Consistency of BIC Model Averaging

Ze Chen[1], Jianqiang Zhang[1], Wangli Xu[1] and Yuhong Yang[2]

[1]*Renmin University of China and* [2]*University of Minnesota*

Supplementary Material

This supplement contains the proofs of Theorems 1 and Theorem 2.

S1 Proof of Theorem 1

The following lemma is presented in order to prove Theorems 1 and 2, and the specific proof of the lemma can be obtained in Luo and Chen (2013).

Lemma 1. *Let* $C_j = 2j\{\log p^* + \log(j \log p^*)\}$*, as* $p^* \to \infty$*, for any* $J \le p^*$*,*

$$\sum_{j=1}^{J} \binom{p^*}{j} P(\chi_j^2 > C_j) \to 0,$$

where χ_j^2 *is a chi-square random variable with degrees of freedom* j*.*

Without loss of generality, we assume $\sigma^2 = 1$. In the remainder of the paper, we assume X_M contains a p^*-dimensional vector of ones. Write $A \subsetneq B$ if $A \subset B$ and $A \neq B$. For notational clarity, let $\mathcal{M}_0 \overset{\text{def}}{=} \{M \in \mathcal{M} : M_0 \subsetneq M\}$ and $\mathcal{M}_1 \overset{\text{def}}{=} \{M \in \mathcal{M} : M_0 \not\subset M\}$. Further, we split

$\mathcal{M}_0$ into $\mathcal{M}_{0,1} \stackrel{\text{def}}{=} \{M \in \mathcal{M}_0 : |M| \leq k|M_0|\}$ and $\mathcal{M}_{0,2} \stackrel{\text{def}}{=} \{M \in \mathcal{M}_0 : k|M_0| < |M| \leq (p^*)^\alpha \wedge (Cn/\log p^*)\}$. Similarly, $\mathcal{M}_1$ can be split into $\mathcal{M}_{1,1} \stackrel{\text{def}}{=} \{M \in \mathcal{M}_1 : |M| \leq k|M_0|\}$ and $\mathcal{M}_{1,2} \stackrel{\text{def}}{=} \{M \in \mathcal{M}_1 : k|M_0| < |M| \leq (p^*)^\alpha \wedge (Cn/\log p^*)\}$. Let $m \stackrel{\text{def}}{=} |M|$ and $m_0 \stackrel{\text{def}}{=} |M_0|$. According to the definition of the w_M in (2.2), we have $w_M/w_{M_0} = \exp(-T_1 - T_2)$, where $T_1 \stackrel{\text{def}}{=} (n/2)\log(RSS_M/RSS_{M_0})$ and $T_2 \stackrel{\text{def}}{=} 2^{-1}(m-m_0)\log n + \psi(m-m_0)(1+\log p^* - \log m) - \psi m_0 \log(m/m_0) + 2\psi \log((m+2)/(m_0+2))$. We only need to show that T_1+T_2 converges to infinity uniformly for all $M \in \mathcal{M}$ s.t. $M \neq M_0$ in order to prove the equality in (2.3), and two scenarios, $M \in \mathcal{M}_1$ and $M \in \mathcal{M}_0$, are considered for certification.

We first prove that $T_1 + T_2$ converges to infinity under the scenario $M \in \mathcal{M}_1$. It is notable that RSS_{M_0} follows the chi-square distribution with degrees of freedom $n - m_0$ and we can obtain $RSS_{M_0} = n(1 + o_p(1))$ by the assumption $m_0 \log p^* = o(n)$ which implies that $m_0 = o(n)$. Let $H_M \stackrel{\text{def}}{=} X_M(X_M^{\mathrm{T}}X_M)^{-1}X_M^{\mathrm{T}}$, $\mu \stackrel{\text{def}}{=} X_{M_0}\beta_{M_0}$ and $\Delta_M \stackrel{\text{def}}{=} \mu^{\mathrm{T}}(I - H_M)\mu$, the term $RSS_M - RSS_{M_0}$ can be rewritten as

$$RSS_M - RSS_{M_0} = \Delta_M + 2\mu^{\mathrm{T}}(I - H_M)\epsilon - \epsilon^{\mathrm{T}}H_M\epsilon + \epsilon^{\mathrm{T}}H_{M_0}\epsilon. \qquad \text{(S.1)}$$

Below, we will prove T_1+T_2 converges to infinity separately under $M \in \mathcal{M}_{1,1}$ and $M \in \mathcal{M}_{1,2}$. We first show that $RSS_M - RSS_{M_0} = \Delta_M(1 + o_p(1))$ holds uniformly for all $M \in \mathcal{M}_{1,1}$.

Consider the term $2\mu^{\mathrm{T}}(I-H_M)\epsilon$ in (S.1) and write $Z_M=\mu^{\mathrm{T}}(I-H_M)\epsilon/\sqrt{\Delta_M}$. By the properties of the multivariate normal distribution, we have $Z_M\sim N(0,1)$. Let $\mathcal{M}_1^j \overset{\text{def}}{=} \{M : M\in\mathcal{M}_1, |M|=j\}$ be the set of size j from $\mathcal{M}_1$. Put $L=[(p^*)^\alpha\wedge(Cn/\log p^*)]$, where $[x]$ denotes the largest integer not exceeding x. By Lemma 1 and the Bonferroni inequality,

$$P\Big(\max_{M\in\mathcal{M}_1}|Z_M/\sqrt{C_m}|>1\Big)\le\sum_{j=1}^{L}\sum_{M\in\mathcal{M}_1^j}P(Z_M^2>C_j)$$
$$<\sum_{j=1}^{L}\binom{p^*}{j}P(\chi_1^2>C_j)<\sum_{j=1}^{L}\binom{p^*}{j}P(\chi_j^2>C_j)\to 0.$$

Therefore, $|\mu^{\mathrm{T}}(I-H_M)\epsilon|=\sqrt{\Delta_M}\,|Z_M|\le(\Delta_M C_m)^{1/2}(1+o_p(1))$ uniformly over $\mathcal{M}_1$. From Conditions 2−3, we deduce that

$$\begin{aligned}\lim_{n\to\infty}\min_{M\in\mathcal{M}_{1,1}}\left\{\frac{\Delta_M}{m_0\log p^*}\right\}&\ge\frac{n}{m_0\log p^*}\lambda_{\min}\Big(\frac{1}{n}X_{M\cup M_0}^{\mathrm{T}}X_{M\cup M_0}\Big)\|\beta_{M_0\setminus M}\|_2^2\\&\ge\frac{n}{m_0\log p^*}\lambda_{\min}\Big(\frac{1}{n}X_{M\cup M_0}^{\mathrm{T}}X_{M\cup M_0}\Big)\min\{\|\beta_{\mathcal{I}_j}\|_2^2:\mathcal{I}_j\subset M_0\}\\&\ge c_1\Big(\frac{n}{m_0\log p^*}\Big)^{\varepsilon}\lambda_{\min}\Big(\frac{1}{n}X_{M\cup M_0}^{\mathrm{T}}X_{M\cup M_0}\Big)\to\infty,\end{aligned}$$

where $M_0\setminus M$ refers to all indices that are in set M_0 but not in set M and $\beta_{M_0\setminus M}$ denotes the vector consisting of the components of β with indices in $M_0\setminus M$. This gives $m_0\log p^*=o(\Delta_M)$ uniformly over $\mathcal{M}_{1,1}$. Since $C_m=O(m_0\log p^*)$ uniformly over $\mathcal{M}_{1,1}$, it follows that $|\mu^{\mathrm{T}}(I-H_M)\epsilon|=o_p(\Delta_M)$ uniformly over $\mathcal{M}_{1,1}$.

For the term $\epsilon^{\mathrm{T}} H_M \epsilon$ in (S.1), invoking Lemma 1, we have

$$P\Big(\bigcup_{M\in\mathcal{M}_1}\{\epsilon^{\mathrm{T}} H_M\epsilon > C_m\}\Big) \le \sum_{j=1}^{L}\sum_{M\in\mathcal{M}_1^j} P(\epsilon^{\mathrm{T}} H_M\epsilon > C_j) < \sum_{j=1}^{L}\binom{p^*}{j} P(\chi_j^2 > C_j) \to 0.$$

Consequently, $\epsilon^{\mathrm{T}} H_M \epsilon \le C_m\left(1+o_p(1)\right) = O(m_0 \log p^*) = o_p\left(\Delta_M\right)$ uniformly over $\mathcal{M}_{1,1}$. In addition, $\epsilon^{\mathrm{T}} H_{M_0}\epsilon = m_0(1+o_p(1)) = o_p\left(\Delta_M\right)$ since $\epsilon^{\mathrm{T}} H_{M_0}\epsilon$ is a random variable that follows chi-square distribution with degrees of freedom m_0.

According to the aforementioned conclusions that $|\mu^{\mathrm{T}}(I-H_M)\epsilon| = o_p(\Delta_M)$ and $\epsilon^{\mathrm{T}} H_M\epsilon = o_p\left(\Delta_M\right)$ uniformly over $\mathcal{M}_{1,1}$, we have $RSS_M - RSS_{M_0} = \Delta_M\left(1+o_p(1)\right)$ uniformly over $\mathcal{M}_{1,1}$ and correspondingly,

$$\log\Big(\frac{RSS_M}{RSS_{M_0}}\Big) = \log\Big(1+\frac{RSS_M - RSS_{M_0}}{RSS_{M_0}}\Big) = \log\Big(1+\frac{\Delta_M}{n}(1+o_p(1))\Big)$$

uniformly over $\mathcal{M}_{1,1}$. For any $K>0$, under the assumption $m_0\log p^* = o(n)$,

$$\begin{aligned} T_1 &= \frac{n}{2}\log\Big(1+\frac{\Delta_M}{n}(1+o_p(1))\Big) \ge \frac{n}{2}\log\Big(1+\frac{Km_0\log p^*}{n}(1+o_p(1))\Big) \\ &= \frac{n}{2}\Big(\frac{Km_0\log p^*}{n}\Big)(1+o_p(1)) = \frac{Km_0\log p^*}{2}(1+o_p(1)) \end{aligned} \tag{S.2}$$

uniformly over $\mathcal{M}_{1,1}$.

For T_2, under the assumptions in Theorem 1 and $M\in\mathcal{M}_{1,1}$, we obtain

$$\begin{aligned} &\frac{(m-m_0)\log n}{2m_0\log p^*} \ge -\frac{\eta}{2}, \quad -\psi\frac{\log\left(m/m_0\right)}{\log p^*} \ge -\psi\frac{\log k}{\log p^*} = o(1), \\ &\psi\frac{(m-m_0)\left(1+\log p^* - \log m\right)}{m_0\log p^*} > -\psi(1+o(1)), \end{aligned}$$

and $2\psi\log((m+2)/(m_0+2))/(m_0\log p^*) \to 0$ as $n \to \infty$.

As a result, $\min_{M\in\mathcal{M}_{1,1}} T_2 \ge (-\eta/2-\psi)(1+o_p(1))m_0\log p^*$. Putting this together with (S.2), we have

$$\min_{M\in\mathcal{M}_{1,1}} (T_1+T_2) \ge (K/2-\eta/2-\psi)(1+o_p(1))m_0\log p^*. \qquad \text{(S.3)}$$

Choosing $K > 2\psi+\eta$, we conclude that $\min_{M\in\mathcal{M}_{1,1}}(T_1+T_2) \to \infty$. Further, $\max_{M\in\mathcal{M}_{1,1}} w_M/w_{M_0} = \max_{M\in\mathcal{M}_{1,1}} \exp(-T_1-T_2) \xrightarrow{P} 0$.

Now, we consider the proof under the case $M \in \mathcal{M}_{1,2}$. As $n \to \infty$, we can obtain from (S.1) and a elementary calculation that

$$\begin{aligned} RSS_M - RSS_{M_0} &\ge (\Delta_M - 2(\Delta_M C_m)^{1/2} - C_m)(1+o_p(1)) + \epsilon^{\mathrm{T}} H_{M_0}\epsilon \\ &\ge -4m(1+(\alpha\wedge\eta))(1+o_p(1))\log p^* \end{aligned}$$

uniformly over $\mathcal{M}_{1,2}$. Note that $x\log(1+1/x)$ is strictly increasing for $x<-1$, and so we can derive

$$\begin{aligned} T_1 &= \frac{n}{2}\log\left(1+\frac{RSS_M - RSS_{M_0}}{RSS_{M_0}}\right) \\ &\ge \frac{n}{2}\log\left(1-\frac{4m(1+(\alpha\wedge\eta))}{n}(1+o_p(1))\log p^*\right) \\ &\ge \frac{\log(1-4C(1+(\alpha\wedge\eta)))}{2C} m\log p^*(1+o_p(1)) \\ &\ge \frac{k\log(1-4C(1+(\alpha\wedge\eta)))}{2C(k-1)}(m-m_0)\log p^*(1+o_p(1)) \qquad \text{(S.4)} \end{aligned}$$

uniformly over $\mathcal{M}_{1,2}$ as $n \to \infty$ when $0 < C < 1/(4(1+(\alpha\wedge\eta)))$. Next, we turn to dealing with T_2. Under the assumptions in Theorem 1, it is straightforward to show that

$$\frac{m_0}{(m-m_0)\log p^*}\log\left(\frac{m}{m_0}\right)\to 0,\quad \frac{\log(m+2)-\log(m_0+2)}{(m-m_0)\log p^*}\to 0,$$

$$\text{and } \frac{\log n}{2\log p^*}+\psi\frac{1+\log p^*-\log m}{\log p^*}\geq \frac{\log m}{2\log p^*}+\psi\frac{1+\log p^*-\log m}{\log p^*}$$
$$\geq((\alpha\wedge\eta)/2+\psi(1-(\alpha\wedge\eta)))(1+o(1))$$

as $n\to\infty$. Hence,

$$T_2\geq((\alpha\wedge\eta)/2+\psi(1-(\alpha\wedge\eta)))(1+o_p(1))(m-m_0)\log p^* \tag{S.5}$$

uniformly over $\mathcal{M}_{1,2}$. Combining (S.4) and (S.5), we can derive that

$$\min_{M\in\mathcal{M}_{1,2}}(T_1+T_2)\geq\left(\frac{\alpha\wedge\eta}{2}+\psi(1-(\alpha\wedge\eta))+\frac{k\log(1-4C(1+(\alpha\wedge\eta)))}{2C(k-1)}\right)$$
$$\times(m-m_0)\log p^*(1+o_p(1)). \tag{S.6}$$

Thus, if we have

$$\psi>\frac{k\log(1-4C(1+(\alpha\wedge\eta)))}{2C(k-1)((\alpha\wedge\eta)-1)}-\frac{\alpha\wedge\eta}{2(1-(\alpha\wedge\eta))}, \tag{S.7}$$

we can obtain $\min_{M\in\mathcal{M}_{1,2}}(T_1+T_2)\to\infty$ as $n\to\infty$.

Below we prove that T_1+T_2 tends to infinity uniformly for all $M\in\mathcal{M}_0$. Note that $RSS_{M_0}-RSS_M\sim\chi^2_{m-m_0}$. Let $\mathcal{M}_0^j\overset{\text{def}}{=}\{M:M\in\mathcal{M}_0,|M|=j\}$. Recall that $C_j=2j\{\log p^*+\log(j\log p^*)\}$. Now, invoking Lemma 1 and the Bonferroni inequality, we have

$$
\begin{aligned}
& P\Big(\bigcup_{(m_0+1)\le j\le L} \Big\{ \bigcup_{M\in\mathcal{M}_0^j} \{(RSS_{M_0} - RSS_M) \ge C_{j-m_0}\}\Big\}\Big) \\
\le & \sum_{j=m_0+1}^{L} P\Big(\bigcup_{M\in\mathcal{M}_0^j} \{(RSS_{M_0} - RSS_M) \ge C_{j-m_0}\}\Big) \\
\le & \sum_{j=m_0+1}^{L} \binom{p^*-m_0}{j-m_0} P(\chi^2_{j-m_0} \ge C_{j-m_0}) < \sum_{j=1}^{L} \binom{p^*}{j} P(\chi^2_j \ge C_j) \to 0.
\end{aligned}
$$

This implies that $RSS_{M_0} - RSS_M \le C_{m-m_0}\,(1+o_p(1))$ uniformly over $\mathcal{M}_0$. Recall that $\mathcal{M}_{0,1} = \{M \in \mathcal{M}_0 : |M| \le k|M_0|\}$ and $\mathcal{M}_{0,2} = \{M \in \mathcal{M}_0 : k|M_0| < |M| \le (p^*)^\alpha \wedge (Cn/\log p^*)\}$. Similarly to before, we divide the proof into two cases: $M \in \mathcal{M}_{0,1}$ and $M \in \mathcal{M}_{0,2}$.

For $M \in \mathcal{M}_{0,1}$, note that $C_{m-m_0} = o(n)$ and $RSS_M = RSS_{M_0} - (RSS_{M_0} - RSS_M) = n(1+o_p(1))$ uniformly over $\mathcal{M}_{0,1}$, we have

$$
\begin{aligned}
T_1 &= -\frac{n}{2}\log\Big(1 + \frac{RSS_{M_0} - RSS_M}{RSS_M}\Big) \ge -\frac{n}{2}\Big(\frac{RSS_{M_0} - RSS_M}{RSS_M}\Big) \\
&\ge -\frac{C_{m-m_0}}{2}\,(1+o_p(1)) \ge -\,(m-m_0)\,(1+o_p(1))\log p^* \\
&\quad \times \Big[1 + \frac{\log((k-1)m_0\log p^*)}{\log p^*}\Big] \ge -\,(m-m_0)\,(1+\delta)\,(1+o_p(1))\log p^*
\end{aligned}
\tag{S.8}
$$

uniformly over $\mathcal{M}_{0,1}$. Moreover, under the assumptions in Theorem 1, it is straightforward to check for T_2 that

$$
\frac{m_0}{(m-m_0)\log p^*}\log\Big(\frac{m}{m_0}\Big) \to 0, \quad \frac{\log(m+2) - \log(m_0+2)}{(m-m_0)\log p^*} \to 0,
$$

$$
\text{and } \frac{\log n}{2\log p^*} + \psi\frac{1+\log p^* - \log m}{\log p^*} \ge (\frac{\delta}{2} + \psi(1-\delta))(1+o(1))
$$

as $n \to \infty$. This leads to $T_2 \geq (\delta/2 + \psi(1-\delta))(1 + o_p(1))(m - m_0) \log p^*$ uniformly over $\mathcal{M}_{0,1}$. Combining this with (S.8), we obtain

$$\min_{M \in \mathcal{M}_{0,1}} (T_1 + T_2) \geq (\delta/2 + \psi(1-\delta) - (1+\delta))(1 + o_p(1))(m - m_0) \log p^*. \quad \text{(S.9)}$$

Clearly, $m > m_0$ for all $M \in \mathcal{M}_0$. As $n \to \infty$, $p^* \to \infty$, whenever

$$\psi > (1 + \delta/2)/(1 - \delta), \quad \text{(S.10)}$$

then $\min_{M \in \mathcal{M}_{0,1}} (T_1 + T_2) \to \infty$.

For $M \in \mathcal{M}_{0,2}$, we can see from (S.8) that

$$\begin{aligned} T_1 &\geq -\frac{n}{2}\Big(\frac{RSS_{M_0} - RSS_M}{RSS_M}\Big) \\ &\geq \frac{-(m - m_0)(1 + (\alpha \wedge \eta))(1 + o_p(1)) \log p^*}{1 - 2(m - m_0)(1 + (\alpha \wedge \eta))(1 + o_p(1)) \log p^*/n} \end{aligned} \quad \text{(S.11)}$$

as $n \to \infty$. In addition, the conclusion for T_2 can be drawn by the same argument as in the proof of (S.5). Combining this with (S.11), we can also derive that

$$\begin{aligned} \min_{M \in \mathcal{M}_{0,2}} (T_1 + T_2) \geq &\Big(\frac{\alpha \wedge \eta}{2} + \psi(1 - (\alpha \wedge \eta)) - \frac{1 + (\alpha \wedge \eta)}{1 - 2C(1 + (\alpha \wedge \eta))}\Big) \\ &\times (m - m_0) \log p^* (1 + o_p(1)). \end{aligned} \quad \text{(S.12)}$$

Further, if we have the following condition

$$\psi > \frac{(1 + (\alpha \wedge \eta))/(1 - 2C(1 + (\alpha \wedge \eta))) - (\alpha \wedge \eta)/2}{(1 - (\alpha \wedge \eta))}, \quad \text{(S.13)}$$

then $\min_{M\in\mathcal{M}_{0,2}}(T_1+T_2)\to\infty$ as $n\to\infty$. It should be noted that (S.10) and (S.13) are automatically satisfied when (S.7) holds due to $\delta<(\alpha\wedge\eta)$. Therefore, when ψ satisfies (S.7), the conclusion (2.3) follows.

Next, on the basis of the above conclusions, we can prove that BIC-p weighting is consistent. For each given candidate model $M_i\in\mathcal{M}_0$, we have $|M_i\nabla M_0|=|M_i\setminus M_0|=|M_i|-|M_0|$. Besides, for a given candidate model $M_i\in\mathcal{M}_1$, we have $|M_i\nabla M_0|=|M_i\setminus M_0|+|M_0\setminus M_i|\le|M_i|+|M_0|$.

Since

$$\begin{aligned}\sum_{i=1}^{N}w_i|M_i\nabla M_0| &= \sum_{M\in\mathcal{M}_1}w_M|M\nabla M_0|+\sum_{M\in\mathcal{M}_0}w_M|M\nabla M_0|\\ &\le \sum_{M\in\mathcal{M}_1}\frac{w_M}{w_{M_0}}|M\nabla M_0|+\sum_{M\in\mathcal{M}_0}\frac{w_M}{w_{M_0}}|M\nabla M_0|, \qquad \text{(S.14)}\end{aligned}$$

we only need to show that the two terms in (S.14) converge to 0 in probability as n tends to infinity. The first term in (S.14) can be written as

$$\sum_{M\in\mathcal{M}_1}\frac{w_M}{w_{M_0}}|M\nabla M_0|=\sum_{M\in\mathcal{M}_{1,1}}\frac{w_M}{w_{M_0}}|M\nabla M_0|+\sum_{M\in\mathcal{M}_{1,2}}\frac{w_M}{w_{M_0}}|M\nabla M_0|\overset{\text{def}}{=}T_{1,1}+T_{1,2}.$$

Applying (S.3) and the fact that $|M\nabla M_0|\le(k+1)m_0$ for $M\in\mathcal{M}_{1,1}$ yields

$$\begin{aligned}T_{1,1} &< (k+1)m_0\sum_{j=1}^{km_0}\binom{p^*}{j}\Big(\max_{M\in\mathcal{M}_1^j}\frac{w_M}{w_{M_0}}\Big)\\ &< (k+1)m_0\sum_{j=1}^{km_0}\exp\{j\log p^*-(K/2-\eta/2-\psi)(1+o_p(1))m_0\log p^*\}\\ &< k(k+1)m_0^2\exp\{-m_0\log p^*(K/2-\eta/2-\psi-k)(1+o_p(1))\}\xrightarrow{P}0,\end{aligned}$$

by choosing sufficiently large K. Now combining (S.6) and the assumption in Theorem 1 that

$$\psi > \frac{k\log(1-4C(1+(\alpha\wedge\eta)))}{2C(k-1)((\alpha\wedge\eta)-1)} + \frac{k/(k-1)-(\alpha\wedge\eta)/2}{1-(\alpha\wedge\eta)}, \tag{S.15}$$

we obtain

$$\begin{aligned} T_{1,2} <& \frac{2Cn}{\log p^*}\sum_{j=km_0+1}^{L}\binom{p^*}{j}\Big(\max_{M\in\mathcal{M}_1^j}\frac{w_M}{w_{M_0}}\Big) \\ <& \frac{2Cn}{\log p^*}\sum_{j=km_0+1}^{L}\exp\{-j\log p^*(((\alpha\wedge\eta)/2+\psi(1-(\alpha\wedge\eta)) \\ &+(2C(k-1))^{-1}k\log(1-4C(1+(\alpha\wedge\eta))))(1-1/k)-1)(1+o_p(1))\} \xrightarrow{P} 0. \end{aligned}$$

The second term in (S.14) can be handled in much the same way, which can be rewritten as

$$\sum_{M\in\mathcal{M}_0}\frac{w_M}{w_{M_0}}|M\nabla M_0| = \sum_{M\in\mathcal{M}_{0,1}}\frac{w_M}{w_{M_0}}|M\nabla M_0| + \sum_{M\in\mathcal{M}_{0,2}}\frac{w_M}{w_{M_0}}|M\nabla M_0| \stackrel{\text{def}}{=} T_{0,1}+T_{0,2}.$$

We first consider the term $T_{0,1}$. Write $\Omega(\psi)=\psi(1-\delta)-\delta/2-2$, a constant independent of n. Noting the condition $\psi > (2+\delta/2)/(1-\delta)$, we have $\Omega(\psi) > 0$. Using (S.9) gives

$$\binom{p^*}{j-m_0}\Big(\max_{M\in\mathcal{M}_0^j}\frac{w_M}{w_{M_0}}\Big) < \exp\{-(j-m_0)\Omega(\psi)\log p^*(1+o_p(1))\}$$

uniformly for all $M\in\mathcal{M}_0$. When $j\geq m_0+r+1$ with $r=[3/\Omega(\psi)]$, we have

$$\exp\{-(j-m_0)\Omega(\psi)\log p^*(1+o_p(1))\} < \exp\{-3\log p^*(1+o_p(1))\}. \tag{S.16}$$

As a consequence, we obtain

$$\sum_{M\in\mathcal{M}_{0,1}} \frac{w_M}{w_{M_0}}|M\nabla M_0| < \sum_{j=m_0+1}^{km_0} \binom{p^*}{j-m_0}\Big(\max_{M\in\mathcal{M}_0^j} \frac{w_M}{w_{M_0}}\Big)(j-m_0) \le T_{0,1,1}+T_{0,1,2}$$

where

$$T_{0,1,1} \stackrel{\text{def}}{=} \sum_{j=m_0+1}^{m_0+r} \exp\{-(j-m_0)\log(p^*)\Omega(\psi)(1+o_p(1)\}(j-m_0),$$

$$\text{and } T_{0,1,2} \stackrel{\text{def}}{=} \sum_{j=m_0+r+1}^{km_0} \exp\{-(j-m_0)\log(p^*)\Omega(\psi)(1+o_p(1)\}(j-m_0).$$

Combining this with the inequality in (S.16), we have

$$T_{0,1,1} \le r^2 \exp\{-\log(p^*)\Omega(\psi)(1+o_p(1))\} \xrightarrow{P} 0,$$

$$\text{and } T_{0,1,2} < (k-1)^2 m_0^2 \exp\{-3\log(p^*)(1+o_p(1)) \xrightarrow{P} 0$$

as $n \to \infty$. For the term $T_{0,2}$, combining (S.12) and the assumption in Theorem 1, we obtain

$$\begin{aligned} T_{0,2} &< \frac{2Cn}{\log p^*} \sum_{j=km_0+1}^{L} \binom{p^*}{j-m_0}\Big(\max_{M\in\mathcal{M}_0^j} \frac{w_M}{w_{M_0}}\Big) \\ &< \frac{2Cn}{\log p^*} \sum_{j=km_0+1}^{L} \exp\{-(j-m_0)\log p^*((\alpha\wedge\eta)/2+\psi(1-(\alpha\wedge\eta)) \\ &\quad -(1+(\alpha\wedge\eta))/(1-2C(1+(\alpha\wedge\eta)))-1)(1+o_p(1))\} \xrightarrow{P} 0. \end{aligned}$$

Overall, when ψ satisfies (S.15), the conclusion (2.4) follows. □

S2 Proof of Theorem 2

First note that

$$\frac{1}{m_0}\sum_{i=1}^{N} w_i|M_i\nabla M_0| = \sum_{M\in\mathcal{M}_0}\frac{w_M}{m_0}|M\nabla M_0| + \sum_{M\in\mathcal{M}_1}\frac{w_M}{m_0}|M\nabla M_0| \overset{\text{def}}{=} I_0 + I_1.$$

In the following proofs, we will prove that I_0 and I_1 converge to 0 in probability.

When we consider the term I_0, it is worth noting that Condition 2 is not applied while we prove $\sum_{M\in\mathcal{M}_0}(w_M/w_{M_0})|M\nabla M_0|$ converges to 0 in probability in the proof of Theorem 1. Hence, without Condition 2 in this theorem, $I_0 < \sum_{M\in\mathcal{M}_0}(w_M/w_{M_0})|M\nabla M_0| \xrightarrow{P} 0$ still holds.

In order to show that I_1 converges to 0 in probability, we need to further split the set $\mathcal{M}_{1,1}$ into multiple subsets. For $i = 1, \dots, p$ and $c_1 > 0$, we define

$$\mathcal{I}_i^L \overset{\text{def}}{=} \begin{cases} \mathcal{I}_i^0 & \text{if } \|\beta_{\mathcal{I}_i^0}\|_2^2/|\mathcal{I}_i^0| \geq c_1\left(|M_0|\log(p^*)/n\right)^{\kappa}, \\ \emptyset & \text{otherwise.} \end{cases}$$

Hence, $M_0^L \overset{\text{def}}{=} \bigcup_{i=1}^{p}\mathcal{I}_i^L$ is the set with indices of larger coefficients. The set $M_0^M \overset{\text{def}}{=} M_0 \setminus (M_0^L \cup M_0^S)$ includes indices of medium size coefficients. For $\mathcal{I}_i^0 \subset M_0^M$, we have $c_2|M_0|\log(p^*)/n \leq \|\beta_{\mathcal{I}_i^0}\|_2^2/|\mathcal{I}_i^0| < c_1(|M_0|\log(p^*)/n)^{\kappa}$, where c_1 and $c_2 > 0$. Clearly, $M_0 = M_0^L \cup M_0^M \cup M_0^S$ and $|M_0^L| + |M_0^M| + |M_0^S| = |M_0|$. By Condition 4, we have $|M_0^S|/|M_0| \leq \xi_n$, where $\{\xi_n\}$ is a nonnegative sequence converging to zero as $n \to \infty$. Let $\{\vartheta_n\}$ and $\{\zeta_n\}$ be strictly positive sequences converging to zero such that $\vartheta_n|M_0| \to \infty$ as $n \to \infty$ and $\zeta_n > \tau(1-\tau)^{-1}(\xi_n + \vartheta_n)$ for each n, where $\tau \in (0,1)$. For example, we can take $\vartheta_n = |M_0|^{-1/2}$ and $\zeta_n = 2\tau(1-\tau)^{-1}(\xi_n + \vartheta_n)$ for the given $\{\xi_n\}$. Let $\mathcal{M}_{11} \overset{\text{def}}{=} \{M \in \mathcal{M}_{1,1} : M_0^L \not\subset M\}$, $\mathcal{M}_{12} \overset{\text{def}}{=} \{M \in \mathcal{M}_{1,1} : |M \cap M_0^M| \leq |M_0^M| - \vartheta_n|M_0|\}$ and $\mathcal{M}_{13} \overset{\text{def}}{=} \mathcal{M}_{1,1} \setminus (\mathcal{M}_{11} \cup \mathcal{M}_{12}) = \{M \in \mathcal{M}_{1,1} : M_0^L \subset M \text{ and } |M \cap M_0^M| > |M_0^M| - \vartheta_n|M_0|\}$. Each model in $\mathcal{M}_{11}$ misses at least one larger coefficient and each model in $\mathcal{M}_{12}$ leaves out $\vartheta_n|M_0|$ indices in M_0^M at least. And for each model in $\mathcal{M}_{13}$, it contains all larger coefficients and many indices in M_0^M. Let $\mathcal{M}_{131} \overset{\text{def}}{=} \{M \in \mathcal{M}_{13} : (|M|-|M_0|)/|M_0| \leq \zeta_n\}$ and $\mathcal{M}_{132} \overset{\text{def}}{=} \{M \in \mathcal{M}_{13} : (|M|-|M_0|)/|M_0| > \zeta_n\}$. Since $\mathcal{M}_1 = \mathcal{M}_{11} \cup \mathcal{M}_{12} \cup \mathcal{M}_{131} \cup \mathcal{M}_{132} \cup \mathcal{M}_{1,2}$, we have $I_1 \leq I_{11} + I_{12} + I_{131} + I_{132} + I_{1,2}$, where

$$I_{11} \overset{\text{def}}{=} \sum_{M \in \mathcal{M}_{11}} \frac{w_M}{m_0}|M \nabla M_0|, I_{12} \overset{\text{def}}{=} \sum_{M \in \mathcal{M}_{12}} \frac{w_M}{m_0}|M \nabla M_0|, I_{131} \overset{\text{def}}{=} \sum_{M \in \mathcal{M}_{131}} \frac{w_M}{m_0}|M \nabla M_0|,$$

$$I_{132} \overset{\text{def}}{=} \sum_{M \in \mathcal{M}_{132}} \frac{w_M}{m_0}|M \nabla M_0| \text{ and } I_{1,2} \overset{\text{def}}{=} \sum_{M \in \mathcal{M}_{1,2}} \frac{w_M}{m_0}|M \nabla M_0|.$$

Furthermore, we only need to prove that each of the above five terms converges to 0 in probability.

When considering the term I_{11}, we need to know that there exists $\mathcal{I}_j \subset M_0^L$ such that $\mathcal{I}_j \not\subset M$ for $M \in \mathcal{M}_{11}$, that is, M does not contain all indices of the larger coefficients. By Condition 3, recall that

$$\frac{\Delta_M}{m_0 \log p^*} \geq \frac{n}{m_0 \log p^*} \lambda_{\min}\Big(\frac{1}{n} X_{M\cup M_0}^{\mathrm{T}} X_{M\cup M_0}\Big) \|\beta_{M_0 \setminus M}\|_2^2, \tag{S.17}$$

we obtain

$$\frac{\Delta_M}{m_0 \log p^*} \geq c_1 \lambda_{\min}\Big(\frac{1}{n} X_{M\cup M_0}^{\mathrm{T}} X_{M\cup M_0}\Big) \Big(\frac{n}{m_0 \log p^*}\Big)^{1-\kappa} \to \infty,$$

since $1-\kappa$ is strictly positive. Under this situation, by a similar manner as in the proof of Theorem 1, we can obtain that $\sum_{M\in\mathcal{M}_{11}} (w_M/w_{M_0})|M \nabla M_0| \xrightarrow{P} 0$. Furthermore, we have $I_{11} < \sum_{M\in\mathcal{M}_{11}} (w_M/w_{M_0})|M \nabla M_0| \xrightarrow{P} 0$.

If $M \in \mathcal{M}_{12}$, we have the following inequality

$$\begin{aligned} \frac{\Delta_M}{m_0 \log p^*} &\geq c_2 \lambda_{\min}\Big(\frac{1}{n} X_{M\cup M_0}^{\mathrm{T}} X_{M\cup M_0}\Big) |M_0 \setminus M| \\ &\geq c_2 \lambda_{\min}\Big(\frac{1}{n} X_{M\cup M_0}^{\mathrm{T}} X_{M\cup M_0}\Big) (\vartheta_n m_0) \to \infty. \end{aligned}$$

by combining Condition 3 and (S.17). Hence, it follows immediately that $I_{12} < \sum_{M\in\mathcal{M}_{12}} (w_M/w_{M_0})|M \nabla M_0| \xrightarrow{P} 0$.

For the term I_{131}, it should be noted that $\sum_{M\in\mathcal{M}_{131}} w_M \leq \sum_{M\in\mathcal{M}} w_M = 1$. Moreover, by the definition of $\mathcal{M}_{131}$ and Condition 4, we obtain $-(\xi_n + \vartheta_n) \leq (|M| - |M_0|)/|M_0| \leq \zeta_n$ for all $M \in \mathcal{M}_{131}$. Therefore, we have $I_{131} \leq (\xi_n + \vartheta_n + \zeta_n) \sum_{M\in\mathcal{M}_{131}} w_M \to 0$.

Now, we turn to the term I_{132}. For $M \in \mathcal{M}_{132}$, let $S \stackrel{\text{def}}{=} M \cap M_0^M$ be the set that contains the indices of the medium size coefficients in M and denote by $\mathcal{M}_{132}^j$ the set that contains all the models in $\mathcal{M}_{132}$ that are of size j.

Under this scenario, $(M_0^L \cup S) \subset M$ and the number of candidate models in $\mathcal{M}_{132}^j$ is less than $m_j' \stackrel{\text{def}}{=} p^*!/(j^*!(p^*-j^*)!)(|M_0^M|)!/((\vartheta_n m_0)!(|M_0^M|-\vartheta_n m_0)!)$, where $j^* \stackrel{\text{def}}{=} j - |M_0^L| - |M_0^M| + \vartheta_n m_0$. Next, we first prove that $w_M \xrightarrow{P} 0$ for $M \in \mathcal{M}_{132}$.

Recall that $w_M/w_{M_0} = \exp(-T_1 - T_2)$ and $RSS_M - RSS_{M_0} = \Delta_M + 2\mu^{\mathrm{T}}(I - H_M)\epsilon + \epsilon^{\mathrm{T}} H_{M_0}\epsilon - \epsilon^{\mathrm{T}} H_M \epsilon$, where $\mu^{\mathrm{T}}(I - H_M)\epsilon = \sqrt{\Delta_M} Z_M$ and $Z_M \sim N(0,1)$. Then we write

$$I_{132}^1 \stackrel{\text{def}}{=} P\Big(\bigcup_{m_0(1+\zeta_n) \le j \le k m_0} \big\{ \max\{|Z_M| : M \in \mathcal{M}_{132}^j\} \ge ((2-\tau)C_{j^*})^{1/2} \big\} \Big)$$
$$= P\Big(\bigcup_{m_0(1+\zeta_n) \le j \le k m_0} \big\{ \max\{Z_M^2 : M \in \mathcal{M}_{132}^j\} \ge (2-\tau)C_{j^*} \big\} \Big).$$

Using similar argument in the proof of Theorem 1 in Luo and Chen (2013) and $\zeta_n > \tau(1-\tau)^{-1}(\xi_n + \vartheta_n)$, we can derive

$$\begin{aligned} I_{132}^1 &< \sum_{j=m_0(1+\zeta_n)}^{km_0} \sum_{M \in \mathcal{M}_{132}^j} P(\chi_{j^*}^2 \ge (2-\tau)C_{j^*}) \\ &< \sum_{j=m_0(1+\zeta_n)}^{km_0} \binom{p^*}{j^*}\binom{|M_0^M|}{\vartheta_n m_0} P(\chi_{j^*}^2 \ge (2-\tau)C_{j^*}) \\ &< \sum_{j=m_0(1+\zeta_n)}^{km_0} \binom{p^*}{j^*}\binom{p^*}{\vartheta_n m_0} P(\chi_{j^*}^2 \ge (2-\tau)C_{j^*}) \to 0. \qquad \text{(S.18)} \end{aligned}$$

Thus, we have $\max\{|Z_M| : M \in \mathcal{M}_{132}^j\} \le ((2-\tau)C_{j^*})^{1/2}(1 + o_p(1))$ uniformly for all $M \in \mathcal{M}_{132}^j$. Moreover, we know that $\epsilon^{\mathrm{T}} H_{M_0}\epsilon - \epsilon^{\mathrm{T}} H_M \epsilon \ge \epsilon^{\mathrm{T}} H_{M_0^L \cup S}\epsilon - \epsilon^{\mathrm{T}} H_M \epsilon$, where $H_{M_0^L \cup S}$ is the projection matrix about $X_{M_0^L \cup S}$ and $X_{M_0^L \cup S}$ denotes an $n \times (|M_0^L| + |S|)$ submatrix of X that is obtained by extracting the columns corresponding to the indices in $M_0^L \cup S$, and $\epsilon^{\mathrm{T}} H_M \epsilon - \epsilon^{\mathrm{T}} H_{M_0^L \cup S} \epsilon \sim \chi^2_{j-|M_0^L|-|S|}$ for $M \in \mathcal{M}_{132}^j$. Similar to (S.18), we have

$$\begin{aligned}\epsilon^{\mathrm{T}} H_{M_0} \epsilon - \epsilon^{\mathrm{T}} H_M \epsilon &\geq -\max\{\epsilon^{\mathrm{T}} H_M \epsilon - \epsilon^{\mathrm{T}} H_{M_0^L \cup S} \epsilon : M \in \mathcal{M}_{132}^j, (M_0^L \cup S) \subset M\} \\ &\geq -2j^*(2-\tau)(1+\delta)(1+o_p(1)) \log p^*, \qquad \text{(S.19)}\end{aligned}$$

uniformly for all $M \in \mathcal{M}_{132}^j$. Furthermore, we also know that $\Delta_M + 2\mu^{\mathrm{T}}(I - H_M)\epsilon = \Delta_M + 2Z_M\sqrt{\Delta_M}$, and since

$$\begin{aligned}\Delta_M + 2Z_M\sqrt{\Delta_M} &\geq \Delta_M - 2(\Delta_M(2-\tau)C_{j^*})^{1/2} \\ &\geq -2j^*(2-\tau)(1+\delta)(1+o_p(1)) \log p^*\end{aligned}$$

uniformly for all $M \in \mathcal{M}_{132}^j$, we can conclude that

$$\begin{aligned}T_1 &= \frac{n}{2} \log \left(1 + \frac{RSS_M - RSS_{M_0}}{RSS_{M_0}}\right) \\ &\geq \frac{n}{2} \log \left(1 - \frac{4(2-\tau)j^*}{RSS_{M_0}}(1+\delta)(1+o_p(1)) \log p^*\right) \\ &= -2j^*(2-\tau)(1+\delta)(1+o_p(1)) \log p^*,\end{aligned}$$

uniformly for all $M \in \mathcal{M}_{132}^j$ by combining (S.19). At the same time, we can calculate $T_2 \geq (j - m_0)(\delta/2 + \psi(1-\delta))(1+o_p(1)) \log p^*$ uniformly for all $M \in \mathcal{M}_{132}^j$.

According to $\zeta_n > \frac{\tau}{1-\tau}(\xi_n + \vartheta_n), 0 < \tau < 1$, we can derive

$$j - m_0 > \zeta_n m_0 > \frac{\tau}{1-\tau}(\xi_n + \vartheta_n) m_0 > \frac{\tau}{1-\tau}(m_0 - |M_0^L| - |M_0^M| + \vartheta_n m_0),$$

further, $j - m_0 > \tau j^*$ can be obtained. Therefore, under the assumption in Theorem 2, which implies that $\psi > (2(2-\tau)(1+\delta) - \tau\delta/2)/(\tau(1-\delta))$ for some τ close to 1, we have

$$\min_{M \in \mathcal{M}_{132}^j} (T_1 + T_2) \geq j^* \log p^* (\frac{\tau\delta}{2} + \psi\tau(1-\delta) - 2(2-\tau)(1+\delta))(1 + o_p(1)) \to \infty. \tag{S.20}$$

It follows that $w_M \leq w_M / w_{M_0} = \exp\{-(T_1 + T_2)\} \xrightarrow{P} 0$ uniformly for all $M \in \mathcal{M}_{132}$.

Now, we prove that the term I_{132} converges to 0 in probability. Clearly,

$$\begin{aligned} I_{132} \leq \sum_{M \in \mathcal{M}_{132}} \frac{w_M}{m_0 w_{M_0}} |M \nabla M_0| \leq \sum_{j=m_0(1+\zeta_n)}^{km_0} m_j^{'} \Big(\max_{M \in \mathcal{M}_{132}^j} \frac{w_M}{m_0 w_{M_0}} \Big) (j + m_0) \\ \leq (k+1) \sum_{j=m_0(1+\zeta_n)}^{km_0} m_j^{'} \Big(\max_{M \in \mathcal{M}_{132}^j} \frac{w_M}{w_{M_0}} \Big). \end{aligned} \tag{S.21}$$

Under the assumption that $\psi > (2 - \tau + 2(2-\tau)(1+\delta) - \tau\delta/2)/(\tau(1-\delta))$ for some τ close to 1, combining (S.20) and (S.21) yields

$$\begin{aligned} I_{132} < \sum_{j=m_0(1+\zeta_n)}^{km_0} \exp\{-j^* \log p^* (\frac{\tau\delta}{2} + \psi\tau(1-\delta) - 2(2-\tau)(1+\delta) - (2-\tau))(1 + o_p(1))\} \\ < km_0 \exp\{-2(1 + o_p(1)) \log p^*\} \xrightarrow{P} 0. \end{aligned}$$

In order to make the condition for ψ easier to be satisfied, we can take $\tau \to 1$, that is, $\psi > (3 + 3\delta/2)/(1-\delta)$.

For $I_{1,2}$, we can derive that $I_{1,2} \leq \sum_{M \in \mathcal{M}_{1,2}} (w_M / w_{M_0}) |M \nabla M_0| \xrightarrow{P} 0$ by a similar manner to the proof of Theorem 1. It is worth noting that the assumption

$$\psi > \frac{k \log\left(1 - 4C(1 + (\alpha \wedge \eta))\right)}{2C(k-1)((\alpha \wedge \eta) - 1)} + \frac{k/(k-1) - (\alpha \wedge \eta)/2}{1 - (\alpha \wedge \eta)}$$

implies that $\psi > (3 + 3\delta/2)/(1-\delta)$ due to $\delta < (\alpha \wedge \eta)$, which completes the proof of Theorem 2. □

References

Luo, S. and Chen, Z. (2013). "Extended BIC for linear regression models with diverging number of relevant features and high or ultrahigh feature spaces." *Journal of Statistical Planning and Inference*, **143**, 494–504.

Ze Chen, Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: chze96@163.com

Jianqiang Zhang, Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: zhangjqs@163.com

Wangli Xu, Center for Applied Statistics, School of Statistics, Renmin University of China, Beijing 100872, China.

E-mail: wlxu@ruc.edu.cn

Yuhong Yang, School of Statistics, University of Minnesota, Minneapolis, MN 55455, USA.

E-mail: yangx374@umn.edu