Supplementary Material: Proofs

The following lemmas prepare for the proof of concentration properties on prior density $\pi_\alpha(\boldsymbol{\beta}_j/\sigma)$ for $j = 1, \ldots, p_n$ on m_n-dimensional vector $\boldsymbol{\beta}_j/\sigma$. As introduced earlier, we consider the shrinkage prior

$$\boldsymbol{\beta}_j|\lambda_j, \sigma^2 \overset{\text{ind}}{\sim} \mathbb{N}_{m_n}(0, \lambda_j\sigma^2\mathbb{I}_{m_n}), \tag{1}$$

$$\lambda_j|\psi_j \overset{\text{ind}}{\sim} \exp(\psi),\ \ \psi_j \overset{\text{iid}}{\sim} \text{gamma}(\alpha, 2). \tag{2}$$

For $j = 1, \ldots, p_n$, we let $p(\lambda_j|\psi_j)$ denote the prior density for λ_j given ψ_j and $p_\alpha(\psi)$ denote the prior density for ψ_j given the hyper-parameter α. In the following lemmas on the multivariate Dirichlet-Laplace (DL) prior, we want to show that it concentrates within a small neighborhood around zero with a dominated and steep peak, while retaining heavy tails away from zero.

Lemma 1 For any positive constant $\mu > 0$ and thresholding value $a_n \leq \sqrt{\frac{d_0'}{m_n}}\epsilon_n/p_n$,

$$1 - \int_{\|\boldsymbol{x}|\leq a_n} \pi_\alpha(\boldsymbol{x})d\boldsymbol{x} \leq p_n^{-(1+\mu)}, \tag{3}$$

if the Dirichlet parameter satisfies $\alpha \asymp p_n^{-(1+\nu)}$ for $\nu > \mu$.

Lemma 2 For a small constant $\eta > 0$ and E defined in assumption $A_2(2)$,

$$-\log\left(\int_{\sum_{j=1}^{p_n-s}\|\boldsymbol{x}_j\|\leq\eta\sqrt{m_n}\epsilon_n} \prod_{j=1}^{p_n-s} \pi_\alpha(\boldsymbol{x}_j)d\boldsymbol{x}_1\ldots d\boldsymbol{x}_{p_n-s}\right) \prec n\epsilon_n^2, \tag{4}$$

$$-\log\left(\inf_{\|\boldsymbol{x}\|\leq\sqrt{m_n}E} \pi_\alpha(\boldsymbol{x})\right) \prec n\epsilon_n^2/s. \tag{5}$$

Lemma 3 The prior density $\pi_\alpha(\boldsymbol{\beta}_j/\sigma)$ satisfies

$$\max_{j\in\xi^*} \sup_{\|\boldsymbol{x}_1\|,\|\boldsymbol{x}_2\|\in[\|\boldsymbol{\beta}_j^*/\sigma\|-\sqrt{m_n}C'\epsilon_n,\|\boldsymbol{\beta}_j^*/\sigma\|+\sqrt{m_n}C'\epsilon_n]} \frac{\pi_\alpha(\boldsymbol{x}_1)}{\pi_\alpha(\boldsymbol{x}_2)} < l_n, \tag{6}$$

for a constant C' and $s\log l_n \prec \log p_n$,

The three lemmas above specify the conditions on prior concentration. Lemma 1 and Lemma 2 establish that the prior probability that the $\mathcal{L}_2$-norm of each basis coefficient larger than a_n is arbitrarily small with the constraint: $a_n \leq \sqrt{d_0'/m_n}\epsilon_n/p_n$. Therefore, the shrinkage prior on $\boldsymbol{\beta}_j/\sigma$ has a dominated and steep peak around zero. Moreover, Lemma 3 controls the variation in the prior density of basis coefficients at points not too close to zero. The shrinkage prior density concentrates in a very small area around zero but also has heavy and sufficiently flat tails. Thus the prior mimics the composition of a continuous distribution and a point mass at zero. The proof of these lemmas follows.

Proof of Lemma 1

For a vector $\boldsymbol{x}$ of dimension m_n, the prior density is defined in (1) and (2). When λ is given, $T \stackrel{\text{def}}{=} \|\boldsymbol{x}\|^2/\lambda \sim \chi^2_{m_n}$. Next, we have

$$\begin{aligned} 1 - \int_{\|\boldsymbol{x}\|\leq a_n} \pi_\alpha(\boldsymbol{x})d\boldsymbol{x} &= \int_0^\infty \int_0^\infty \mathrm{P}\left(T > \frac{a_n^2}{\lambda}\right) p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda \\ &= \int_0^{a_n^2/p_n} \mathrm{P}\left(T > \frac{a_n^2}{\lambda}\right)\left[\int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi\right]d\lambda \\ &\quad + \int_{a_n^2/p_n}^\infty \mathrm{P}\left(T > \frac{a_n^2}{\lambda}\right)\left[\int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi\right]d\lambda \end{aligned} \tag{7}$$

For any constant $\mu > 0$, we can conclude that

$$\int_0^{a_n^2/p_n} \mathrm{P}\left(T > \frac{a_n^2}{\lambda}\right)\left[\int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi\right]d\lambda \leq \mathrm{P}(T > p_n)\mathrm{P}_\alpha(\lambda < a_n^2/p_n) \leq e^{-p_n/2}, \tag{8}$$

where the last inequality is obtained from the tail property of χ^2-distribution.

Next, from Lemma 3.3 in Bhattacharya et al. (2015), it follows that

$$\int_{a_n^2/p_n}^{\infty} \mathrm{P}\left(T > \frac{a_n^2}{\lambda}\right) \left[\int_0^{\infty} p(\lambda|\psi) p_\alpha(\psi) d\psi\right] d\lambda \leq \mathrm{P}_\alpha(\lambda > a_n^2/p_n) \leq C_1 \alpha \log(p_n/a_n^2) \tag{9}$$

for a constant $C_1 > 0$. Combining the inequalities in (7), (8) and (9), we have

$$1 - \int_{\|\boldsymbol{x}\| \leq a_n} \pi_\alpha(\boldsymbol{x}) d\boldsymbol{x} \leq e^{-p_n/2} + C_1 \alpha \log(p_n/a_n^2) \leq p_n^{-(1+\mu)}.$$

for $\alpha \asymp p_n^{-(1+\nu)}$ where $\nu > \mu$.

Proof of Lemma 2

When λ_j is given, the m_n-dimensional vector $\boldsymbol{x}_j$ satisfies that $\|\boldsymbol{x}_j\|/\sqrt{\lambda_j} \overset{\text{iid}}{\sim} \chi_{m_n}$ for $j = 1, \ldots, p_n$. Therefore, for the prior of λ_j defined in (2),

$$\mathrm{E}(\|\boldsymbol{x}_j\|) = \mathrm{E}\big[\mathrm{E}\big(\|\boldsymbol{x}_j\|\big|\lambda_j\big)\big] = \mathrm{E}\big[\sqrt{\lambda_j}\mathrm{E}(\|\boldsymbol{x}_j\|/\sqrt{\lambda_j}|\lambda_j)\big] = \frac{\sqrt{2}\Gamma((m_n+1)/2)}{\Gamma(m_n/2)}\mathrm{E}(\sqrt{\lambda_j}), \tag{10}$$

and then $\mathrm{E}(\sqrt{\lambda_j}) = \mathrm{E}\big[\mathrm{E}(\sqrt{\lambda_j}|\psi_j)\big] = \Gamma(3/2)\mathrm{E}(\sqrt{\psi_j}) = \Gamma(3/2)\frac{\sqrt{2}\Gamma(\alpha+1/2)}{\Gamma(\alpha)}$. Then for a constant $C_2 > 0$ free of α, $\mathrm{E}(\|\boldsymbol{x}_j\|) = C_2\sqrt{m_n}\alpha$. By the Central Limit Theorem,

$$\mathrm{P}\left(\sum_{j=1}^{p_n-s} \frac{\|\boldsymbol{x}_j\|}{\sqrt{m_n}} \leq \eta\epsilon_n\right) = \mathrm{P}\left(\frac{1}{p-s}\sum_{j=1}^{p_n-s} \frac{\|\boldsymbol{x}_j\|}{\sqrt{m_n}} \leq \frac{\eta\epsilon_n}{p-s}\right) \geq 1/2 \tag{11}$$

if $\alpha \leq \frac{C_3\epsilon_n}{p-s}$, which can be verified since $\alpha \asymp p_n^{-(1+\nu)}$ for $\nu > 0$. Therefore, we show that

$$-\log\left(\int_{\sum_{j=1}^{p_n-s}\frac{\|\boldsymbol{x}_j\|}{\sqrt{m_n}} \leq \eta\epsilon_n} \prod_{j=1}^{p_n-s} \pi_\alpha(\boldsymbol{x}_j) d\boldsymbol{x}_1 \ldots d\boldsymbol{x}_{p_n-s}\right) \prec n\epsilon_n^2. \tag{12}$$

Next, we need to prove that $-\log\left(\inf_{\|\boldsymbol{x}\|\le\sqrt{m_n}E}\pi_\alpha(\boldsymbol{x})\right)\prec n\epsilon_n^2/s$. For s_1/s_2 close to zero,

$$\inf_{\|\boldsymbol{x}\|\le\sqrt{m_n}E}\pi_\alpha(\boldsymbol{x})=\int_0^\infty\int_0^\infty(2\pi\lambda)^{-m_n/2}e^{-m_nE^2/2\lambda}p(\lambda|\psi)p_\alpha(\psi)d\lambda d\psi$$
$$\ge(2\pi s_2)^{-m_n/2}e^{-m_nE^2/2s_1}\mathrm{P}_\alpha(s_1<\lambda<s_2). \quad (13)$$

As λ follows prior distribution in (2) for a given hyper-parameter α, we have

$$\mathrm{P}_\alpha(s_1<\lambda<s_2)=\int_0^\infty(e^{-s_1/\psi}-e^{-s_2/\psi})p_\alpha(\psi)d\psi\ge\frac{e^{-s_2/2}}{\Gamma(\alpha)}(s_2-s_1)\int_0^{s_2}\psi^{\alpha-2}e^{-s_2/\psi}d\psi$$
$$\ge\alpha e^{-s_2/2}(s_2-s_1)s_2^{-(1-\alpha)}\int_1^\infty t^{-\alpha}e^{-t}dt$$
$$\ge C_4\alpha(1-s_1/s_2)e^{-s_2/2}$$

for a constant $C_4>0$. Therefore, when we have $m_n/s_1\prec n\epsilon_n^2/s$ and $s_2\prec n\epsilon_n^2/s$, we have

$$-\log\left(\inf_{\|\boldsymbol{x}\|\le\sqrt{m_n}E}\pi_\alpha(\boldsymbol{x})\right)\le\frac{m_n}{2}\log(2\pi s_2)+\frac{m_nE^2}{2s_1}+(1+\nu)\log p_n+s_2/2\prec n\epsilon_n^2/s. \quad (14)$$

Proof of Lemma 3

By assumption, $\min_{j\in\xi^*}\frac{\|\boldsymbol{\beta}_j^*/\sigma^*\|}{\sqrt{m_n}}>\epsilon_n$, $\max_{j\in\xi^*}\frac{\|\boldsymbol{\beta}_j^*/\sigma^*\|}{\sqrt{m_n}}<E$ and

$$\pi_\alpha(\boldsymbol{x})=\int_0^\infty\int_0^\infty\phi(\boldsymbol{x}|\lambda)p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda$$
$$=\int_0^\infty(2\pi\lambda)^{-m_n/2}e^{-\|\boldsymbol{x}\|^2/2\lambda}\int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda$$

where the function of λ in the integrand $f^*(\lambda)\overset{\text{def}}{=}(2\pi\lambda)^{-m_n/2}e^{-\|\boldsymbol{x}\|^2/2\lambda}$ is concave and it is increasing on $(0,\|\boldsymbol{x}\|^2/m_n]$ and decreasing on $(\|\boldsymbol{x}\|^2/m_n,\infty)$. Then we can separate the integral by $\|\boldsymbol{x}\|^2/m_n \log p_n$ so that $f^*(\lambda)$ is increasing until the point $\|\boldsymbol{x}\|^2/m_n$. Since

$$\int_0^{\|\boldsymbol{x}\|^2/m_n \log p_n} f^*(\lambda) \int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda \leq \left(\frac{2\pi\|\boldsymbol{x}\|^2}{m_n \log p_n}\right)^{-m_n/2} e^{-m_n \log p_n/2} \tag{15}$$

$$\begin{aligned}\int_{\|\boldsymbol{x}\|^2/m_n \log p_n}^\infty f^*(\lambda) \int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda \geq & \int_{2\|\boldsymbol{x}\|^2/m_n \log p_n}^{\|\boldsymbol{x}\|^2/m_n} f^*(\lambda) \int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda \\ \geq & (2\pi\|\boldsymbol{x}\|^2/m_n)^{-m_n/2} e^{-m_n \log p_n/4} \mathrm{P}_\alpha\left(\frac{2\|\boldsymbol{x}\|^2}{m_n \log p_n} < \lambda < \frac{\|\boldsymbol{x}\|^2}{m_n}\right) \\ \geq & C_5 (2\pi\|\boldsymbol{x}\|^2/m_n)^{-m_n/2} e^{-m_n \log p_n/4} \alpha e^{-\|\boldsymbol{x}\|^2/2m_n} \end{aligned} \tag{16}$$

for $C_5 > 0$ a constant free of n. The inquality in (16) is obtained from the conlcusion in the proof of Lemma 2. Then for $m_n > n\epsilon_n^2$, we can prove that

$$\begin{aligned} & \frac{\int_0^{\|\boldsymbol{x}\|^2/m_n \log p_n} f^*(\lambda) \int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda}{\int_{\|\boldsymbol{x}\|^2/m_n \log p_n}^\infty f^*(\lambda) \int_0^\infty p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda} \leq \frac{(\log p_n)^{m_n/2} e^{-m_n \log p_n/2}}{C_5 e^{-m_n \log p_n/4} \alpha e^{-\|\boldsymbol{x}\|^2/2m_n}} \\ =& \exp\left\{ -\log C_5 + \frac{m_n}{2} \log\log p_n - \frac{m_n \log p_n}{4} + (1+\nu)\log p_n + \frac{\|\boldsymbol{x}\|^2}{2m_n}\right\} \\ \leq & \exp\{ -C_5' n\epsilon_n^2\} \end{aligned}$$

for a constant $C_5' > 0$. The inequality on the ratio implies that, for a given $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|^2/\sqrt{m_n} \in (\epsilon_n, E)$, the prior density of $\boldsymbol{x}$ has the dominated mass on $\lambda \in (\|\boldsymbol{x}\|^2/m_n \log p_n, \infty)$:

$$\frac{\pi_\alpha(\boldsymbol{x})}{\int_{\|\boldsymbol{x}\|^2/m_n \log p_n}^\infty \int_0^\infty \phi(\boldsymbol{x}|\lambda)p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda} \asymp 1. \tag{17}$$

From (17), we have the following results. For any $\|\boldsymbol{x}_1\|$ and $\|\boldsymbol{x}_2\|$ such that $\|\boldsymbol{x}_1\| < \|\boldsymbol{x}_2\|$, contributions of λ to $\pi_\alpha(\boldsymbol{x}_1)$ and $\pi_\alpha(\boldsymbol{x}_2)$ outside of the interval $\lambda \in (s_3, \infty)$ are negligible, where $s_3 \stackrel{\text{def}}{=} \min\left(\|\boldsymbol{x}_1\|^2/m_n \log p_n, \|\boldsymbol{x}_2\|^2/m_n \log p_n\right)$. Then we have

$$
\begin{aligned}
l_n =& \max_{j\in\xi^*} \sup_{\|\boldsymbol{x}_1\|,\|\boldsymbol{x}_2\|\in\|\boldsymbol{\beta}_j^*/\sigma^*\|\pm C_0\sqrt{m_n}\epsilon_n} \frac{\pi_\alpha(\boldsymbol{x}_1)}{\pi_\alpha(\boldsymbol{x}_2)} \\
\asymp& \max_{j\in\xi^*} \sup_{\|\boldsymbol{x}_1\|,\|\boldsymbol{x}_2\|\in\|\boldsymbol{\beta}_j^*/\sigma^*\|\pm C_0\sqrt{m_n}\epsilon_n} \frac{\int_{s_3}^\infty \int_0^\infty \left(\frac{\phi_\lambda(\boldsymbol{x}_1)}{\phi_\lambda(\boldsymbol{x}_2)}\right)\phi_\lambda(\boldsymbol{x}_2)p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda}{\int_{s_3}^\infty \int_0^\infty \phi_\lambda(\boldsymbol{x}_2)p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda} \\
=& \max_{j\in\xi^*} \sup_{\|\boldsymbol{x}_1\|,\|\boldsymbol{x}_2\|\in\|\boldsymbol{\beta}_j^*/\sigma^*\|\pm C_0\sqrt{m_n}\epsilon_n} \frac{\int_{s_3}^\infty \int_0^\infty e^{\frac{1}{2\lambda}(\|\boldsymbol{x}_2\|^2-\|\boldsymbol{x}_1\|^2)}\phi_\lambda(\boldsymbol{x}_2)p(\lambda|\psi)p_\alpha(\psi)d\psi d\lambda}{\int_{s_3}^\infty \int_0^\infty \phi_\lambda(\boldsymbol{x}_2)\pi(\lambda|\psi)\pi_\alpha(\psi)d\psi d\lambda} \\
\leq& \max_{j\in\xi^*} \sup_{\|\boldsymbol{x}_1\|,\|\boldsymbol{x}_2\|\in\|\boldsymbol{\beta}_j^*/\sigma^*\|\pm C_o\sqrt{m_n}\epsilon_n} \exp\left\{\frac{1}{2s_3}\left(\|\boldsymbol{x}_2\|^2 - \|\boldsymbol{x}_1\|^2\right)\right\} \\
\leq& \max_{j\in\xi^*} \exp\left\{2m_n(\|\boldsymbol{\beta}_j^*/\sigma^*\|/\sqrt{m_n} + C_0\epsilon_n)C_0\epsilon_n/s_3\right\}.
\end{aligned}
$$

Hence $s\log l_n \leq C_6 s m_n E\epsilon_n/s_3$ for constant $C_6 > 0$. In view of $sm_n\epsilon_n \prec \log p_n$, it follows that $s\log l_n \prec \log p_n$.

Proof of Theorem 4.1

The proof of Theorem 4.1 follows the results of Theorem A.1 and Theorem A.2 in Song and Liang (2016), but the difference is that we have additional bias in the true model,

$$
Y = \sum_{j=1}^{p} f_j(X_j) + \sigma\boldsymbol{\varepsilon} = B(X)\boldsymbol{\beta} + \sigma\boldsymbol{\delta} + \sigma\boldsymbol{\varepsilon}, \tag{18}
$$

where each additive function is estimated by m_n-dimensional basis expansion. Therefore, Y includes both bias term and linear regression part analogous to the model in Song and Liang (2016). The linear coefficient $\boldsymbol{\beta} = (\boldsymbol{\beta}_1, \dots \boldsymbol{\beta}_{p_n})^T$ is a $p_n \times m_n$-dimensional basis coefficient vector where a multivariate DL prior will be imposed for each covariate. The true model is the additive non-parametric model $\sum_{j=1}^{p} f_j^*(X_j) + \sigma^*\boldsymbol{\varepsilon}$, where $f_j^* \neq 0$ are κ-smooth functions for $j \in \xi^*$ and $f_j^* = 0$ are zero effects for $j \notin \xi^*$. In our proposed model, we estimate each function by basis expansion $B(X_j)\boldsymbol{\beta}_j$, which imposes an extra bias term $\boldsymbol{\delta}$ in (21) due to approximation error. The magnitude of $\boldsymbol{\delta}$ is bounded by a multiplier of $m_n^{-\kappa}$.

For $B_n = \{\text{At least } \tilde{p} \text{ of } \|\boldsymbol{\beta}_j/\sigma\| \text{ is larger than } a_n\}$ with $m_n\tilde{p} \asymp n\epsilon_n^2/\log p_n$ and $m_n\tilde{p_n} \prec n\epsilon_n^2$, $C_n = \{\|B(X)\boldsymbol{\beta} - \sum_{j=1}^p f_j^*(X_j)\| \geq \sigma^*\epsilon_n\} \cup \{\sigma^2/\sigma^{*2} > (1+\epsilon_n)/(1-\epsilon_n)\} \cup \{\sigma^2/\sigma^{*2} < (1-\epsilon_n)/(1+\epsilon_n)\}$ and $A_n = B_n \cup C_n$, we first consider test function $\phi_n = \max\{\phi'_n, \tilde{\phi}_n\}$,

$$\phi'_n = \bigvee_{\{\xi \supset \xi^*, |\xi| \leq \tilde{p}+s\}} 1\left\{\left|\frac{Y^T(I-H_\xi)Y}{(n-m_n|\xi|)\sigma^{*2}} - 1\right| \geq c'_0\epsilon_n\right\} \tag{19}$$

$$\tilde{\phi}_n = \bigvee_{\{\xi \supset \xi^*, |\xi| \leq \tilde{p}+s\}} 1\left\{\|B(X_\xi)\left(B(X_\xi)^T B(X_\xi)\right)^{-1} B(X_\xi)^T Y - \sum_{j\in\xi} f_j^*(X_j)\| \geq \tilde{c}_0\sigma^*\sqrt{n}\epsilon_n/5\right\} \tag{20}$$

for sufficiently large constants c'_0 and $\tilde{c}_0$, where $H_\xi = B(X_\xi)\left(B(X_\xi)^T B(X_\xi)\right)^{-1} B(X_\xi)^T$ is the hat matrix corresponding to the subgroup ξ, and $\sum_{j\in\xi} f_j^*(X_j)$ is the summation of true non-parametric functions for covariates X_j within subgroup ξ. For any ξ that satisfies $\xi \supset \xi^*$, $m_n|\xi| \leq m_n(\tilde{p}+s) \prec n\epsilon_n^2$,

$$\begin{aligned}
&\mathrm{E}_{(f^*,\sigma^{*2})}1\left\{\left|\frac{Y^T(I-H_\xi)Y}{(n-m_n|\xi|)\sigma^{*2}} - 1\right| \geq c'_0\epsilon_n\right\} \\
=&\mathrm{E}_{(f^*,\sigma^{*2})}1\left\{\left|\frac{\boldsymbol{\varepsilon}^T(I-H_\xi)\boldsymbol{\varepsilon}}{n-m_n|\xi|} - 1 + 2\frac{\boldsymbol{\varepsilon}^T(I-H_\xi)\boldsymbol{\delta}}{n-m_n|\xi|} + \frac{\boldsymbol{\delta}^T(I-H_\xi)\boldsymbol{\delta}}{n-m_n|\xi|}\right| \geq c'_0\epsilon_n\right\} \\
\leq&\mathrm{P}\left(\left|\boldsymbol{\varepsilon}^T(I-H_\xi)\boldsymbol{\varepsilon} - (n-m_n|\xi|)\right| + 2\left|\boldsymbol{\varepsilon}^T(I-H_\xi)\boldsymbol{\delta}\right| + \boldsymbol{\delta}^T(I-H_\xi)\boldsymbol{\delta} \geq (n-m_n|\xi|)c'_0\epsilon_n\right).
\end{aligned}$$

Since the magnitude of basis expansion bias is in the order of $m_n^{-\kappa}$, we have

$$\begin{aligned}
\boldsymbol{\delta}^T(I-H_\xi)\boldsymbol{\delta} \leq& \|\boldsymbol{\delta}\|^2 \asymp nm_n^{-2\kappa} \prec (n-m_n|\xi|)\epsilon_n, \\
2\boldsymbol{\varepsilon}^T(I-H_\xi)\boldsymbol{\delta} \leq& 2\|\boldsymbol{\delta}\| \cdot \|\boldsymbol{\varepsilon}\| \asymp 2\sqrt{n}m_n^{-\kappa}\|\boldsymbol{\varepsilon}\|.
\end{aligned}$$

For normally distributed error $\boldsymbol{\varepsilon}$, given the property of chi-square distribution, we have

$$\mathrm{P}\left(2\left|\boldsymbol{\varepsilon}^T(I-H_\xi)\boldsymbol{\delta}\right| \geq (n-m_n|\xi|)c'_0\epsilon_n\right) \leq \mathrm{P}\left(\|\boldsymbol{\varepsilon}\| \geq c''_0 nm_n^{2\kappa}\epsilon_n^2\right) \prec \exp\{-c'''_0 n\epsilon_n^2\}.$$

Combining with the conclusion in (A.3) of Song and Liang (2016), we can prove that

$$\mathrm{E}_{(f^*,\sigma^{*2})}1\left\{\left|\frac{Y^T(I-H_\xi)Y}{(n-m_n|\xi|)\sigma^{*2}}-1\right|\geq c_0'\epsilon_n\right\}\leq\exp\{-c_3n\epsilon_n^2\}. \tag{21}$$

Next, $f_j^*(X_j)$ can be approximated by B-spline basis expansion $B(X_j)\boldsymbol{\beta}_j^*$ with bias term $\boldsymbol{\delta}$, where $\boldsymbol{\beta}_j^*$ is the B-spline basis coefficients for the true non-parametric function. We have

$$\begin{aligned}\left\|B(X)\beta-\sum_{j=1}^{p_n}f_j^*(X_j)\right\|=&\|B(X)(\boldsymbol{\beta}-\boldsymbol{\beta}^*)-\boldsymbol{\delta}\|\\ \leq&\|B(X_\xi)(\boldsymbol{\beta}_\xi-\boldsymbol{\beta}_\xi^*)\|+\|B(X_{\xi^c})\boldsymbol{\beta}_{\xi^c}\|+\|\boldsymbol{\delta}\|,\end{aligned}$$

where $\|B(X_\xi)(\boldsymbol{\beta}_\xi-\boldsymbol{\beta}_\xi^*)\|\leq\sqrt{n}\|\boldsymbol{\beta}_\xi-\boldsymbol{\beta}_\xi^*\|$, $\|B(X_{\xi^c})\boldsymbol{\beta}_{\xi^c}\|\prec\sqrt{n}a_np_n\prec\sqrt{n}\epsilon_n$ and $\|\boldsymbol{\delta}\|\prec\sqrt{n}\epsilon_n$. By the condition that $m_n(\tilde{p}+s)\prec n\epsilon_n^2$, we have, for some constant c_3',

$$\begin{aligned}&\mathrm{E}_{(f^*,\sigma^{*2})}1\{\|B(X_\xi)\left(B(X_\xi)^TB(X_\xi)\right)^{-1}B(X_\xi)^TY-\sum_{j\in\xi}f_j^*(X_j)\|\geq\tilde{c}_0\sigma^*\sqrt{n}\epsilon_n/5\}\\ =&\mathrm{E}_{(f^*,\sigma^{*2})}1\{\|B(X_\xi)(\boldsymbol{\beta}_\xi-\boldsymbol{\beta}_\xi^*)-\big(\sum_{j\in\xi}f_j^*(X_j)-B(X_\xi)\boldsymbol{\beta}_\xi\big)\|\geq\tilde{c}_0\sigma^*\sqrt{n}\epsilon_n/5\}\\ \leq&\mathrm{E}_{(f^*,\sigma^{*2})}1\{\|\left(B(X_\xi)^TB(X_\xi)\right)^{-1}B(X_\xi)^TY-\boldsymbol{\beta}_\xi\|\geq\tilde{c}_0'\sigma^*\epsilon_n/5\}\\ \leq&\mathrm{E}_{(f^*,\sigma^{*2})}1\{\|B(X_\xi)\left(B(X_\xi)^TB(X_\xi)\right)^{-1}B(X_\xi)^T\boldsymbol{\varepsilon}+\left(B(X_\xi)^TB(X_\xi)\right)^{-1}B(X_\xi)^T\boldsymbol{\delta}\|\geq\tilde{c}_0\epsilon_n/5\}\\ \leq&\exp(-c_3'n\epsilon_n^2),\end{aligned} \tag{22}$$

where the inequality is given by (A.4) of Song and Liang (2016) and the fact that $\|\boldsymbol{\delta}\|\asymp\sqrt{n}m_n^{-\kappa}$. Therefore, combining (21) and (22), we can prove that $\mathrm{E}_{(f^*,\sigma^{*2})}\phi_n\leq\exp(-\tilde{c}_3n\epsilon_n^2)$ for some fixed $\tilde{c}_3$, according to (A.5) in Song and Liang (2016).

For the additive model, the process to prove that $\sup_{(\boldsymbol{\beta},\sigma^2)\in C_n}\mathrm{E}_{\boldsymbol{\beta}}(1-\phi_n)$ is similar to the proof of Theorem A.1 in Song and Liang (2016), by replacing their covariate matrix X with

$B(X)$. Therefore, we have inequalities

$$\mathrm{E}_{(f^*,\sigma^{*2})}\phi_n \leq \exp(-cn\epsilon_n^2), \tag{23}$$

$$\sup_{(\boldsymbol{\beta},\sigma^2)\in C_n} \mathrm{E}_{(\boldsymbol{\beta},\sigma^2)}(1-\phi_n) \leq \exp(-c'n\epsilon_n^2). \tag{24}$$

The next step is to show that

$$\lim_n \mathrm{P}\Big\{\frac{m(D_n)}{L^*(D_n)} \geq \exp(-c_4 n\epsilon_n^2)\Big\} > 1 - \exp\{-c_5 n\epsilon_n^2\}, \tag{25}$$

for any positive c_4, where $m(D_n)$ is the marginal likelihood of the data after integrating out the parameters on their priors and $L^*(D_n)$ is the likelihood under the truth so that

$$\frac{m(D_n)}{L^*(D_n)} = \int \frac{(\sigma^*)^n \exp\{-\|Y - B(X)\boldsymbol{\beta}\|^2/2\sigma^2\}}{\sigma^n \exp\{-\|Y - \sum_{j=1}^{p_n} f_j^*(X_j)\|^2/\sigma^{*2}\}}\pi(\boldsymbol{\beta},\sigma^2)d\boldsymbol{\beta}d\sigma^2.$$

From the proof of Theorem A.1 in Song and Liang (2016), we only need to show that, conditional on $0 \leq \sigma^2 - \sigma^{*2} \leq \eta'\epsilon_n^2$, for some constants c_5 and c_6,

$$\mathrm{P}\left(\pi\left(\|Y - B(X)\boldsymbol{\beta}\|^2/2\sigma^2 < \Big\|Y - \sum_{j=1}^{p_n} f_j^*(X_j)\Big\|^2/\sigma^{*2} + c_5 n\epsilon_n^2\right) \geq e^{-c_6 n\epsilon_n^2}\right) \geq 1 - e^{-c_5 n\epsilon_n^2}$$

For our proposed model, each element in $B(X)$ is smaller than 1 and $\|\boldsymbol{\delta}\|^2 \asymp n m_n^{-2\kappa} \prec n\epsilon_n^2$. By the property of chi-square distribution and normal distribution, with probability larger than $1-\exp\{-c_5 n\epsilon_n^2\}$, we have $\|\boldsymbol{\varepsilon}\|^2 \leq n(1+c_5')$, $\|\boldsymbol{\varepsilon}^T B(X)\|_\infty \leq c_5' n\epsilon_n$ for constant c_5'. Then for model parameters $(\boldsymbol{\beta}, \sigma^2)$, we have

$$
\begin{aligned}
&\{\|Y - B(X)\boldsymbol{\beta}\|^2/2\sigma^2 < \|Y - \left(f_1^*(X_1) + \ldots + f_{p_n}^*(X_{p_n})\right)\|^2/\sigma^{*2} + c_5 n\epsilon_n^2\} \\
=&\{\|\boldsymbol{\varepsilon}\sigma^*/\sigma + B(X)(\boldsymbol{\beta}^* - \boldsymbol{\beta})/\sigma + \boldsymbol{\delta}\sigma^*/\sigma\|^2 < \|\boldsymbol{\varepsilon}\|^2 + 2c_5 n\epsilon_n^2\} \\
\supset&\{\|B(X)(\boldsymbol{\beta}^* - \boldsymbol{\beta})/\sigma\| < \sqrt{\|\boldsymbol{\varepsilon}\|^2 + 2c_5 n\epsilon_n^2} - (\sigma^*/\sigma)\|\boldsymbol{\varepsilon}\| - (\sigma^*/\sigma)\|\boldsymbol{\delta}\|\} \\
\supset&\Big\{\sum_{j=1}^{p_n} \|(\boldsymbol{\beta}_j^* - \boldsymbol{\beta}_j)/\sigma\| \le 2\eta\sqrt{m_n}\epsilon_n\Big\} \\
\supset&\Big\{\|\boldsymbol{\beta}_j/\sigma\| \in \left[\|\boldsymbol{\beta}_j^*/\sigma\| - \eta\sqrt{m_n}\epsilon_n/s, \|\boldsymbol{\beta}_j^*/\sigma\| + \eta\sqrt{m_n}\epsilon_n/s\right] \text{ for all } j \in \xi^*\Big\} \\
&\cap \Big\{\sum_{j \notin \xi^*} \|\boldsymbol{\beta}_j/\sigma\| \le \eta\sqrt{m_n}\epsilon_n\Big\}.
\end{aligned}
$$

for some constant η. The second subset relation is given by Assumption $A_1(3)$ such that the eigenvalues of $B(X)^T B(X)$ are in the same order as n/m_n. By Lemma 1 and Lemma 2,

$$
\begin{aligned}
&\pi_\alpha\Big(\sum_{j \notin \xi^*} \big\|\boldsymbol{\beta}_j/\sigma\big\| \le \eta\sqrt{m_n}\epsilon_n \big| 0 \le \sigma^2 - \sigma^{*2} \le \eta'\epsilon_n^2\Big) \\
=&\int_{\sum_{j=1}^{p_n-s}\|\boldsymbol{x}_j\| \le \eta\sqrt{m_n}\epsilon_n} \prod_{j=1}^{p_n-s} \pi(\boldsymbol{x}_i) d\boldsymbol{x}_1 \ldots d\boldsymbol{x}_{p_n-s} \ge \exp\{-c_6' n\epsilon_n^2\} \\
&\pi_\alpha\Big(\|\boldsymbol{\beta}_j/\sigma\| \in \left[\|\boldsymbol{\beta}_j^*/\sigma\| - \eta\sqrt{m_n}\epsilon_n/s, \|\boldsymbol{\beta}_j^*/\sigma\| + \eta\sqrt{m_n}\epsilon_n/s\right] |\{0 \le \sigma^2 - \sigma^{*2} \le \eta'\epsilon_n^2\}\Big) \\
\ge&\eta\epsilon_n \cdot \inf_{\|\boldsymbol{x}\| \le \sqrt{m_n}E} \pi_\alpha(\boldsymbol{x})/s \ge \exp\{-c_6'' n\epsilon_n^2\}.
\end{aligned}
$$

These verify the inequality in (25).

Finally, from the proof of Theorem A.1 in Song and Liang (2016), for B_n defined above, we have $\mathrm{P}(B_n) < e^{-c_7 n\epsilon_n^2}$ for some constants c_7, given that $\int_{\|\boldsymbol{x}\| > a_n} \pi_\alpha(\boldsymbol{x}) d\boldsymbol{x} \le p_n^{-(1+\mu)}$.

Therefore, combining the results of $Pr(B_n) < e^{-c_7 n\epsilon_n^2}$ with (23), (24) and (25), we have

$$
\mathrm{P}^*\left(\pi\Big(\big\|B(X)\boldsymbol{\beta} - \sum_{j=1}^{p_n} f_j^*(X_j)\big\| \ge \sigma^*\sqrt{n}\epsilon_n \text{ or } |\sigma^2 - \sigma^{*2}| > c_4\epsilon_n \big| X, Y\Big) \ge e^{-c_1 n\epsilon_n^2}\right) \le e^{-c_2 n\epsilon_n^2}.
$$

In conclusion, the results in Theorem 4,1 is proved.

Proof of Theorem 4.2

The proof of Theorem 4.2 follows the proof of Theorem A.3 in Song and Liang (2016). Note that in our proposed model, we have basis expansion matrix $B(X)$ instead of covariate matrix X, the coefficients vector $\boldsymbol{\beta} = (\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_{p_n}^T)^T$ are shrunk within blocks $\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_{p_n}$. Furthermore, our model has an additional bias term $\boldsymbol{\delta}$ due to B-spline basis expansion.

We first define the set $S_1 = \{\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| \leq c_1\epsilon_n,\ |\sigma^2 - \sigma^{*2}| \leq c_2\epsilon_n\}$, and $\underline{\pi}(\boldsymbol{\beta}|\sigma^2) = \inf_{(\boldsymbol{\beta},\sigma^2)\in S_1} \pi(\boldsymbol{\beta}, \sigma^2)/\pi(\sigma^2)$, $\overline{\pi}(\boldsymbol{\beta}|\sigma^2) = \sup_{(\boldsymbol{\beta}_1,\sigma^2)\in S_1} \pi(\boldsymbol{\beta}, \sigma^2)/\pi(\sigma^2)$. The maximum of $\mathcal{L}_2$-norm for each group within coefficient vector is defined as $\|\boldsymbol{\beta}\|_{\max} \triangleq \max \|\boldsymbol{\beta}_j\|$. We can conclude by proof of Theorem A.3 in Song and Liang (2016),

$$\begin{aligned} \mathrm{P}(\xi = \xi^*|X, Y) \gtrsim & \pi\left(\left\|\frac{\boldsymbol{\beta}_{(\xi^*)^c}}{\sigma}\right\|_{\max} \leq a_n\right) \underline{\pi}(\boldsymbol{\beta}_{\xi^*}|\sigma^2)\sqrt{\det(B(X_{\xi^*})^T B(X_{\xi^*}))^{-1}} \left(\sqrt{2\pi}\right)^s \\ & \times \inf_{\|\boldsymbol{\beta}_{(\xi^*)^c}\|_{\max} \leq a_n(\sigma^*+c_2\epsilon_n)} \frac{\Gamma(a_0 + (n-s)/2)}{\left(\mathrm{SSE}(\boldsymbol{\beta}_{(\xi^*)^c})/2 + b_0\right)^{a_0+(n-s)/2}}, \end{aligned} \tag{26}$$

where $\mathrm{SSE}(\boldsymbol{\beta}_{(\xi^*)^c})$ is the SSE when $\boldsymbol{\beta}_{(\xi^*)^c}$ is given and fixed.

Next, we let $\|\boldsymbol{\beta}\|_{\min} \triangleq \min \|\boldsymbol{\beta}_j\|$ be the minimum of $\mathcal{L}_2$-norms for each predictor within coefficient vector. Similarly, we can show that

$$\begin{aligned} \mathrm{P}(\xi = \xi'|X, Y) \lesssim & \ \pi\left(\left\|\frac{\boldsymbol{\beta}_{\xi'\setminus\xi^*}}{\sigma}\right\|_{\min} > a_n, \left\|\frac{\boldsymbol{\beta}_{(\xi')^c}}{\sigma}\right\|_{\max} \leq a_n\right) \overline{\pi}(\boldsymbol{\beta}_{\xi^*}|\sigma^2)\sqrt{\det(B(X_{\xi^*})^T B(X_{\xi^*}))^{-1}} \\ & \times \left(\sqrt{2\pi}\right)^s \sup_{\|\boldsymbol{\beta}_{(\xi')^c}\|_{\max} \leq a_n(\sigma^*+c_2\epsilon_n)} \frac{\Gamma(a_0 + (n-s)/2)}{\left(\mathrm{SSE}(\boldsymbol{\beta}_{(\xi')^c})/2 + b_0\right)^{a_0+(n-s)/2}}, \end{aligned}$$

where $\mathrm{SSE}(\boldsymbol{\beta}_{(\xi')^c})$ is the SSE when $\boldsymbol{\beta}_{(\xi')^c}$ is given and fixed.

Using the fact that the basis expansion bias is $\boldsymbol{\delta}$ such that $\|\boldsymbol{\delta}\|^2 \asymp n m_n^{-2\kappa}$, we can follow the derivation of (A.13) and (A.14) of Song and Liang (2016). With probability larger than

$1 - 4p_n \cdot p_n^{-c_6'}$,

$$
\begin{aligned}
\mathrm{SSE}(\boldsymbol{\beta}_{(\xi^*)^c}) =&(Y - B(X_{(\xi^*)^c})\boldsymbol{\beta}_{(\xi^*)^c})^T(I - H_{\xi^*})(Y - B(X_{(\xi^*)^c})\boldsymbol{\beta}_{(\xi^*)^c}) \\
\leq&\sigma^{*2}\boldsymbol{\varepsilon}^T(I - H_{\xi^*})\boldsymbol{\varepsilon} + \sigma^{*2}\boldsymbol{\delta}^T(I - H_{\xi^*})\boldsymbol{\delta} + \left\|B(X_{(\xi^*)^c})\boldsymbol{\beta}_{(\xi^*)^c}\right\|^2 \\
&+ 2\sigma^{*2}\sqrt{2c_6' n \log p_n} \sum_{j \in (\xi^*)^c} \|\boldsymbol{\beta}_j\|; \qquad (27)
\end{aligned}
$$

$$
\mathrm{SSE}(\boldsymbol{\beta}_{(\xi')^c}) \geq \sigma^{*2}\boldsymbol{\varepsilon}^T(I - H_{\xi'})\boldsymbol{\varepsilon} + \sigma^{*2}\boldsymbol{\delta}^T(I - H_{\xi'})\boldsymbol{\delta} - 2\sigma^*\sqrt{2c_6' n \log p_n} \sum_{j \in (\xi')^c} \|\boldsymbol{\beta}_j\|, \qquad (28)
$$

and with probability $1 - p_n^{-c_6''}$ for any constant c_6'', we have

$$
\boldsymbol{\varepsilon}^T(I - H_{\xi^*})\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}^T(I - H_{\xi'})\boldsymbol{\varepsilon} \leq c_7' m_n |\xi' \setminus \xi^*| \log p_n.
$$

Since $\boldsymbol{\delta}^T(I - H_{\xi^*})\boldsymbol{\delta} - \boldsymbol{\delta}^T(I - H_{\xi'})\boldsymbol{\delta} \leq c_7'' m_n |\xi' \setminus \xi^*| n m_n^{-2\kappa}$, we can conclude that

$$
\begin{aligned}
&\frac{\sup_{\|\boldsymbol{\beta}_{(\xi^*)^c}\|_{\max} \leq a_n(\sigma^* + c_2\epsilon_n)} (\mathrm{SSE}(\boldsymbol{\beta}_{(\xi^*)^c})/2 + b_0)^{a_0 + (n-s)/2}}{\inf_{\|\boldsymbol{\beta}_{(\xi')^c}\|_{\max} \leq a_n(\sigma^* + c_2\epsilon_n)} (\mathrm{SSE}(\boldsymbol{\beta}_{(\xi')^c})/2 + b_0)^{a_0 + (n-s)/2}} \\
\leq& \exp\{c_8' m_n |\xi' \setminus \xi^*| \log p_n + c_8'' m_n |\xi' \setminus \xi^*| n m_n^{-2\kappa}\}. \qquad (29)
\end{aligned}
$$

If $m_n \succ (\frac{n}{\log p_n})^{1/2\kappa}$, we have $n m_n^{-2\kappa} \prec \log p_n$, so that the right hand side of (29) is bounded by $c_8 m_n |\xi' \backslash \xi^*| \log p_n$ for some constant $c_8 > 0$. Given (6) in Lemma 3, $\overline{\pi}(\boldsymbol{\beta}_{\xi^*}|\sigma^2)/\underline{\pi}(\boldsymbol{\beta}_{\xi^*}|\sigma^2) \leq l_n^s$. We can proceed with

$$
\begin{aligned}
\frac{\mathrm{P}(\xi = \xi'|X, Y)}{\mathrm{P}(\xi = \xi^*|X, Y)} \leq& l_n^s \frac{\sigma^* + c_2\epsilon_n}{\sigma^* - c_2\epsilon_n}[p_n^{-(1+\mu)}/(1 - p_n^{-(1+\mu)})]^{|\xi' \backslash \xi^*|} \exp\{c_8 m_n |\xi' \setminus \xi^*| \log p_n\} \\
\leq& l_n^s \frac{\sigma^* + c_2\epsilon_n}{\sigma^* - c_2\epsilon_n} p_n^{-\mu' m_n |\xi' \backslash \xi^*|},
\end{aligned}
$$

with proper choice of the constants.

Therefore, with $\min_{j \in \xi^*} \|\boldsymbol{\beta}_j\| \succ \sqrt{m_n}\epsilon_n$, following the proof of Theorem A.2 in Song and

Liang (2016), we can prove that

$$\sum_{\xi' \supsetneq \xi^*} \mathrm{P}\left(\xi = \xi' | X, Y\right) \leq \pi\left(\|\boldsymbol{\beta} - \boldsymbol{\beta}^*\| > \left| \min_{j \in \xi^*} \|\boldsymbol{\beta}_j\| - a_n \sigma \right| \Big| X, Y \right) \leq e^{-cn\epsilon_n^2}$$

In conclusion, we have $\mathrm{P}\Big\{\pi\big(\xi(a_n) = \xi^* | X, Y\big) > 1 - p_n^{-\mu''}\Big\} > 1 - p_n^{-\mu'}$ for some constants $\mu', \mu'' > 0$.