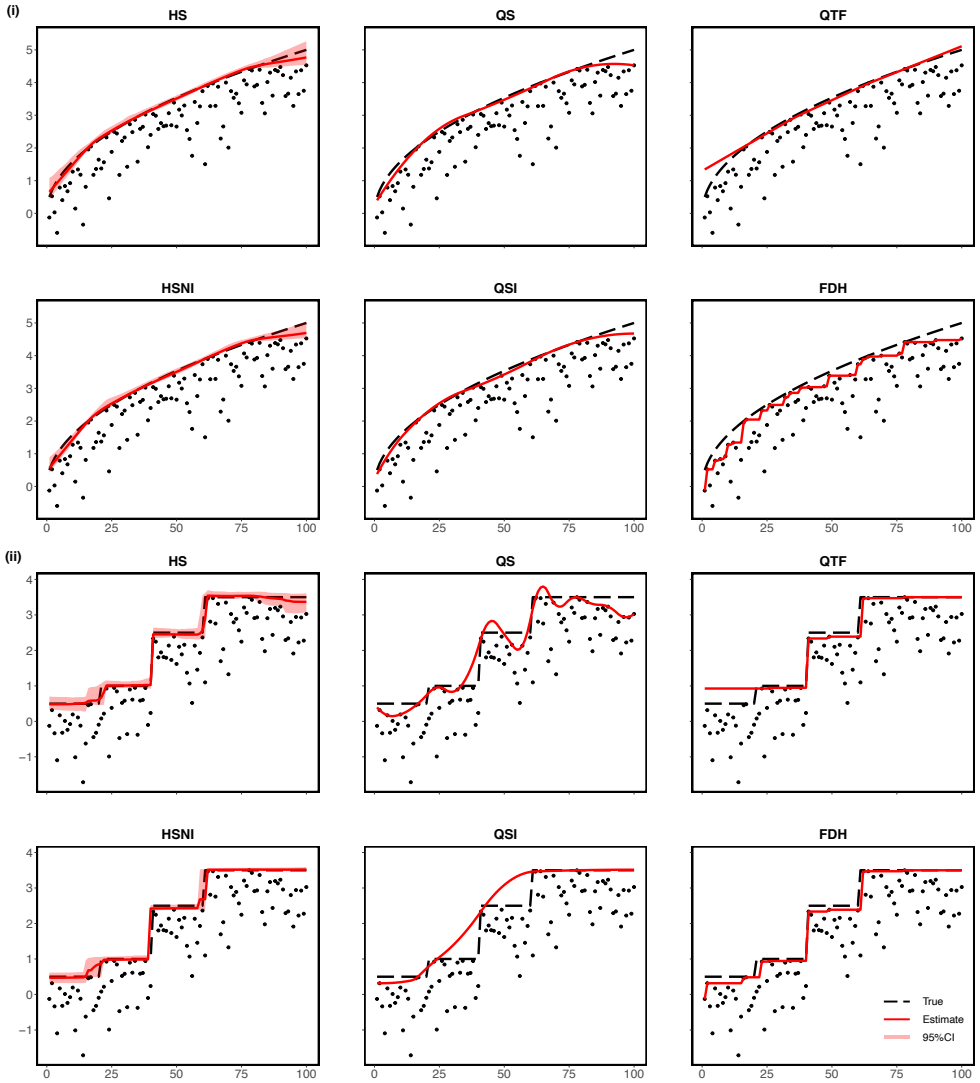

Figure 5.1: One-shot simulation results of six methods for two scenarios under the noise (b). For (i) Sqrt scenario (top panels) and (ii) PC scenario (bottom panels), the resulting estimates of HS, HSNI, QS, QSI, QTF, and FDH are shown.

Table 5.1: RMSE averaged over 100 Monte Carlo replications and their standard deviation (shown in parenthesis). The best score is bolded.

(i) Sqrt				
	(a)	(b)	(c)	(d)
HS	**0.041** (0.013)	0.072 (0.027)	0.122 (0.050)	0.257 (0.141)
Lap	0.134 (0.026)	0.265 (0.058)	0.528 (0.100)	0.347 (0.088)
Nor	0.049 (0.011)	0.078 (0.025)	**0.121** (0.050)	0.149 (0.031)
HSNI	**0.041** (0.014)	**0.070** (0.028)	0.131 (0.075)	0.174 (0.047)
LapNI	0.090 (0.023)	0.153 (0.039)	0.308 (0.083)	0.154 (0.043)
NorNI	0.047 (0.012)	0.077 (0.026)	0.122 (0.056)	0.138 (0.027)
QS	0.083 (0.043)	0.133 (0.088)	0.231 (0.173)	0.120 (0.055)
CS	0.078 (0.043)	0.130 (0.082)	0.248 (0.179)	0.125 (0.062)
QSI	0.064 (0.026)	0.094 (0.044)	0.166 (0.091)	**0.108** (0.052)
CSI	0.060 (0.027)	0.092 (0.042)	0.171 (0.087)	0.109 (0.050)
QTF	0.145 (0.037)	0.136 (0.041)	0.170 (0.053)	0.445 (0.042)
FDH	0.209 (0.025)	0.310 (0.042)	0.488 (0.100)	0.357 (0.080)
LFDH	0.170 (0.033)	0.363 (0.120)	0.876 (0.322)	0.952 (0.402)

(ii) PC				
	(a)	(b)	(c)	(d)
HS	**0.060** (0.026)	0.131 (0.041)	0.237 (0.058)	0.276 (0.163)
Lap	0.168 (0.025)	0.297 (0.051)	0.583 (0.116)	0.395 (0.112)
Nor	0.163 (0.016)	0.221 (0.027)	0.282 (0.050)	0.265 (0.040)
HSNI	0.061 (0.035)	**0.120** (0.064)	**0.230** (0.077)	**0.174** (0.070)
LapNI	0.122 (0.017)	0.193 (0.030)	0.318 (0.069)	0.235 (0.038)
NorNI	0.273 (0.023)	0.292 (0.032)	0.291 (0.033)	0.310 (0.038)
QS	0.227 (0.034)	0.320 (0.059)	0.459 (0.147)	0.349 (0.084)
CS	0.240 (0.034)	0.334 (0.059)	0.468 (0.159)	0.364 (0.089)
QSI	0.260 (0.079)	0.333 (0.076)	0.377 (0.059)	0.352 (0.068)
CSI	0.286 (0.083)	0.364 (0.054)	0.365 (0.063)	0.364 (0.061)
QTF	0.207 (0.029)	0.251 (0.050)	0.370 (0.223)	0.273 (0.057)
FDH	0.141 (0.032)	0.239 (0.068)	0.414 (0.108)	0.302 (0.089)
LFDH	0.350 (0.079)	0.602 (0.163)	1.297 (0.403)	1.293 (0.412)

Table 5.2: Average lengths (AL) and coverage probabilities (CP) of 95% credible intervals averaged over 100 Monte Carlo replications.

(i) Sqrt								
	(a)		(b)		(c)		(d)	
	AL	CP	AL	CP	AL	CP	AL	CP
HS	0.200	0.990	0.342	0.983	0.731	0.980	0.989	0.979
Lap	0.444	0.931	0.885	0.925	1.747	0.925	1.588	0.975
Nor	0.175	0.969	0.271	0.961	0.620	0.972	0.853	0.990
HSNI	0.183	0.987	0.287	0.973	0.413	0.939	0.637	0.955
LapNI	0.261	0.919	0.419	0.890	0.778	0.860	0.638	0.972
NorNI	0.160	0.960	0.219	0.920	0.343	0.901	0.430	0.956
(ii) PC								
	(a)		(b)		(c)		(d)	
	AL	CP	AL	CP	AL	CP	AL	CP
HS	0.211	0.980	0.438	0.969	0.948	0.958	0.780	0.975
Lap	0.564	0.942	1.024	0.941	1.927	0.935	1.616	0.961
Nor	0.583	0.944	0.813	0.939	1.259	0.957	0.978	0.929
HSNI	0.121	0.936	0.210	0.911	0.380	0.834	0.340	0.938
LapNI	0.301	0.872	0.472	0.854	0.814	0.844	0.682	0.901
NorNI	0.329	0.698	0.404	0.700	0.533	0.744	0.494	0.714

5.3.2 Simulation (II): Estimation of piecewise monotone boundary

In this subsection, we give an additional simulation study to show the usefulness of the nearly isotonic constraint. Although the monotone constraint $\theta_1 \leq \cdots \leq \theta_n$ is not robust against structural misspecification, we show that the proposed nearly isotonic method provides a reasonable estimate even if the monotonicity is partially violated. To verify the robustness of the proposed method, we consider a piecewise monotone function as a true function in the data generating process $y_i = f(x_i) + \varepsilon_i$ for $x_i = 1, \ldots, 100$. To this end, we consider the following piecewise sigmoid function (see also Meyer and Woodroofe, 2000; Minami, 2020):

$$\begin{aligned} f(x) = 1 &+ 4\exp(32x - 8)/(1 + \exp(16x - 8)) \cdot 1_{[0,50]}(x) \\ &+ 4\exp(16(2x - 1) - 8)/(1 + \exp(16(2x - 1) - 8)) \cdot 1_{[51,100]}(x). \end{aligned}$$

The true function is monotone except for a jumping point at $x_i = 50$, then the scenario is reasonable to compare the HS, HSNI, and other isotonic constraint methods. As noise distributions, we also assume the half-normal distribution $\mathrm{HN}(0, \sqrt{\pi/2}/\sigma)$ with (a) $\sigma = 0.5$, (b) $\sigma = 1$ and (c) $\sigma = 2$ as in Subsection 5.3.1, and report the averaged values of RMSE, AL, and CP. We compare the proposed HS and HSNI methods as well as the frequentist boundary spline methods with/without the isotonic constraint (QS, QSI, CS, and CSI). The number of generated posterior samples is the same as that of Subsection 5.3.1, and we set $k = 1$ and $\eta = 500$ for the proposed methods.

We show one-shot simulation results in Figure 5.2 under case (b). From the figure, it is observed that the QS and CS methods can not capture well the jump point at $x = 50$. The QSI and CSI methods do not work at all because of model misspecification. In contrast to such methods, the proposed HS and HSNI methods provide smoother boundary trends and their estimates seem to be comparable. The result shows that the proposed HSNI method is more robust against structural misspecification than that of assuming a completely monotone constraint.

From Table 5.3, it is observed that the HSNI method has the smallest RMSE. For uncertainty quantification, we report average lengths and coverage probabilities of credible intervals in Table 5.4. The proposed methods have reasonable coverage probabilities except for the HSNI under (c). Therefore, the proposed nearly isotonic

method can reasonably estimate a piecewise monotone trend as well as a completely monotonic trend.

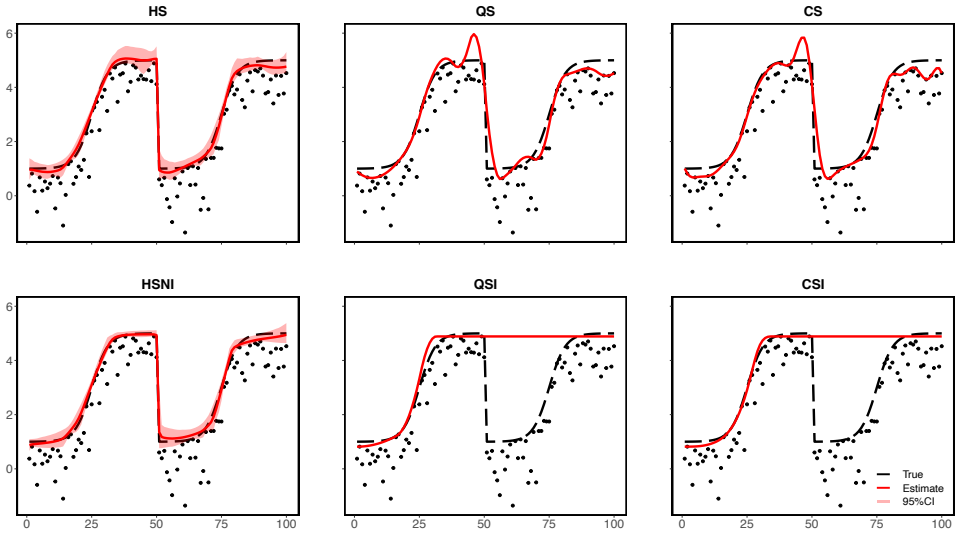

Figure 5.2: One-shot simulation results of six methods (HS, HSNI, QS, QSI, CS, and CSI) for the piecewise sigmoid scenario under the noise (b).

Table 5.3: RMSE averaged over 100 Monte Carlo replications and their standard deviation (shown in parenthesis). The best score is bolded.

	(a)	(b)	(c)
HS	0.089 (0.026)	0.194 (0.065)	0.408 (0.083)
HSNI	**0.082** (0.019)	**0.178** (0.064)	**0.375** (0.094)
QS	0.467 (0.074)	0.515 (0.093)	0.659 (0.110)
CS	0.478 (0.065)	0.536 (0.076)	0.706 (0.103)
QSI	1.826 (0.017)	1.809 (0.030)	1.790 (0.052)
CSI	1.826 (0.019)	1.808 (0.031)	1.794 (0.055)

5.3.3 Sensitivity analysis for selecting of η

Since the sigmoid function $\sigma_\eta(x)$ defined by (5.7) converges in the sense of L_1-convergence to the indicator function $1_{[0,\infty)}(x)$ as $\eta \to \infty$, we may select a moderate large η in practice. We here check the sensitivity of the point estimates of θ for various values of η using the same simulated dataset as Subsection 5.3.1. We considered three values $\eta \in \{100, 200, 500\}$. The boxplots of RMSE for the two scenarios (i) Sqrt and (ii) PC are provided in Figure 5.3, where the noise distribution is half-

Table 5.4: Average lengths (AL) and coverage probabilities (CP) of 95% credible intervals averaged over 100 Monte Carlo replications.

	(a)		(b)		(c)	
	AL	CP	AL	CP	AL	CP
HS	0.363	0.967	0.645	0.961	1.205	0.933
HSNI	0.311	0.959	0.495	0.927	0.812	0.868

normal with the standard deviation $\sigma = 1$. From these figures, we can observe that the results of RMSE for each η do not change very much.

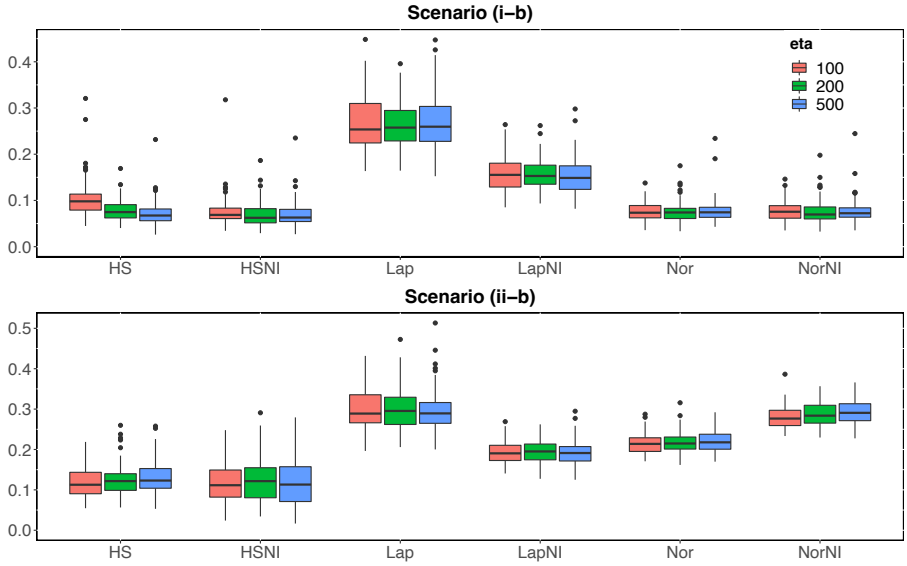

Figure 5.3: Sensitivity analysis of η under scenarios (i-b) and (ii-b).

5.3.4 Efficiency of sampling

We evaluate the efficiency of the proposed Gibbs sampler under the same simulation setting as in Subsection 5.3.1. We here adopt scenario (i-b) in Subsection 5.3.1 and employ the HS method.

The rejection sampler proposed by Botev (2017) is known as an efficient sampling method from the tMVN distribution, and it is interesting to compare the proposed sampler with Botev's one. However, the algorithm gets stack even if the dimension n is equal to 50 in our model because the acceptance rate becomes very low in high-dimension (see also Souris et al., 2018). Pakman and Paninski (2014) also proposed an excellent sampling algorithm based on Hamiltonian Monte Carlo, and

the algorithm can be easily implemented by using their R-package `tmg`. However, it is known that their method has the following problems: 1) Leaf-frog steps with careful tuning are necessary to obtain good mixing; 2) The algorithm often fails to produce answers and has high-computational cost (see also Souris et al., 2018; Ray et al., 2020). For these reasons, we consider a coordinate-wise sampler to obtain samples from the tMVN distribution as a competitor. Since the full conditional distribution of parameter vector $\theta = (\theta_1, \ldots, \theta_n)$ is the tMVN distribution, we can easily derive the full conditional distribution of θ_i given $\theta_{-i} = (\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_n)$ as one-dimensional truncated normal distribution (e.g. Okano et al., 2024). By using the coordinate-wise sampler, we can construct the Gibbs sampler in our model without approximating the indicator function. Although the run-time of the coordinate-wise sampler is faster than that of the proposed Gibbs sampler, especially for high-dimensional situations, we show that the proposed method is more efficient in terms of the effective sample size (ESS) through a simple simulation study.

We report the result of the sampling efficiency of two methods in Table 5.5. For a low dimensional case such as $n = 50$, they are comparable. On the other hand, it is observed that the larger dimension, the larger difference between them. For the coordinate-wise sampler, although run-time is relatively faster than the proposed method, the ESS is relatively small against the proposed one, especially for high-dimension. The run-time of the proposed method increases rapidly with dimension, but the ESS is relatively better than that of the coordinate-wise sampler. Interestingly, we can observe that the ESS of the proposed method tends to increase with dimension. As an example, we show the sample path and autocorrelation plot of parameter θ_{50} which is a specific location in Figure 5.4. It indicated that the autocorrelation does not rapidly decay for the coordinate-wise sampler, while the proposed method has reasonable mixing and autocorrelation.

5.4 Real data examples

We apply the proposed methods to two real data examples.

5.4.1 Production activity of air traffic controllers

We consider an efficient frontier estimation that corresponds to the production activity of the 37 European air traffic controllers (Mouchart and Simar, 2002). The

Table 5.5: CPU times (in seconds) and the mean of effective sample sizes (ESS) for $\theta_1, \ldots, \theta_n$ under scenario (i-b) (in Subsection 5.3.1) with HS method, averaged over 10 runs which are generated 2000 posterior samples after a burn-in period of 500 and after thinning the chain by 5.

	$n = 50$		$n = 100$		$n = 200$		$n = 300$	
	Time	ESS	Time	ESS	Time	ESS	Time	ESS
Proposed	11.17	69.75	47.73	92.59	289.14	112.19	910.72	112.07
Coordinate-wise	12.12	11.83	32.99	11.98	128.34	9.35	343.50	9.57

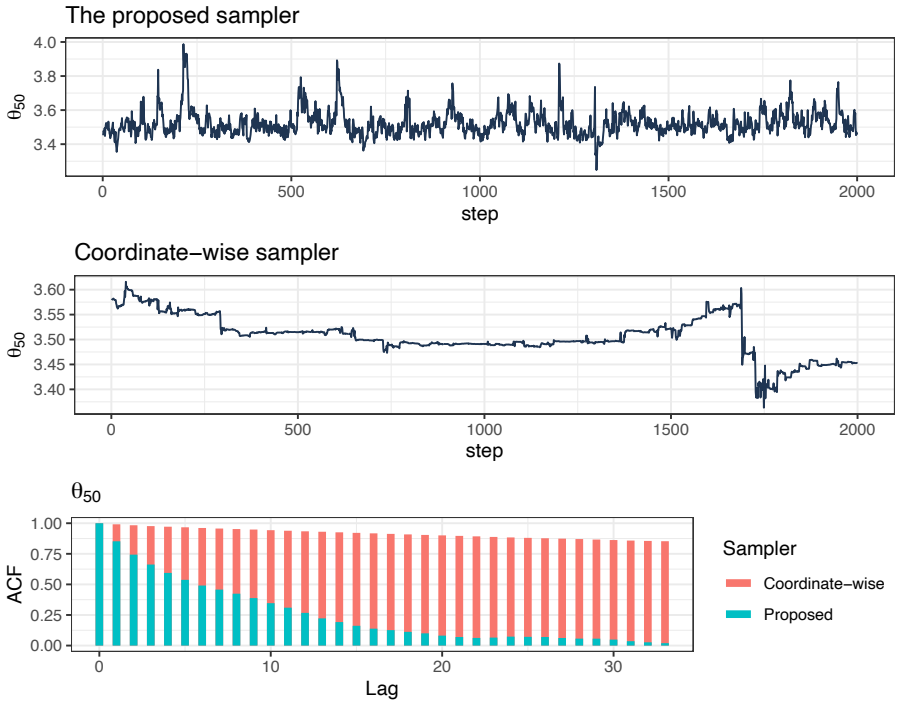

Figure 5.4: Trace plots and autocorrelations of posterior samples of θ_{50} under scenario (i-b) (in Subsection 5.3.1) with HS method. We used 2000 posterior samples after a burn-in period of 500 and after thinning the chain by 5.

data was also analyzed by Daouia et al. (2016), and they applied their boundary spline methods. We can obtain the data in `R` package `npbr`. From the scatter plot of the data in Figure 5.5, we observe that the assumption of a monotone boundary seems to be reasonable. The x and y-axis indicate the input (an aggregate factor of a different kind of labor) and output (an aggregate factor of the activity produced, based on the number of controlled air movements, the number of controlled flight hours, etc.) variables. We applied the proposed HS and HSNI methods compared with existing boundary quadratic spline with isotonic constraint (QSI) and LFDH methods. We note that Brantley's quantile trend filtering method we used in the previous section can not be applied to irregular grid data. To handle such an irregular grid, we need to use the proposed trend filtering methods with an adjusted difference matrix defined in Remark 3.2.1. To avoid numerical instability of the matrix $D^\top U^{-1} D$ in the proposed methods, we employed a transformation of the input variables as $1000 \times x$. We set the order of trend filtering as $k = 1$. We generated 10500 samples (burn-in 500) and then saved the 5th scanned samples.

The results are shown in Figure 5.5. Although the proposed HS method provides an almost monotonic point estimate, there is observed a decreasing trend between 2000 and 4000. On the other hand, the proposed HSNI method provides a reasonable point estimate of monotone boundary and uncertainty quantification. The average length of 95% credible intervals made by the HSNI method was 0.692, which was considerably smaller than the 0.936 produced by the HS method. Although the point estimates using the QSI and HSNI methods seem to be comparable, it is observed that the QSI and LFDH methods tend to give over-fitting estimates to data. In particular, the LFDH method can not estimate a smooth monotone boundary.

5.4.2 Global warming data

Global warming is one of the important issues in the world. Although the attention to the problem is often focused on the prediction of future climate change, it is also important to look back and explore the processes of past climate change. We apply the proposed method to estimate the past trend of annual temperature anomalies. The data is available from `R` package `CVXR`, and it includes the global monthly and annual temperature anomalies relative to the mean of 1960–1990 during 1850–2015 (the sample size is 166). The data is equally spaced non-stationary time series data, we can also observe that the monotonicity assumption is partially violated from the

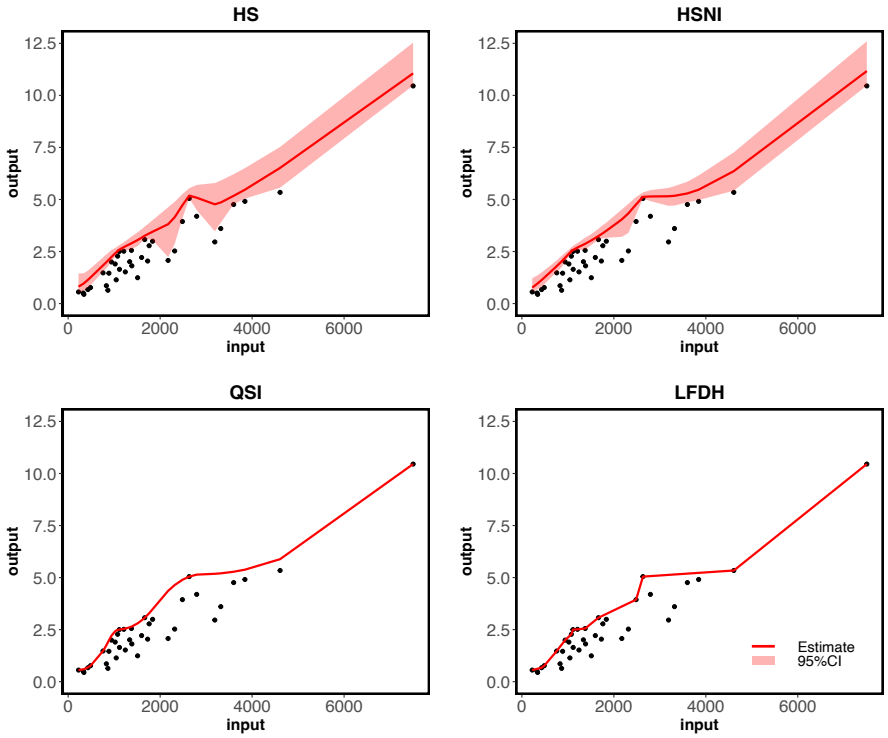

Figure 5.5: Estimated boundary curves for the proposed methods (HS and HSNI) and existing frequentist methods (QSI and LFDH) for air controllers data.

scatter plot in Figure 5.6. There exist a few outliers in 1877 and 1878, and they are reported as unexpected climate change (see e.g. Aceituno et al., 2009). Hence, the use of the nearly isotonic constraint may be useful. For the data, Tibshirani et al. (2011) applied the nearly isotonic regression to estimate the mean trend. In our analysis, we are interested in the estimation of the (upper and lower) boundary trends not in the mean trend. Estimating boundary trends is useful to clarify the variability of extreme values, and we can obtain the range of variability as a by-product.

In this analysis, we applied the proposed HS and HSNI methods. We generated 10500 samples (burn-in 500) and then saved the 5th scanned samples. As competitors, we employ the unconstraint quadratic spline (QS) and quantile trend filtering (QTF). We set the quantile levels of the QTF method as 0.99 and 0.01. For the HS, HSNI, and QTF methods, we assume that the order of trend filtering is $k = 1$.

The results of point estimates and credible intervals are shown in Figure 5.6. It is observed that the QS method gives a slightly overfitted estimate of the data and quantile trend filtering tends to induce over-shrinkage. On the other hand, the proposed HS and HSNI methods provide smoother and more locally adaptive boundary

trend estimates. The HSNI method provides almost monotone boundary trend estimates, while for the upper boundary, the monotonicity was violated from 1878 to 1920. Furthermore, average distances between upper and lower point estimates which is defined by $\sum_{i=1}^{166} |\hat{\theta}_{\text{upper},i} - \hat{\theta}_{\text{lower},i}|/166$ were 0.328 (HS), 0.439 (HSNI), 0.269 (QS), and 0.431 (QTF), respectively. In terms of the average distance, the proposed HSNI method is quite similar to the QTF method, and it is also considered that the QS method overfits the data and underestimates the extent to which data exists. For uncertainty quantification, average lengths of 95% credible intervals made by the HS method were 0.100 (lower) and 0.117 (upper), which are smaller than 0.126 (lower) and 0.164 (upper) by the HSNI method, respectively. At first glance, this result appears as if the assumption of shape constraint was not a reasonable one, but it also appears to successfully capture the uncertainty in the estimation of the boundary trend. For example, in Figure 5.6, the 95% credible intervals for the upper boundary become wider during 1978-1920, and the lower and upper limits of the intervals give the trends like the HS method and monotone trend, respectively. Hence, it indicates that the HSNI method provides a reasonable uncertainty evaluation to some extent. As we observed in Section 5.3, if the potential boundary is monotone or not monotone, the resulting 95% credible intervals of the HSNI method tend to be narrower than those of the HS method, unlike this result.

5.5 Concluding remarks

In this chapter, we proposed a Bayesian boundary trend filtering using the truncated multivariate normal working likelihood and global-local shrinkage priors. Using the approximation of the indicator function in the truncated multivariate normal likelihood, an efficient Gibbs sampling algorithm to sample from posterior distribution was also constructed.

We close this chapter by considering some future directions. Although we employ the truncated multivariate normal distribution as a working likelihood, it may lead to undesirable inference when the model is misspecified. To overcome this problem, we need to consider a kind of calibration method to obtain the correct coverage probability of credible interval (e.g. Syring and Martin, 2019; Onizuka et al., 2024a). Another important issue is "robustness" against outliers. However, it is well-known that the robust estimation of boundary curve is not easy unlike mean curve (e.g.

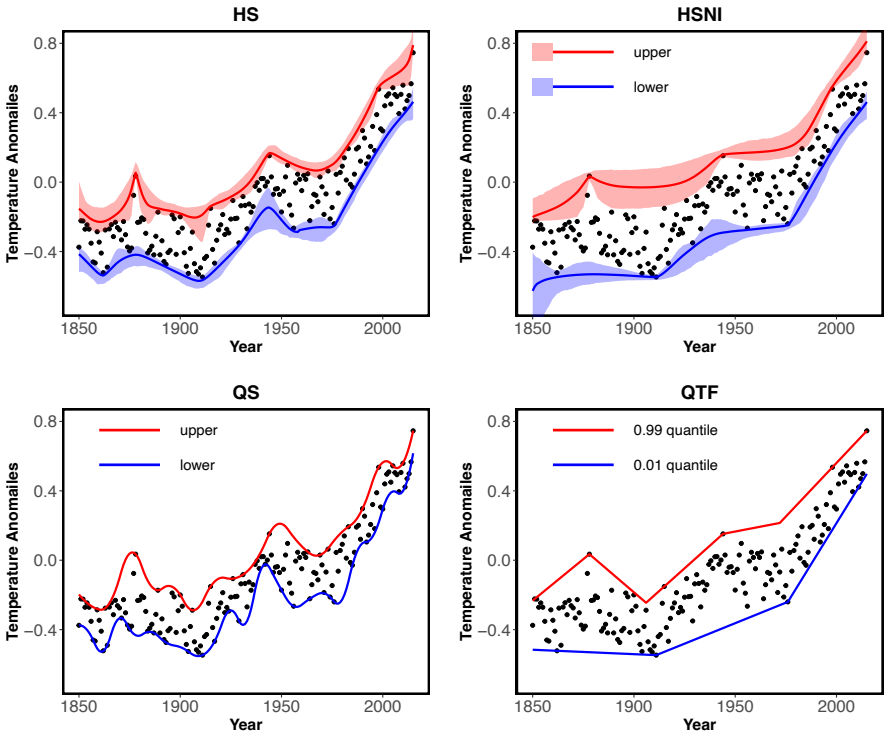

Figure 5.6: Estimated boundary curves for the proposed methods (HS and HSNI) and existing frequentist methods (QS and QTF) for global warming data.

Daouia and Simar, 2005; Daouia and Ruiz-Gazen, 2006; Daouia et al., 2021). In our framework, the scale constant η in the sigmoid function may play an important role in controlling the boundary constraint. Developing a suitable selection criterion of η in the presence of outliers will be interesting future work.

Chapter 6

Conclusions and Discussions

In this thesis, we proposed several locally adaptive Bayesian smoothing methods for not only time series data but also spatial data. The proposed methods were inspired by ℓ_1 trend filtering which was originally proposed by Kim et al. (2009). By assuming different types of working likelihoods, we considered to estimate various statistics such as quantiles and boundaries. Introducing global-local shrinkage priors, the locally adaptive smoothing was achieved, and the performance of the proposed method was illustrated through simulation and real data analysis. It is a crucial point that all proposed methods we presented in this thesis can be implemented by efficient algorithms such as MFVB and Gibbs sampler by using mixture representations or approximations of likelihood function.

There are several future directions for this study. First, it is interesting to apply other working likelihoods to trend filtering. For example, although we employ the asymmetric Laplace likelihood in Chapters 3 and 4, it would be possible to extend the framework to the extended asymmetric Laplace likelihood which is a more flexible distribution considering an additional shape parameter (e.g. Yan and Kottas, 2017). Moreover, since the proposed methods do not work well to estimate extremal quantiles, it is also important to extend the proposed methods to smoothing for extremal quantiles (e.g. Chernozhukov, 2005). For boundary trend filtering, it is considered to use heavy-tailed distribution (e.g. Desgagné, 2015; Desgagné and Gagnon, 2019; Hamura et al., 2022a) to deal with outliers. It is also useful to propose a locally adaptive stochastic frontier model by adding the non-negative technical inefficiency term. Since the stochastic frontier model (e.g. Aigner et al., 1977) assumes the error distribution which is the sum of truncated normal distribution and Gaussian distribution, it is associated with boundary trend filtering. Second, for the quantile

estimation, there is a crossing problem when we use asymmetric Laplace likelihood in general. Crossing means the absence of monotonicity with respect to quantile levels in the estimation of conditional and structural quantile functions. The non-crossing Bayesian quantile smoothing methods have been developed in recent years (e.g. Wang and Cai, 2023), and it is expected to extend the proposed quantile trend filtering to Bayesian non-crossing smoothing. Furthermore, it is also possible to establish the non-crossing smoothing focusing on the extremal quantiles. Third, although we only modeled smoothing without covariates, it may be important to consider the covariates. Since Sadhanala and Tibshirani (2019) considered an extension of trend filtering to additive models to handle covariates, such an extension of our proposed model will be also expected. Finally, it is important to prove some theoretical properties such as posterior consistency for the proposed models. In particular, it is a challenging issue for the Bayesian boundary trend filtering due to its non-regularity.

References

Aceituno, P., M. d. R. Prieto, M. E. Solari, A. Martínez, G. Poveda, and M. Falvey (2009). The 1877–1878 el niño episode: associated impacts in south america. *Climatic Change 92*, 389–416.

Aigner, D., C. K. Lovell, and P. Schmidt (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics 6*(1), 21–37.

Balke, N. S. (1993). Detecting level shifts in time series. *Journal of Business & Economic Statistics 11*(1), 81–92.

Balocchi, C., S. K. Deshpande, E. I. George, and S. T. Jensen (2023). Crime in philadelphia: Bayesian clustering with particle optimization. *Journal of the American Statistical Association 118*(542), 818–829.

Balocchi, C. and S. T. Jensen (2019). Spatial modeling of trends in crime over time in philadelphia. *The Annals of Applied Statistics 13*(4), 2235–2259.

Barata, R., R. Prado, and B. Sansó (2022). Fast inference for time-varying quantiles via flexible dynamic models with application to the characterization of atmospheric rivers. *The Annals of Applied Statistics 16*(1), 247–271.

Blei, D. M., A. Kucukelbir, and J. D. McAuliffe (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association 112*(518), 859–877.

Botev, Z. I. (2017). The normal law under linear restrictions: simulation and estimation via minimax tilting. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 79*(1), 125–148.

Braga, A. A. (2001). The effects of hot spots policing on crime. *The Annals of the American Academy of Political and Social Science 578*(1), 104–125.

Brantley, H. L., J. Guinness, and E. C. Chi (2020). Baseline drift estimation for air quality data using quantile trend filtering. *The Annals of Applied Statistics 14*(2), 585–604.

Carvalho, C. M., N. G. Polson, and J. G. Scott (2010). The horseshoe estimator for sparse signals. *Biometrika 97*(2), 465–480.

Castillo-Mateo, J., J. Asín, A. C. Cebrián, A. E. Gelfand, and J. Abaurrea (2023). Spatial quantile autoregression for season within year daily maximum temperature data. *The Annals of Applied Statistics 17*(3), 2305–2325.

Chernozhukov, V. (2005). Extremal quantile regression. *The Annals of Statistics 33*(2), 806–839.

Cobb, G. W. (1978). The problem of the nile: Conditional solution to a changepoint problem. *Biometrika 65*(2), 243–251.

Daouia, A., J.-P. Florens, and L. Simar (2021). Robustified expected maximum production frontiers. *Econometric Theory 37*(2), 346–387.

Daouia, A., T. Laurent, and H. Noh (2017). npbr: a package for nonparametric boundary regression in r. *Journal of Statistical Software 79*, 1–43.

Daouia, A., H. Noh, and B. U. Park (2016). Data envelope fitting with constrained polynomial splines. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 78*(1), 3–30.

Daouia, A. and A. Ruiz-Gazen (2006). Robust nonparametric frontier estimators: qualitative robustness and influence function. *Statistica Sinica*, 1233–1253.

Daouia, A. and L. Simar (2005). Robust nonparametric estimators of monotone boundaries. *Journal of Multivariate Analysis 96*(2), 311–331.

Datta, J. and D. B. Dunson (2016). Bayesian inference on quasi-sparse count data. *Biometrika 103*(4), 971–983.

Deprins, D., L. Simar, and H. Tulkens (1984). Measuring labor-efficiency in post offices. In *The Performance of Public Enterprises: Concepts and Measurements*, pp. 243–267.

Desgagné, A. (2015). Robustness to outliers in location–scale parameter model using log-regularly varying distributions. *The Annals of Statistics 43*(4), 1568–1595.

Desgagné, A. and P. Gagnon (2019). Bayesian robustness to outliers in linear regression and ratio estimation. *Brazilian Journal of Probability and Statistics 32*(2), 205–221.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. SIAM.

Eilers, P. H. and R. X. De Menezes (2005). Quantile smoothing of array cgh data. *Bioinformatics 21*(7), 1146–1153.

Fasiolo, M., S. N. Wood, M. Zaffran, R. Nedellec, and Y. Goude (2021). Fast calibrated additive quantile regression. *Journal of the American Statistical Association 116*(535), 1402–1412.

Faulkner, J. R., A. F. Magee, B. Shapiro, and V. N. Minin (2020). Horseshoe-based bayesian nonparametric estimation of effective population size trajectories. *Biometrics 76*(3), 677–690.

Faulkner, J. R. and V. N. Minin (2018). Locally adaptive smoothing with markov random fields and shrinkage priors. *Bayesian analysis 13*(1), 225.

Hall, P. and B. U. Park (2002). New methods for bias correction at endpoints and boundaries. *The Annals of Statistics 30*(5), 1460–1479.

Hall, P., B. U. Park, and S. E. Stern (1998). On polynomial estimators of frontiers and boundaries. *Journal of Multivariate Analysis 66*(1), 71–98.

Hall, P. and L. Simar (2002). Estimating a changepoint, boundary, or frontier in the presence of observation error. *Journal of the American statistical Association 97*(458), 523–534.

Hamura, Y., K. Irie, and S. Sugasawa (2021). Robust hierarchical modeling of counts under zero-inflation and outliers. *arXiv preprint arXiv:2106.10503*.

Hamura, Y., K. Irie, and S. Sugasawa (2022a). Log-regularly varying scale mixture of normals for robust regression. *Computational Statistics & Data Analysis 173*, 107517.

Hamura, Y., K. Irie, and S. Sugasawa (2022b). On global-local shrinkage priors for count data. *Bayesian Analysis 17*(2), 545–564.

Hamura, Y., T. Onizuka, S. Hashimoto, and S. Sugasawa (2024). Sparse bayesian inference on gamma-distributed observations using shape-scale inverse-gamma mixtures. *Bayesian Analysis 19*(1), 77–97.

Heng, Q., H. Zhou, and E. C. Chi (2023). Bayesian trend filtering via proximal markov chain monte carlo. *Journal of Computational and Graphical Statistics*, 1–12.

Kim, S.-J., K. Koh, S. Boyd, and D. Gorinevsky (2009). ℓ_1 trend filtering. *SIAM review 51*(2), 339–360.

Koenker, R. and G. Bassett Jr (1978). Regression quantiles. *Econometrica: journal of the Econometric Society*, 33–50.

Kowal, D. R., D. S. Matteson, and D. Ruppert (2019). Dynamic shrinkage processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 81*(4), 781–804.

Kozumi, H. and G. Kobayashi (2011). Gibbs sampling methods for bayesian quantile regression. *Journal of Statistical Computation and Simulation 81*(11), 1565–1578.

Liechty, M. W., J. C. Liechty, and P. Müller (2009). The shadow prior. *Journal of Computational and Graphical Statistics 18*(2), 368–383.

Liu, F., S. Chakraborty, F. Li, Y. Liu, and A. C. Lozano (2014). Bayesian regularization via graph laplacian. *Bayesian Analysis 9*(2), 449–474.

Lum, K. and A. E. Gelfand (2012). Spatial quantile multiple regression using the asymmetric laplace process. *Bayesian Analysis 7*(2), 235–258.

Madrid Padilla, O. H. and S. Chatterjee (2022). Risk bounds for quantile trend filtering. *Biometrika 109*(3), 751–768.

Makalic, E. and D. F. Schmidt (2015). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters 23*(1), 179–182.

Meyer, M. and M. Woodroofe (2000). On the degrees of freedom in shape-restricted regression. *The annals of Statistics 28*(4), 1083–1104.

Minami, K. (2020). Estimating piecewise monotone signals. *Electronic Journal of Statistics 14*, 1508–1576.

Mouchart, M. and L. Simar (2002). Efficiency analysis of air controllers: first insights. *Consulting report 202*.

Nychka, D., R. Furrer, J. Paige, and S. Sain (2017). *fields: Tools for spatial data*. R package version 9.6.

Nychka, D., G. Gray, P. Haaland, D. Martin, and M. O'connell (1995). A nonparametric regression approach to syringe grading for quality improvement. *Journal of the American Statistical Association 90*(432), 1171–1178.

Oh, H.-S., D. Nychka, T. Brown, and P. Charbonneau (2004). Period analysis of variable stars by robust smoothing. *Journal of the Royal Statistical Society: Series C (Applied Statistics) 53*(1), 15–30.

Okano, R., Y. Hamura, K. Irie, and S. Sugasawa (2024). Locally adaptive bayesian isotonic regression using half shrinkage priors. *Scandinavian Journal of Statistics 51*(1), 109–141.

Onizuka, T., S. Hashimoto, and S. Sugasawa (2024a). Fast and locally adaptive bayesian quantile smoothing using calibrated variational approximations. *Statistics and Computing 34*(1), 15.

Onizuka, T., S. Hashimoto, and S. Sugasawa (2024b). Locally adaptive spatial quantile smoothing: Application to monitoring crime density in tokyo. *Spatial Statistics 59*, 100793.

Onizuka, T., F. Iwashige, and S. Hashimoto (2024). Bayesian boundary trend filtering. *Computational Statistics & Data Analysis 191*, 107889.

Pakman, A. and L. Paninski (2014). Exact hamiltonian monte carlo for truncated multivariate gaussians. *Journal of Computational and Graphical Statistics 23*(2), 518–542.

Park, T. and G. Casella (2008). The bayesian lasso. *Journal of the American Statistical Association 103*(482), 681–686.

Politsch, C. A., J. Cisewski-Kehe, R. A. Croft, and L. Wasserman (2020). Trend filtering–i. a modern statistical tool for time-domain astronomy and astronomical spectroscopy. *Monthly Notices of the Royal Astronomical Society 492*(3), 4005–4018.

Polson, N. G. and J. G. Scott (2012). On the half-cauchy prior for a global scale parameter. *Bayesian Analysis 7*(4), 887 – 902.

Polson, N. G. and J. G. Scott (2013). Data augmentation for non-gaussian regression models using variance-mean mixtures. *Biometrika 100*(2), 459–471.

Polson, N. G., J. G. Scott, and J. Windle (2013). Bayesian inference for logistic models using pólya–gamma latent variables. *Journal of the American statistical Association 108*(504), 1339–1349.

Ramdas, A. and R. J. Tibshirani (2016). Fast and flexible admm algorithms for trend filtering. *Journal of Computational and Graphical Statistics 25*(3), 839–858.

Ray, P., D. Pati, and A. Bhattacharya (2020). Efficient bayesian shape-restricted function estimation with constrained gaussian process priors. *Statistics and Computing 30*(4), 839–853.

Reich, B. J., M. Fuentes, and D. B. Dunson (2011). Bayesian spatial quantile regression. *Journal of the American Statistical Association 106*(493), 6–20.

Reiß, M. and J. Schmidt-Hieber (2020a). Nonparametric bayesian analysis of the compound poisson prior for support boundary recovery. *The Annals of Statistics 48*(3), 1432–1451.

Reiß, M. and J. Schmidt-Hieber (2020b). Posterior contraction rates for support boundary recovery. *Stochastic Processes and their Applications 130*(11), 6638–6656.

Reiß, M. and L. Selk (2017). Efficient estimation of functionals in nonparametric boundary models. *Bernoulli 23*(2), 1022–1055.

Robertson, T., F. T. Wright, and R. L. Dykstra (1988). *Order restricted statistical inference*, Volume 229. Wiley.

Roualdes, E. A. (2015). Bayesian trend filtering. *arXiv preprint arXiv:1505.07710*.

Rue, H. and L. Held (2005). *Gaussian Markov Random Fields: Theory and Applications*. CRC Press.

Sadhanala, V. and R. J. Tibshirani (2019). Additive models with trend filtering. *The Annals of Statistics 47*(6), 3032–3068.

Selk, L., C. Tillier, and O. Marigliano (2022). Multivariate boundary regression models. *Scandinavian Journal of Statistics 49*(1), 400–426.

Smith, R. L. (1994). Nonregular regression. *Biometrika 81*(1), 173–183.

Souris, A., A. Bhattacharya, and D. Pati (2018). The soft multivariate truncated normal distribution with applications to bayesian constrained estimation. *arXiv preprint arXiv:1807.09155*.

Sriram, K. (2015). A sandwich likelihood correction for bayesian quantile regression based on the misspecified asymmetric laplace density. *Statistics & Probability Letters 107*, 18–26.

Sriram, K., R. Ramamoorthi, and P. Ghosh (2013). Posterior consistency of bayesian quantile regression based on the misspecified asymmetric laplace density. *Bayesian Analysis 8*(2), 479–504.

Syring, N. and R. Martin (2019). Calibrating general posterior credible regions. *Biometrika 106*(2), 479–486.

Taddy, M. A. (2010). Autoregressive mixture models for dynamic spatial poisson processes: Application to tracking intensity of violent crime. *Journal of the American Statistical Association 105*(492), 1403–1417.

Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *The Annals of statistics 42*(1), 285–323.

Tibshirani, R. J., H. Hoefling, and R. Tibshirani (2011). Nearly-isotonic regression. *Technometrics 53*(1), 54–61.

Tibshirani, R. J. and J. Taylor (2011). The solution path of the generalized lasso. *The Annals of Statistics 39*(3), 1335–1371.

Tran, M.-N., T.-N. Nguyen, and V.-H. Dao (2021). A practical tutorial on variational bayes. *arXiv preprint arXiv:2103.01327*.

Wakayama, T. and S. Sugasawa (2023). Trend filtering for functional data. *Stat 12*(1), e590.

Wakayama, T. and S. Sugasawa (2024). Functional horseshoe smoothing for functional trend estimation. *Statistica Sinica 34*(3).

Wand, M. P., J. T. Ormerod, S. A. Padoan, and R. Frühwirth (2011). Mean field variational bayes for elaborate distributions. *Bayesian Analysis 6*(4), 847–900.

Wang, Q. and Z. Cai (2023). A composite bayesian approach for quantile curve fitting with non-crossing constraints. *Communications in Statistics-Theory and Methods*, 1–25.

Wang, Y.-X., J. Sharpnack, A. Smola, and R. Tibshirani (2015). Trend filtering on graphs. In *Artificial Intelligence and Statistics*, pp. 1042–1050. PMLR.

Xu, X. and M. Ghosh (2015). Bayesian variable selection and estimation for group lasso. *Bayesian Analysis 10*(4), 909–936.

Yamada, H. (2022). Trend extraction from economic time series with missing observations by generalized hodrick–prescott filters. *Econometric Theory 38*(3), 419–453.

Yan, Y. and A. Kottas (2017). A new family of error distributions for bayesian quantile regression. *arXiv preprint arXiv:1701.05666*.

Yano, K., R. Kaneko, and F. Komaki (2021). Minimax predictive density for sparse count data. *Bernoulli 27*(2), 1212–1238.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters 54*(4), 437–447.