Bayesian Quantile Trend Filtering on Graphs using Shrinkage Priors

Takahiro Onizuka\(^1\), Shintaro Hashimoto\(^1\) and Shonosuke Sugasawa\(^2\)

\(^1\)Department of Mathematics, Hiroshima University
\(^2\)Center for Spatial Information Science, The University of Tokyo

Abstract

Quantiles are useful characteristics of random variables that can provide substantial information of distributions compared with commonly used summary statistics such as means. In this paper, we propose a Bayesian quantile trend filtering method to estimate non-stationary trend of quantiles on graphs. We introduce general shrinkage priors for graph differences to induce locally adaptive Bayesian inference on trends. Introducing so-called shadow priors with multivariate truncated distribution for local scale parameters and mixture representation of the asymmetric Laplace distribution, we provide a simple Gibbs sampling algorithm to generate posterior samples. We also develop variational Bayes approximation to quickly compute point estimates (e.g. posterior means). The numerical performance of the proposed method is demonstrated through simulation study with time series data, application of quantile regression and robust spatial quantile smoothing.

Key words: Markov chain Monte Carlo; Markov random fields; shrinkage prior; variational Bayes
1 Introduction

Smoothing or trend estimation for data on general graph (e.g. time series and spatial data) is an important statistical method to investigate characteristics of the 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, 2021) and environmetrics (e.g. Brantley et al., 2020) among others. For estimating underlying trend, the $\ell_1$ trend filtering (Kim et al., 2009; Tibshirani, 2014) is known to be a powerful tool that can flexibly capture local abrupt changes in trend, compared with spline methods. The $\ell_1$ 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 were also proposed (e.g. Ramdas and Tibshirani, 2016). Due to such advantages in terms of flexibility and computation, extensions of the original trend filtering have been considered. For example, Wang et al. (2015) proposed trend filtering for spatial data using graph difference operator, Wakayama and Sugasawa (2021) extended the scalar trend filtering approach to the case with functional data. However, most of the studies focused on estimating mean trend under homogeneous variance structure, and these methods may not work well in the presence of outliers or data with heterogeneous variance. In addition, we are often interested in estimating other quantities such as quantiles instead of mean. Recently, Brantley et al. (2020) proposed the quantile version of trend filtering (QTF) for time series data in frequentist perspective, and they demonstrated QTF provides reasonable estimates of the baseline even under the presence of outliers.

The main difficulty in applying the frequentist trend filtering is that uncertainty quantification is not straightforward. Moreover, the frequentist formulation includes a tuning parameters in the regularization, but the data-dependent selection of the tuning parameter is not obvious especially under quantile smoothing. A natural alternative is employing Bayesian formulation for the trend filtering by introducing priors on trend. A simple Bayesian counterpart is a combination of the Gaussian likelihood and Laplace
prior distribution for differences between parameters (e.g., Roualdes, 2015). However, it is well-known that the shrinkage based on Laplace prior often causes over-shrinkage problem due to the tail of Laplace distribution. Recently, Faulkner and Minin (2018) proposed a more advanced version of Bayesian trend filtering via shrinkage priors such as horseshoe prior (Carvalho et al., 2010). Kowal et al. (2019) also considered the Bayesian trend filtering model based on dynamic shrinkage process. In spite the effectiveness of the existing Bayesian methods, the existing approach suffers from mainly two problems: 1) The current methods cannot be applied to quantile smoothing. 2) The current methods only focus on time series data; thereby it cannot handle smoothing of data on general graphs such as spatial data.

To overcome the issues, we extend the Bayesian trend filtering for time series data to Bayesian quantile trend filtering on general graph including spatial neighboring structures. 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 to construction of the prior distribution on the graph difference. In particular, we consider the horseshoe 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 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. Moreover, as a fast computation algorithm, we also propose mean-field variational Bayes (MFVB) approximation of the posterior, which enables us to quickly compute the posterior summary (e.g. posterior means). Since the proposed method is based on a general graph structure, it is not only the generalization of Faulkner and Minin (2018) or Brantley et al. (2020), but also enables the quantile smoothing for spatial data. We demonstrate the usefulness and wide applicability of proposed methods through extensive simulation studies and real data examples.

The paper is organized as follows: In Section 2, we propose a new Bayesian trend filtering method to estimate quantiles, and construct an efficient posterior sampling al-
algorithm based on Gibbs sampling and variational Bayesian approximation. In Section 3, we illustrate some simulation studies to compare the performance of proposed methods. In Section 4, we apply the proposed methods to some real data examples of both time series and spatial data. Additional numerical results are provided in the Supplementary Material.

2 Bayesian quantile trend filtering

2.1 Trend filtering via optimization

Let \( y_i = \theta_i + \varepsilon_i \) \( (i = 1, \ldots, n) \) be a sequence model, where \( y_i \) is an observation, \( \theta_i \) is a true location and \( \varepsilon_i \) is a noise. The estimate of \( \ell_1 \) trend filtering (Kim et al., 2009) is given by solving the optimization problem

\[
\hat{\theta} = \arg \min_{\theta} \| y - \theta \|^2_2 + \lambda \| D^{(k+1)} \theta \|_1,
\]

where \( y = (y_1, \ldots, y_n)^\top \), \( \theta = (\theta_1, \ldots, \theta_n)^\top \), \( D^{(k+1)} \) is a \((n-k-1) \times n\) difference operator matrix of order \( k+1 \), and \( \lambda > 0 \) is a tuning constant. Depending on the different order \( k \), we can express various smoothing such as piecewise constant, linear, quadratic and so on (Tibshirani, 2014). A fast and efficient optimization algorithm for the problem (1) is also proposed by Ramdas and Tibshirani (2016).

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

\[
\hat{\theta} = \arg \min_{\theta} \rho_p(y - \theta) + \lambda \| D^{(k+1)} \theta \|_1,
\]

where \( \lambda > 0 \) is a tuning constant and \( \rho_p(\cdot) \) is a check loss function given by

\[
\rho_p(r) = \sum_{i=1}^n r_i \{ p - 1 (r_i < 0) \}, \quad 0 < p < 1.
\]

To solve the problem (2), Brantley et al. (2020) proposed a parallelizable alternating direction method of multipliers (ADMM) algorithm, and they also proposed the selection of smoothing parameter \( \lambda \) using a modified criterion based on the extended Bayesian
information criterion.

2.2 Shrinkage priors on graph differences

To estimate the quantile trend, we consider the following model:

\[ y_i = \theta_i + \varepsilon_i, \quad \varepsilon_i \sim \text{AL}(p, \sigma^2), \quad i = 1, \ldots, n, \tag{3} \]

where \( \theta_i \) and \( \sigma^2 \) are unknown parameters and \( p \) is a fixed quantile level. Here \( \text{AL}(p, \sigma^2) \) denotes the asymmetric Laplace distribution:

\[ f_{\theta}(x) = \frac{p(1-p)}{\sigma^2} \exp \left\{ -p \left( \frac{x}{\sigma^2} \right) \right\}, \]

where \( p \) is a fixed constant which characterizes the quantile level and \( \sigma^2 \) (not \( \sigma \)) is a scale parameter.

Suppose that \( \theta_1, \ldots, \theta_n \) are on general graphs (including the standard trend filtering as a linear chain graph). 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 note that the graph represents time series data when \( G \) is a linear chain graph. We assume that \( |V| = n \) and \( |E| = m \).

For \( k = 0 \), if \( e_\ell = (i, j) \in V \), then \( D^{(1)} \) has \( \ell \)-th row

\[ D_{\ell}^{(1)} = (0, \ldots, 0, \frac{1}{i}, 0 \ldots, 0, -1, 0, \ldots, 0), \tag{4} \]

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

\[ D^{(k+1)} = \begin{cases} \left( D^{(1)} \right)^\top D^{(k)} \quad \text{for odd } k, \\ D^{(1)} D^{(k)} \quad \text{for even } k. \end{cases} \tag{5} \]

Note that 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 \( k \). If we consider the linear chain graph

\[ G = (V, E), \quad V = \{1, \ldots, n\}, \quad E = \{(i, i+1) \mid i = 1, \ldots, n-1\}, \]
the graph difference operator $D^{(1)}$ is equal to the usual difference operator used in Kim et al. (2009).

Let $D$ be a $m \times n$ full-rank matrix representing general difference operator on a graph, and we consider flexible shrinkage priors on $D\theta$. We here assume that $m \geq n$ since the number of edges is typically larger than that of nodes. When $m$ is smaller than $n$ as in a linear chain graph, $D$ can be transformed to $n \times n$ non-singular matrix.

We consider the prior $D\theta | \tau^2, \sigma^2, w \sim N_n(0, \tau^2 \sigma^2 W)$ with a diagonal covariance matrix $W = \text{diag}(w_1^2, \ldots, w_m^2)$, where $w = (w_1, \ldots, w_m)$ represents local shrinkage parameters for each element in $D\theta$ and $\tau^2$ is a global shrinkage parameter. When $m = n$, the prior can be rewritten as

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

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 $\theta$ is given by

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

(6)

Now, we consider the prior for $w$. The standard approach is using 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 (6). Alternatively, we consider the following joint prior:

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

(7)

where $\pi(w_i)$ is a proper univariate distribution, and $\underline{w}$ and $\bar{w}$ are very small and large constants such as $\underline{w} = 10^{-10}$ and $\bar{w} = 10^{10}$. Note that the prior (7) is truncated on the region $(\underline{w}, \bar{w})^m$, which ensures that the prior (7) is proper since $|D^\top W^{-1}D|^{-1/2}|W|^{-1/2}$ is finite on the truncated region. The prior (7) is a kind of “shadow priors”, and such
priors are used to improve mixing of MCMC (e.g. [Liechty et al., 2009]) or to construct tractable full conditional distributions (e.g. [Liu et al., 2014; Xu and Ghosh, 2015]). As shown later, the resulting full conditional distributions of $w$ are familiar forms under well-known priors for local shrinkage parameters.

As univariate distribution $\pi(w_i)$ in (7), we consider two types of distributions, $w_i \sim \text{Exp}(1/2)$ and $w_i \sim C^+(0,1)$. These priors are motivated from 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 $\tau \sim C^+(0, C_\tau)$.

To handle the case that there are multiple observation per grid point, we will consider the following model

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

where $\theta(x)$ is a quantile in the location $x$, and $N_i$ is the number of data per each location $x_i$. It is a natural generalization of the sequence model (3) (see also Heng et al., 2022).

**2.3 Posterior computation**

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 $\epsilon_{ij} \sim \text{AL}(p, \sigma^2)$, we have the following argumentation

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

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 $\theta$ in (6), the full conditional distributions of $z_i$ and $\theta$ are given by

$$
\theta \mid y, z, \sigma^2, \tau^2 \sim N_n \left( A^{-1} B, \sigma^2 A^{-1} \right),
$$

$$
z_{ij} \mid y_{ij}, \theta_i, \sigma^2 \sim \text{GIG} \left( \frac{1}{2}, \frac{(y_{ij} - \theta_i)^2}{\tau^2 \sigma^2}, \psi \frac{\sigma^2}{\tau^2}, \psi \frac{\sigma^2}{\tau^2} + 2 \right), \quad i = 1, \ldots, n, \ j = 1, \ldots, N_i,
$$

where

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

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

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

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

$$
\tau^2 \mid \theta, w, \sigma^2, \xi \sim \text{IG} \left( \frac{n + 1}{2}, \frac{1}{2\sigma^2} \theta^\top D^\top W^{-1} D \theta + \frac{1}{\xi} \right), \quad \xi \mid \tau^2 \sim \text{IG} \left( \frac{1}{2}, \frac{1}{\tau^2} + 1 \right),
$$

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

where $N$ is the number of total data and $\xi$ is an augmented parameter for $\tau^2$. 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.

- **(Normal-type prior)** For Normal-type prior, in the prior (7), $W = I_m$, where $I_m$ is $m \times m$ identity matrix. So, for Normal type prior, $\theta$, $z_i$, $\sigma^2$, $\tau^2$ and $\xi$ are the all parameters and the full conditional distributions of these parameters are equal to the previous equation with $W = I_m$.

- **(Laplace-type prior)** The full conditional distributions of $\theta$, $z_i$ and $\sigma^2$ has already been mentioned. For Laplace-type prior, we give $\tau^2 = 1$ and $w_i \mid \gamma^2 \sim \text{IG} \left( \frac{1}{2}, \frac{1}{\gamma^2} + 1 \right)$.
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 \(\text{IG}(\gamma^2 | 1/2, 1/\nu)\) and \(\text{IG}(\nu | 1/2, 1/a^2)\), then \(\gamma \sim C^+(0,a)\), the full conditional distributions of \(w_i\), \(\gamma^2\) and \(\nu\) are given by

\[
\begin{align*}
  w^2_i \mid \theta, \sigma^2, \gamma^2, \nu &\sim \text{GIG}_{[w, \overline{w}]} \left(1, \frac{\eta_i^2}{\sigma^2}, \gamma^2\right), \\
  \gamma^2 \mid w, \nu &\sim \text{GIG} \left(m - \frac{1}{2}, \frac{2}{\nu}, \sum_{i=1}^{m} w_i^2\right), \\
  \nu \mid \gamma^2 &\sim \text{IG} \left(1/2, \frac{1}{\gamma^2} + 1\right),
\end{align*}
\]

where \(\text{GIG}_{[w, \overline{w}]}(a, b, c)\) is truncated generalized inverse Gaussian distribution on the region \([w, \overline{w}]\) and \(\eta_i = (D\theta)_i\).

-(**Horseshoe-type prior**) The full conditional distributions of \(\theta, z_i, \sigma^2\) and \(\tau^2\) has already been mentioned. For Horseshoe-type prior, \(w_i \sim C^+(0,1)\). By using the representation that \(w^2_i \mid \nu_i \sim \text{IG}(1/2, 1/\nu_i)\) and \(\nu_i \sim (1/2, 1)\), the full conditional distributions of \(w_i\) and \(\nu_i\) are given by

\[
\begin{align*}
  w^2_i \mid \theta, \sigma^2, \gamma^2, \nu &\sim \text{IG}_{[w, \overline{w}]} \left(1, \frac{1}{\nu_i} + \frac{\eta_i^2}{2\sigma^2\tau^2}\right), \\
  \nu_i \mid w_i &\sim \text{IG} \left(1/2, \frac{1}{w_i^2} + 1\right),
\end{align*}
\]

where \(\text{IG}_{[w, \overline{w}]}(a, b)\) is truncated inverse Gamma distribution on the region \([w, \overline{w}]\) and \(\eta_i = (D\theta)_i\).

### 2.4 Variational Bayes approximation

The MCMC algorithm could be computationally intensive when the number of edges is large. For more quick computation of the posterior distribution, we here derive variational Bayes approximation (e.g. Blei et al., 2017; Tran et al., 2021) of the joint posterior. The idea of the variational Bayes method is to approximate a intractable posterior distribution by more simple probability distribution by using optimization method, and it is different from MCMC method since the variational Bayes method does not need the sampling from the posterior distributions. In particular, we employ the mean-field variational Bayes (MFVB) approximation algorithms that requires the forms of full conditional distributions as given in Section 2.3.
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 \( p(\theta | y) \), that is,

\[
q^* = \arg\min_{q \in \mathcal{Q}} \text{KL}(q(\theta || p(\theta | y))) = \arg\min_{q \in \mathcal{Q}} \int q(\theta) \log \frac{q(\theta)}{p(\theta | y)} d\theta.
\] (8)

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

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

where \( \theta_{-k} \) denotes the parameters other than \( \theta_k \) and \( E_{\theta_{-k}}[\cdot] \) denotes the expectation under the probability density given parameters except for \( \theta_k \). Such a form of approximating is known as MFVB approximation. If the full conditional distribution of \( \theta_k \) has a familiar form, the above formula is easy to compute. According to the Gibbs sampling algorithm in Section 2.3, we use 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

\[
q(\theta) \sim N_n(B^{-1}C, E_{\sigma^2}B^{-1}), \quad q(z_{ij}) \sim \text{GIG}\left(\frac{1}{2}, a_{z_{ij}}, b_{z_{ij}}\right),
\]

\[
q(\sigma^2) \sim \text{IG}\left(\frac{n + 3N}{2} + a_{\sigma}, A_{\sigma^2}\right), \quad q(\tau^2) \sim \text{IG}\left(\frac{n + 1}{2}, a_{\tau^2}\right), \quad q(\xi) \sim \text{IG}\left(\frac{1}{2}, E_{1/\tau^2} + 1\right).
\]

To obtain the variational parameters in each distribution, we can update the parameters
as follows:

\[ B = \frac{1}{t^2} \text{diag} \left( \sum_{j=1}^{N_1} E_{1/z_{1j}}, \ldots, \sum_{j=1}^{N_n} E_{1/z_{nj}} \right) + E_{1/t^2} D^T \text{diag}(E_{1/u_{1}^2}, \ldots, E_{1/u_{m}^2}) D, \]

\[ C = \frac{1}{t^2} (C^* - \psi \mathbf{1}_n), \quad C^* = \left( \sum_{j=1}^{N_1} y_{1j} E_{1/z_{1j}}, \ldots, \sum_{j=1}^{N_n} y_{nj} E_{1/z_{nj}} \right)^T, \]

\[ E_{\eta_i^2} = d_i^T (E_{\sigma_i^2} B^{-1} + B^{-1} C^T B^{-1}) d_i, \quad E_{\theta_i} = E_{\theta_i}^T (E_{\sigma_i^2} B^{-1} + B^{-1} C^T B^{-1}) \]

\[ a_{z_{ij}} = \frac{1}{t^2} (y_{ij}^2 - 2y_{ij} (B^{-1} C)_i + E_{\eta_i^2}) \frac{n + 3N + 2a_{\sigma}}{2A_{\sigma^2}}, \quad b_{z_{ij}} = \left( \frac{\psi^2}{t^2} + 2 \right) \frac{n + 3N + 2a_{\sigma}}{2A_{\sigma^2}}, \]

\[ E_{z_{ij}} = \frac{\sqrt{a_{z_{ij}}} K_{3/2} (\sqrt{a_{z_{ij}}} b_{z_{ij}})}{\sqrt{b_{z_{ij}}} K_{1/2} (\sqrt{a_{z_{ij}}} b_{z_{ij}})}, \quad E_{1/z_{ij}} = \frac{\sqrt{b_{z_{ij}}} K_{3/2} (\sqrt{a_{z_{ij}}} b_{z_{ij}})}{\sqrt{a_{z_{ij}}} K_{1/2} (\sqrt{a_{z_{ij}}} b_{z_{ij}})} - \frac{1}{a_{z_{ij}}}, \]

where \( e_i \) is an unit vector that the \( i \)th component is 1, \( d_i^T \) is the \( i \)th row of difference matrix \( D \), and \( K_c(\cdot) \) is the modified Bessel function of the second kind with order \( c \).

The details under each prior are provided as follows:

- **(Normal-type prior)** In \([\text{[a]}]\), if we assume that \( E_{1/u_i^2} = 1 \) for \( i = 1, \ldots, m \), then we can derive the variational distributions under Normal-type prior.

- **(Laplace-type prior)** Since variational distributions under Laplace-type prior hold \( E_{1/t^2} = 1 \) and \( E_{1/\theta} = 0 \), the variational distributions of \( \theta \), \( z_i \) and \( \sigma^2 \) are the same as those of \([\text{[b]}]\). The variational distributions for \( u_i^2 \), \( \gamma^2 \) and \( \nu \) are given by

\[ q(u_i^2) \sim \text{GIG}_{[\nu, \nu]} \left( \frac{1}{2} a_{u_i^2}, E_{\eta_i^2} \right), \]

\[ q(\gamma^2) \sim \text{GIG} \left( m - \frac{1}{2}, 2E_{1/\nu}, \sum_{i=1}^{m} E_{u_i^2} \right), \quad q(\nu) \sim \text{IG} \left( \frac{1}{2}, E_{1/\gamma^2} + 1 \right), \]

11
where

\[ a_{w^2}^2 = \frac{2n + a_{\sigma}}{A_{\sigma}^2} E_{q^2_i}, \quad E_{q^2_i} = \frac{G(w) - G(w)}{E(w) - F(w)} \sqrt{a_{w^2}^2 K_{3/2}}(\sqrt{a_{w^2}^2 E_{q^2_i}}) \]

\[ E_{1/w^2_i} = \frac{H(w) - H(w)}{F(w) - F(w)} \sqrt{a_{w^2}^2 K_{1/2}}(\sqrt{a_{w^2}^2 E_{q^2_i}}) \]

\[ E_{\eta^2} = \frac{\sqrt{2E_{1/\nu}K_{m+1/2}}(\sqrt{2E_{1/\nu} \sum_{i=1}^{m} E_{w_i^2}})}{\sqrt{\sum_{i=1}^{m} E_{w_i^2}K_{m-1/2}}(\sqrt{2E_{1/\nu} \sum_{i=1}^{m} E_{w_i^2}})} \]

\[ E_{1/\gamma^2} = \frac{\sqrt{2E_{1/\nu}K_{m-1/2}}(\sqrt{2E_{1/\nu} \sum_{i=1}^{m} E_{w_i^2}})}{\sqrt{\sum_{i=1}^{m} E_{w_i^2}K_{m+1/2}}(\sqrt{2E_{1/\nu} \sum_{i=1}^{m} E_{w_i^2}})} \]

\[ E_{1/\nu^2} = \frac{1}{2(E_{1/\gamma^2} + 1)}, \]

and \( F, G \) and \( H \) are cumulative distribution functions of \( \text{GIG}(1/2, a_{w^2}, E_{\eta^2}) \), \( \text{GIG}(3/2, a_{w^2}, E_{\eta^2}) \) and \( \text{GIG}(-1/2, a_{w^2}, E_{\eta^2}) \) respectively.

**Horseshoe-type prior** The variational distributions for \( w^2_i \) and \( \nu_i \) are given by

\[ q(w^2_i) \sim \text{IG}(1, a_{w^2}), \quad q(\nu_i) \sim \text{IG} \left( \frac{1}{2}, E_{1/w^2_i} + 1 \right), \]

where

\[ a_{w^2}^2 = E_{1/\nu^2} + \frac{12n + a_{\sigma}}{A_{\sigma}^2} E_{1/\eta^2}, \]

\[ E_{1/w^2_i} = \frac{G(w) - G(w)}{F(w) - F(w)} \frac{1}{a_{w^2}^2} \]

\[ E_{1/\nu^2} = \frac{1}{2(E_{1/w^2_i} + 1)}, \]

and \( F \) and \( G \) are cumulative distribution functions of \( \text{IG}(1, a_{w^2}) \) and \( \text{IG}(2, a_{w^2}) \) respectively.

### 2.5 Weighted edge

We follow Heng et al. (2022) to give some discussion about extension to the proposed method to the situation where edge has some weight. It is equal to that the locations of data \( x = (x_1, \ldots, x_n) \) have the ordering \( x_1 < x_2 < \cdots < x_n \) and \( \omega_j = x_{j+1} - x_j \) is not constant. This issue is related to nonparametric quantile regression. When the
locations $x \in \mathbb{R}^n$ are irregular and strictly increasing, Heng et al. (2022) proposed an adjusted difference operator for $k \geq 1$ (see also Yamada, 2021).

For vertexes $x \in V = \mathbb{R}^n$ for a graph $(V,E)$, if there exists a weight $\omega_\ell > 0$ on the edge $e_\ell \in E$ between $x_i$ and $x_j$, we define the second order adjusted graph difference operator as

$$D^{(2)*} = (D^{(1)})^\top \text{diag} \left(\frac{1}{\omega_1}, \ldots, \frac{1}{\omega_m}\right) D^{(1)},$$

where $D^{(1)}$ is the usual graph difference operator. The operator (10) is thought of a generalization of the usual second order graph difference operator. Although Wang et al. (2015) mentioned that the weighted graph difference operator can be defined by using the weighted incidence matrix, they did not define it. Hence, we do not discuss the adjusted graph difference operator for $k \geq 2$ here. As an example of a graph with weighted edge, we will deal with data with irregular spacing in Section 4.2.

3 Simulation studies

We illustrate performance of the proposed method through two types of graph-structure (linear chain graphs and 2-D lattice graphs). For linear chain graphs, we compare the performance of the proposed method under three priors. On the other hand, we apply the proposed method to robust estimation for signals on 2-D lattice graphs.

3.1 Linear chain graphs

To compare performance of shrinkage priors, we consider the following data generating processes which are also considered in Faulkner and Minin (2018) and Brantley et al. (2020). We assume that the data generating process is

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

where $f(x)$ and $\varepsilon(x)$ are a true function and a noise function, respectively. First, we consider the following two true functions: (PC) Piecewise constant $f(x) = 2.5$ ($1 \leq x \leq 20$); $= 1.0$ ($21 \leq x \leq 40$); $= 3.5$ ($41 \leq x \leq 60$); $= 1.5$ ($61 \leq x \leq 100$) and (VS)
Varying smoothness \( f(x) = 2 + \sin(4x - 2) + 2 \exp(-30(4x - 2)^2) \). Since (PC) has three change-points, we use this function to assess the ability to catch a constant line and jumping structure. (VS) is smooth and it has a rapid change (see also Faulkner and Minin (2018)). Hence, this situation are reasonable to compare the shrinkage effect of the proposed methods and to show the ability of trend filtering to adapt to localized change. Second, we consider the following three scenarios for noise:

(I) Gaussian noise: \( \varepsilon(x) \sim N(0, ((1 + x^2)/4)^2) \).

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

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

Note that the three scenarios represent the heterogeneous variance. For each scenario, true quantile trends are summarized in Figure 1. True quantile trends were computed from the quantiles of point-wise noise distributions. We nest introduce the details of simulations. We used the six quantile trend filtering methods proposed in Section 2.3 and 2.4. Hereafter, the three priors are often denoted by HS, Lap and Norm, respectively. To compare with frequentist method, we use the quantile trend filtering based on the ADMM algorithm proposed by Brantley et al. (2020), where the tuning parameter of quantile trend filtering is determined by eBIC. Furthermore, we consider \( k \)th order quantile trend filtering for \( k = 0, 1, 2 \). We note that \( k = 0, 1, 2 \) correspond to piecewise constant, piecewise linear, and piecewise quadratic, respectively. We generated 5000 posterior samples by using Gibbs sampler given in Section 2.3, and then only every 10th scan was saved (thinning). As criteria to measure performances of estimators, we adopt mean squared error (MSE), mean absolute deviation (MAD), mean credible interval width (MCIW) and coverage probability (CP) which are defined by

\[
\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (\theta_i - \hat{\theta}_i)^2, \quad \text{MAD} = \frac{1}{n} \sum_{i=1}^{n} |\theta_i - \hat{\theta}_i|,
\]

\[
\text{MCIW} = \frac{1}{n} \sum_{i=1}^{n} \hat{\theta}_{97.5,i} - \hat{\theta}_{2.5,i}, \quad \text{CP} = \frac{1}{n} \sum_{i=1}^{n} I(\hat{\theta}_{2.5,i} \leq \theta_i^* \leq \hat{\theta}_{97.5,i}),
\]

\(^1\)R package is available in https://github.com/halleybrantley/detrendr
respectively, where $\hat{\theta}_{\alpha,i}$ represent the $\alpha\%$ posterior quantiles of $\theta_i$ and $\theta^*_i$ are true quantiles at location $x_i$. These values were averaged over 100 replications of simulating datasets. We report results of simulation studies for each scenario as follows. In the main manuscript, we show results of $k = 0$ for scenario (PC) and $k = 1$ for scenario (VS). Other simulation results are provided in Supplementary Material.

![Graphs](image)

**Figure 1: True quantile trends.**

**Piecewise constant.** We summarize numerical results for (PC) with $k = 0$ in Table 1. Point estimates via MCMC under horseshoe prior performed the best in all cases, and the variational Bayes methods under normal prior performed the worst in terms of MSE and MAD. Point estimates via ADMM also performed well under Gaussian and Beta noises. For uncertainly quantification, all MCMC methods have reasonable coverage probabilities for each quantile, while variational Bayes methods seem to be far away from nominal coverage. We also show one-shot simulation results under Gaussian noise in Figures 2 and 3.

**Varying smoothness.** From Figure 2, MCMC methods under horseshoe prior performed well relative to the other priors in terms of point estimates, while variational Bayes methods under horseshoe prior also provided comparable point estimates. Although
Table 1: Average values of MSE, MAD, MCIW and CP based on 100 replications for piecewise constant with \( k = 0 \). The minimum values of MSE and MAD are represented in bold.

|           | (I) Gauss | (II) Beta | (III) Mixed Normal |
|-----------|-----------|-----------|-------------------|
| MSE       | 0.013     | 0.099     | 0.014             |
|           | 0.002     | 0.003     | 0.005             |
|           | 0.046     | 0.037     | 0.044             |
| MCMC-HS   | 0.043     | 0.037     | 0.043             |
| MCMC-Lap  | 0.074     | 0.062     | 0.075             |
| MCMC-Norm | 0.027     | 0.022     | 0.028             |
| VB-HS     | 0.037     | 0.033     | 0.037             |
| VB-Norm   | 0.096     | 0.071     | 0.094             |
| ADMM      | 0.040     | 0.045     | 0.102             |
| MAD       | 0.086     | 0.072     | 0.087             |
|           | 0.025     | 0.038     | 0.052             |
|           | 0.157     | 0.138     | 0.152             |
| MCMC-HS   | 0.160     | 0.147     | 0.159             |
| MCMC-Lap  | 0.191     | 0.180     | 0.191             |
| MCMC-Norm | 0.119     | 0.101     | 0.119             |
| VB-HS     | 0.148     | 0.138     | 0.147             |
| VB-Norm   | 0.203     | 0.182     | 0.201             |
| ADMM      | 0.147     | 0.168     | 0.190             |
| MCIW      | 0.492     | 0.453     | 0.487             |
|           | 0.151     | 0.200     | 0.263             |
|           | 0.802     | 0.753     | 0.800             |
| MCMC-HS   | 0.821     | 0.817     | 0.820             |
| MCMC-Lap  | 0.914     | 0.916     | 0.910             |
| MCMC-Norm | 0.256     | 0.268     | 0.253             |
| VB-HS     | 0.567     | 0.604     | 0.566             |
| VB-Norm   | 0.712     | 0.757     | 0.711             |
| CP        | 0.965     | 0.975     | 0.956             |
|           | 0.978     | 0.960     | 0.945             |
|           | 0.934     | 0.959     | 0.947             |
| MCMC-HS   | 0.958     | 0.971     | 0.956             |
| MCMC-Lap  | 0.933     | 0.946     | 0.935             |
| VB-HS     | 0.840     | 0.744     | 0.617             |
| VB-Norm   | 0.859     | 0.890     | 0.860             |
Figure 2: One-shot simulation results under piecewise constant and Gauss noise. The order of trend filtering is $k = 0$ for all methods.

Figure 3: One-shot simulation results under piecewise constant and Gauss noise. The order of trend filtering is $k = 0$ for all methods.
MCMC methods derive reasonable coverage probabilities under Gauss and Beta noises, they are slightly worse under Mixed normal noise. We also show one-shot simulation results under Gaussian noise in Figures 4 and 5.

Figure 4: One-shot simulation results under varying smoothness and Gauss noise. The order of trend filtering is $k = 1$ for all methods.

3.2 2-D lattice graphs

Next, we show simulation studies for data on 2-D lattice graph. In particular, we apply the proposed method to robust graph signal denoising problem by using Bayesian median trend filtering which is a special case of the proposed method.

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 $G$. We generate a noisy data from the model $y_i = \theta_i + \varepsilon_i$ ($i = 1, \ldots, 100$), where we assume that the parameter $\theta_i$ has the following two block structure:

$$\theta_i = 5 \quad \text{(center), \quad and \quad } \theta_i = 0 \quad \text{(other)}.$$

and $\varepsilon_i$ is independently generated from a mixture distribution $0.95N(0, 1) + 0.05N(\mu, 1)$, where $\mu$ is the magnitude of outliers. Such data generation is widely used in the robust graph signal denoising problem, and contaminated mixture distribution is also used in robust statistics. We consider $\mu = 5$ (Scenario 1) and 10 (Scenario 2). The true
Table 2: Average values of MSE, MAD, MCIW and CP based on 100 replications for varying smoothness with $k = 1$. The minimum values of MSE and MAD are represented in bold.

|       | (I) Gauss |       | (II) Beta |       | (III) Mixed Normal |       |
|-------|-----------|-------|-----------|-------|-------------------|-------|
| MSE   |           |       |           |       |                   |       |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |                   |       |
| MCMC-HS | 0.027 0.016 0.018 | 0.003 0.004 0.006 | 0.089 0.050 0.046 |       |                   |       |
| MCMC-Lap | 0.058 0.027 0.027 | 0.005 0.006 0.009 | 0.108 0.070 0.060 |       |                   |       |
| MCMC-Norm | 0.085 0.037 0.030 | 0.007 0.007 0.010 | 0.139 0.093 0.069 |       |                   |       |
| VB-HS  | 0.021 0.018 0.021 | 0.004 0.005 0.008 | 0.051 0.040 0.048 |       |                   |       |
| VB-Lap | 0.064 0.027 0.027 | 0.004 0.006 0.009 | 0.112 0.072 0.065 |       |                   |       |
| VB-Norm | 0.120 0.050 0.035 | 0.014 0.007 0.011 | 0.158 0.106 0.082 |       |                   |       |
| ADMM  | 0.097 0.039 0.040 | 0.087 0.017 0.022 | 0.111 0.069 0.073 |       |                   |       |
| MAD   |           |       |           |       |                   |       |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |                   |       |
| MCMC-HS | 0.109 0.091 0.100 | 0.035 0.044 0.056 | 0.186 0.156 0.159 |       |                   |       |
| MCMC-Lap | 0.144 0.120 0.125 | 0.045 0.056 0.070 | 0.203 0.181 0.185 |       |                   |       |
| MCMC-Norm | 0.159 0.129 0.131 | 0.052 0.059 0.073 | 0.214 0.194 0.195 |       |                   |       |
| VB-HS  | 0.108 0.098 0.106 | 0.038 0.050 0.065 | 0.164 0.147 0.165 |       |                   |       |
| VB-Lap | 0.150 0.120 0.126 | 0.042 0.055 0.071 | 0.210 0.188 0.194 |       |                   |       |
| VB-Norm | 0.179 0.142 0.139 | 0.060 0.061 0.076 | 0.225 0.210 0.210 |       |                   |       |
| ADMM  | 0.176 0.132 0.144 | 0.120 0.078 0.098 | 0.219 0.186 0.203 |       |                   |       |
| MCIW  |           |       |           |       |                   |       |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |                   |       |
| MCMC-HS | 0.467 0.449 0.452 | 0.190 0.220 0.236 | 0.642 0.654 0.649 |       |                   |       |
| MCMC-Lap | 0.543 0.552 0.551 | 0.272 0.295 0.328 | 0.685 0.728 0.749 |       |                   |       |
| MCMC-Norm | 0.518 0.556 0.557 | 0.299 0.310 0.340 | 0.608 0.684 0.734 |       |                   |       |
| VB-HS  | 0.226 0.240 0.227 | 0.108 0.128 0.133 | 0.318 0.334 0.321 |       |                   |       |
| VB-Lap | 0.321 0.374 0.358 | 0.183 0.214 0.219 | 0.401 0.460 0.467 |       |                   |       |
| VB-Norm | 0.272 0.359 0.368 | 0.205 0.240 0.243 | 0.331 0.400 0.442 |       |                   |       |
| CP    |           |       |           |       |                   |       |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |                   |       |
| MCMC-HS | 0.914 0.943 0.920 | 0.961 0.944 0.920 | 0.868 0.903 0.887 |       |                   |       |
| MCMC-Lap | 0.894 0.927 0.914 | 0.966 0.953 0.928 | 0.864 0.888 0.880 |       |                   |       |
| MCMC-Norm | 0.875 0.909 0.903 | 0.956 0.947 0.921 | 0.838 0.859 0.859 |       |                   |       |
| VB-HS  | 0.629 0.713 0.636 | 0.789 0.736 0.633 | 0.591 0.665 0.593 |       |                   |       |
| VB-Lap | 0.702 0.799 0.746 | 0.904 0.878 0.795 | 0.661 0.710 0.683 |       |                   |       |
| VB-Norm | 0.644 0.747 0.727 | 0.858 0.872 0.804 | 0.608 0.644 0.640 |       |                   |       |
Figure 5: One-shot simulation results under varying smoothness and Gauss noise. The order of trend filtering is $k = 1$ for all methods.
structures are shown in upper left panel of Figures 6 and 7. One-shot simulated noisy datasets are also shown in upper middle panel of Figures 6 and 7. We now compare the proposed BQTF with $p = 0.5$ (median filtering) under three priors and original Bayesian trend filtering under horseshoe prior (denoted by BTF-HS). To the best of our knowledge, although Bayesian trend filtering on graph (not quantile filtering) has not been proposed yet, we can construct the Gibbs sampler in a similar way to Section 2.3. We generated 5000 posterior samples and then only every 10th scan was saved. The order of trend filtering is $k = 1$ for all methods. To evaluate performance of estimates, we adopt MSE and MAD defined by Section 3.1. These values were averaged over 100 replications of simulating datasets.

One shot simulation results for both scenarios are reported in Figures 6 and 7. We can see that non-robust BTF-HS is strongly affected by outliers at one location since the BTF is based on the Gaussian likelihood. It can be also seen that the BTF in Scenario 2 is more affected by outliers than that of Scenario 1. On the other hand, the proposed three methods provide reasonable estimate for baseline trend for both scenarios. The results of MSE and MAD for both scenarios are shown in Figure 8. From Figure 8, although the proposed methods outperform the BTF-HS, BQTF with Laplace prior is slightly better among proposed methods.

4 Real data analysis

4.1 Nile data

We apply the proposed methods to famous Nile data (e.g. Cobb, 1978; Balke, 1993). The data contains measurements of annual flow of the river Nile from 1871 to 1970, and we can find that there is an apparent change-point near 1898. We now set $k = 0$, and compare the proposed methods via MCMC under three priors. For the each case, we generate 20000 posterior samples after discarding the first 5000 posterior samples as burn-in. The resulting estimates of quantiles and the corresponding 95% credible intervals are shown in Figure 9. From the results, the horseshoe prior seems to capture the piecewise constant structures well. However, the remaining two priors do not capture the change points well, and the fluctuation ranges of the estimates are relatively large.
Figure 6: One-shot simulation results for 2-D lattice graphs with $\mu = 5$.

Figure 7: One-shot simulation results for 2-D lattice graphs with $\mu = 10$. 

Figure 8: Boxplots of MSE (Left) and MAD (Right) for Scenario 1 ($\mu = 5$) and Scenario 2 ($\mu = 10$).

Figure 9: Point estimates and 95% credible intervals for Nile data.
4.2 Munich rent data

The proposed methods can be also applied to a linear chain graph data with weighted edge. We use the Munich rent data which includes the value of rent per square meter and floor space in Munich (see also Rue and Held 2005, Faulkner and Minin 2018, Heng et al. 2022). The data has multiple observation per location and 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, where for a location \( x_i \), \( y_i \) is multiple observation per location: \( y_i = (y_{i1}, \ldots, y_{iN_i}) \in \mathbb{R}^{N_i} \), and this data is unequally spaced, that is, \( x_{j+1} - x_j \) is not constant. The data contains 2035 observations and floor space (or location) has 134 distinct values. We now apply the second order adjusted difference operator defined in Section 2.5 to the proposed quantile trend filtering methods, and we calculate estimates of three quantile levels using MCMC under the normal, Laplace and horseshoe priors. We generated 20000 posterior samples after discarding the first 5000 posterior samples as burn-in. The results are reported in Figure 10. All three methods give the reasonable baseline estimates and uncertainly quantification as well as that of the Bayesian mean trend filtering shown in the supplementary material of Faulkner and Minin (2018). From Figure 10 all three models can estimate the decreasing baseline trend, that is, houses with small floor size have greater effect on rent. We can also see that the credible intervals are wider for large value of floor size because the data are less in such region.

![Figure 10: Point estimates and 95% credible intervals for Munich rent data.](image)

---

2The Munich rent data is available in https://github.com/jrfaulkner/spmrf
4.3 Tokyo crime data

We apply the proposed methods to spatial data analysis. We use “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 data contains the number of the crime in Tokyo, and we use the violent crime data in particular. We use the number of violent crimes at from some 7 wards in Tokyo in 2017 whose the sample size is $n = 697$ and the number of edges are 4023. Since for locations $x_i$ and $x_j$, $\theta_i$ and $\theta_j$ are considered to be connected if $d(x_i, x_j) < \varepsilon$ for some $\varepsilon$, we can easily construct the edge between $\theta_i$ and $\theta_j$. The data is shown in the left panel of Figure 11 and we can see some outliers. The data is also used in the paper Hamura et al. (2021) to find the hot spot in the presence of zero-inflations and outliers. In this section, the goal is to estimate the baseline and trend of the number of crimes, not detect hot spot. We adopt the Bayesian median trend filtering which is a special case of the proposed quantile trend filtering methods, and compare performances with the Bayesian mean trend filtering (denoted by BTF for short) which is not robust against outliers. In both methods, we employ the horseshoe prior. We generated 20,000 posterior samples and then only every 20th scan was saved for each method. The order of trend filtering is $k = 1$. Posterior means after logarithmic transformation are shown in Figure 11. Since original data contains some outliers, standard Bayesian trend filtering for mean seems to be highly affected by the outliers. On the other hand, the proposed median trend filtering is more robust than the mean trend filtering and provides spatially smoothed estimates of the baseline trend. For uncertainly quantification, average lengths of 95% posterior credible intervals for BTF and BQTF were 0.566 and 0.206, respectively. Hence, BQTF can provide efficient interval estimation without being much affected by potential outliers, whereas BTF produces much longer interval estimation due to outliers.
Figure 11: Estimated median trends for Tokyo crime data. Original data (Left), Bayesian trend filtering (Center) and Bayesian quantile trend filtering with $p = 0.5$ (Right).

5 Concluding remarks

In this paper, we proposed a Bayesian quantile trend filtering method. The proposed methods enable us to estimate quantile trend for not only time series data but also spatial data on graph. Although many quantile smoothing methods are proposed in frequentist perspective, the advantage of the proposed Bayesian methods are capable of full probabilistic uncertain quantification by using MCMC samples from the posterior distribution. We provided two-types of algorithm to estimate the posterior distributions, that is, the Gibbs sampler and mean-field variational Bayes approximation under three types of continuous shrinkage priors. Through simulation studies, it is shown that estimates under the horseshoe prior provide more locally adaptive smoothing of the quantile trend than other Laplace and normal priors in many scenarios. In computational perspectives, variational methods give accurate point estimates as well as MCMC. However, the uncertainly quantification is quite worse than that of MCMC. Hence, it might be useful when we want to obtain only point estimates like frequentist optimization methods.

There are several future directions for this paper. First, it should be proved some theoretical results for Bayesian quantile trend filtering such as the posterior consistency under misspecified asymmetric Laplace likelihood, the valid uncertainly quantification, and the posterior contraction rate (see also Sriram et al., 2013; Sriram, 2015; Banerjee...
Furthermore, since the proposed methods do not work well to estimate extremal quantiles such as $p = 0.95, 0.05$, it is also important to extend the proposed methods to smoothing for extremal quantiles (e.g. Chernozhukov 2005).

Acknowledgement

This work is partially supported by Japan Society for Promotion of Science (KAKENHI) grant numbers 21K13835 and 21H00699.

References

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

Banerjee, S. (2021). Horseshoe shrinkage methods for bayesian fusion estimation. *arXiv preprint arXiv:2102.07378*.

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.

Brantley, H. L., J. Guinness, and E. C. Chi (2020). Baseline drift estimation for air quality data using quantile trend filtering. *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.

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.

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

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

Heng, Q., H. Zhou, and E. C. Chi (2022). Bayesian trend filtering via proximal markov chain monte carlo. *arXiv preprint arXiv:2201.00092*.

Kim, S.-J., K. Koh, S. Boyd, and D. Gorinevsky (2009). $\ell_1$ trend filtering. *SIAM review* 51(2), 339–360.

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.

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.
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.

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.

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.

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

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 (2021). Locally adaptive smoothing for functional data. *arXiv preprint arXiv:2104.02456*.

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. (2021). Trend extraction from economic time series with missing observations by generalized hodrick–prescott filters. *Econometric Theory*, 1–35.

Yu, K. and R. A. Moyeed (2001). Bayesian quantile regression. *Statistics & Probability Letters* 54(4), 437–447.
Supplementary Materials for “Bayesian Quantile Trend Filtering on Graphs using Shrinkage Priors”

This Supplementary Material provides sample paths of the MCMC algorithm applied to Tokyo crime data and additional simulation results.

S1 Sampling efficiency of the MCMC algorithm in the application of Tokyo crime data

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 analysis of Tokyo crime data. We use the same posterior samples in Section 4.3, that is, every 20th value of $\theta_i$ from a Markov chain of length 20,000. Sample paths and autocorrelations of the posterior samples in 12 selected areas are provided in Figures S1 and S2, respectively. Mixing properties of the proposed algorithm seems satisfactory, and autocorrelations rapidly decay.

S2 Additional simulation results

We provide numerical results for other scenarios in Section 3.1. Results for piecewise constant with $k = 1$ and $k = 2$ are reported in Tables S1 and S2, respectively. We also show the results for varying smoothness with $k = 0$ and $k = 2$ in Tables S3 and S4, respectively.
Figure S1: Sample paths of the posterior samples of $\theta_i$ of BQTF in 12 areas under horseshoe prior applied to Tokyo crime data.
Figure S2: Autocorrelations of the posterior samples of $\theta_i$ of BQTF in 12 areas under horseshoe prior applied to Tokyo crime data.
Table S1: Average values of MSE, MAD, MCIW and CP based on 100 replications for picewise constant with \( k = 1 \). The minimum values of MSE and MAD are represented in bold.

|          | (I) Gauss |         | (II) Beta |         | (III) Mixed Normal |         |
|----------|-----------|---------|-----------|---------|--------------------|---------|
|          | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |        |                    |         |
| MSE      |           |         |           |         |                    |         |
| MCMC-HS  | 0.076 0.059 0.077 | 0.004 0.005 0.013 | 0.145 0.127 0.143 |        |                    |         |
| MCMC-Lap | 0.111 0.087 0.108 | 0.075 0.052 0.085 | 0.150 0.131 0.148 |        |                    |         |
| MCMC-Norm| 0.118 0.093 0.115 | 0.088 0.061 0.097 | 0.154 0.135 0.154 |        |                    |         |
| VB-HS    | 0.082 0.057 0.083 | 0.050 0.033 0.063 | 0.132 0.109 0.133 |        |                    |         |
| VB-Lap   | 0.120 0.091 0.115 | 0.077 0.051 0.089 | 0.158 0.134 0.159 |        |                    |         |
| VB-Norm  | 0.133 0.104 0.132 | 0.103 0.067 0.114 | 0.165 0.142 0.169 |        |                    |         |
| ADMM     | 0.153 0.110 0.148 | 0.125 0.086 0.140 | 0.190 0.153 0.188 |        |                    |         |
| MAD      |           |         |           |         |                    |         |
| MCMC-HS  | 0.175 0.157 0.175 | 0.035 0.047 0.069 | 0.265 0.249 0.260 |        |                    |         |
| MCMC-Lap | 0.210 0.193 0.208 | 0.128 0.122 0.151 | 0.270 0.254 0.267 |        |                    |         |
| MCMC-Norm| 0.216 0.200 0.215 | 0.143 0.133 0.164 | 0.274 0.259 0.272 |        |                    |         |
| VB-HS    | 0.169 0.145 0.170 | 0.078 0.082 0.110 | 0.244 0.226 0.247 |        |                    |         |
| VB-Lap   | 0.218 0.198 0.215 | 0.125 0.120 0.153 | 0.276 0.258 0.277 |        |                    |         |
| VB-Norm  | 0.231 0.212 0.231 | 0.155 0.142 0.179 | 0.282 0.266 0.286 |        |                    |         |
| ADMM     | 0.247 0.213 0.246 | 0.166 0.156 0.198 | 0.308 0.277 0.308 |        |                    |         |
| MCIW     |           |         |           |         |                    |         |
| MCMC-HS  | 0.663 0.651 0.668 | 0.220 0.258 0.330 | 0.829 0.825 0.824 |        |                    |         |
| MCMC-Lap | 0.705 0.705 0.703 | 0.465 0.482 0.505 | 0.858 0.851 0.852 |        |                    |         |
| MCMC-Norm| 0.693 0.697 0.695 | 0.475 0.490 0.511 | 0.839 0.831 0.831 |        |                    |         |
| VB-HS    | 0.290 0.308 0.289 | 0.140 0.171 0.176 | 0.389 0.411 0.386 |        |                    |         |
| VB-Lap   | 0.457 0.490 0.458 | 0.302 0.342 0.331 | 0.568 0.596 0.556 |        |                    |         |
| VB-Norm  | 0.459 0.488 0.460 | 0.329 0.368 0.353 | 0.563 0.585 0.548 |        |                    |         |
| CP       |           |         |           |         |                    |         |
| MCMC-HS  | 0.903 0.922 0.904 | 0.971 0.958 0.938 | 0.822 0.837 0.823 |        |                    |         |
| MCMC-Lap | 0.854 0.871 0.858 | 0.900 0.903 0.885 | 0.816 0.831 0.815 |        |                    |         |
| MCMC-Norm| 0.836 0.852 0.841 | 0.872 0.876 0.864 | 0.802 0.812 0.803 |        |                    |         |
| VB-HS    | 0.633 0.728 0.648 | 0.857 0.822 0.724 | 0.579 0.623 0.562 |        |                    |         |
| VB-Lap   | 0.710 0.760 0.718 | 0.838 0.841 0.782 | 0.665 0.707 0.652 |        |                    |         |
| VB-Norm  | 0.690 0.732 0.688 | 0.798 0.810 0.760 | 0.653 0.689 0.635 |        |                    |         |
Table S2: Average values of MSE, MAD, MCIW and CP based on 100 replications for piecewise constant with $k = 2$. The minimum values of MSE and MAD are represented in bold.

|       | (I) Gauss | (II) Beta | (III) Mixed Normal |
|-------|-----------|-----------|--------------------|
| **MSE** |           |           |                    |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |
| MCMC-HS | 0.146 0.119 0.150 | 0.087 0.059 0.112 | 0.168 0.148 0.167 |
| MCMC-Lap | 0.141 0.115 0.140 | 0.107 0.070 0.117 | 0.167 0.148 0.167 |
| MCMC-Norm | 0.148 0.125 0.151 | 0.128 0.083 0.135 | 0.173 0.155 0.175 |
| VB-HS | **0.123 0.095 0.119** | **0.053 0.040 0.071** | **0.164 0.144 0.163** |
| VB-Lap | 0.147 0.116 0.147 | 0.105 0.066 0.119 | 0.172 0.149 0.174 |
| VB-Norm | 0.154 0.131 0.159 | 0.147 0.087 0.154 | 0.179 0.158 0.186 |
| ADMM | 0.144 0.110 0.142 | 0.134 0.077 0.143 | 0.172 **0.144 0.176** |
| **MAD** |           |           |                    |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |
| MCMC-HS | 0.247 0.230 0.253 | 0.144 0.134 0.181 | 0.288 0.272 0.286 |
| MCMC-Lap | 0.240 0.226 0.242 | 0.163 0.145 0.183 | 0.288 0.272 **0.284** |
| MCMC-Norm | 0.250 0.236 0.254 | 0.183 0.163 0.201 | 0.294 0.279 0.293 |
| VB-HS | **0.228 0.204 0.223** | **0.104 0.103 0.139** | **0.284 0.270 0.286** |
| VB-Lap | 0.247 0.230 0.250 | 0.159 0.142 0.185 | 0.291 0.274 0.294 |
| VB-Norm | 0.257 0.244 0.265 | 0.197 0.170 0.218 | 0.300 0.282 0.305 |
| ADMM | 0.248 0.226 0.249 | 0.184 0.159 0.210 | 0.301 0.277 0.306 |
| **MCIW** |           |           |                    |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |
| MCMC-HS | 0.609 0.614 0.620 | 0.439 0.452 0.487 | 0.745 0.740 0.743 |
| MCMC-Lap | 0.643 0.645 0.648 | 0.450 0.462 0.488 | 0.776 0.767 0.770 |
| MCMC-Norm | 0.616 0.612 0.620 | 0.449 0.475 0.488 | 0.748 0.738 0.738 |
| VB-HS | 0.314 0.337 0.316 | 0.184 0.211 0.213 | 0.391 0.416 0.392 |
| VB-Lap | 0.397 0.422 0.397 | 0.276 0.312 0.300 | 0.494 0.517 0.481 |
| VB-Norm | 0.384 0.402 0.383 | 0.279 0.316 0.300 | 0.475 0.493 0.460 |
| **CP** |           |           |                    |
|       | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75 |
| MCMC-HS | 0.753 0.769 0.757 | 0.885 0.882 0.839 | 0.738 0.760 0.743 |
| MCMC-Lap | 0.784 0.795 0.789 | 0.850 0.853 0.831 | 0.754 0.770 0.758 |
| MCMC-Norm | 0.750 0.755 0.750 | 0.822 0.828 0.808 | 0.726 0.748 0.732 |
| VB-HS | 0.541 0.608 0.548 | 0.795 0.776 0.662 | 0.502 0.549 0.491 |
| VB-Lap | 0.603 0.638 0.591 | 0.768 0.780 0.692 | 0.582 0.622 0.563 |
| VB-Norm | 0.574 0.604 0.557 | 0.713 0.724 0.639 | 0.553 0.600 0.533 |
Table S3: Average values of MSE, MAD, MCIW and CP based on 100 replications varying smoothness with $k = 0$. The minimum values of MSE and MAD are represented in bold.

|        | (I) Gauss | (II) Beta | (III) Mixed Normal |
|--------|-----------|-----------|-------------------|
| **MSE** |           |           |                   |
|         | 0.25      | 0.5       | 0.75              | 0.25  | 0.5  | 0.75  | 0.25  | 0.5  | 0.75  |
| MCMC-HS | 0.029     | 0.022     | 0.025             | 0.008 | 0.008 | 0.010 | 0.060 | 0.048| 0.049 |
| MCMC-Lap | 0.031   | 0.026     | 0.030             | 0.010 | 0.009 | 0.011 | 0.063 | 0.051| 0.061 |
| MCMC-Norm | 0.036  | 0.027     | 0.030             | 0.010 | 0.012 | 0.015 | 0.077 | 0.063| 0.074 |
| VB-HS | 0.039     | 0.034     | 0.039             | 0.007 | 0.009 | 0.011 | 0.073 | 0.052| 0.054 |
| VB-Lap | 0.035     | 0.026     | 0.028             | 0.045 | 0.008 | 0.011 | 0.117 | 0.073| 0.059 |
| VB-Norm | 0.079    | 0.033     | 0.028             | 0.104 | 0.084 | 0.081 | 0.141 | 0.117| 0.115 |
| ADMM | 0.113     | 0.094     | 0.088             |       |       |       |       |       |       |

|        | (I) Gauss | (II) Beta | (III) Mixed Normal |
|--------|-----------|-----------|-------------------|
| **MAD** |           |           |                   |
|         | 0.25      | 0.5       | 0.75              | 0.25  | 0.5  | 0.75  | 0.25  | 0.5  | 0.75  |
| MCMC-HS | 0.120     | 0.108     | 0.116             | 0.056 | 0.062 | 0.073 | 0.171 | 0.154| 0.163 |
| MCMC-Lap | 0.136    | 0.123     | 0.134             | 0.060 | 0.066 | 0.076 | 0.195 | 0.177| 0.195 |
| MCMC-Norm | 0.141   | 0.127     | 0.136             | 0.067 | 0.069 | 0.078 | 0.196 | 0.178| 0.194 |
| VB-HS | 0.147     | 0.134     | 0.147             | 0.064 | 0.074 | 0.090 | 0.210 | 0.188| 0.206 |
| VB-Lap | 0.134     | 0.121     | 0.129             | 0.057 | 0.067 | 0.078 | 0.190 | 0.171| 0.181 |
| VB-Norm | 0.158    | 0.128     | 0.130             | 0.094 | 0.067 | 0.077 | 0.206 | 0.183| 0.185 |
| ADMM | 0.206     | 0.192     | 0.209             | 0.149 | 0.156 | 0.198 | 0.246 | 0.222| 0.235 |

|        | (I) Gauss | (II) Beta | (III) Mixed Normal |
|--------|-----------|-----------|-------------------|
| **MCIW** |           |           |                   |
|         | 0.25      | 0.5       | 0.75              | 0.25  | 0.5  | 0.75  | 0.25  | 0.5  | 0.75  |
| MCMC-HS | 0.957     | 0.587     | 0.604             | 0.340 | 0.366 | 0.399 | 0.758 | 0.745| 0.782 |
| MCMC-Lap | 0.713    | 0.706     | 0.714             | 0.385 | 0.395 | 0.418 | 0.987 | 0.973| 0.998 |
| MCMC-Norm | 0.716   | 0.715     | 0.723             | 0.421 | 0.420 | 0.437 | 0.954 | 0.949| 0.983 |
| VB-HS | 0.267     | 0.290     | 0.270             | 0.155 | 0.175 | 0.167 | 0.362 | 0.382| 0.360 |
| VB-Lap | 0.474     | 0.514     | 0.489             | 0.261 | 0.291 | 0.291 | 0.618 | 0.672| 0.660 |
| VB-Norm | 0.462    | 0.549     | 0.545             | 0.311 | 0.368 | 0.367 | 0.548 | 0.634| 0.662 |

|        | (I) Gauss | (II) Beta | (III) Mixed Normal |
|--------|-----------|-----------|-------------------|
| **CP** |           |           |                   |
|         | 0.25      | 0.5       | 0.75              | 0.25  | 0.5  | 0.75  | 0.25  | 0.5  | 0.75  |
| MCMC-HS | 0.941     | 0.956     | 0.944             | 0.974 | 0.967 | 0.939 | 0.910 | 0.928| 0.930 |
| MCMC-Lap | 0.961    | 0.973     | 0.962             | 0.978 | 0.972 | 0.946 | 0.954 | 0.971| 0.955 |
| MCMC-Norm | 0.957   | 0.973     | 0.962             | 0.980 | 0.974 | 0.951 | 0.948 | 0.959| 0.952 |
| VB-HS | 0.558     | 0.640     | 0.550             | 0.735 | 0.697 | 0.583 | 0.517 | 0.599| 0.520 |
| VB-Lap | 0.851     | 0.900     | 0.868             | 0.910 | 0.905 | 0.852 | 0.832 | 0.878| 0.852 |
| VB-Norm | 0.846    | 0.906     | 0.900             | 0.896 | 0.952 | 0.914 | 0.807 | 0.853| 0.842 |
Table S4: Average values of MSE, MAD, MCIW and CP based on 100 replications varying smoothness with \( k = 2 \). The minimum values of MSE and MAD are represented in bold.

|          | (I) Gauss | (II) Beta | (III) Mixed Normal |
|----------|-----------|-----------|--------------------|
| MSE      |           |           |                    |
|          | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75     |
| MCMC-HS  | 0.037 0.025 0.023 | 0.002 0.004 0.007 | 0.086 0.058 0.059 |
| MCMC-Lap | 0.070 0.040 0.034 | 0.005 0.007 0.010 | 0.129 0.088 0.074 |
| MCMC-Norm| 0.118 0.063 0.043 | 0.020 0.008 0.011 | 0.140 0.104 0.091 |
| VB-HS    | **0.025** 0.019 0.021 | 0.003 0.005 0.007 | 0.086 0.063 0.059 |
| VB-Lap   | 0.067 0.034 0.033 | 0.005 0.006 0.010 | 0.112 0.072 0.065 |
| VB-Norm  | 0.150 0.080 0.050 | 0.029 0.008 0.013 | 0.178 0.133 0.109 |
| ADMM     | 0.035 0.027 0.033 | 0.011 0.008 0.013 | 0.075 0.057 0.071 |

|          | (I) Gauss | (II) Beta | (III) Mixed Normal |
|----------|-----------|-----------|--------------------|
| MAD      |           |           |                    |
|          | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75     |
| MCMC-HS  | 0.126 0.111 0.114 | **0.033** 0.045 0.059 | 0.196 0.172 0.180 |
| MCMC-Lap | 0.158 0.133 0.137 | 0.047 0.058 0.073 | 0.212 0.195 0.202 |
| MCMC-Norm| 0.179 0.152 0.149 | 0.065 0.062 0.078 | 0.220 0.206 0.219 |
| VB-HS    | **0.113** 0.102 0.110 | 0.034 0.047 0.061 | **0.192** 0.177 0.181 |
| VB-Lap   | 0.157 0.131 0.137 | 0.044 0.057 0.073 | 0.210 0.188 0.194 |
| VB-Norm  | 0.196 0.167 0.160 | 0.071 0.064 0.082 | 0.235 0.227 0.239 |
| ADMM     | 0.137 0.126 0.137 | 0.058 0.062 0.082 | 0.210 0.186 0.207 |

|          | (I) Gauss | (II) Beta | (III) Mixed Normal |
|----------|-----------|-----------|--------------------|
| MCIW     |           |           |                    |
|          | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75     |
| MCMC-HS  | 0.483 0.465 0.462 | 0.184 0.218 0.254 | 0.664 0.678 0.675 |
| MCMC-Lap | 0.517 0.530 0.528 | 0.251 0.273 0.309 | 0.661 0.702 0.721 |
| MCMC-Norm| 0.452 0.515 0.529 | 0.285 0.284 0.319 | 0.579 0.628 0.675 |
| VB-HS    | 0.238 0.252 0.238 | 0.104 0.125 0.131 | 0.313 0.342 0.334 |
| VB-Lap   | 0.295 0.340 0.327 | 0.161 0.191 0.196 | 0.401 0.460 0.467 |
| VB-Norm  | 0.218 0.291 0.316 | 0.166 0.202 0.206 | 0.283 0.330 0.367 |

|          | (I) Gauss | (II) Beta | (III) Mixed Normal |
|----------|-----------|-----------|--------------------|
| CP       |           |           |                    |
|          | 0.25 0.5 0.75 | 0.25 0.5 0.75 | 0.25 0.5 0.75     |
| MCMC-HS  | 0.890 0.912 0.891 | 0.965 0.945 0.913 | 0.850 0.885 0.862 |
| MCMC-Lap | 0.870 0.892 0.871 | 0.946 0.929 0.899 | 0.854 0.867 0.848 |
| MCMC-Norm| 0.829 0.856 0.839 | 0.929 0.918 0.891 | 0.809 0.814 0.795 |
| VB-HS    | 0.639 0.705 0.631 | 0.813 0.747 0.648 | 0.547 0.602 0.577 |
| VB-Lap   | 0.651 0.738 0.679 | 0.860 0.834 0.742 | 0.661 0.710 0.683 |
| VB-Norm  | 0.539 0.635 0.637 | 0.798 0.806 0.721 | 0.537 0.550 0.522 |