Empirical priors and posterior concentration in a piecewise polynomial sequence model

Chang Liu[1], Ryan Martin[1] and Weining Shen[2]

[1]*Department of Statistics, North Carolina State University*

[2]*Department of Statistics, University of California, Irvine*

Supplementary Material

This file is organized as follows. We provide technical proofs in Section S1. Section S2 gives a detailed description of the simulation settings in the main paper. Section S3 gives additional numerical results for the main paper, and Section S4 discusses the posterior sampling algorithm.

S1 Proofs

S1.1 Preliminary results

Before getting to the proofs of the main theorems, we first present some preliminary lemmas that helps to construct the proofs. First, we define

$$A_n = \{\theta \in \Theta_n(K) : \|\theta - \theta^\star\|_n^2 > M_n \varepsilon_n(\theta^\star)\}.$$

For our theoretical analysis, it will help to rewrite the posterior distribution $\Pi^n(A_n)$ as the ratio $\Pi^n(A_n) = N_n(A_n)/D_n$. The numerator and denominator are

$$N_n(A_n) = \sum_B \pi_n(B) \int_{A_n \cap \Theta^B} R_n(\theta^B)^\alpha \, \Pi_n(d\theta^B \mid B),$$

$$D_n = \sum_B \pi_n(B) \int_{\Theta^B} R_n(\theta^B)^\alpha \, \Pi_n(d\theta^B \mid B),$$

where $R_n(\theta) = L_n(\theta)/L_n(\theta^\star)$ is the likelihood ratio, Θ_B is the column space of Z^B, i.e., Θ_B consists of all n-dimensional vectors θ^B such that $\theta^B = Z^B\beta^B$. For a properly chosen matrix Z, $N_n(A_n)$ and D_n can be rewritten in terms of β^B, given B,

$$N_n(A_n) = \sum_B \pi_n(B) \int_{\{\beta^B \in \mathbb{R}^{|B|(K+1)}: Z^B\beta^B \in A_n\}} R_n(Z^B\beta^B)^\alpha \, \tilde{\pi}_n(\beta^B \mid B) \, d\beta^B, \tag{S1.1}$$

$$D_n = \sum_B \pi_n(B) \int_{\mathbb{R}^{|B|(K+1)}} R_n(Z^B\beta^B)^\alpha \, \tilde{\pi}_n(\beta^B \mid B) \, d\beta^B. \tag{S1.2}$$

For the given Z and $B_{\theta^\star}$, we let $\beta^\star$ be such that $\theta^\star = Z^{B_{\theta^\star}}\beta^\star$ and let $\beta_s^\star$ be its s^{th} $(K+1)$-dimensional component. We also abbreviate $B_{\theta^\star}$ by $B^\star$ and $\varepsilon_n^2(\theta^\star)$ by ε_n^2. As discussed above, $B^\star$ may not be unique, but certain features of $B^\star$ are determined, in particular, its size $|B^\star|$ which, in turn, determines other features like $\pi_n(B^\star)$, etc.

Lemma 1. There exists a constant $c = c(\alpha, \sigma^2, K)$ such that $D_n \gtrsim \pi_n(B^\star)e^{-c|B^\star|}$ for all sufficiently large n.

Proof. Given that D_n is a sum of non-negative terms, it is straightforward to have

$$D_n > \pi_n(B^\star) \int_{\mathbb{R}^{|B^\star|(K+1)}} R_n(Z^{B^\star}\beta^{B^\star})^\alpha \, \tilde{\pi}_n(\beta^{B^\star} \mid B^\star) \, d\beta^{B^\star}.$$

The integral in the right-hand side of the above inequality can be further written as,

$$\prod_{s=1}^{|B^\star|} \int e^{-\frac{\alpha}{2\sigma^2}\{\|Y_{B^\star(s)} - Z_{B^\star(s)}\beta_s^{B^\star}\|^2 - \|Y_{B^\star(s)} - Z_{B^\star(s)}\beta_s^\star\|^2\}} \mathsf{N}(\beta_s^{B^\star} \mid \hat{\beta}_s^{B^\star}, v(Z_{B^\star(s)}^\top Z_{B^\star(s)})^{-1}) \, d\beta_s^{B^\star}.$$

Direct calculation shows that the above quantity equals

$$e^{\frac{\alpha}{2\sigma^2}\|Z_{B^\star(s)}(\hat{\beta}_s^{B^\star} - \beta_s^\star)\|^2} \big(1 + \tfrac{\alpha v}{\sigma^2}\big)^{-\frac{(K+1)|B^\star|}{2}} \geq \big(1 + \tfrac{\alpha v}{\sigma^2}\big)^{-\frac{(K+1)|B^\star|}{2}}.$$

Therefore, $D_n > \pi_n(B^\star) e^{-c|B^\star|}$, where $c = \frac{(K+1)}{2}\log(1+\frac{\alpha v}{\sigma^2}) > 0$. □

Lemma 2. Take $q > 1$ such that $\alpha q < 1$. Then $\mathsf{E}_{\theta^*}\{N_n(A_n)\} \lesssim e^{-M_n r n \varepsilon_n^2}$, for all large n, where $r = \alpha(1-q\alpha)/2\sigma^2$.

Proof. Towards an upper bound, we interchange expectation with the finite sum over B and the integral over β^B, the latter step justified by Tonelli's theorem, so that

$$\mathsf{E}_{\theta^\star}\{N_n(A_n)\} = \sum_B \pi_n(B) \int_{\{\beta^B : Z^B\beta^B \in A_n\}} \mathsf{E}_{\theta^\star}\{R_n(Z^B\beta^B)^\alpha\, \tilde{\pi}_n(\beta^B \mid B)\}\, d\beta^B. \tag{S1.3}$$

Next, we work with each of the B-dependent integrands separately. For $q > 1$ such that $\alpha q < 1$, define the Hölder conjugate $p = q/(q-1) > 1$. Then Hölder's inequality gives

$$\mathsf{E}_{\theta^\star}\{R_n(Z^B\beta^B)^\alpha\, \tilde{\pi}_n(\beta^B \mid B)\} \le \mathsf{E}_{\theta^\star}^{1/q}\{R_n(Z^B\beta^B)^{\alpha q}\}\, \mathsf{E}_{\theta^\star}^{1/p}\{\tilde{\pi}_n(\beta^B \mid B)^p\}.$$

On the set A_n, since $\alpha q < 1$, the first term above is uniformly bounded by $e^{-M_n r n \varepsilon_n^2}$. To see this, note that, for a general $\theta \in A_n$, if p_θ^n denotes the joint density of Y under Eq (1.1) in the main paper, and $D_{\alpha q}$ the Rényi αq-divergence of one normal distribution from another (e.g., Van Erven and Harremos, 2014, p. 3800), then

$$\mathsf{E}_{\theta^\star}\{R_n(Z^B\beta^B)^{\alpha q}\} = \int \{p_{\theta^B}^n(y)\}^{\alpha q}\, \{p_{\theta^\star}^n(y)\}^{1-\alpha q}\, dy = e^{-\frac{\alpha q(1-\alpha q)n}{2\sigma^2}\|\theta^B - \theta^\star\|_n^2}.$$

Then for the second term in the upper bound above, using prior in Eq (2.7) of the main paper, we can have

$$\mathsf{E}_{\theta^\star}^{1/p}\{\tilde{\pi}_n(\beta^B \mid B)^p\} = \prod_{s=1}^{|B|} \mathsf{E}_{\theta^*}^{1/p}\{\mathsf{N}^p(\beta_s^B \mid \hat{\beta}_s^B, v(Z_{B(s)}^\top Z_{B(s)})^{-1})\}, \tag{S1.4}$$

which is equivalent to

$$\prod_{s=1}^{|B|} \frac{|Z_{B(s)}^\top Z_{B(s)}|^{\frac{1}{2}}}{(2\pi v)^{\frac{K+1}{2}}}\, \mathsf{E}_{\theta^\star}^{1/p}(e^{-\frac{p\sigma^2}{2v}V_s}), \tag{S1.5}$$

where

$$V_s = \sigma^{-2}\|P_{B(s)}(Z_{B(s)}\beta_s^B - Y_{B(s)})\|^2,$$

is distributed as a non-central chi-square with $(K+1)$ degrees of freedom and non-centrality parameter

$$\lambda_s = \sigma^{-2}\|P_{B(s)}(Z_{B(s)}\beta_s^B - \theta^\star_{B(s)})\|^2,$$

with $P_{B(s)} = Z_{B(s)}(Z_{B(s)}^\top Z_{B(s)})^{-1}Z_{B(s)}^\top$. Using the familiar formula for the moment generating function of non-central chi-square, we have

$$\mathsf{E}_{\theta^*}^{1/p}(e^{-\frac{p\sigma^2}{2v}V_s}) = (1+p\sigma^2/v)^{-(K+1)/(2p)}e^{-\frac{1}{2(v/\sigma^2+p)}\lambda_s} \tag{S1.6}$$

For the integral

$$\int_{\mathbb{R}^{|B|(K+1)}} \mathsf{E}_{\theta^\star}^{1/p}\{\tilde{\pi}_n(\beta^B \mid B)^p\}\, d\beta^B,$$

if we plug (S1.4), (S1.5), and (S1.6) into the integrand, then it simplfies as

$$\prod_{s=1}^{|B|} \frac{|Z_{B(s)}^\top Z_{B(s)}|^{\frac{1}{2}}}{(2\pi v)^{\frac{K}{2}}}(1+\tfrac{p\sigma^2}{v})^{-\frac{K}{2p}}$$
$$\times \int_{\mathbb{R}^{K+1}} \mathsf{N}(\beta_s^B \mid (Z_{B(s)}^\top Z_{B(s)})^{-1}Z_{B(s)}^\top\theta^\star_{B(s)}, (\tfrac{v}{\sigma^2}+p)(Z_{B(s)}^\top Z_{B(s)})^{-1})\, d\beta_s^B.$$

By direct calculation, this can be written as $\zeta^{|B|}$, where $\zeta = \big\{\frac{(1+p\sigma^2/v)^{1/q}}{\sigma^2}\big\}^{(K+1)/2}$. Therefore, we have

$$\mathsf{E}_{\theta^\star}\{N_n(A_n)\} \le e^{-M_n rn\varepsilon_n^2}\sum_B \zeta^{|B|}\pi_n(B) = e^{-M_n rn\varepsilon_n^2}\sum_{b=1}^n \zeta^b f_n(b).$$

Given that $f_n(b) \propto n^{-\lambda(b-1)}$ and ζ is a positive constant, the summation term in the above upper bound is uniformly bounded in n, proving the claim. □

S1.2 Proof of Theorem 1

By Lemma 1, for sufficiently large n, we have

$$\mathsf{E}_{\theta^\star}\{\Pi^n(A_n)\} \le \frac{e^{c|B^\star|}}{\pi_n(B^\star)}\mathsf{E}_{\theta^\star}\{N_n(A_n)\}.$$

Plug in the bound from Lemma 2 to get

$$\mathsf{E}_{\theta^\star}\{\Pi^n(A_n)\} \lesssim e^{c|B^\star| - M_n r n\varepsilon_n^2} \frac{\binom{n-1}{|B^\star|-1}}{f_n(B^\star)}.$$

On the one hand, if $|B_{\theta^\star}| = 1$, then both $n\varepsilon_n^2$ and the ratio in the above display are constant. Therefore, the upper bound is $\lesssim e^{-GM_n} \to 0$ for a constant G and $M_n \to \infty$. On the other hand, if $|B_{\theta^\star}| \geq 2$, then $n\varepsilon_n^2$ is diverging, and we take $M_n \equiv M$ a fixed constant. Also, using the formula for $f_n(|B^\star|)$ in Eq (2.5) of the main paper and the standard bound, $\binom{n}{b} \leq e^{b\log(en/b)}$, on the binomial coefficient, we get

$$\mathsf{E}_{\theta^\star}\{\Pi^n(A_n)\} \lesssim \exp\{-Mrn\varepsilon_n^2 + n\varepsilon_n^2 + \lambda|B^\star|\log n + c|B^\star|\}. \tag{S1.7}$$

The exponent on the right-hand side can be rewritten as

$$-n\varepsilon_n^2\Big(Mr - 1 - \frac{\lambda|B^\star|\log n}{n\varepsilon_n^2} - \frac{c|B^\star|}{n\varepsilon_n^2}\Big).$$

Since $|B^\star| = o(n)$, if we let $n\varepsilon_n^2 = |B^\star|\log n$, the two ratios inside the parentheses in the above display are upper bounded by a constant for sufficiently large n. Therefore, for a sufficiently large M, there exists a constant $G > 0$, depending on the constants M, r, λ, and $c = c(\alpha, \sigma^2, K)$, such that the right-hand side of (S1.7) can be written as $e^{-Gn\varepsilon_n^2} \to 0$.

S1.3 Proof of Theorem 2

It follows from Jensen's inequality that $\|\hat{\theta} - \theta^\star\|_n^2 \leq \int \|\theta - \theta^\star\|_n^2 \, \Pi^n(d\theta)$. So it suffices to bound the expectation of the upper bound. Towards this, write $\mathbb{R}^n$ as $A \cup A^c$, where $A = A_n$ is as defined above. Then

$$\mathsf{E}_{\theta^\star} \int \|\theta - \theta^\star\|_n^2 \, \Pi^n(d\theta) \leq M_n\varepsilon_n^2 + \mathsf{E}_{\theta^\star} \int_A \|\theta - \theta^\star\|_n^2 \, \Pi^n(d\theta). \tag{S1.8}$$

That remaining integral can be expressed as a ratio of numerator to denominator, where the denominator D_n is just as in Lemma 1 and the numerator $\widetilde{N}_n(A)$ is

$$\begin{aligned}\widetilde{N}_n(A) &= \int_A \|\theta - \theta^\star\|_n^2 \, R_n(\theta)^\alpha \, \Pi_n(d\theta) \\ &= \sum_B \pi_n(B) \int_{A \cap \Theta_B} \|\theta^B - \theta^\star\|_n^2 \, R_n(\theta^B)^\alpha \, \pi_n(d\theta^B \mid B).\end{aligned}$$

Take expectation of the numerator to the inside of the integral and apply Hölder's inequality just like in the proof of Lemma 2. This gives the following upper bound on each B-specific integral:

$$\int_{A \cap \Theta_B} \|\theta - \theta^\star\|_n^2 e^{-h\|\theta - \theta^\star\|^2} \mathsf{E}_{\theta^\star}^{1/p}\{\pi_n(d\theta^B \mid B)^p\},$$

where $h > 0$ is a constant that depends only on α, σ^2, and the Hölder constant $q > 1$. Since the function $x \mapsto xe^{-hx}$ is eventually monotone decreasing, for sufficiently large M we get a trivial upper bound on the above display, i.e.,

$$M_n \varepsilon_n^2 e^{-M_n h n \varepsilon_n^2} \int_{\Theta_B} \mathsf{E}_{\theta^\star}^{1/p}\{\pi_n(d\theta^B \mid B)^p\}.$$

The same argument as above bounds the remaining integral by $\zeta^{|B|}$, and the prior $\pi_n(B)$ takes care of that contribution. In the case where $|B^\star| = 1$, where $n\varepsilon_n^2$ is bounded, choosing $M_n \to \infty$ will take care of the bound on D_n from Lemma 1. Similarly, for cases when $n\varepsilon_n^2 \to \infty$, we can use a sufficiently large constant $M_n \equiv M$ to take care of the lower bound on D_n. Therefore, the second term on the right-hand side of (S1.8) is also bounded by a multiple of $M_n \varepsilon_n^2$.

S1.4 Proof of Theorem 3

Following arguments similar to those in the proof of Theorem 1, we get

$$\mathsf{E}_{\theta^\star} \pi_n(\{B : |B| > C|B^\star|\}) \lesssim \frac{e^{c|B^\star|}}{\pi_n(B^\star)} \sum_{b=C|B^\star|+1}^{n} \zeta^b f_n(b).$$

From the formula (2.5) for f_n in the main paper, factor out a common $(\zeta n^{-\lambda})^{C|B^\star|}$ from the summation, which will be the dominant term. Indeed, like in the proof of Theorem 1, the ratio in the above display is of order $\exp\{n\varepsilon_n^2 + \lambda|B^\star|\log n\}$. Then the right-hand side above is of order

$$\exp\{n\varepsilon_n^2 + \lambda|B^\star|\log n + C|B^\star|\log\zeta - C\lambda|B^\star|\log n\}.$$

The exponent can be written as

$$-|B^\star|\log n\Big(C\lambda - \lambda - \frac{C\log\zeta}{\log n} - \frac{n\varepsilon_n^2}{|B^\star|\log n}\Big).$$

Since $n\varepsilon_n^2 \le |B^\star|\log n$, it is clear that, if C is strictly larger than $1+\lambda^{-1}$, then the term in parentheses is bigger than some constant $G > 0$ for all large n.

S1.5 Proof of Theorem 4

Choose and fix any $B^\star \in \mathbb{B}^\star$. For a generic configuration B, we have

$$\begin{aligned}
\pi^n(B) &\le \pi^n(B)/\pi^n(B^\star)\\
&= \tfrac{\pi_n(B)}{\pi_n(B^\star)}(1+\tfrac{v\alpha}{\sigma^2})^{(K+1)(|B^\star|-|B|)/2} e^{-\frac{\alpha}{2\sigma^2}\{\sum_{s=1}^{|B|}\|(I-P_{B(s)})Y_{B(s)}\|^2 - \sum_{s=1}^{|B^\star|}\|(I-P_{B^\star(s)})Y_{B^\star(s)}\|^2\}}\\
&= \tfrac{\pi_n(B)}{\pi_n(B^\star)}(1+\tfrac{v\alpha}{\sigma^2})^{(K+1)(|B^\star|-|B|)/2} e^{-\frac{\alpha}{2\sigma^2}\{\|(I_n-P^B)Y\|^2 - \|(I_n-P^{B^\star})Y\|^2\}} \qquad \text{(S1.9)}
\end{aligned}$$

where $P^B = Z^B(Z^{B\top}Z^B)^{-1}Z^{B\top}$, $P^{B^\star} = Z^{B^\star}(Z^{B^\star\top}Z^{B^\star})^{-1}Z^{B^\star\top}$ and Z^B is defined in Eq (2.3) in the main paper. If B is a refinement of $B^\star$, then column space of $Z^{B^\star}$ is a subset of the column space of Z^B, i.e., $\mathcal{C}(Z^{B^\star}) \subseteq \mathcal{C}(Z^B)$ and therefore, $P^B - P^{B^\star}$ is idempotent of rank $(K+1)(|B|-|B^\star|) > 0$. Thus, (S1.9) can be rewritten as,

$$\pi^n(B) \le \frac{\pi_n(B)}{\pi_n(B^\star)}\Big(1+\frac{v\alpha}{\sigma^2}\Big)^{(K+1)(|B|-|B^\star|)/2} e^{\frac{\alpha}{2\sigma^2}V}, \qquad \text{(S1.10)}$$

where $V = Y^\top(P^B - P^{B^\star})Y$ and V/σ^2 is distributed as a central chi-square with $(K+1)(|B| - |B^\star|)$ degrees of freedom. From the chi-square moment generating function and $\alpha < 1$, we get

$$\mathsf{E}_{\theta^\star}\{\pi^n(B)\} \leq \frac{\pi_n(B)}{\pi_n(B^\star)}\psi^{|B|-|B^\star|},$$

where $\psi = \psi(\alpha, v, \sigma, K)$ is a positive constant. For a suitable constant C as in Theorem 3, write $\mathcal{B}_n = \{B : B \sqsupset B^\star,\, |B| \leq C|B^\star|\}$. Then

$$\{B : B \sqsupset B^\star\} \subseteq \mathcal{B}_n \cup \{B : |B| > C|B^\star|\}.$$

Since the right-most event has vanishing Π^n-probability by Theorem 3, it follows that we can focus just on the event $\mathcal{B}_n$, and

$$\mathsf{E}_{\theta^\star}\pi^n(\mathcal{B}_n) = \sum_{B\in\mathcal{B}_n}\mathsf{E}_{\theta^\star}\pi^n(B) \leq \sum_{B\in\mathcal{B}_n}\frac{\pi_n(B)}{\pi_n(B^\star)}\psi^{|B|-|B^\star|}.$$

Plug in the prior for B—which only depends on $|B|$—and simplify:

$$\mathsf{E}_{\theta^\star}\{\Pi^n(\theta : B_\theta \in \mathcal{B}_n)\} \leq \sum_{b=b^\star+1}^{Cb^\star}\frac{\binom{n-1}{b^\star-1}\binom{n-b^\star}{b-b^\star}}{\binom{n-1}{b-1}}(\psi n^{-\lambda})^{b-b^\star},$$

where $b^\star = |B^\star|$. From

$$\frac{\binom{n-1}{b^\star-1}\binom{n-b^\star}{b-b^\star}}{\binom{n-1}{b-1}} = \binom{b-1}{b^\star-1} \leq b^{b-b^\star},$$

and the assumption that $b^\star = o(n^\lambda)$, we get that the summation on the right-hand side above is upper-bounded by

$$\sum_{b=b^\star+1}^{Cb^\star}(\psi b n^{-\lambda})^{b-b^\star} \leq \sum_{b=b^\star+1}^{Cb^\star}(C\psi b^\star n^{-\lambda})^{b-b^\star} \lesssim e^{-\lambda b^\star \log n}, \quad \text{for all large } n.$$

This argument can be duplicated for any $B^\star \in \mathbb{B}^\star$, and there are at most $O(|B^\star|)$ many equivalent block configurations, so we get

$$\mathsf{E}_{\theta^\star}\pi^n(\{B : B \sqsupset B^\star \text{ for some } B^\star \in \mathbb{B}^\star\}) \lesssim |B^\star|e^{-\lambda|B^\star|\log n} \to 0.$$

S1.6 Proof of Theorem 5

Choose and fix any $B^\star \in \mathbb{B}^\star$. In light of Theorem 4, it suffices to show that

$$\mathsf{E}_{\theta^\star} \pi^n(\{B : B \not\supseteq B^\star\}) \to 0.$$

For a generic $B \not\supseteq B^\star$, according to (S1.9) we have

$$\pi^n(B) \leq \frac{\pi_n(B)}{\pi_n(B^\star)} \Big(1 + \frac{v\alpha}{\sigma^2}\Big)^{(K+1)(|B|-|B^\star|)/2} e^{\frac{\alpha}{2\sigma^2} Y^\top (P^B - P^{B^\star})Y}.$$

We proceed with a proof for the piecewise constant ($K = 0$) case first, then describe how the general K case is the same. Let θ be a piecewise constant signal corresponding to the block configuration B. Then we can rewrite θ as $X\eta$, where X is an $n \times n$ lower triangular matrix with unit entries and

$$\eta = (\theta_1, \theta_2 - \theta_1, \theta_3 - \theta_2, \dots, \theta_n - \theta_{n-1})^\top.$$

It is easy to show that η is sparse. Let $J = \{1\} \cup \{j : \eta_j \neq 0\}$, then $|J| = |B|$. Let η_J be the $|J|$-vector containing the particular entries with their indices in J and X_J be the columns of X corresponding to J. Then we can also write $\theta = X_J \eta_J$. Hence we can reformulate model (1.1) in the main paper as

$$Y = X\eta^\star + \sigma\xi, \quad \xi \sim \mathsf{N}(0, I). \tag{S1.11}$$

Under this formulation, recovering block structure $B^\star$ is equivalent to identifying the non-zero coefficients in $\eta^\star$, i.e., recovering $J^\star$. One basic observation is that P^B is equal to $P_J = X_J(X_J^\top X_J)^{-1} X_J^{-1}$. Then we can rewrite $Y^\top(P^B - P^{B^\star})Y$ as,

$$-\|(I - P_J)X\eta^\star\|^2 - 2\sigma\xi^\top(I - P_J)X\eta^\star + \sigma^2\xi^\top(P_J - P_{J^\star})\xi.$$

In addition, because $(P_{J^\star} - P_{J\cap J^\star})$ is positive definite, the right-most quadratic form above can be bounded as follows,

$$\xi^\top(P_J - P_{J^\star})\xi = \xi^\top(P_J - P_{J\cap J^\star})\xi - \xi^\top(P_{J^\star} - P_{J\cap J^\star})\xi \leq \xi^\top(P_J - P_{J\cap J^\star})\xi.$$

Therefore, $Y^\top(P^B - P^{B^\star})Y$ can be bounded above by,

$$-\|(I-P_J)X\eta^\star\|^2 - 2\sigma\xi^\top(I-P_J)X\eta^\star + \xi^\top(P_J - P_{J\cap J^\star})\xi.$$

Note that the second and third terms in the above upper bound follow normal and chi-square distributions respectively, and additionally $(I-P_J)(P_J - P_{J\cap J^\star}) = 0$ implies independence. Hence, using normal and chi-square moment generating functions we can have,

$$\mathsf{E}_{\theta^\star}[e^{\frac{\alpha}{2\sigma^2}Y^\top(P^B-P^{B^\star})Y}] \leq (1-\alpha)^{-\frac{1}{2}(|J^\star|-|J\cap J^\star|)} e^{-\frac{\alpha(1-\alpha)}{2\sigma^2}\|(I-P_J)X\eta^\star\|^2}.$$

Since

$$\|(I-P_J)X\eta^\star\|^2 = \|(I-P_J)X_{J^\star\cap J^c}\eta^\star_{J^\star\cap J^c}\|,$$

if we let

$$\delta_n = \min_{j\in J^\star\cap\{j>1\}} |\theta^\star_j - \theta^\star_{j-1}|,$$

then

$$\begin{aligned}\|(I-P_J)X\eta^\star\|^2 &\geq \lambda_{\min}\big(X^\top_{J^\star\cap J^c}X_{J^\star\cap J^c}\big)\|\eta^\star_{J^\star\cap J^c}\|^2 \\ &\geq \lambda_{\min}\big(X^\top_{J^\star}X_{J^\star}\big)\delta_n^2(|J^\star| - |J^\star\cap J|).\end{aligned}$$

Then we let

$$\gamma_n = \min_{j,j'\in J^\star, j\neq j'} |j-j'|,$$

according to Lemma 3 below, $\|(I-P_J)X\eta^\star\|^2$ can be further lower bounded by,

$$\gamma_n\delta_n^2(|J^\star| - |J^\star\cap J|)/4.$$

Therefore, if we let

$$\gamma_n\delta_n^2 \geq \frac{4M\sigma^2}{\alpha(1-\alpha)}\log n,$$

then

$$\mathsf{E}_{\theta^\star}\{\pi^n(B)\} \leq \frac{\pi_n(B)}{\pi_n(B^\star)}\phi^{|B^\star|-|B|}(\omega n^{-M})^{|J^\star|-|J\cap J^\star|}, \tag{S1.12}$$

where $\omega = (1-\alpha)^{-1/2}$ and $\phi = (1+\frac{v\alpha}{\sigma^2})^{-1/2}$. Plug in the expressions for $\pi_n(B)$ and $\pi_n(B^\star)$ from Eq (2.5) and then sum over all B such that $B \not\sqsupseteq B^\star$ to get

$$\mathsf{E}_{\theta^\star}\pi^n(\{B : B \not\sqsupseteq B^\star\})\}) \le \sum_{b=1}^{Cb^\star}\sum_{t=1}^{b\wedge b^\star} \frac{\binom{b^\star}{t}\binom{n-b^\star}{b-t}\binom{n}{b^\star}}{\binom{n}{b}} (\phi n^\lambda)^{b^\star-b}(\omega n^{-M})^{b^\star-t},$$

where $b^\star = |B^\star|$ and the first sum is restricted to $b \le Cb^\star$ by Theorem 3. The ratio of binomial coefficients can be bounded as

$$\frac{\binom{b^\star}{t}\binom{n-b^\star}{b-t}\binom{n}{b^\star}}{\binom{n}{b}} = \binom{b}{t}\binom{n-b}{b^\star-t} \le (n^3)^{b\vee b^\star-t}.$$

Plug in this bound and split the sum over b into two cases: $b \le b^\star$ and $b > b^\star$. For the first case, we have

$$\sum_{b=1}^{b^\star}\sum_{t=1}^{b}(\phi n^\lambda)^{b^\star-b}(n^3)^{b^\star-t}(\omega n^{-M})^{b^\star-t} \lesssim \sum_{b=1}^{b^\star}(\phi n^{3+\lambda-M})^{b^\star-b},$$

and the right-hand side vanishes since $M > 3+\lambda$. Similarly, for the second case

$$\sum_{b=b^\star+1}^{Cb^\star}\sum_{t=1}^{b^\star}(\phi n^\lambda)^{b^\star-b}(n^3)^{b^\star-t}(\omega n^{-M})^{b^\star-t} \lesssim \omega n^{3-M}\sum_{b=b^\star+1}^{Cb^\star}(n^{3-\lambda})^{b-b^\star},$$

and if $\lambda \ge 3$, then the sum is dominated by term n^{3-M}. In either case, the upper bound vanishes—actually the upper bound is $O(n^{-1})$ because $M > 4+\lambda$—which proves the claim for the particular $B^\star \in \mathbb{B}^\star$. The above argument is not specific to any $B^\star$, so if we repeat the above argument sum over all such $B^\star \in \mathbb{B}^\star$, then we get

$$\sum_{B^\star\in\mathbb{B}^\star}\mathsf{E}_{\theta^\star}\pi^n(\{B : B \not\sqsupseteq B^\star\}) \lesssim |B^\star| n^{-1} \to 0.$$

Finally, since

$$1-\mathsf{E}_{\theta^\star}\pi^n(\mathbb{B}^\star) = \sum_{B^\star\in\mathbb{B}^\star}\mathsf{E}_{\theta^\star}\pi^n(\{B : B \sqsupset B^\star\}) + \sum_{B^\star\in\mathbb{B}^\star}\mathsf{E}_{\theta^\star}\pi^n(\{B : B \not\sqsupseteq B^\star\}),$$

and first term on the right-hand side vanishes by Theorem 4 and the second term vanishes by the argument above, we conclude that

$$\mathsf{E}_{\theta^\star}\pi^n(\mathbb{B}^\star) \to 1, \quad n \to \infty.$$

Next, we show that for general piecewise polynomial $K \geq 1$, $\mathsf{E}_{\theta^\star}[\Pi^n(\theta : B_\theta \not\supseteq B^\star)]$ can be bounded in a similar fashion. We define

$$X^{(K)} = \begin{pmatrix} I_K & \\ & L_{n-K} \end{pmatrix}, \quad K = 1, \ldots, n-1,$$

where L_{n-K} is an $(n-K) \times (n-K)$-dimensional lower triangular matrix with unit entries.

Now, let's consider a generic degree-K piecewise polynomial signal θ with underlying block configuration B, then θ can be written as $X\eta$, where $X = L_n X^{(1)} \cdots X^{(K)}$ and,

$$\eta = \big((\Delta^0\theta)_1, (\Delta^1\theta)_1, \ldots, (\Delta^K\theta)_1, \Delta^{K+1}\theta\big),$$

with Δ^K being the K^{th}-order difference operator defined in Section 3.3 of the main paper. Note that η is also sparse here, and if we let $J = \{1, \ldots, K+1\} \cup \{j : \eta_j \neq 0\}$, then $|J| = (K+1)|B|$. Therefore, a similar result to (S1.12) can be obtained,

$$\pi(B) \leq \frac{\pi_n(B)}{\pi_n(B^\star)} \phi^{(K+1)(|B|-|B^\star|)/2} (\omega n^{-M})^{|J^\star|-|J\cap J^\star|}.$$

Then based on Lemma 3, using recursion, rest of the proofs can follow similar arguments in the $K = 0$ case with

$$\delta_n = \min_{j \in J^\star \cap \{j > K+1\}} |\eta_j|,$$

and

$$\gamma_n \delta_n^2 \geq \frac{4^{K+1} M \sigma^2}{\alpha(1-\alpha)} \log n.$$

S1.7 An eigenvalue bound

Lemma 3. Consider $J = \{j_1, \ldots, j_s\} \subset \{1, \ldots, n\}$, let X be an $n \times n$-dimensional lower triangular matrix with unit entries and X_S be the $n \times s$-dimensional sub-matrix of X with the columns corresponding to J, define

$$\gamma = \min_{1 \leq \ell < k \leq s} |j_\ell - j_k|,$$

then the smallest eigenvalue of $X_S^\top X_S$ satisfies

$$\lambda_{\min}(X_S^\top X_S) > \gamma/4.$$

Proof. Without loss of generality, here and throughout we assume that $j_1 < \cdots < j_s$. It is straightforward to observe that,

$$X_S^\top X_S = \begin{pmatrix} a_1 & a_2 & \vdots & a_s \\ a_2 & a_2 & \vdots & \vdots \\ \cdots & \cdots & . & \vdots \\ a_s & \cdots & \cdots & a_s \end{pmatrix},$$

with $a_i = n + 1 - j_i$, $i = 1, \ldots, s$. According to Lemma 3 in Qian and Jia (2016), the inverse of $X_S^\top X_S$ is tridiagonal, i.e.,

$$(X_S^\top X_S)^{-1} = \begin{pmatrix} r_{11} & r_{12} & & & & \\ r_{21} & r_{22} & r_{23} & & & \\ & r_{32} & r_{33} & r_{34} & & \\ & & \ddots & \ddots & \ddots & \\ & & & r_{s-1,s-2} & r_{s-1,s-1} & r_{s-1,s} \\ & & & & r_{s,s-1} & r_{s,s} \end{pmatrix}$$

where

$$r_{ij} = \begin{cases} \frac{1}{a_1 - a_2} & i = j = 1 \\ -\frac{1}{a_{j-1} - a_j} & i = j - 1 \\ -\frac{1}{a_j - a_{j+1}} & i = j + 1 \\ \frac{a_{j-1} - a_{j+1}}{(a_{j-1} - a_j)(a_j - a_{j+1})} & 1 < i = j < s \\ \frac{a_{s-1}}{(a_{s-1} - a_s)(a_s)} & i = j = s \\ 0 & \text{otherwise.} \end{cases}$$

For any vector $u \in \mathbb{R}^{s\times 1}$,

$$\begin{aligned} u^\top (X_S^\top X_S)^{-1} u &= \sum_{i=1}^{s} r_{ii} u_i^2 + \sum_{i=1}^{s-1} r_{i,i+1} u_i u_{i+1} + \sum_{i=1}^{s-1} r_{i+1,i} u_i u_{i+1} \\ &\le \max_i |r_{ii}| \sum_{i=1}^{s} u_i^2 + (\max_i |r_{i,i+1}| + \max_i |r_{i+1,i}|) \sum_{i=1}^{s-1} |u_i u_{i+1}| \end{aligned}$$

Because $\sum_{i=1}^{s-1} |u_i u_{i+1}| \le \frac{1}{2} \sum_{i=1}^{s-1} (u_i^2 + u_{i+1}^2) \le \sum_{i=1}^{s} u_i^2$,

$$\begin{aligned} u^\top (X_S^\top X_S)^{-1} u &\le \max_i |r_{ii}| + \max_i |r_{i,i+1}| + \max_i |r_{i+1,i}|) \sum_{i=1}^{s} u_i^2 \\ &\le \frac{4}{\gamma} \sum_{i=1}^{s} u_i^2. \end{aligned}$$

Thus, $\lambda_{\max}\{(X_S^\top X_S^\top)^{-1}\} \le 4/\gamma$, and therefore, $\lambda_{\min}(X_S^\top X_S^\top) \ge \gamma/4$. □

S2 Simulation setting

We describe six different models for the true signal $\theta^\star$ in the simulation study. The underlying truth and the simulated data are depicted in Figure S1.

Model 1. *Piecewise constant model:*

$$\theta_i^\star = f_1(i), \quad \textit{for} \quad i = 1, \dots, n, \quad n = 497,$$

where f_1 is a piecewise constant function with $|B^\star| = 7$ as in Frick et al. (2014), p. 561; see, also, Fryzlewicz (2014), Appendix B(2). See Figure S1(a). The data is generated by

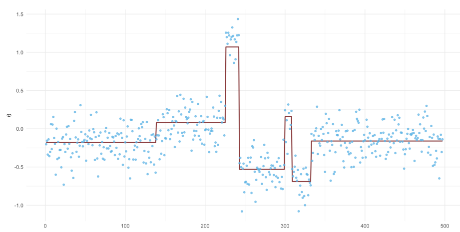
$$Y_i \overset{ind}{\sim} \mathsf{N}(\theta_i^\star, 0.04), \quad i = 1, \dots, n, \quad n = 497.$$

Model 2. *Piecewise constant model:*

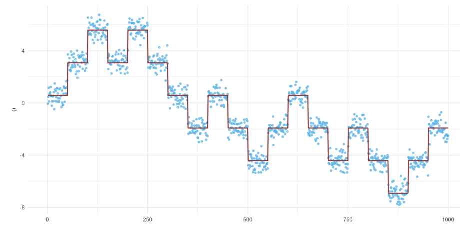
$$\theta_i^\star = f_2(i), \quad \textit{for} \quad i = 1, \dots, n, \quad n = 1000,$$

where f_2 is a piecewise constant function with equal block lengths and jump sizes. Here, $\delta(\theta^\star) = 2.5$, $\gamma(\theta^\star) = 50$ *and* $|B^\star| = 20$. *See Figure S1(b). Then the data is generated by*

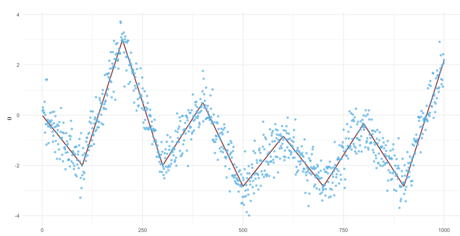
$$Y_i \overset{ind}{\sim} \mathsf{N}(\theta_i^\star, 0.25), \quad i = 1, \dots, n, \quad n = 1000.$$



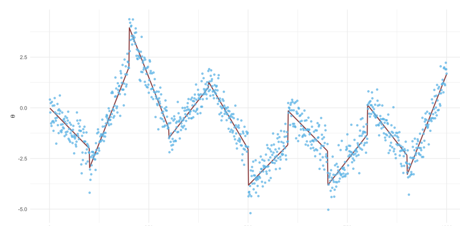
(a) Model 1



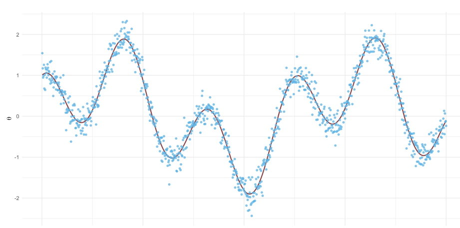
(b) Model 2



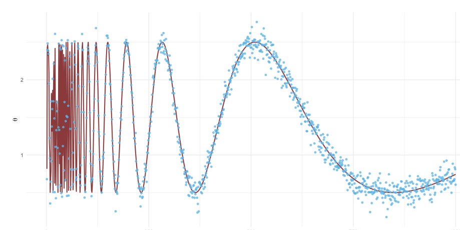
(c) Model 3



(d) Model 4



(e) Model 5



(f) Model 6

Figure S1: Plots of the signal and a representative data set from each of the six models.

Model 3. *Piecewise linear model with continuous mean:*

$$\theta_i^\star = f_3(i), \quad \textit{for} \quad i = 1, \ldots, n, \quad n = 1000,$$

where f_3 *is a piecewise linear function with continuous means (signal only with a change in its slope);* $|B^\star| = 10$*; see Figure S1(c). Then the data is generated by*

$$Y_i \overset{ind}{\sim} \mathsf{N}(\theta_i^\star, 0.25), \quad i = 1, \ldots, n, \quad n = 1000.$$

Model 4. *Piecewise linear model with jumps in mean:*

$$\theta_i^\star = f_4(i), \quad \textit{for} \quad i = 1, \ldots, n, \quad n = 1000,$$

where f_4 *is a piecewise linear function with signal both changing in its slope and intercept;* $|B^\star| = 10$*; see Figure S1(d). Then the data is generated by*

$$Y_i \overset{ind}{\sim} \mathsf{N}(\theta_i^\star, 0.25), \quad i = 1, \ldots, n, \quad n = 1000.$$

Model 5. *Trigonometric wave:*

$$\theta_i^\star = \sin(i/100) + \cos(i/33), \quad \textit{for} \quad i = 1, \ldots, n, \quad n = 1000,$$

see Figure S1(e). Then the data is generated by

$$Y_i \overset{ind}{\sim} \mathsf{N}(\theta_i^\star, 0.04), \quad i = 1, \ldots, n, \quad n = 1000.$$

Model 6. *Doppler wave as in Tibshirani (2014), p. 298:*

$$\theta_i^\star = \sin(4n/i) + 1.5, \quad \textit{for} \quad i = 1, \ldots, n, \quad n = 1000,$$

see Figure S1(f). Then the data is generated by

$$Y_i \overset{ind}{\sim} \mathsf{N}(\theta_i^\star, 0.04), \quad i = 1, \ldots, n, \quad n = 1000.$$

S3 Additional numerical results

We provide additional numerical results for simulation and data analysis sections.

Method	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Empirical Bayes	1.22	21.96	14.79	20.07	3.61	10.64
$\lambda = 1.0$	(0.07)	(1.01)	(0.47)	(0.54)	(0.09)	(0.25)
Empirical Bayes	0.82	9.60	10.79	**15.30**	2.85	8.04
$\lambda = 0.50$	(0.05)	(0.67)	(0.33)	(0.41)	(0.09)	(0.15)
Empirical Bayes	**0.70**	**8.09**	13.06	**15.44**	2.84	**7.35**
$\lambda = 0.20$	(0.03)	(0.52)	(0.44)	(0.54)	(0.07)	(0.14)
Trend Filtering	1.52	22.94	**6.85**	29.99	**1.26**	45.22
(cross-validation)	(0.04)	(0.55)	(0.18)	(0.29)	(0.04)	(0.70)

Table S1: Squared error loss between $\hat{\theta}$ and $\theta^\star$ across 100 replications. Tuning parameter of trend filtering is selected by 5-fold cross-validation.

Method	$\mathsf{P}(\hat{B} = B^\star)$	$\mathsf{P}(\hat{B} \supset B^\star)$	Hamming	Hausdorff	$\mathsf{E}\lvert\hat{B}\rvert$
Empirical Bayes $\lambda = 1.0$	0.03	0.03	4.14 (0.19)	3.86 (0.40)	**7.00** (0.00)
Empirical Bayes $\lambda = 0.50$	0.11	0.11	3.13 (0.18)	2.73 (0.31)	**7.00** (0.00)
Empirical Bayes $\lambda = 0.20$	**0.19**	0.19	**2.39** (0.15)	**2.57** (0.41)	**7.01** (0.01)
Trend Filtering (one std error)	0.05	**0.28**	3.27 (0.17)	17.99 (3.53)	8.77 (0.13)

Table S2: Structure recovery results for Model 1 ($|B^\star| = 7$) across 100 replications. The tuning parameter of trend filtering is chosen by one-standard-error rule.

S4 Posterior sampling algorithm

The posterior sampling can be achieved by the following steps.

1. At iteration t, given current block partition $B^{(t)}$, sample $B' \sim q(\cdot \mid B^{(t)})$.

2. Sample $U \sim \mathsf{Unif}(0, 1)$, let $B^{(t+1)} = B'$, if

$$U \le \min\Big\{1, \frac{\pi^n(B')\, q(B^{(t)} \mid B')}{\pi^n(B^{(t)})\, q(B' \mid B^{(t)})}\Big\};$$

   otherwise, let $B^{(t+1)} = B^{(t)}$.

3. Given $B^{(t+1)}$, obtain $\beta^{B^{(t+1)}} = (\beta_1^{B^{(t+1)}}, \ldots, \beta_{|B|}^{B^{(t+1)}})^\top$ via sampling

$$\beta_s^{B^{(t+1)}} \sim \mathsf{N}_{K+1}(\hat{\beta}_s^{B^{(t+1)}}, \tfrac{\sigma^2 v}{\sigma^2 + \alpha v}\{Z_{B^{(t+1)}(s)}^\top Z_{B^{(t+1)}(s)}\}^{-1}),$$

   and then set $\theta^{B^{(t+1)}} = Z^{B^{(t+1)}} \beta^{B^{(t+1)}}$.

Method	$\mathsf{P}(\hat{B} = B^\star)$	$\mathsf{P}(\hat{B} \supset B^\star)$	Hamming	Hausdorff	$\mathsf{E}\lvert\hat{B}\rvert$
Empirical Bayes $\lambda = 1.0$	0.02	0.02	5.80 (0.44)	2.20 (0.56)	**20.00** (0.00)
Empirical Bayes $\lambda = 0.50$	**0.52**	0.54	**1.32** (0.24)	**0.58** (0.12)	20.04 (0.03)
Empirical Bayes $\lambda = 0.20$	0.14	0.74	2.26 (0.27)	13.78 (1.38)	21.58 (0.19)
Trend Filtering (one std error)	0.00	**0.96**	25.41 (0.63)	20.76 (0.58)	45.36 (0.63)

Table S3: Structure recovery results for Model 2 ($|B^\star| = 20$) across 100 replications. The tuning parameter of trend filtering is chosen by one-standard-error rule.

Repeating this process M times and discarding the first m burn-in iterations, we obtain a sample of $(B^{(m+1)}, \theta^{B^{(m+1)}})$, $\ldots$, $(B^{(M)}, \theta^{B^{(M)}})$ from the joint posterior $\pi^n(B, \theta^B)$. Then posterior mean of θ can be approximated by $(M-m)^{-1}\sum_{i=m+1}^{M}\theta^{B^{(i)}}$. Credible sets for any real-valued function $g(\theta)$ of θ can be obtained by obtaining quantiles of the samples $g(\theta^{B^{(i)}})$, $i = m+1, \ldots, M$. For block configuration recovery, the maximum a posteriori (MAP) estimator for B can be readily found by evaluating π^n, up to the normalizing constant, using formula (2.12), for each Monte Carlo sample $B^{(i)}$ and returning the maximizer. For simplicity, we use a symmetric proposal distribution $q(B' \mid B)$ for the above algorithm, i.e., in each iteration, there is 0.5 probability for a “jump location” to vanish and 0.5 probability for a non-jump location to become a “jump”.

Bibliography

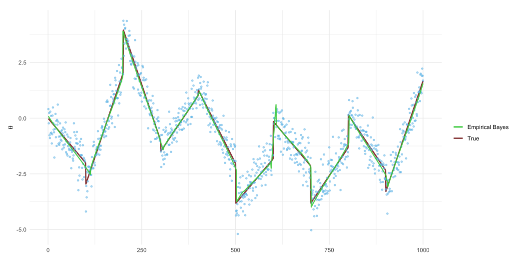
Frick, K., A. Munk, and H. Sieling (2014). Multiscale change point inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(3), 495–580.

Fryzlewicz, P. (2014). Wild binary segmentation for multiple change-point detection. *The Annals of Statistics 42*(6), 2243–2281.

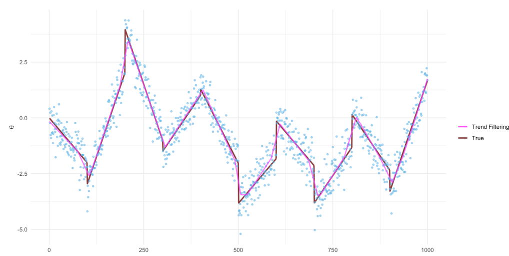
Qian, J. and J. Jia (2016). On stepwise pattern recovery of the fused lasso. *Computational Statistics & Data Analysis 94*, 221–237.

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

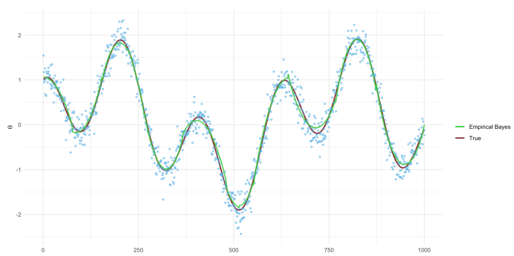
Van Erven, T. and P. Harremos (2014). Rényi divergence and Kullback-Leibler divergence. *IEEE Transactions on Information Theory 60*(7), 3797–3820.



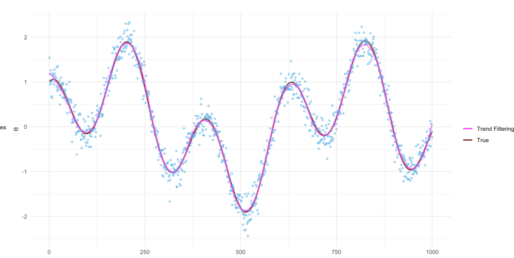
(a) Model 4: Empirical Bayes



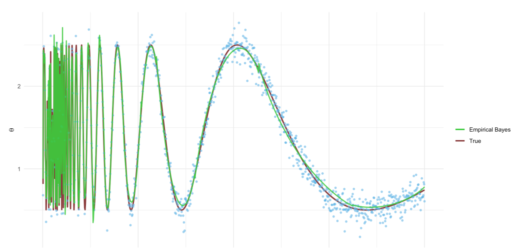
(b) Model 4: Trend filtering



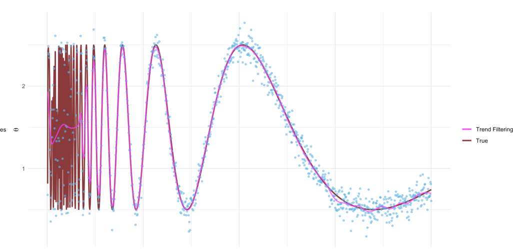
(c) Model 5: Empirical Bayes



(d) Model 5: Trend filtering



(e) Model 6: Empirical Bayes



(f) Model 6: Trend filtering

Figure S2: Plots of the empirical Bayes and trend filtering estimates of the signal for representative cased under Models 4–6.

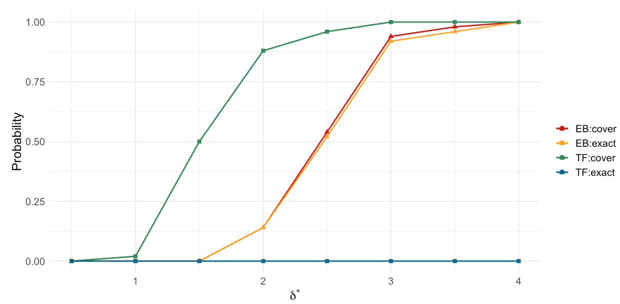
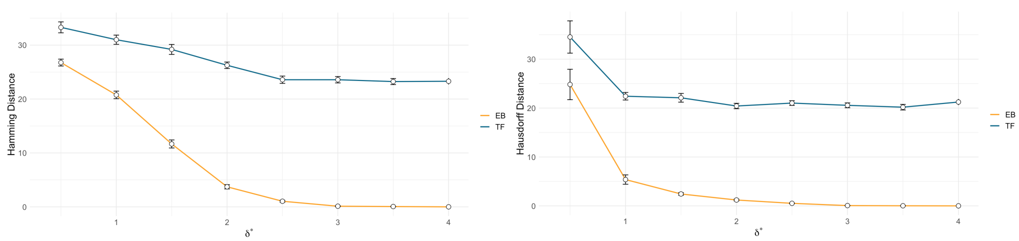


Figure S3: Plots of $\delta^\star \mapsto \mathsf{P}(\hat{B} = B^\star)$ (exact) and $\delta^\star \mapsto \mathsf{P}(\hat{B} \supset B^\star)$ (cover) in Model 2, with $\delta^\star$ ranging over $[0.5, 4.0]$, all the other conditions fixed, for the emprical Bayes (EB) and trend filtering (TF) methods.



(a) Plot of $\delta^\star$ vs. Hamming Distance

(b) Plot of $\delta^\star$ vs. Hausdorff Distance

Figure S4: Results of Model 2 with $\delta^\star$ ranging over $[0.5, 4.0]$, with all the other conditions fixed, for the empirical Bayes (EB) and trend filtering (TF) methods.

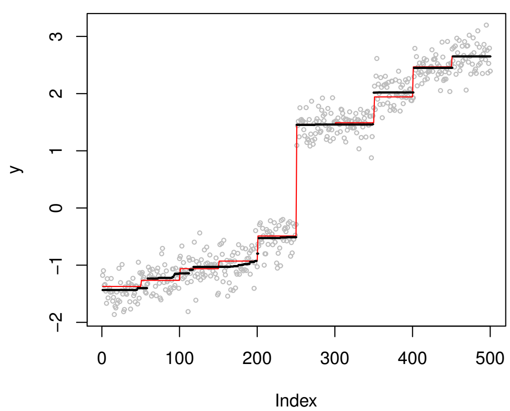


Figure S5: Simulated data (gray) from a distribution with a monotone piecewise constant mean (red) and the projection posterior mean (black).