Fast and locally adaptive Bayesian quantile smoothing using calibrated variational approximations

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

Received: 17 November 2022 / Accepted: 10 October 2023
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2023

Abstract
Quantiles are useful characteristics of random variables that can provide substantial information on distributions compared with commonly used summary statistics such as means. In this study, we propose a Bayesian quantile trend filtering method to estimate the non-stationary trend of quantiles. We introduce general shrinkage priors to induce locally adaptive Bayesian inference on trends and mixture representation of the asymmetric Laplace likelihood. To quickly compute the posterior distribution, we develop calibrated mean-field variational approximations to 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.

Keywords Calibration · Shrinkage prior · Trend filtering · Variational Bayes · Nonparametric quantile regression · Model misspecification

1 Introduction
Smoothing or trend estimation is an important statistical method to investigate 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) and environmetrics (e.g. Brantley et al. 2020) among others. For estimating underlying trends, 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 trends, 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 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, Brantley et al. (2020) proposed a quantile version of trend filtering (QTF) by adding a \(\ell_1\) penalization to the well-known check loss function, but the literature on quantile trend filtering remains scarce.

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 particular, Roualdes (2015) and Faulkner...
and Minin (2018) proposed the use of shrinkage priors for differences between parameters and Kowal et al. (2019) also considered a Bayesian formulation based on a dynamic shrinkage process under Gaussian likelihood to estimate the mean trend. However, the existing approach suffers mainly from three problems: (1) The current methods cannot be applied to quantile smoothing, (2) The posterior computation could be computationally intensive for a large sample size, and (3) The current methods cannot avoid the bias induced by model misspecification, therefore the resulting credible interval may not be valid.

In this study, we first proposed a Bayesian quantile trend filtering method. To this end, we employed the asymmetric Laplace distribution as a working likelihood (Yu and Moyeed 2001; Sriram et al. 2013). Combining the data augmentation strategy by Kozumi and Kobayashi (2011), we then constructed an efficient Gibbs sampling algorithm and a mean-field variational Bayes (MFVB) algorithm under two types of shrinkage priors; Laplace (Park and Casella 2008) and horseshoe (Carvalho et al. 2010) priors. It is well-known that the variational Bayes method enables the quick calculation of point estimates, while the MFVB algorithm tends to provide narrower credible intervals than that of Gibbs sampling (e.g. Blei et al. 2017). Additionally, the (possibly) misspecified asymmetric Laplace likelihood may produce invalid credible intervals. To overcome such problems, we proposed a new simulation-based calibration algorithm for variational posterior distribution which is expected to provide fast and valid credible intervals. We demonstrated the usefulness of the proposed methods through extensive simulation studies and real data examples.

The remainder of the paper is structured as follows: In Sect. 2, we formulate a Bayesian quantile trend filtering method and provide Gibbs sampling and variational Bayes algorithms. In Sect. 3, we describe the main proposal of this paper, a new calibration algorithm for approximating posterior distribution with variational Bayes approximation. In Sect. 4, we illustrate simulation studies to compare the performance of the proposed methods. In Sect. 5, we apply the proposed methods to real data examples. Concluding remarks are presented in Sect. 6. Additional information on the proposed algorithms and numerical experiments are provided in the Supplementary Material. R code implementing the proposed methods is available at GitHub repository (https://github.com/Takahiro-Onizuka/BQTF-VB).

2 Bayesian quantile trend filtering

2.1 Trend filtering via optimization

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

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

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 forth (see e.g. Tibshirani 2014). A fast and efficient optimization algorithm for the problem (1) was proposed by Ramdas and Tibshirani (2016).

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

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

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 also proposed the selection of smoothing parameters $\lambda$ using a modified criterion based on the extended Bayesian information criterion.

2.2 Bayesian formulation and shrinkage priors for differences

To conduct Bayesian inference of the quantile trend, we often use the following model:

$$y_i = \theta_{pi} + \epsilon_i, \quad \epsilon_i \sim \text{AL}(p, \sigma^2), \quad i = 1, \ldots, n,$$

where $\theta_{pi}$ and $\sigma^2$ are unknown parameters, $p$ is a fixed quantile level, and $\text{AL}(p, \sigma^2)$ denotes the asymmetric Laplace distribution with the density function

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

where $p$ is a fixed constant which characterizes the quantile level, $\sigma^2$ (not $\sigma$) is a scale parameter, and $\rho_p(\cdot)$ is a check loss function defined by (3). However, we often handle multiple observations per grid point in practice. Hereafter, we considered the following model which accounts for multiple observations per grid point:

$$y_{ij} = \theta_p(x_i) + \epsilon_{ij},$$
\[ \varepsilon_{ij} \sim \text{AL}(p, \sigma^2), \quad i = 1, \ldots, n, \quad j = 1, \ldots, N_i, \]  
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 (4) (see also Heng et al. 2022). Note that the model (5) 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 dropped the subscript \( p \) from \( \theta \) for the remainder of the paper.

We then introduce shrinkage priors on differences. We define the \( (k + 1) \)th order difference operator \( D \) as

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

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 that is defined by

\[
D_n^{(1)} = \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & -1 & 0 \end{pmatrix} \in \mathbb{R}^{(n-1) \times n}, \quad D_n^{(k+1)} = D_n^{(1)} D_n^{(k)}.
\]

We consider flexible shrinkage priors on \( D\theta \), and the priors are represented by

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

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

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

Note that since \( (D\theta)_i = \theta_i \sim N(0, w_i^2) \) for \( i = 1, \ldots, k+1 \), \( w_i \) is independent of \( \theta_i \). For this reason, we assumed the conjugate inverse gamma distribution \( IG(a_{w_i}, b_{w_i}) \) for \( w_i^2 \). For \( i = k+2, \ldots, n \), we considered two types of distribution; \( w_i \sim \text{Exp}(1/2) \) and \( w_i \sim C^+(0, 1) \). These priors 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 \( \sigma^2 \sim IG(a_{\sigma}, b_{\sigma}) \) and \( \tau \sim C^+(0, C_{\tau}) \). The default choice of hyperparameters is \( a_{\sigma} = b_{\sigma} = 0.1 \) and \( C_{\tau} = 1 \).

**Remark 1** We follow Tibshirani (2014) for discussion about an extension to the proposed method for the situation where data is observed at an irregular grid. It is equal so that the

locations of data \( x = (x_1, \ldots, x_n) \) have the ordering \( x_1 < x_2 < \cdots < x_n \) and \( d_j = x_{j+1} - x_j \) is not constant. This issue is related to nonparametric quantile regression. When the locations \( x \in \mathbb{R}^n \) are irregular and strictly increasing, Tibshirani (2014) proposed an adjusted difference operator for \( k \geq 1 \)

\[
D_n^{(k+1)} = D_n^{(1)} \text{ diag}(\frac{k}{x_{k+1} - x_1}, \ldots, \frac{k}{x_n - x_{n-k}}) D_n^{(k)},
\]

where \( D_n^{(1)} = D_n^{(1)} \) and when \( x_1 = 1, x_2 = 2, \ldots, x_n = n, D_n^{(k+1)} \) is equal to \( D_n^{(k+1)} \) (see also Heng et al. 2022). The matrix \( D \) is given by

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

where \( O \) is a zero matrix.

### 2.3 Gibbs sampler

We first derived a Gibbs sampler by using 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 stochastic expression:

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

where \( u_{ij} \sim N(0, 1) \) and \( z_{ij} \sim \text{Exp}(1/\sigma^2) \) 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

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

Under the prior (6), the full conditional distributions of \( \theta \) and \( z_{ij} \) are given by

\[
\theta \mid y, z, \sigma^2, \gamma^2 \sim N_n \left( A^{-1} B, \sigma^2 A^{-1} \right), \quad z_{ij} \mid y_{ij}, \theta_i, \sigma^2 \sim \text{GIG} \left( \frac{1}{2}, \frac{(y_{ij} - \theta_i)^2}{\tau^2 \sigma^2}, \frac{\psi^2}{\tau^2 \sigma^2} + \frac{2}{\sigma^2} \right),
\]

\( i = 1, \ldots, n,\; j = 1, \ldots, N_i \),

respectively, where

\[
A = D^\top W^{-1} D + \frac{1}{\tau^2} \text{ diag} \left( \sum_{j=1}^{N_i} z_{ij}^{-1}, \ldots, \sum_{j=1}^{N_i} z_{nj}^{-1} \right),
\]
The full conditional distribution of local shrinkage parameter $w_i$ is given by

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

and $\text{GIG}(\alpha, \beta, \gamma)$ denotes the generalized inverse Gaussian distribution. The full conditional distribution of the scale parameter $\sigma^2$ is given by

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

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

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

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

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

$$\xi | \tau^2 \sim \text{IG} \left( \frac{1}{2}, \frac{1}{\tau^2} + 1 \right),$$

where $\xi$ is an augmented parameter for $\tau^2$. For $i = 1, \ldots, k + 1$, we assumed the prior $\text{IG}(\alpha_{w_i}, \beta_{w_i})$ for $w_i$, and then the full conditional distribution of $w_i$ is given by

$$w_i^2 | \theta, \sigma^2, \gamma^2 \sim \text{GIG} \left( \frac{1}{2} + a_{w_i} \frac{\eta_i^2}{2\sigma^2} + b_{w_i} \right), \quad i = 1, \ldots, k + 1.$$

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

$$w_i^2 | \theta, \sigma^2, \gamma^2 \sim \text{GIG} \left( \frac{1}{2} \left( \frac{\eta_i^2}{\sigma^2} \right), \gamma^2 \right),$$

$$\gamma^2 | w, \nu \sim \text{GIG} \left( n - k - \frac{3}{2}, \frac{2}{\nu}, \sum_{i=k+2}^n w_i^2 \right),$$

$$\nu | \gamma^2 \sim \text{IG} \left( \frac{1}{2}, \frac{1}{\gamma^2} + 1 \right).$$

- **(Horseshoe-type prior)** The full conditional distributions of $\theta_i, z_i$, $\sigma^2$ and $\tau^2$ were already derived. For Horseshoe-type prior, we assume $w_i \sim C^+(0, 1)$ for $i = k + 2, \ldots, n$. By using the representation $w_i^2 | \nu_i \sim \text{IG}(1/2, 1/\nu_i)$ and $\nu_i \sim \text{IG}(1/2, 1)$, the full conditional distributions of $w_i$ and $\nu_i$ are, respectively, given by

$$w_i^2 | \theta, \sigma^2, \tau^2, \nu_i \sim \text{IG} \left( 1, \frac{1}{\nu_i} + \frac{\eta_i^2}{2\sigma^2 \tau^2} \right),$$

$$\nu_i | w_i \sim \text{IG} \left( \frac{1}{2}, \frac{1}{w_i^2} + 1 \right).$$

### 2.4 Variational Bayes approximation

The MCMC algorithm presented in Sect. 2.3 can be computationally intensive when the sample size is large. For the quick computation of the posterior distribution, we derived the 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 an intractable posterior distribution by using a simpler probability distribution. Note that 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 Sect. 2.3.

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

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

If $\theta$ 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

$$q(\theta_k) \propto \exp(E_{\theta_{-k}}[\log p(y, \theta)]),$$

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 approximation is...
known as the MFVB approximation. If the full conditional distribution of $\theta_i$ has a familiar form, the above formula is easy to compute. According to the Gibbs sampling algorithm in Sect. 2.3, 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

$$q(\theta) \sim N_n(A^{-1}B, (E_{1/\sigma^2}A)^{-1}), \quad q(z_{ij}) \sim \text{GIG} \left( \frac{1}{2}, \alpha_{z_{ij}}, \beta_{z_{ij}} \right),$$

$$q(\sigma^2) \sim \text{IG} \left( \frac{n + 3N}{2} + a_{\sigma}, \alpha_{\sigma} \right), \quad q(\tau^2) \sim \text{IG} \left( \frac{n - k}{2}, \alpha_{\tau} \right),$$

$$q(\xi) \sim \text{IG} \left( \frac{1}{2}, E_{1/\tau^2} + 1 \right).$$

(9)

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

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

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 $w_i^2$ ($i = k + 2, \ldots, n$), $\gamma^2$ and $\nu$ are given by

  $$q(w_i^2) \sim \text{GIG} \left( \frac{1}{2}, \alpha_{w_i^2}, E_{\gamma^2} \right),$$

  $$q(\gamma^2) \sim \text{GIG} \left( n - k - \frac{3}{2}, 2E_{1/\nu}, \sum_{i=k+2}^{n} E_{w_i^2} \right),$$

  $$q(\nu) \sim \text{IG} \left( \frac{1}{2}, E_{1/\nu^2} + 1 \right),$$

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

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

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 $i$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 $c$ in Algorithms 1 and 2.

### 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 $\theta_i$ is represented by the normal distribution $N(\mu_i, \Sigma_{ii})$, where the mean $\mu_i$ and variance $\Sigma_{ii}$ are defined in Sect. 2.4. 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 $\theta$ were given by (6) and (7), respectively. Here we add a common (non-random) scale parameter $\lambda$, and then replace (6) and (7) with

$$p(y_{ij} \mid \theta_i, z_{ij}, \sigma^2) = \frac{\exp \left( -\frac{(y_{ij} - \theta_i)^2}{2\sigma^2} \right)}{2\sigma^2},$$

$$p(\theta \mid \sigma^2, \tau, w) = \frac{\exp \left( -\frac{1}{2\lambda^2} \theta^T D^{-1} W^{-1} D \theta \right)}{2\lambda},$$

respectively. Based on these representations, the variational posterior of $\theta$ is given by

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

The constant $\lambda$ 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 $\lambda$ is given locally for each $\theta_i$ (i.e. $\lambda_i$), then the variational posterior of $\theta_i$ 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 $B$
Algorithm 1 — Variational Bayes approximation under Laplace prior.

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

1. Cycle the following:

(i) $A \leftarrow \sum_{j=1}^{N_i} E_{1/ (z_{ij})} + D^T \tilde{W}^{-1} D,$

$B \leftarrow \frac{1}{\nu} \left( C - \psi_{1 \nu} \right), \quad C \leftarrow \left( \sum_{j=1}^{N_i} y_{1j} E_{1/ (z_{ij})} \right) \top,$

$\tilde{W}^{-1} \leftarrow \text{diag} \left( E_{1/ w_{i1}^2}, \ldots, E_{1/ w_{i2}^2}, E_{1/ \tau^2} E_{1/ w_{i2}^2}, \ldots, E_{1/ \tau^2} E_{1/ w_{i2}^2} \right),$ 

$E_{\theta} \leftarrow (A^{-1} B), \quad E_{\psi} \leftarrow e_1 (E_{\sigma^2}^{-1} A^{-1} + A^{-1} B B^T A^{-1}) e_1,$

$E_{\nu} \leftarrow d_1 (E_{\sigma^2}^{-1} A^{-1} + A^{-1} B B^T A^{-1}) d_1 \quad (i = 1, \ldots, n),$

$\alpha_{a^2} \leftarrow \frac{1}{2 \nu^2} \sum_{i=1}^{N_i} \left( y_{ij}^2 E_{1/ (z_{ij})} - 2 \psi_{1} y_{ij} + \psi_{1}^2 E_{1/ (z_{ij})} - 2 (E_{1/ (z_{ij})} y_{ij} - \psi_{1}) E_{\theta} + E_{\psi} E_{1/ (z_{ij})} \right)$

$+ \frac{1}{2} \sum_{i=1}^{N_i} E_{\nu} - \frac{1}{2} \sum_{i=k+2}^{n} E_{\nu}^2 E_{1/ (z_{ij})} + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{N_i} E_{z_{ij}} + b_{\sigma},$

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

$E_{\alpha_{ij}} \leftarrow \frac{1}{2} \left( y_{ij}^2 - 2 \psi_{1} y_{ij} + E_{\theta} + E_{\psi} E_{1/ \tau^2} \right) E_{1/ \tau^2}, \quad \beta_{z_{ij}} \leftarrow \left( \frac{\psi_{1}^2}{\nu^2} + 2 \right) E_{1/ \tau^2},$

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

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

$E_{1/ w_{i1}^2} \leftarrow (1 + 2a_{w_{i1}})/(E_{\psi} E_{1/ \tau^2} + 2b_{w_{i1}}) \quad (i = 1, \ldots, n, k + 1)$

(ii) $\alpha_{w_{i1}^2} \leftarrow E_{1/ \sigma^2} E_{\nu}^2, \quad E_{w_{i1}^2} \leftarrow \sqrt{\alpha_{w_{i1}^2} K_{3/2}(\sqrt{a_{a^2} E_{\tau^2}})}$/

$\sqrt{E_{\tau^2} K_{1/2}(\sqrt{a_{w_{i1}^2} E_{\tau^2}})} \quad (i = k + 2, \ldots, n),$

$E_{w_{i2}^2} \leftarrow \sqrt{E_{\nu}^2 K_{3/2}(\sqrt{a_{a^2} E_{\tau^2}})} - \frac{1}{a_{w_{i2}^2}} \quad (i = k + 2, \ldots, n),$

$E_{\nu} \leftarrow \sqrt{2 E_{1/ \nu} K_{n-k-3/2}(\sqrt{2 E_{1/ \nu} \sum_{i=k+2}^{n} E_{w_{i2}^2}})}$

$\sqrt{\sum_{i=k+2}^{n} E_{w_{i2}^2} K_{n-k-3/2}(\sqrt{2 E_{1/ \nu} \sum_{i=k+2}^{n} E_{w_{i2}^2}})} \quad \left( E_{1/ \nu} \leftarrow \frac{1}{2 E_{1/ \nu} + 1} \right),$

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

After we obtain the optimal value of $A$ using Algorithm 3, we use $q(\theta) \sim N_A(\mu^A, \Lambda \Sigma_A^*)$ as the calibrated variational posterior distribution. We then construct the calibrated credible interval of $\theta_i$ ($i = 1, \ldots, n$) by calculating the quantile of the marginal distribution of $N_A(\mu^A, \Lambda \Sigma_A^*)$. Here we used the residual bootstrap method (e.g. Efron 1982) to obtain boot-
Remark 2: In Algorithm 3, we employed the residual bootstrap using 50% 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 \( \mu_p \) is \( p \)-th quantile trend estimate. However, since our aim is 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.

4 Simulation studies

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

4.1 Simulation setting

To compare the performance of the proposed methods, we considered the following data generating processes (see also Faulkner and Minin 2018; Brantley et al. 2020): We assumed that the data generating process was

\[ y_i = f(x_i) + \varepsilon(x_i), \quad i = 1, \ldots, 100, \]
where \( f(x) \) is a true function and \( \varepsilon(x) \) is a noise function. First, we considered the following two true functions:

- **Piecewise constant (PC)**

\[
f(x) = 2.5 \cdot I(1 \leq x \leq 20) + I(21 \leq x \leq 40) + 3.5 \cdot I(41 \leq x \leq 60) + 1.5 \cdot I(61 \leq x \leq 100)
\]

- **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 catch a constant trend and jumping structure. The second scenario (VS) is smooth and has a rapid change near \( x = 50 \). Hence, the scenario 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: \( \varepsilon(x) \sim N(0, \{(1 + x^2)/4\}^2) \).

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

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

For each scenario, simulated true quantile trends are summarized in Figure S1 of the Supplementary Materials. 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 \( \mathtt{R} \) 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 Sect. 2.3, 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

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

\[
\text{MCIW} = \frac{1}{n} \sum_{i=1}^{n} I(\hat{\theta}_{0.025,i} < \theta^*_i \leq \hat{\theta}_{0.975,i}), \quad \text{CP} = \frac{1}{n} \sum_{i=1}^{n} I(\hat{\theta}_{0.025,i} \leq \theta^*_i \leq \hat{\theta}_{0.975,i}),
\]

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

### 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 1 and 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 1 and 2, respectively. From Table 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 Fig. 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.

**Varying smoothness.** The results for the (VS) scenario are reported in Tables 3 and 4. From Table 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...
Table 1  Average values of MSE and MAD based on 100 replications for piecewise constant with $k = 0$

|        | MSE       | MAD       |
|--------|------------|-----------|
| (I) Gauss |           |           |
| MCMC-HS | 0.046     | 0.013     |
| VB-HS | 0.094     | 0.026     |
| MCMC-Lap | 0.105     | 0.040     |
| VB-Lap | 0.091     | 0.033     |
| ADMM | 0.384     | 0.045     |
| (II) Beta |           |           |
| MCMC-HS | 0.004     | 0.003     |
| VB-HS | 0.001     | 0.009     |
| MCMC-Lap | 0.013     | 0.015     |
| VB-Lap | 0.059     | 0.011     |
| ADMM | 0.793     | 0.011     |
| (III) Mixed normal |           |           |
| MCMC-HS | 0.089     | 0.029     |
| VB-HS | 0.247     | 0.067     |
| MCMC-Lap | 0.191     | 0.094     |
| VB-Lap | 0.137     | 0.069     |
| ADMM | 0.315     | 0.183     |

The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively.

Table 2  Average values of MCIW and CP based on 100 replications for piecewise constant with $k = 0$

|        | MCIW       | CP       |
|--------|------------|----------|
| (I) Gauss |           |           |
| MCMC-HS | 0.549     | 0.447     |
| VB-HS | 1.034     | 0.620     |
| MCMC-Lap | 0.117     | 0.205     |
| VB-Lap | 0.775     | 0.764     |
| ADMM | 0.139     | 0.822     |
| (II) Beta |           |           |
| MCMC-HS | 0.113     | 0.137     |
| VB-HS | 0.282     | 0.202     |
| MCMC-Lap | 0.034     | 0.069     |
| VB-Lap | 0.377     | 0.326     |
| ADMM | 0.871     | 0.730     |
| (III) Mixed normal |           |           |
| MCMC-HS | 1.536     | 0.870     |
| VB-HS | 1.129     | 1.135     |
| MCMC-Lap | 1.573     | 1.155     |
| VB-Lap | 0.548     | 0.778     |

The CP values above 90% are represented in bold.
One-shot simulation results under piecewise constant and Gauss noise. The order of trend filtering is $k = 0$ for all methods.

### Table 3

|                  | 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 |
| 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 |
| 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 |

The minimum values and second smallest values of MSE and MAD are represented in bold and italics respectively.
Table 4  Average values of MCIW and CP based on 100 replications for varying smoothness with $k = 1$

|         | MCIW     | CP       |         | MCIW     | CP       |
|---------|----------|----------|---------|----------|----------|
| (I) Gauss |          |          |         |          |          |
| MCMC-HS | 0.05     | 0.25     | 0.5     | 0.75     | 0.95     |
| CVB-HS  | 1.010    | 0.592    | 0.496   | 0.572    | 1.097    |
| VB-HS   | 0.124    | 0.196    | 0.209   | 0.195    | 0.118    |
| MCMC-Lap| 0.545    | 0.571    | 0.562   | 0.553    | 0.524    |
| CVB-Lap | 1.335    | 0.890    | 0.602   | 0.821    | 1.240    |
| VB-Lap  | 0.207    | 0.326    | 0.369   | 0.348    | 0.227    |
| (II) Beta |          |          |         |          |          |
| MCMC-HS | 0.162    | 0.178    | 0.213   | 0.241    | 0.260    |
| CVB-HS  | 0.445    | 0.259    | 0.209   | 0.347    | 0.724    |
| VB-HS   | 0.051    | 0.085    | 0.103   | 0.104    | 0.065    |
| MCMC-Lap| 0.251    | 0.248    | 0.282   | 0.306    | 0.304    |
| CVB-Lap | 0.522    | 0.362    | 0.268   | 0.458    | 0.673    |
| VB-Lap  | 0.108    | 0.164    | 0.199   | 0.201    | 0.136    |
| (III) Mixed normal |          |          |         |          |          |
| MCMC-HS | 0.600    | 0.636    | 0.654   | 0.655    | 0.666    |
| CVB-HS  | 1.271    | 0.854    | 0.747   | 0.850    | 1.540    |
| VB-HS   | 0.172    | 0.286    | 0.303   | 0.284    | 0.178    |
| MCMC-Lap| 0.727    | 0.755    | 0.765   | 0.768    | 0.755    |
| CVB-Lap | 1.780    | 1.214    | 0.891   | 1.249    | 1.972    |
| VB-Lap  | 0.279    | 0.400    | 0.458   | 0.455    | 0.324    |

The CP values above 90% are represented in bold.

![Fig. 2](image-url)  One-shot simulation results under varying smoothness and Gauss noise. The order of trend filtering is $k = 1$ for all methods.
Table 5  Average values of raw computing time and effective sample size per unit time based on 100 replications for all scenarios

| Scenario    | Computation time (second) | ESS (per second) |
|-------------|---------------------------|------------------|
| (PC) Piecewise constant |                          |                  |
| 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 32            | 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              | 1305 1564 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 |                          |                  |
| 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       | 13 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 |

methods had slightly worse coverage probabilities. In particular, the MCMC-HS and MCMC-Lap under the mixed normal noise, which is a relatively high degree of misspecification, appeared to be far from the nominal coverage rate of 0.95. Although the MCICW 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 Fig. 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 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.
5 Real data analysis

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 $(\alpha, \beta) = (1, 3)$ as hyperparameters in the inverse gamma prior to $\sigma^2$. The resulting estimates of quantiles and the corresponding 95% credible intervals are shown in Fig. 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 Nile data are

|                   | Nile data (Sect. 5.1) | Munich rent data (Sect. 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      |

![Fig. 3 Point estimates and 95% credible intervals for Nile data](image-url)
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 6). This is consistent with the simulation results in Sect. 4.2. In Table 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.

### 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. 2022). 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 \( x_j \), the response \( y_j \) has multiple observations per location, that is, \( y_j = (y_{j1}, \ldots, y_{jN_j})^\top \in \mathbb{R}^{N_j} \). 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 Sect. 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 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 gsreg 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 Fig. 4. The frequentist smoothing spline method is denoted by “Spline” in Fig. 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 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 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.

### 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 numerical 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 more accurate point and interval estimates. However, the estimation of low/high quantile tends to be unstable, and if the model

| Table 7 Effective sample size per unit time for real data examples | Nile data (Sect. 5.1) | Munich rent data (Sect. 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      6       5       | 4      7       7      6      5      |
| CVB-HS               | 47,619  52,632 47,619 55,556 48,780 | 4029  6233  1943  6124  2949      |
is misspecified, 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 drawbacks 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.

In our study, we focused on the use of asymmetric Laplace likelihood, but it would be possible to extend the framework to extended asymmetric Laplace (e.g. Yan and Kottas 2017). In the setting, it is possible to develop a similar computation algorithm to obtain posterior distribution and a valid credible interval, which will be left to future research. 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 paper was no longer applicable, therefore a detailed investigation extends the scope of this study.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11222-023-10327-y.
Acknowledgements The authors would like to thank the Associate Editor and a reviewer for their valuable comments and suggestions to improve the quality of this article. This work was supported by the JST, the establishment of university fellowships towards the creation of science technology innovation, grant number JPMJFS2129. This work was partially supported by the Japan Society for Promotion of Science (KAKENHI), grant numbers 21K13835 and 21H00699.

Author Contributions All authors wrote the main manuscript text.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

Balke, N.S.: Detecting level shifts in time series. J. Bus. Econ. Stat. 11(1), 81–92 (1993)
Blei, D.M., Kucukelbir, A., McAuliffe, J.D.: Variational inference: a review for statisticians. J. Am. Stat. Assoc. 112(518), 859–877 (2017)
Brantley, H.L., Guinness, J., Chi, E.C.: Baseline drift estimation for air quality data using quantile trend filtering. Ann. Appl. Stat. 14(2), 585–604 (2020)
Carvalho, C.M., Polson, N.G., Scott, J.G.: The horseshoe estimator for sparse signals. Biometrika 97(2), 465–480 (2010)
Cobb, G.W.: The problem of the Nile: conditional solution to a change-point problem. Biometrika 65(2), 243–251 (1978)
Efron, B.: The Jackknife, the Bootstrap and Other Resampling Plans. SIAM (1982)
Eilers, P.H., De Menezes, R.X.: Quantile smoothing of array CGH data. Bioinformatics 21(7), 1146–1153 (2005)
Faulkner, J.R., Minin, V.N.: Locally adaptive smoothing with Markov random fields and shrinkage priors. Bayesian Anal. 13(1), 225 (2018)
Faulkner, J.R., Magee, A.F., Shapiro, B., et al.: Horseshoe-based Bayesian nonparametric estimation of effective population size trajectories. Biometrics 76(3), 677–690 (2020)
Heng, Q., Zhou, H., Chi, E.C.: Bayesian trend filtering via proximal Markov chain Monte Carlo (2022). arXiv preprint arXiv:2201.00092
Kim, S.J., Koh, K., Boyd, S., et al.: ℓ1 trend filtering. SIAM Rev. 51(2), 339–360 (2009)
Kowal, D.R., Matteson, D.S., Ruppert, D.: Dynamic shrinkage processes. J. Roy. Stat. Soc. Ser. B (Stat. Methodol.) 81(4), 781–804 (2019)
Kozumi, H., Kobayashi, G.: Gibbs sampling methods for Bayesian quantile regression. J. Stat. Comput. Simul. 81(11), 1565–1578 (2011)
Makalic, E., Schmidt, D.F.: A simple sampler for the horseshoe estimator. IEEE Signal Process. Lett. 23(1), 179–182 (2015)
Nychka, D., Furrer, R., Paige, J., et al.: Fields: tools for spatial data. R package version 9.6.0 (2017)
Nychka, D., Gray, G., Haaland, P., et al.: A nonparametric regression approach to syringe grading for quality improvement. J. Am. Stat. Assoc. 90(432), 1171–1178 (1995)
Oh, H.S., Nychka, D., Brown, T., et al.: Period analysis of variable stars by robust smoothing. J. Roy. Stat. Soc.: Ser. C (Appl. Stat.) 53(1), 15–30 (2004)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (eg. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.