Bayesian Spectral Analysis Regression with Shape Restrictions

Peter J. Lenk and Taeryon Choi

University of Michigan and Korea University

Supplementary Material

The online appendix provides technical details. Section 1 gives the equations for the integrated, mean–centered, cosine basis. Section 2 describes the MCMC algorithms. Section 3 proves the theorems. Section 4 provides additional information about the simulation studies, and includes simulations for BSARS and Spike–and–Slab priors.

S1 Integrated, Mean–centered Cosine Basis

We use the cosine basis functions on $[0,1]$: $\varphi_0(x) = 1$ and $\varphi_j(x) = \sqrt{2}\cos(\pi j x)$ for $j \geq 0$. The cosine basis functions are integrated and mean–centered in Section 3.1. Define:

$$\varphi^a_{j,k}(x) = \int_0^x \varphi_j(s)\varphi_k(s)ds - \int_0^1 \int_0^s \varphi_j(t)\bar{\varphi}_k(t)dt\, ds \text{ for } j,k \geq 0.$$

The integrals of the cosine basis functions for BSARM are:

$$\begin{aligned}
\varphi^a_{0,0}(x) &= x - 0.5 \\
\varphi^a_{0,j}(x) &= \varphi^a_{j,0}(x) = \frac{\sqrt{2}}{\pi j}\sin(\pi j x) - \frac{\sqrt{2}}{(\pi j)^2}\left[1 - \cos(\pi j)\right] \text{ for } j \geq 1, \\
\varphi^a_{j,j}(x) &= \frac{\sin(2\pi j x)}{2\pi j} + x - 0.5 \text{ for } j \geq 1, \\
\varphi^a_{j,k}(x) &= \frac{\sin\left[\pi(j+k)x\right]}{\pi(j+k)} + \frac{\sin\left[\pi(j-k)x\right]}{\pi(j-k)} \\
&\quad - \frac{1-\cos[\pi(j+k)]}{[\pi(j+k)]^2} - \frac{1-\cos[\pi(j-k)]}{[\pi(j-k)]^2}, \quad \text{for } j \neq k \text{ and } j,k \geq 1.
\end{aligned}$$

Define:

$$\varphi^b_{j,k}(x) = \int_0^x \int_0^s \varphi_j(t)\varphi_k(t)dt\, ds - \int_0^1 \int_0^x \int_0^s \varphi_j(t)\varphi_k(t)dt\, ds\, dx.$$

For the cosine basis the $\varphi^b_{j,k}$ for BSARMC are:

$$\begin{aligned}
\varphi^b_{0,0}(x) &= \frac{3x^2-1}{6} \\
\varphi^b_{0,j}(x) &= \varphi^b_{j,0}(x) = -\frac{\sqrt{2}}{(\pi j)^2}\cos(\pi j x) \text{ for } j \geq 1 \\
\varphi^b_{j,j}(x) &= -\frac{\cos(2\pi j x)}{(2\pi j)^2} + \frac{3x^2-1}{6} \text{ for } j \geq 1 \\
\varphi^b_{j,k}(x) &= -\frac{\cos[\pi(j+k)x]}{[\pi(j+k)]^2} - \frac{\cos[\pi(j-k)x]}{[\pi(j-k)]^2}, \quad \text{for } j \neq k \text{ and } j,k \geq 1.
\end{aligned}$$

S2 Supplemental Material: MCMC

S2.1 MCMC for BSARM and BSARMC

This supplementary material summarizes the MCMC algorithm for the monotone and concave/convex models in Section 3.1 of the paper. The differences among the models are the integrated basis functions φ^a, φ^b, and φ^c and the linear term α for convex or concave models. In the appendix, we will use ϕ generically for φ^a, φ^b, and φ^c and $\boldsymbol{\Phi}_J(x)$ for the matrices of integrated basis functions $\boldsymbol{\Phi}^a_J(x)$, $\boldsymbol{\Phi}^b_J(x)$, and $\boldsymbol{\Phi}^c_J(x)$. The supplementary material displays the algorithm for one f. Multiple f's merely add a set of loops.

We use the notation:

$$\begin{aligned}
\boldsymbol{y} &= (y_1,\ldots,y_n)^\intercal \text{ and } \boldsymbol{W} = (\boldsymbol{w}_1,\ldots,\boldsymbol{w}_n) \\
\boldsymbol{x} &= (x_1,\ldots,x_n)^\intercal \text{ and } \boldsymbol{\epsilon} = (\epsilon_1,\ldots,\epsilon_n)^\intercal \\
\boldsymbol{\theta}_J &= (\theta_0,\theta_1,\ldots,\theta_J)^\intercal \text{ and } \boldsymbol{\theta}_{J-0} = (\theta_1,\ldots,\theta_J)^\intercal \\
f_J(x) &= \delta\boldsymbol{\theta}_J^\intercal\boldsymbol{\Phi}_J(x) + \alpha(x-0.5) \text{ and } \boldsymbol{f}_J = (f_J(x_1),\ldots,f_J(x_n))^\intercal \\
\boldsymbol{\Gamma} &= J \times J \text{ diagonal matrix with } \exp(-j\gamma) \text{ as } (j,j) \text{ element.}
\end{aligned}$$

Note that $\boldsymbol{\theta}_{J-0}$ is $\boldsymbol{\theta}_J$ without θ_0. The combined models can be written as:

$$\boldsymbol{Y} = \boldsymbol{W}\boldsymbol{\beta} + \boldsymbol{f}_J + \boldsymbol{\epsilon}$$

We use the following notation for the MCMC algorithm. We use superscripts (m) and $*$ to distinguish current and candidate values for the focal parameters in the full conditional distributions. Parameters that are conditional in the full conditional distributions will not have superscripts (m) or $*$. These parameters are set to their current values on iteration m. We also refer to these conditional parameters as "Rest." After initializing the parameters, the MCMC algorithm has the following steps.

Generate $\boldsymbol{\theta}_J$

We generate $\boldsymbol{\theta}_J$ by random walk Metropolis where the random walk variance is proportional to the model variance of $\boldsymbol{\theta}_J$. We found that the performance of the algorithm depends on the proportionality factor η, and good values of η depend on the true f and the number of observations. The random walk Metropolis algorithm with fixed η is standard. We could let the user pick η by trial and error but that plan puts a large burden on the user. Instead, we propose an adaptive Metropolis (Haario et al. (2001)) method to make the algorithm more robust and less reliant on the user specification of η. We assume that η follows a random walk from an inverse gamma distribution, and the proposal distribution for η adapts to past values of η. We use a Metropolis step to accept or reject candidate values of η along with those of $\boldsymbol{\theta}_J$. If we integrate over η, we are essentially generating $\boldsymbol{\theta}_J$ from a random walk with multivariate T error terms, and the random walk variance adapts to previous values.

At iteration m of the MCMC, the current value of $\boldsymbol{\theta}_J$ and η are $\boldsymbol{\theta}_J^{(m)}$ and $\eta^{(m)}$. At iteration $m+1$, we draw a candidate value η^* from an inverse gamma distribution, $IG(a_T, b_{T,m})$. The shape parameter a_T is constant over iterations, and it is larger than 2 so that the variance of η exists. In the empirical examples, we use $a_T = 3$. The scale parameter $b_{T,m}$ adapts to previous draws of η. The mean of η at iteration m is: $E(\eta_m) = b_{T,m}/(a_T - 1)$. Given the mean, the scale parameter is: $b_{T,m} = (a_T - 1)E(\eta_m)$. We specify the mean as a convex combination with equal weights of the current value of η and a rolling estimate of the mean $\bar{\eta}_m$ based on previous draws: $E(\eta_m) = 0.5\eta^{(m)} + 0.5\bar{\eta}_m$. In empirical studies, we found the convex combination works better than $E(\eta_m) = \eta^{(m)}$ or $E(\eta_m) = \bar{\eta}_m$. We initialize η_0 and $\bar{\eta}_0$ to small constants, say 0.01. We then run the MCMC with fixed $\bar{\eta}_0$ for B_0 iterations. In the examples, $B_0 = 1000$. After B_0 iterations, we allow $\bar{\eta}_m$ and $b_{T,m}$ to adapt to previous draws of η.

At iteration $m+1$ we generate candidates θ_j* and η^*:

$$\begin{aligned}
\eta^* &\sim IG(a_T, b_{m,T}) \\
\theta_0^* &\sim N(\theta_0^{(m)}, 5.66\eta^*\sigma v_{\theta_0}^2) I(\theta_0 \geq 0) \\
\theta_j^* &\sim N(\theta_j^{(m)}, 5.66\eta^*\sigma\tau^2 \exp(-j\gamma)) \text{ for } j \geq 1.
\end{aligned}$$

The constant 5.66 was recommended in Haario et al. (2001). We use the inverse CDF transform to generate draws from univariate, truncated distributions.

The jump probability is given by:

$$\begin{aligned}
\log \alpha(\boldsymbol{\theta}_J^*, \boldsymbol{\theta}_J^{(m)}) &= \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\beta} - \boldsymbol{f}_J^{(m)})^\intercal(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\beta} - \boldsymbol{f}_J^{(m)}) \\
&- \frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\beta} - \boldsymbol{f}_J^{*})^\intercal(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\beta} - \boldsymbol{f}_J^{*}) \\
&+ \frac{(\theta_0^{(m)2} - \theta_0^{*2})}{2\sigma v_{\theta_0}^2} + \frac{1}{2\sigma\tau^2}\left(\boldsymbol{\theta}_{J-0}^{(m)\intercal}\Gamma^{-1}\boldsymbol{\theta}_{J-0}^{(m)} - \boldsymbol{\theta}_{J-0}^{*\intercal}\Gamma^{-1}\boldsymbol{\theta}_{J-0}^{*}\right) \\
&+ \ln\left[1 - \text{ZCDF}\left(-\frac{\theta_0^{(m)}}{\sqrt{\sigma}V_0}\right)\right] - \ln\left[1 - \text{ZCDF}\left(-\frac{\theta_0^{(*)}}{\sqrt{\sigma}V_0}\right)\right] \\
&- (a_T - 1)\ln[\eta^{(m)}] - b_{T,*}/\eta^{(m)} + (a_T - 1)\ln[\eta^*] + b_{T,m}/\eta^*
\end{aligned}$$

where ZCDF is the standard normal cumulative distribution function, and $b_T^* = (a_T - 1)(0.5\eta^* + 0.5\bar{\eta}_m)$. If the candidates are accepted, then we update the running mean of η:

$$\bar{\eta}_{m+1} = \bar{\eta}_m + (\eta_{m+1} - \bar{\eta}_m)/N$$

where N is the number of times that η^* is accepted after the initial B_0 iterations. If the candidates are rejected, then we do not update the running mean of η: $\bar{\eta}_{m+1} = \bar{\eta}_m$. Because the shape parameter a_T is fixed across iterations, adapting the inverse gamma distribution to the mean is equivalent to adapting it to the variance because the variance is proportional to the square of the mean.

As a general rule of thumb, one is better off starting $\bar{\eta}_0$ too small than too large. The Inverse Gamma distribution has a "dead zone" around 0 where its density is essentially zero. This dead zone is to the left of the mean. If $\bar{\eta}_0$ is much larger than the optimal value for η, then the optimal value will be in the dead zone, and the adaptive algorithm will have trouble marching towards η. Conversely, if the optimal value of η is outside the dead zone, then the adaptive algorithm quickly converges on it.

We also monitor the process to ensure that the acceptance rate has reasonable values. We run the MCMC B_0 iterations before adaptation, and B_0 iterations after adaptation where B_0. If the proportion of times that $\boldsymbol{\theta}_J$ is accepted in the adaptation period is between 0.3 and 0.6, then we use the current values of the parameters to start the actual MCMC. If the proportion is less than 0.3, the jumps in the random walk for η are too large, and we reduce the $\bar{\eta}_0$ by a factor of 10 and restart the procedure. If the proportion is greater than 0.6, then the jumps are too small, and we increase $\bar{\eta}_0$ by a factor of 10 and restart the procedure.

Generate τ^2

For the T Smoother, the prior distribution for τ^2 is inverse Gamma, which is also its full conditional distribution:

$$\begin{aligned}
\tau^2|\text{Rest} &\sim IG\left(\frac{r_{n,\tau}}{2}, \frac{s_{n,\tau}}{2}\right) \\
r_{n,\tau} &= r_{0,\tau} + J \text{ and } s_{n,\tau} = s_{0,\tau} + \frac{1}{\sigma}\boldsymbol{\theta}_{J-0}^{\intercal}\Gamma^{-1}\boldsymbol{\theta}_{J-0}.
\end{aligned}$$

For the Lasso Smoother, the prior distribution for τ^2 is Exponential with rate u_0. Its full conditional distribution is:

$$\begin{aligned} p(\tau^2|\text{ Rest}) &\sim \tau^{-J}\exp\left(-u_0\tau^2 - \frac{s}{2\tau^2}\right) \\ s &= \frac{1}{\sigma}\boldsymbol{\theta}_{J-0}^{\intercal}\Gamma^{-1}\boldsymbol{\theta}_{J-0} \end{aligned}$$

We use slice sampling (Damien et al. (1999) and Neal (2003)) to generate τ^2. Slice sampling introduces an auxiliary uniform random variable U into the full conditional distribution to break it into two factors:

$$p(U,\tau^2|\text{ Rest}) \propto I\left[0 \le U \le \exp\left(-u_0\tau^2\right)\right]\tau^{-J}\exp\left(-\frac{s}{2\tau^2}\right)$$

Integration of $p(U,\tau^2|\text{ Rest})$ over U gives $p(\tau^2|\text{ Rest})$. The right–hand factor is proportional to the inverse Gamma distribution for τ^2. Given $\tau^{(m)2}$ at iteration m, $U^{(m+1)}$ is generated from a unform distribution:

$$U^{(m+1)} \sim U\big(0, \exp\left[-u_0\tau^{(m)2}\right]\big).$$

Given $U^{(m+1)}$, the draw of τ^2 on iteration $m+1$ is constrained above by:

$$\tau^2 < -\log\left(U^{(m+1)}\right)/u_0.$$

Then τ^2 given $U^{(\mathbf{m+1})?}$ is generated from a truncated inverse Gamma distribution:

$$\tau^2 \sim IG\left(\frac{J-2}{2}, \frac{s}{2}\right) I\left[\tau^2 < -\log(U^{(m+1)})/u_0\right],$$

or τ^{-2} is generated from a truncated Gamma distribution.

Generate γ

The full conditional density of $\gamma > 0$ is given by

$$\begin{aligned} p(\gamma|\text{ Rest }) &\propto \exp\left\{w_J\gamma - \frac{1}{\tau^2}\sum_{j=1}^{J} c_j\exp(\gamma j)\right\} \qquad \text{(S2.1)} \\ w_J &= \frac{1}{2}\sum_{j=0}^{J} j - w_0 \text{ and } c_j = \frac{\theta_j^2}{2\sigma\tau^2}. \end{aligned}$$

We use slice sampling to generate γ, as in Lenk (1999), which introduces J latent variables $(U_1,\ldots,U_J)$ from uniform distributions:

$$p(U_1,\ldots,U_J,\gamma|\text{ Rest }) \propto \prod_{j=1}^{J} I\big(0 < U_j < \exp\left\{-c_j\exp(j\gamma)\right\}\big)\exp\left\{w_J\gamma\right\}.$$

Slice sampling first generates $U_j^{(m+1)}$ given $\gamma^{(m)}$:

$$U_j^{(m+1)} \sim U\big(0, \exp\left\{-c_j \exp(j\gamma^{(m)})\right\}\big) \text{ for } j = 1, \ldots J$$

Given these $U_j^{(m+1)}$, $\gamma^{(m+1)}$ has to satisfy constraint:

$$U_j^{(m+1)} \leq \exp\left\{-c_j \exp(j\gamma^{(m+1)})\right\} \text{ for } j = 1, \ldots, J$$

or

$$0 \leq \gamma^{(m+1)} \leq b \text{ where } b = \min\left\{\frac{1}{j}\log\left[\frac{1}{c_j}\log\left(U_j^{(m+1)}\right)\right]\right\}.$$

The full conditional distribution of γ given $\{U_j^{(m+1)}\}$is:

$$p(\gamma) \propto \exp(w_J\gamma) \text{ for } 0 \leq \gamma \leq b.$$

Using the inverse CDF, we obtain:

$$\gamma^{(m+1)} = b + \frac{1}{w_J}\log\left[U + \exp(-w_J b)(1-U)\right] \text{ where } U \sim U(0,1).$$

Generate σ^2

The scale-invariant priors for the spectral coefficients break the conditional conjugacy of the inverse Gamma distribution for σ^2. Its full conditional distribution is:

$$\begin{aligned}
p(\sigma^2|\text{ Rest }) &\propto \exp\left(-c_J/\sigma\right)\left(\sigma^2\right)^{-\left(\frac{r_{n,\sigma}}{2}+1\right)}\exp\left(-\frac{s_{n,\sigma}}{2\sigma^2}\right)\\
c_J &= \frac{1}{2v_{\theta_0}^2}\theta_0^2 + \frac{1}{2\tau^2}\boldsymbol{\theta}_{J-0}^{\intercal}\Gamma^{-1}\boldsymbol{\theta}_{J-0}\\
r_{n,\sigma} &= r_{0,\sigma} + n + p + (J+1)/2 + \dim(\boldsymbol{\alpha})\\
s_{n,\sigma} &= s_{0,\sigma} + \left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\beta} - \boldsymbol{f}_J\right)^{\intercal}\left(\boldsymbol{y} - \boldsymbol{W}\boldsymbol{\beta} - \boldsymbol{f}_J\right)\\
&+ \left(\boldsymbol{\beta} - \boldsymbol{m}_{0,\beta}\right)^{\intercal}\boldsymbol{V}_{0,\beta}^{-1}\left(\boldsymbol{\beta} - \boldsymbol{m}_{0,\beta}\right)\\
&+ \left(\alpha - m_{0,\alpha}\right)^2/v_{0,\alpha}^2.
\end{aligned}$$

where the last line is absent for monotone restriction and present for convex or concave restrictions. We use a uniform variable U in the full conditional of σ^2:

$$\begin{aligned}
p(U, \sigma^2|\text{ Rest }) &\propto I\left[0 \leq U \leq \exp\left(-c_J/\sigma\right)\right]\\
&\times \left(\sigma^2\right)^{-\left(\frac{r_{n,\sigma}}{2}+1\right)}\exp\left(-\frac{s_{n,\sigma}}{2\sigma^2}\right).
\end{aligned}$$

Given $\sigma^{(m)}$ on iteration m, we generate:

$$U^{(m+1)} \sim U\left[0, \exp\left(-c_J/\sigma^{(m)}\right)\right]$$

Given U, draws of σ^2 must satisfy:

$$\left(c_J/\log(U^{(m+1)})\right)^2 \leq \sigma^2.$$

We generate $\sigma^{2,(m+1)}$ given $U^{(m+1)}$ from:

$$\sigma^2 \sim IG\left(\frac{r_{n,\sigma}}{2}, \frac{s_{n,\sigma}}{2}\right) I\left[\left(c_j/\log\left(U^{(m+1)}\right)\right)^2 \leq \sigma^2\right].$$

Generate β

If $p > 0$ and there are covariates $\boldsymbol{W}$, then the full conditional distribution of $\boldsymbol{\beta}$ is:

$$\begin{aligned}
\boldsymbol{\beta}|\text{Rest} &\sim N(\boldsymbol{m}_{n,\beta}, \sigma^2 \boldsymbol{V}_{n,\beta}) \\
\boldsymbol{V}_{n,\beta} &= \left(\boldsymbol{V}_{0,\beta}^{-1} + \boldsymbol{W}^\intercal \boldsymbol{W}\right)^{-1} \\
\boldsymbol{m}_{n,\beta} &= \boldsymbol{V}_{n,\beta}\left[\boldsymbol{V}_{0,\beta}^{-1}\boldsymbol{m}_{0,\beta} + \boldsymbol{W}^\intercal\left(\boldsymbol{y} - \boldsymbol{f}_J\right)\right].
\end{aligned}$$

Generate α

With convex/concave restrictions the full conditional distribution of α is:

$$\begin{aligned}
\alpha|\text{Rest} &\sim N(m_{n,\alpha}, \sigma^2 v_{n,\alpha}^2) I(\delta\alpha > 0) \\
v_{n,\alpha}^2 &\sim \left[v_{0,\alpha}^{-2} + (\boldsymbol{x} - \bar{x}\mathbf{1})^\intercal(\boldsymbol{x} - \bar{x}\mathbf{1})\right]^{-1} \\
m_{n,\alpha_0} &\sim v_{n,\alpha}^2\left[v_{0,\alpha}^{-2} m_{0,\alpha} + (\boldsymbol{x} - \bar{x}\mathbf{1})^\intercal\left(\boldsymbol{y} - \boldsymbol{W}^\intercal\boldsymbol{\beta} - \boldsymbol{f}_J + \alpha^{(m)}(\boldsymbol{x} - \bar{x}\mathbf{1})\right)\right].
\end{aligned}$$

S2.2 MCMC for BSARU and BSARS

This supplementary material summarizes the MCMC algorithm for the U–Shape and S–Shape restrictions of Section 3.2 of the paper. Other than ω and ψ, all of the parameter are generated as those in BSARM and BSARMC. There is a slight complication in computing f_J because the integrals of $Z_J^2 h$ do not have closed-form expressions. Given the spectral coefficients, we compute $Z_J^2 h$ over a fine grid on $[0, 1]$ and apply Simpson's Rule to the integrals at the grid points. To compute the integrals at an observation x_i, we first find the largest grid point $\tilde{x}_i$ that is smaller than x_i, and add the area of the trapezoid between $\tilde{x}_i$ and x_i to the definite integral up to $\tilde{x}_i$. Also, we use the double integral at the grid points to approximate ξ.

We use truncated random walks to generate candidate values for ω and ψ.

$$\begin{aligned}
\omega^* &\sim N(\omega^{(m)}, V_1^2) I(0 \leq \omega \leq 1) \text{ where } V_1^2 \sim IG(a_T, b_T) \\
\psi^* &\sim N(\psi^{(m)}, V_2^2) I(\psi > 0) \text{ where } V_2^2 \sim IG(a_T, b_T)
\end{aligned}$$

The jump probability is:

$$\begin{aligned}
\log\alpha\big[\big(\omega^*,\psi^*\big),\big(\omega^{(m)},\psi^{(m)}\big)\big] &= \frac{1}{2\sigma^2}\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{(m)}\big)^\intercal\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{(m)}\big)\\
&- \frac{1}{2\sigma^2}\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{*}\big)^\intercal\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{*}\big)\\
&+ \frac{1}{2v_{0,\omega}^2}(\omega^{(m)}-m_{0,\omega})^2-\frac{1}{2v_{0,\omega}^2}(\omega^{*}-m_{0,\omega})^2\\
&+ \frac{1}{2v_{0,\psi}^2}(\psi^{(m)}-m_{0,\psi})^2-\frac{1}{2v_{0,\psi}^2}(\psi^{*}-m_{0,\psi})^2\\
&+ \log\Big[\text{ZCDF}\Big(\{1-\omega^{(m)}\}/V_1\Big)-\text{ZCDF}\Big(-\omega^{(m)}/V_1\Big)\Big]\\
&- \log\left[\text{ZCDF}\left(\{1-\omega^{*}\}/V_1\right)-\text{ZCDF}\left(-\omega^{*}/V_1\right)\right]\\
&+ \log\Big[1-\text{ZCDF}\Big(-\psi^{(m)}/V_2\Big)\Big]-\log\left[1-\text{ZCDF}\left(-\psi^{*}/V_2\right)\right]
\end{aligned}$$

In some examples, the MCMC seems to become stuck when ω is at one of the endpoints of the domain of x. To avoid this, we randomly generate ω from its prior distribution 10% of the time, and modify the jump probability:

$$\begin{aligned}
\log\alpha\big[\big(\omega^*,\psi^*\big),\big(\omega^{(m)},\psi^{(m)}\big)\big] &= \frac{1}{2\sigma^2}\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{(m)}\big)^\intercal\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{(m)}\big)\\
&- \frac{1}{2\sigma^2}\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{*}\big)^\intercal\big(\boldsymbol{y}-\boldsymbol{W\beta}-\boldsymbol{f_J}^{*}\big)\\
&+ \frac{1}{2v_{0,\psi}^2}(\psi^{(m)}-m_{0,\psi})^2-\frac{1}{2v_{0,\psi}^2}(\psi^{*}-m_{0,\psi})^2\\
&+ \log\Big[1-\text{ZCDF}\Big(-\psi^{(m)}/V_2\Big)\Big]-\log\left[1-\text{ZCDF}\left(-\psi^{*}/V_2\right)\right]
\end{aligned}$$

The Metropolis algorithms for ω and ψ are sensitive to the mean of the inverse Gamma distribution of the proposal distribution. We found that increasing or decreasing this mean depending on the acceptance probabilities in test runs improves the MCMC.

S2.3 MCMC for Spike–and–Slab Prior

The MCMC for the Spike–and–Slab Prior adds several steps to the previous MCMC algorithms. All of the parameters are generated as before with the following changes. Recall that $\tilde{Z}$, the latent Z, has a Gaussian process, while the Z used in f truncates $\tilde{Z}$: $Z(x)=\tilde{Z}(x)$ if $\tilde{Z}^2>\chi$ and $Z(x)=0$ if $\tilde{Z}^2<\chi$. We make the following modification in the Metropolis algorithm for generating the spectral coefficients. After obtaining the candidate draw $\boldsymbol{\theta}_J^*$, compute the candidate latent $\tilde{Z}^*(x)=\sum\theta_j^*\varphi(x)$, and preform the truncation to obtain the candidate Z^*. Use the candidate Z^* in the model of f to compute the likelihood function the jump probability.

We also generate the truncation parameter χ with a random walk Metropolis using truncated normal distributions on the non-negative numbers.

S3 Supplemental Material: Proofs of Theorems

S3.1 Proof of Theorem 1

Proof of Theorem 1 is based on the argument that characterizes the differentiable function with the nonnegative first order derivative in terms of the integral representation using the piecewise continuous function, similar to those from Shively et al. (2009) and Ramsay (1998).

If $f(x) \in C^{(q)+}$, then let $u(x) = \sqrt{f^{(q)}(x)}$ and set $a = f(0)$. Then, it follows that $u(x) \in C^1[0,1]$, which yields the representation. Conversely, if $f^{(q-1)}(x) = a + \int_0^x u^2(t)dt$ and $u(x) \in C^1[0,1]$, then it follows that $f^{(q)}(x) = u^2(x) \geq 0$, and that it is piecewise continuous.

S3.2 Proof of Theorem 2

Proof of posterior consistency in Theorem 2 adapts general theorems of Choi and Schervish (2007) and Shively et al. (2009). The idea for proving Theorem 2 is to bound the numerator and denominator of the posterior probability of $W_{\epsilon,n}$, $\Pi\{W_{\epsilon,n}^C|(x_1,Y_1),\ldots,(x_n,Y_n)\}$ separately in the following way :

Let $\vartheta_0 = (f_0, \sigma_0)$ and let $P_{\vartheta_0}^n$ and $p_{\vartheta_0}^n$ stand for the joint distribution and the probability density of $Y^n = (Y_1, Y_2, \ldots, Y_n)$ when the true model of Y^n is nonparametric regression parametrized by ϑ, and ϑ_0 is the true value of ϑ. The posterior distribution $\Pi\{W_{\epsilon,n}^C|(x_1,Y_1),\ldots,(x_n,Y_n\}$ is written by $\Pi(W_{\epsilon,n}^C|Y^n) = \frac{J_{W_{\epsilon,n}^C}^{\vartheta_0}(Y^n)}{J^{\vartheta_0}(Y^n)}$, where

$$J_{W_{\epsilon,n}^C}^{\vartheta_0}(Y^n) = \int_{W_{\epsilon,n}^C} \frac{p_\vartheta^n(Y^n)}{p_{\vartheta_0}^n(Y^n)}\pi(d\vartheta), \quad \text{and} \quad J^{\vartheta_0}(Y^n) = \int \frac{p_\vartheta^n(Y^n)}{p_{\vartheta_0}^n(Y^n)}\pi(d\vartheta).$$

Prior positivity of neighborhoods

To obtain the exponentially increasing lower bound of $J^{\vartheta_0}(Y^n)$, from the sufficient condition (A1) of Theorem 1 in Choi and Schervish (2007), we verify that ϑ_0 is in the Kullback-Leibler support of the prior Π, and that Π assigns a positive probability on a specific set in which conditions for applying law of large numbers are satisfied.

For this purpose, define $\Lambda_i(\vartheta_0, \vartheta) = \log \frac{p_{\vartheta_0}(Y_i)}{p_\vartheta(Y_i)}$, where $p_\vartheta(y_i)$ is a normal density with mean $f(x_i)$ and variance σ^2. Also define the mean of Λ_i, i.e. the Kullback-Leibler divergence between $p_{\vartheta_0}(y_i)$ and $p_\vartheta(y_i)$ and the variance of Λ_i, $K_i(\vartheta_0, \vartheta) = \mathrm{E}_{\vartheta_0}(\Lambda_i(\vartheta_0, \vartheta))$

and $V_i(\vartheta_0, \vartheta) = \mathrm{Var}_{\vartheta_0}(\Lambda_i(\vartheta_0, \vartheta))$. Then, direct calculations shows that

$$\begin{aligned} K_i(\vartheta_0; \vartheta) &= \frac{1}{2}\log\frac{\sigma^2}{\sigma_0^2} - \frac{1}{2}\left(1 - \frac{\sigma_0^2}{\sigma^2}\right) + \frac{1}{2}\frac{[f_0(x_i) - f(x_i)]^2}{\sigma^2}, \text{ and} \\ V_i(\vartheta_0, \vartheta) &= 2\left[-\frac{1}{2} + \frac{1}{2}\frac{\sigma_0^2}{\sigma^2}\right]^2 + \left[\frac{\sigma_0^2}{\sigma^2}[f(x_i) - f_0(x_i)]\right]^2. \end{aligned}$$

Let $u(x)$ and $u_0(x)$ be two piecewise continuous functions on $[0,1]$, i.e. $u(x), u_0(x) \in C^1[0,1]$ that corresponds to $f(x)$ and $f_0(x)$ respectively, $f(x) = \int_0^x u^2(t)dt$ and $f_0(x) = \int_0^x u_0^2(t)dt$ as given in Theorem 1. Define

$$C_\delta = \left\{(u, \sigma) \ : \ \|u - u_0\|_\infty < \delta \ , \left|\frac{\sigma}{\sigma_0} - 1\right| < \delta\right\},$$

where $\|\cdot\|_\infty$ denotes the supremum norm. Then, it follows from the integral representation in Theorem 1 that for every $\epsilon > 0$, there exists $\delta > 0$ such that $\forall\ (u, \sigma) \in C_\delta$, $K_i(\vartheta_0, \vartheta) < \epsilon$ for all i. and that $(ii)\ \sum_{i=1}^{\infty} \frac{V_i(\vartheta_0, \vartheta)}{i^2} < \infty$, $\forall\ (u, \sigma) \in C_\delta$, assuming $u(t)$ and $u_0(t)$ are uniformly bounded. Hence, if the prior, Π assigns positive probability to C_δ for each $\delta > 0$, then the prior positivity condition of Choi and Schervish (2007) holds. Note that the nonparametric prior for f we considered based on cosine series is a Gaussian process prior as described in Section 2.1 using the Karhunen-Lóeve expansion and Mercer's theorem), with a reproducing kernel Hilbert space. Thus, it follows from Choi and Schervish (2007) and van der Vaart and van Zanten (2008) that for every $\delta > 0$, $\Pi(C_\delta) > 0$.

Existence of tests

To obtain the upper bound of the numerator $J^{\vartheta_0}_{W^C_{\epsilon,n}}(Y^n)$, $W^C_{\epsilon,n}$ and ϑ_0 need to be strongly separated Choi and Ramamoorthi (2008), which can be formulated in terms of uniformly consistent tests (see, e.g. Choi and Schervish (2007) and Ghosal and van der Vaart (2007).

To verify (A2) of Theorem 1 in Choi and Schervish (2007), we first construct a sieve and then construct a test for each element of the sieve. Let $M_n = O(n^{\alpha_1})$, where $\frac{2\delta + 1}{2} < \alpha_1 < 1$ for some $0 < \delta < 1/2$ and define $\Theta_n = \Theta_{1n} \times \mathbb{R}^+$, where $\Theta_{1n} = \{u(\cdot) \ : \ \|u\|_\infty < M_n\}$. Note that by the integral representation of $f(x)$ as in Theorem 1, we need a sieve only based on the supremum norm of $u(x)$ (*cf.* Choi and Schervish (2007)).

The nth test is constructed by combining a collection of tests, one for each of finitely many elements of Θ_n. Note that there exists a constant K' such that the ϵ-covering number $N(\epsilon, \Theta_{1n}, \|\cdot\|_\infty)$ of Θ_{1n} satisfies $\log N(\epsilon, \Theta_n, \|\cdot\|_\infty) \leq \frac{K' M_n}{\epsilon}$, which follows from Theorem 2.7.1. of van der Vaart and Wellner (1996). When the values of the covariate

in $[0,1]$ are equally spaced, there exist test functions $\{\Phi_n\}$ such that $\sum_{n=1}^{\infty} \mathrm{E}_{\theta_0}\Phi_n < \infty$ and $\sup_{\theta \in W_{\epsilon,n}^C \bigcap \Theta_n} \mathrm{E}_\theta(1-\Phi_n) \leq \exp(-C*n)$ for a constant $C* > 0$, which follows from Theorem 2 of Choi and Schervish (2007). Note that when the empirical norm Q_n is used, it is obvious that there exists $\gamma > 0$ such that $\sum_{i=1}^n |f(x_i) - f_0(x_i)| > \gamma n$, which is essential in the construction of uniformly consistent tests (Choi and Schervish (2007)). The remaining thing is to show that Θ_n^C has an exponentially small prior probability, which also follows from Section 6 of Choi and Schervish (2007).

Hellinger neighborhood

As shown in Shively et al. (2009), the numerator $J^{\vartheta_0}_{W^C_{\epsilon,n}}(Y^n)$ is upper-bounded by the properties of maximum likelihood estimators under monotone constraints using Theorem 3 in Walker and Hjort (2001). This upper-bound holds as long as the regression function is monotone increasing which is applied to the proposed BSAR with shape restriction. Alternatively, the uniformly consistent tests can also be constructed as in Choi and Schervish (2007). Since the exponentially increasing lower bound of the denominator $J^{\vartheta_0}_{W^C_{\epsilon,n}}(Y^n)$ is obtained regardless of the neighborhood under consideration as shown in C.1., it is easy to see that posterior consistency with H_ϵ also holds.

S3.3 Proof of Theorem 3

Proof of Theorem 3 is based on the argument that characterizes the sample paths of the q^{th} derivative to be almost surely piecewise continuous and has a unique root at ω, using a similar reasoning to Theorem 1 with L'Hospital's Theorem.

By the definitions of h, f_1 and f_2, it is obvious that $f_1^{(1)}(x)$ and $f_2^{(2)}(x)$ are continuous and have unique zeros at $x = \omega$. Conversely, define

$$g(x) = \begin{cases} \frac{f^{(q)}(x)}{h(x)}, & 0 < x < 1,\ x \neq \omega \\ \frac{-2f^{(q+1)}(\omega)}{\psi}, & x = \omega. \end{cases} \tag{S3.2}$$

Then, $g(x)$ is continuous for all $0 < x < 1$, $x \neq \omega$ by definition and is also continuous at $x = \omega$, which follows from the differentiability of $f^{(q)}(x)$ and $h(x)$ at $x = \omega$ as well as L'Hospital's theorem. In addition, it is obvious that $g(x)$ is positive for all $0 < x < 1$ by construction since $f^{(q)}(x)$ is assumed to have a unique zero at $x = \omega$ from being positive-to-negative, which implies that $f^{(q+1)}(x)$ is negative at $x = \omega$, and $h(x)$ is positive if $x < \omega$ and negative if $x > \omega$. Thus, as in the same argument of Theorem 1, there exists $u(x) \in C^1[0,1]$ such that $u(x) = \sqrt{g(x)}$. Hence, it follows that $f^{(q)}(x) = u^2(x)h(x)$ and that $f^{(q)}(\omega) = -u^2(\omega)\psi/2$, which corresponds to $g(\omega)$ in (S3.2).

S4 Simulation Studies

S4.1 Details for Comparing Models

The models for the first simulation study are:

$$\begin{aligned}
\text{Linear: } Y &= x + \epsilon && \text{(S4.3)}\\
\text{Sinusoid: } Y &= 2\pi x + \sin(2\pi x) + \epsilon && \text{(S4.4)}\\
\text{Expo: } Y &= \exp(6x - 3) + \epsilon && \text{(S4.5)}\\
\text{QuadCos: } Y &= 16x^2 - \frac{4}{\pi^2}\cos(2\pi x) - \frac{1}{\pi^2}\cos(4\pi x) \\
&\quad -\frac{32}{9\pi^2}\cos(3\pi x) - \frac{32}{\pi^2}\cos(\pi x) + \frac{365}{9\pi^2} + \epsilon && \text{(S4.6)}\\
\text{LogX: } Y &= \log(1 + 10x) + \epsilon && \text{(S4.7)}
\end{aligned}$$

where $\epsilon \sim N(0, 1)$, and the values of X are equally spaced on 0 to 1. The true f mean–centers the functions given in (S4.3–S4.7), and adds the integral to the intercept. The first two functions (S4.3) and (S4.4) are increasing, the next two functions (S4.5)-(S4.6) are increasing and convex, whereas `LogX` in (S4.7) is increasing and concave. BSARMC sets $\alpha = 0$ to focus attention on the spectral representation. After an initial period to identify satisfactory variances for the adaptive Metropolis algorithm, the MCMC chain ran for 11,000 iterations, and the last 1000 were used for estimation.

S4.2 S–Shaped Functions and BSARS

To test BSARS we consider a cubic function (`CubicX`) that has an inflection point $\omega = -1$:

$$Y = (x + 1)^3 + \epsilon \text{ where } \epsilon \sim N(0, \sigma^2) \text{ and } x \in [-5, 5]. \quad \text{(S4.8)}$$

For this simulation, the low information condition is $n = 100$ and $\sigma = 10$, and the high information condition is $n = 200$ and $\sigma = 5$. We fit the data based on BSARS. Numerical comparison of BSARS is made with BSAR and BSARM with Gamma Priors and BBPM with monotone restriction. In Table 1 BSARS outperforms the other approaches according to average RMISE. The proposed BSARS fits the true inflection point -1 reasonably well.

S4.3 Spike–and–Slab Prior

We next compare the Gamma Prior to the Spike-and-Slab Prior for BSARM and BSARMC. We simulated 100 observations from $y_i = 20 + f(x_i) + \epsilon_i$ where $x_i \sim U(0, 10)$; $\epsilon_i \sim N(0, 40)$ and

$$f(x) = \begin{cases} -24.15 & \text{if } 0 \le x \le 2 \\ \exp[.7(x - 2)] - 25.15 & \text{if } 2 < x \le 7 \\ 7.96 + 23.18x & \text{if } 7 < x \le 10. \end{cases}$$

Table 1: Simulation results for `CubicX`

Average RMISE	BSARS	BSARM	BRSM	BBPM
$n = 100$, $\sigma = 10$	**2.359**	2.638	2.772	2.933
(s.e.)	(0.0917)	(0.0899)	(0.0948)	(0.0982)
$n = 200$, $\sigma = 5$	**1.039**	1.504	1.253	1.091
(s.e.)	(0.0438)	(0.0953)	(0.0310)	(0.0306)
Inflection Point ($\omega = -1$)	BSARS	BSARM	BRSM	BBPM
$n = 100$, $\sigma = 10$	-0.8300	NA	NA	NA
(s.e.)	(0.0389)	NA	NA	NA
$n = 200$, $\sigma = 5$	-0.9270	NA	NA	NA
(s.e.)	(0.0330)	NA	NA	NA

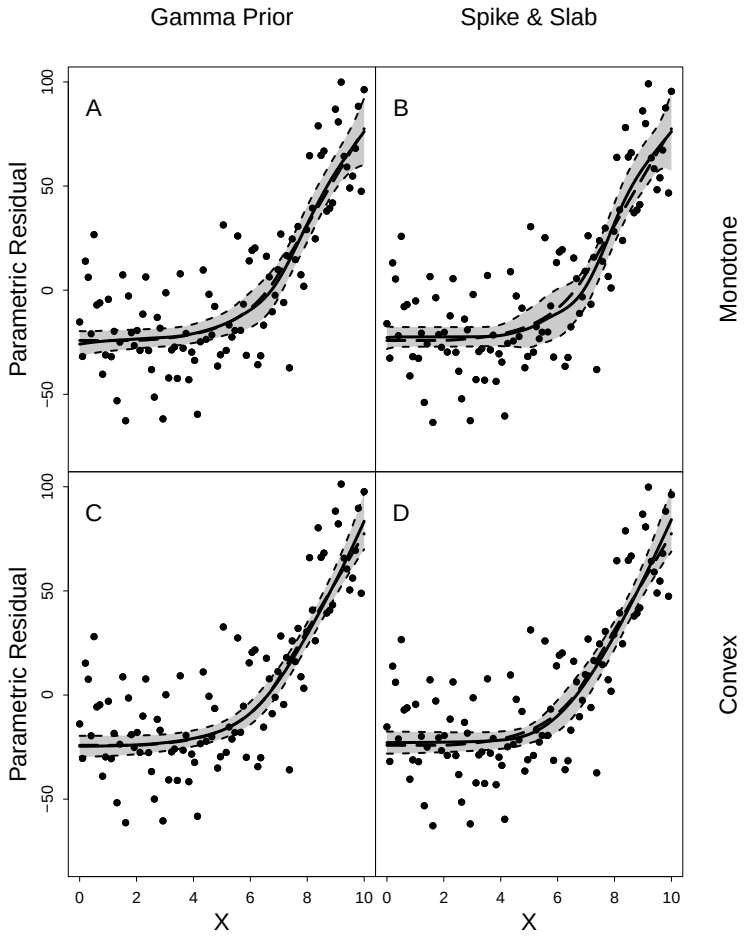

Figure 1: Gamma versus Spike-and-Slab Priors. Dots are parametric residuals, solid lines are posterior means, and shaded areas between dashed lines are 95% credible intervals. Panel A: BSARM with Gamma Prior. Panel B: BSARM with Spike-and-Slab Prior. Panel C: BSARMC with Gamma Prior. Panel D: BSARMC with Spike-and-Slab Prior.

The true f is constant from 0 to 2, exponential from 2 to 7, and linear from 7 to 10. f is on the boundary of BSARM and BSARMC when $0 < x < 2$, and it is on the boundary of BSARMC when $7 < x < 10$. It is not on the boundary of either function space between $2 < x < 7$. Figure (1) plots the true function and its four estimators. Each of the models contains the true function within its 95% HPD intervals. Table (2) presents fit statistics and estimates for BSARM and BSARMC along with BSAR. The largest improvement in the LIL and RMISE results from going from the unconstrained BSAR to the constrained models. Spike-and-Slab BSARM has the largest LIL, followed by Gamma BSARM, Spike-and-Slab BSARMC, and Gamma BSARMC. The RMISE is better for the Gamma Priors than the Spike-and-Slab Priors when using monotone and convex restrictions. The R-Squares and estimated model parameters are nearly the same. In this example, the Spike-and-Slab Prior helps recover the boundary conditions.

Table 2: Comparing Gamma to Spike-and-slab Priors

	BSAR	BSARM Gamma	BSARM Spike & Slab	BSARMC Gamma	BSARMC Spike & Slab
LIL	-585.0068	-546.6553	-537.1292	-554.8863	-544.3778
RMISE	17.9629	3.9932	8.4022	4.0332	5.9708
R-Square	0.7374	0.7057	0.7131	0.7008	0.7031
σ	20.5229	20.6560	20.5539	20.5393	20.5493
STD DEV	1.5919	1.4606	1.4518	1.4425	1.4623
Intercept	18.9562	19.1642	19.0543	19.0794	19.0252
STD DEV	2.0971	2.1706	2.0709	2.0559	2.0558
Cutoff			14.7185		4.3785
STD DEV			3.5100		2.7632

S4.4 Bayesian Hypothesis Testing Details

Our simulation study considers five true models: one for each set of constraints and use the Gamma Prior of Sections 3.1 and 3.2. The simulation models are:

$$\text{M1: } Y_i = 200(x_i - 0.1)(x_i - 0.6)(x_i - 0.8) + \epsilon_i \tag{S4.9}$$

$$\text{M2: } Y_i = 10\frac{\exp[15(x_i - 0.4)]}{\exp[15(x_i - 0.4)] + 1} + \exp[20(x_i - 0.9)] + \epsilon_i \tag{S4.10}$$

$$\text{M3: } Y_i = 20\exp[5(x_i - 0.9)] + \epsilon_i \tag{S4.11}$$

$$\text{M4: } Y_i = 20\exp\left[-15(x_i - 0.7)^2\right] + \epsilon_i \tag{S4.12}$$

$$\text{M5: } Y_i = 10\frac{\exp[10(x_i - 0.4)]}{\exp[10(x_i - 0.4)] + 1} + \epsilon_i \tag{S4.13}$$

where $\epsilon_i \sim N(0,1)$ and $x \sim U(0,1)$. Equation (S4.9) is only consistent with the unrestricted BSAR. Equation (S4.10) is increasing and consistent with the monotone BSARM and the unrestricted BSAR. Equation (S4.11) is increasing and convex and consistent with monotone and convex BSARMC, monotone BSARM, and unrestricted BSAR. Equation (S4.12) is inverted U shaped and consistent with U-shaped BSARU and unrestricted BSAR. Equation (S4.13) is S–shaped and consistent with S–shaped

BSARS, monotone BSARM, and unrestricted BSAR. Also, BSARU and BSARS can mimic monotonic models by sending ω towards the end point 0 or 1.

The simulations have a low information condition with 50 observations and a high information condition with 200 observations. Fifty data sets were generated for each function and information condition. BSAR, BSARM, BSARMC, BSARU, and BSARS are fitted to each of the simulated data sets. The simulations did not include the integration parameter α, and ψ was set to 1000 for the U and S shaped functions. The simulations used the same prior parameters. The MCMC used a burn-in period of 20,000 iterations and an estimation period of 10,000 iterations. We use Gelfand and Dey (1994) to approximate the log-integrated likelihood (LIL). Tables (3) to (7) report the mean and standard deviation of the LIL for the 50 simulated data sets in each condition, and the proportion of times that the LIL was maximum for each model.

Table 3: LIL Simulation: Unrestricted Function M1

50 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-168.114	-192.767	-201.994	-220.503	-209.099
LIL: STD DEV	5.254	6.365	5.664	133.533	22.954
LIL: Choice	1.000	0.000	0.000	0.000	0.000
RMISE: Mean	0.598	2.407	3.850	3.231	4.137
RMISE: STD DEV	0.088	0.213	0.286	0.433	0.266
200 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-501.123	-655.949	-701.773	-690.688	-710.554
LIL: STD DEV	9.502	14.521	13.855	14.869	14.639
LIL: Choice	1.000	0.000	0.000	0.000	0.000
RMISE: Mean	0.339	2.257	3.229	2.918	3.337
RMISE: STD DEV	0.044	0.100	0.149	0.227	0.274

Table 4: LIL Simulation: Monotone Function M2

50 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-159.346	-145.007	-158.733	-159.162	-153.302
LIL: STD DEV	5.711	7.663	6.920	9.929	6.697
LIL: Choice	0.000	0.840	0.020	0.020	0.120
RMISE: Mean	0.549	0.398	1.273	0.649	1.091
RMISE: STD DEV	0.086	0.104	0.102	0.326	0.119
200 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-492.962	-480.329	-550.351	-499.800	-526.203
LIL: STD DEV	6.718	9.639	11.235	19.697	14.076
LIL: Choice	0.040	0.920	0.000	0.040	0.000
RMISE: Mean	0.336	0.221	1.185	0.300	0.779
RMISE: STD DEV	0.049	0.044	0.036	0.175	0.078

Table 5: LIL Simulation: Monotone and Convex Function M3

50 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-161.931	-142.352	-140.618	-152.399	-147.947
LIL: STD DEV	4.992	6.254	7.096	6.953	6.345
LIL: Choice	0.000	0.380	0.540	0.000	0.080
RMISE: Mean	0.509	0.336	0.318	0.376	0.374
RMISE: STD DEV	0.083	0.095	0.096	0.082	0.094
200 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-490.013	-474.379	-465.555	-488.631	-485.809
LIL: STD DEV	12.056	14.699	13.081	11.787	16.645
LIL: Choice	0.000	0.140	0.740	0.020	0.100
RMISE: Mean	0.315	0.200	0.171	0.237	0.199
RMISE: STD DEV	0.043	0.054	0.050	0.046	0.056

Table 6: LIL Simulation: U–Shaped Function M4

50 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-160.708	-203.013	-220.222	-159.440	-203.926
LIL: STD DEV	7.150	6.466	8.740	10.097	5.721
LIL: Choice	0.360	0.000	0.000	0.640	0.000
RMISE: Mean	0.442	3.389	5.167	0.442	3.669
RMISE: STD DEV	0.096	0.406	0.419	0.097	0.390
200 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-490.729	-705.068	-776.406	-509.745	-720.649
LIL: STD DEV	10.657	13.355	11.097	13.133	13.397
LIL: Choice	0.940	0.000	0.000	0.060	0.000
RMISE: Mean	0.256	3.133	4.913	0.246	3.164
RMISE: STD DEV	0.041	0.199	0.218	0.055	0.197

Table 7: LIL Simulation: S–Shaped Function M5

50 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-215.488	-173.730	-215.466	-179.728	-178.649
LIL: STD DEV	16.206	10.601	9.961	8.123	11.989
LIL: Choice	0.000	0.440	0.000	0.160	0.400
RMISE: Mean	0.337	0.340	2.227	0.379	0.381
RMISE: STD DEV	0.088	0.091	0.133	0.085	0.114
200 Observations	BSAR	BSARM	BSARMC	BSARU	BSARS
LIL: Mean	-569.465	-535.112	-690.668	-580.578	-576.267
LIL: STD DEV	24.934	15.919	16.561	242.663	215.160
LIL: Choice	0.020	0.540	0.000	0.260	0.180
RMISE: Mean	0.182	0.188	2.117	0.253	0.168
RMISE: STD DEV	0.052	0.040	0.063	0.236	0.051

References

Choi, T. and R. V. Ramamoorthi (2008). Remarks on consistency of posterior distributions. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, Volume 3 of Inst. Math. Stat. Collect., pp. 170-186. Beachwood, OH: Inst. Math. Statist.

Choi, T. and M. J. Schervish (2007). On posterior consistency in nonparametric re-

gression problems. *J. Multivariate Anal.* **98**, 1969-1987.

Damien, P., J. Wakefield, and S. Walker (1999). Gibbs sampling for Bayesian non-conjugate and hierarchical models by using auxiliary variables. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 331-344.

Ghosal, S. and van der Vaart (2007). Convergence rates of posterior distributions for noniid observations. *Ann. Statist.* **35**, 192-223.

Haario, H., E. Saksman, and J. Tamminen (2001). An adaptive metropolis algorithm. *Bernoulli* **7**, 223-242.

Lenk, P. J. (1999). Bayesian inference for semiparametric regression using a fourier representation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **61**, 863-879.

Neal, R. M. (2003). Slice sampling. *Ann. Statist.* **31**, 705767. With discussions and a rejoinder by the author.

Shively, T. S., T. W. Sager, and S. G. Walker (2009). A Bayesian approach to non-parametric monotone function estimation. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71**, 159-175.

van der Vaart, A. W. and J. H. van Zanten (2008). Reproducing kernel Hilbert spaces of Gaussian priors. In *Pushing the limits of contemporary statistics: contributions in honor of Jayanta K. Ghosh*, Volume 3 of Inst. Math. Stat. Collect., pp. 170-186. Beachwood, OH: Inst. Math. Statist.

van der Vaart, A. W. and J. A. Wellner (1996). Weak convergence and empirical processes. Springer Series in Statistics. Springer-Verlag, New York. With applications to statistics.

Walker, S. and N. L. Hjort (2001). On Bayesian consistency. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **63**, 811-821.