SUPPLEMENT TO “EMPIRICAL BAYES ORACLE UNCERTAINTY QUANTIFICATION FOR REGRESSION”

BY EDUARD BELITSER AND SUBHASHIS GHOSAL*

VU Amsterdam and North Carolina State University

In this supplement, we elaborate on some points and background information related to the paper “Empirical Bayes oracle uncertainty quantification for regression” and provide some proofs not presented in the main body of the paper. In what follows we use the notations and cross-references to numbered elements (like equations, sections) from the main paper.

1. Empirical Bayes representation. In the model indexed by $I$, let the prior mean for $\theta(I)$ be denoted by $\mu(I)$, where $\mu_I(I) = 0$. Now the marginal density $\pi_{I,\mu}(Y)$ of $Y$ is given by

\[ \pi_{I,\mu}(Y) : Y|I \sim N_n(X_I\mu(I),\sigma^2(I + \kappa H_I)). \]

(S1)

The corresponding posterior distribution of $\theta$ is described hierarchically by

\[ \pi_{I,\mu}(\theta|Y) : \theta|Y \sim N|I| \left( \mu(I) + \frac{\kappa \hat{\theta}_I(I)}{\kappa + 1}, \frac{\kappa \sigma^2}{\kappa + 1}(X_I^tX_I)^{-1} \right) \otimes \delta_{|Ic|}, \]

(S2)

\[ \pi_{\mu}(I|Y) = \frac{\lambda_I \pi_{I,\mu}(Y)}{\sum_{J \in \mathcal{I}} \lambda_J \pi_{J,\mu}(Y)}, \quad I \in \mathcal{I}. \]

(S3)

Hence the posterior density of $\theta$ given $Y$ is

\[ \pi_{\mu}(\theta|Y) = \sum_{I \in \mathcal{I}} \pi_{I,\mu}(\theta|Y)\pi_{\mu}(I|Y). \]

(S4)

Further, the marginal of $Y$ is given by $\pi_{\mu}(Y) = \sum_{I \in \mathcal{I}} \lambda_I \pi_{I,\mu}(Y)$. To choose the family of hyper-parameters $\mu = \mu_{|I} = (\mu(I) : I \in \mathcal{I})$ of the prior distribution, we follow the empirical Bayes approach of maximizing the marginal density $\pi_{\mu}(Y)$ with respect to $\mu$. This boils down to maximizing $\pi_{I,\mu}(Y)$ defined by (S1) with respect to $\mu(I)$ for every $I$.

*Research is partially supported by National Science Foundation (NSF) grant number DMS-1510238. A visit to VU Amsterdam is supported by visitor STAR grant from Netherlands Organisation for Scientific Research (NWO).
separately, which in turn reduces to minimizing $(Y - X_I \mu(I))' (\sigma^2(I + \kappa H_I))^{-1} (Y - X_I \mu(I))$. The solution is given by the weighted least square estimator: for each $I \in \mathcal{I}$,

$$\hat{\mu}(I) = (X_I' (\sigma^2(I + \kappa H_I))^{-1} X_I)^{-1} X_I' (\sigma^2(I + \kappa H_I))^{-1} Y = (X_I' X_I)^{-1} X_I' Y = \hat{\theta}_I(I).$$

The second equality holds because for a symmetric invertible matrix $B$ such that $BA = bA$ for some $b \neq 0$, $(A'B^{-1}A)^{-1}A'B^{-1} = (A'A)^{-1}A'$.

The choices $\mu(I) = \hat{\mu}(I)$ substituted in $\pi(\mu|Y)$ and $\pi_{I,\mu}(\theta|Y)$ lead to the empirical Bayes marginal and posterior distributions studied in the paper.

### 2. Computation.

The proposed EBMS (empirical Bayes model selection) method can be implemented by using Monte Carlo sampling. Clearly the main challenge is the computation of $\hat{I}$ in (2.10), since once $\hat{I}$ is computed, posterior samples for $\theta$ are easily obtained from (2.9). Since the number of possible index sets is exponential in $p$, evaluation of the marginal likelihood of $I$ is generally difficult. We propose an evaluation of the marginal posterior $\pi(I|Y)$ of $I$ given the data by the following simulated annealing technique (see Kirkpatrick et al. [2]).

Recall that the simulated annealing technique finds a global maximizer of a positive function $f$ on a domain through stochastic iterations. Like the Metropolis-Hastings algorithm, given a state $x$, it proposes a move to a neighboring state $x'$. If $f(x') > f(x)$, the move is accepted. If $f(x') \leq f(x)$, the move is accepted with some probability $p(x'|x,T)$ depending on the function $f$, the proposed value $x'$, the current value $x$ and a “temperature” parameter which goes to zero as the iteration progresses making $p(x'|x,T)$ more and more “peaked” around $x$. While various choices are possible, a common idea is to choose $p(x'|x,T)$ to be proportional to $(f(x')/f(x))^{1/T}$, and let $T$ “cool down” slowly with the number $t$ of iteration step like $1/\log(1 + t)$.

In our context, the function to be maximized over $I \in \mathcal{I}$ is the empirical Bayes posterior model probability $\hat{\pi}(I|Y)$ which is proportional to $\lambda_I \hat{\pi}_I(Y)$ (see (2.10)), where $\hat{\pi}_I(Y)$ is given by (2.7). Therefore, given an initial state $I \in \mathcal{I}$ and a proposed state $I' \in \mathcal{I}$, it suffices to efficiently compute $[(\lambda_{I'} \hat{\pi}_{I'}(Y))/ (\lambda_I \hat{\pi}_I(Y))]^{1/T}$, which is easily obtained once

$$\log \hat{\pi}_{I'}(Y) - \log \hat{\pi}_I(Y) = \frac{1}{2\sigma^2} Y' (H_{I'} - H_I) Y - \frac{1}{2} \log(1 + \kappa)(|I'| - |I|)$$

is computed. Clearly, the main challenge is the matrix inversion involved in the computation of the projection matrix. We avoid repeated fresh matrix inversions through the recurrence relations discussed below.

We start with an initial state which can be taken to be $\{1\}$, say. Given a current state $I \in \mathcal{I}$, we randomly choose $i \in \{1, \ldots, p\}$, and propose a new state (i) $I' = I \cup \{i\}$ if $i \notin I$ and $I' \in \mathcal{I}$ and (ii) $I' = I \setminus \{i\}$ if $i \in I$. Instead of choosing $i$ randomly, an alternative is to choose $i$ cyclically from $\{1, 2, \ldots, p\}$. Note that $I \cup \{i\} \notin \mathcal{I}$ if and only if $(1 - H_i)X_i = 0$, and hence the condition can be easily checked. The proposed move will be rejected immediately.
if $I \cup \{i\} \not\in \mathcal{I}$. Suppose that we have computed $(X_I X_I)^{-1}$ (and hence $H_I$ also). To develop a recursion for computing $(X'_I X'_I)^{-1}$ (and hence $H'_I$), we consider cases (i) and (ii).

For (i), we can write $X'_I = (X_I, X_i)$. For an invertible block matrix $M = \begin{pmatrix} A & b \\ b' & c \end{pmatrix}$,

$$M^{-1} = \begin{pmatrix} A^{-1} + A^{-1}bb'A^{-1}/k & -A^{-1}b/k \\ -b'A^{-1}/k & 1/k \end{pmatrix},$$

where $k = c - b'A^{-1}b \neq 0$ since $M$ is invertible. The inverse of $X'_I X'_I = \begin{pmatrix} X'_I X_I & X'_I X_i \\ X'_I X_i & X'_i X_i \end{pmatrix}$, is then easily obtained in terms of $(X_I X_I)^{-1}$, $X'_I X_i$ and $\|X_i\|^2$, and we can verify that

$$H'_I = H_I + \frac{1}{X'_i (1 - H_I) X_i} (I - H_I) X_i X'_i (1 - H_I),$$

so that (S6) can easily be evaluated in this case.

For the proposed move (ii), we have $X_I = (X'_I, X_i)$. Denote $X'_I X_I = \begin{pmatrix} X'_I X'_I & X'_I X_i \\ X'_I X_i & X'_i X_i \end{pmatrix} = \begin{pmatrix} A & b \\ b' & c \end{pmatrix} = M$ and $(X'_I X'_I)^{-1} = M^{-1} = \begin{pmatrix} D & e \\ e' & f \end{pmatrix}$ which is already computed. The goal is to obtain $(X'_I X'_I)^{-1} = A^{-1}$. The block-matrix inversion formula (S7) yields the equations $A^{-1} + A^{-1}bb'A^{-1}/k = D$, $-A^{-1}b/k = e$ and $f = 1/k$. The solutions are $k = 1/f$, $b = -Ae/f$ and $A^{-1} = D - ee'/f$. Thus $(X'_I X'_I)^{-1} = D - ee'/f$ is easily obtained and hence $H'_I$ too.

If the alternative EBMA (empirical Bayes model averaging) posterior (2.12) is to be computed, instead of maximizing the posterior distribution of $I$ given $Y$, one has to sample from it. This can be done by the Metropolis-Hastings algorithm based on the same proposed moves and a slight modification of the acceptance probabilities. In fact, if we fix the temperature parameter $T$ to 1 instead of decaying it slowly, then the successive iterations give Metropolis-Hastings samples from the posterior distribution of $I$ given $Y$, which can be collected after a sufficient burn-in. However, typically the Metropolis chain needs to be run longer compared with the number of iterations in simulated annealing. This is because, in the latter method, the probability function gets more peaked and hence the iteration stabilizes more quickly.

3. Proof of Theorem 3. First we establish one auxiliary lemma.

**Lemma 3.** If $\tau > [4\kappa(e + 1) + 2e \log(1 + \kappa)]/(e - 2)$, then there exist some constants $\tilde{C}_1, \tilde{C}_2 > 0$ such that for any $\theta_0 \in \mathbb{R}^p$, any statistic $\tilde{\theta} \in \mathbb{R}^p$ and any $\delta \in (0, \sqrt{\kappa}/e \sqrt{\kappa + 1}]$,

$$E_{\theta_0} \tilde{\pi}(\|X\theta - X\tilde{\theta}\|^2 \leq \delta^2 \sigma^2 |I_0^\tau(\theta_0)||Y)$$
\[
\leq \frac{C_1 \delta}{|I_o^\tau(\theta_0)|^{1/2}} + \frac{\delta}{[\log(ep/|I_o^\tau(\theta_0)|)]^{1/2}} + C_2 \delta \left( \frac{\log(1/\delta) + \frac{1}{2} \log \log(ep/|I_o^\tau(\theta_0)|)}{\log(ep/|I_o^\tau(\theta_0)|)} \right)^{1/2} \\
\leq (C_1 + 2C_2)\delta \left( \log(1/\delta) \right)^{1/2}.
\]

**Proof of Lemma 3.** For brevity, we suppress the dependence of \(I_o^\tau = I_o^\tau(\theta_0)\) on \(\theta_0\). Since under \(\tilde{\pi}_I(\cdot | Y)\), \(X\theta \sim N_n(H_I Y, e^{\alpha I_o^\tau}H_I)\) and \(H_I\) is symmetric and idempotent, the posterior variation of \(X\theta\) under \(\tilde{\pi}_I(\cdot | Y)\) can be represented by \(X\theta = H_I Y + \sqrt{\frac{\alpha I_o^\tau}{\kappa + 1}}H_I Z\), where \(Z \sim N_n(0, I) \triangleq P_Z\). Hence by Anderson’s lemma (cf., for example, Ibragimov and Has’minskii [1]) which ensures that the concentration probability of a multivariate normal variable is maximal when centered, we obtain

\[
\tilde{\pi}_I(\|X\theta - X\bar{\theta}\|^2 \leq \delta^2 \sigma^2 |I_o^\tau| |Y|) = P_Z \left( \|H_I Y + \sqrt{\frac{\alpha I_o^\tau}{\kappa + 1}}H_I Z - X\bar{\theta}\|^2 \leq \delta^2 \sigma^2 |I_o^\tau| \right)
\leq P_Z \left( \frac{\alpha I_o^\tau}{\kappa + 1} \|H_I Z\|^2 \leq \delta^2 \sigma^2 |I_o^\tau| \right).
\]

As \(\|H_I Z\|^2 \sim \chi^2_{|I|}\) and the standard normal density in \(\mathbb{R}^{|I|}\) is uniformly bounded by \((2\pi)^{-|I|/2}\), the expression above can be bounded by \((2\pi)^{-|I|/2} \Lambda \{ w : \|w\| \leq \delta \left( \frac{\kappa + 1}{\kappa} |I_o^\tau| \right)^{1/2} \}\), where \(\Lambda\) stands for the Lebesgue measure. The unit ball in dimension \(k\) has Lebesgue measure given by \(\pi^{k/2}/\Gamma(1 + k/2) \leq e^{(\pi k)^{-1/2}}(2\pi e/k)^{k/2}\), which leads to the bound

\[
\tilde{\rho}_I \triangleq \tilde{\pi}_I(\|X\theta - X\bar{\theta}\|^2 \leq \delta^2 \sigma^2 |I_o^\tau| |Y|) \leq \frac{e}{\sqrt{\pi |I|}} \left( \frac{e^{(\kappa + 1)\delta^2 I_o^\tau}}{\kappa |I|} \right)^{|I|/2}.
\]

To establish the claim of the lemma, we bound the quantity \(E_{\theta_0} \sum_I \tilde{\rho}_I \tilde{\pi}(I|Y)\) by using the relation \((S8)\) and Lemma 2 with \(\rho = e^{-1}\). Note that by the assumption on \(\tau\) in Lemma 3, the condition required in Lemma 2 holds. Consider the cases \(e^{-\alpha |I_o^\tau| \log(ep/|I_o^\tau|)} \leq \delta/|\log(ep/|I_o^\tau|)|^{1/2}\) and \(e^{-\alpha |I_o^\tau| \log(ep/|I_o^\tau|)} > \delta/|\log(ep/|I_o^\tau|)|^{1/2}\) separately.

If \(e^{-\alpha |I_o^\tau| \log(ep/|I_o^\tau|)} \leq \delta/|\log(ep/|I_o^\tau|)|^{1/2}\), using \((S8)\) and Lemma 2 with \(\rho = e^{-1}\) yields

\[
E_{\theta_0} \sum_I \tilde{\rho}_I \tilde{\pi}(I|Y) \\
\leq \sum_{I: |I| \geq |I_o^\tau|/e} \frac{e}{\sqrt{\pi |I|}} \left( \frac{e^{(\kappa + 1)\delta^2 I_o^\tau}}{\kappa |I|} \right)^{|I|/2} E_{\theta_0} \tilde{\pi}(I|Y) + E_{\theta_0} \tilde{\pi}(I|Y) \\
\leq \frac{C_1 \delta}{\sqrt{|I_o^\tau|}} \sum_{I: |I| \geq |I_o^\tau|/e} \left( \frac{e^{(\kappa + 1)\delta^2 I_o^\tau}}{\kappa |I|} \right)^{|I|/2} - \alpha |I_o^\tau| \log(ep/|I_o^\tau|) \\
\leq \frac{C_1 \delta}{\sqrt{|I_o^\tau|}} + \frac{\delta}{\sqrt{\log(ep/|I_o^\tau|)}},
\]

where \(C_1 = e^{5/2}[\kappa + 1]/(\pi \kappa)^{1/2}\).
For the other case \( |I_o^r| < \log \left( \delta^{-1} \log^{1/2}(ep/|I_o^r|) \right) / \left[ \log \left( \log(ep/|I_o^r|) \right) \right] \), the relation (S10) implies that \( \bar{p} \leq \exp^{-1/2} \left[ C_2 \delta^2 \log \left( \delta^{-1} \log^{1/2}(ep/|I_o^r|) \right) / \left( \log(ep/|I_o^r|) \right) \right]^{1/2} |I|^{|I|-1/|I|+1/2} \), where \( C_2 = e^{(\kappa + 1)/\alpha} \). Since \( \bar{p}(I|Y) \) sums to 1, \( E_{\theta_0} \sum_I \bar{p}I \bar{p}(I|Y) \) is bounded by

\[
C_3 \delta \left( \log \left( \delta^{-1} \log^{1/2}(ep/|I_o^r|) \right) / \log(ep/|I_o^r|) \right)^{1/2} \sum_{I \in \mathcal{I}} \left( C_2 \delta^2 \log \left( \delta^{-1} \log^{1/2}(ep/|I_o^r|) \right) / \left( \log(ep/|I_o^r|) \right) \right)^{1/2} E_{\theta_0} \tilde{p} \bar{p}(I|Y),
\]

where the constants are \( C_3 = e(C_2/\pi)^{1/2} \), \( C_4 = C_3 \max\{C_5^{(k-1)/2}k^{-(k+1)/2} : k = 1, 2, \ldots \} \) and \( C_5 = C_2(1+\max\{\delta^2 \log(1/\delta) : \delta \in [0, 1]\}) = C_2[1+1/(2e)] \). Combining the last bound with (S9), the result follows by choosing \( \bar{C}_1 = C_1, \bar{C}_2 = C_4 \).

\[\square\]

PROOF OF THEOREM 3. First notice that, by applying Markov’s inequality to the second claim of Theorem 1, we obtain that for any \( \theta_0 \in \mathbb{R}^p \) and \( M > 0 \)

(S10) \( P_{\theta_0}(\|X \theta_0 - X \hat{\theta}\| \geq M r(\theta_0)) \leq C M^{-2}. \)

Next, we write \( r_\tau^2(\theta) = A^2(\theta, \tau) \sigma^2 |I_o^r(\theta)| \), where \( A^2(\theta, \tau) = (b_r(\theta) + \tau) \log(ep/|I_o^r(\theta)|) \). By replacing \( \delta \) by \( \delta A(\theta_0, \tau) \), the bound in Lemma 3 can be restated as

\[
E_{\theta_0} \tilde{p}(\|X \theta - X \hat{\theta}\| \leq \delta r(\theta_0) |Y|) \leq \bar{C}_3 \delta A(\theta_0, \tau) \left[ \log \left( \delta A(\theta_0, \tau) \right) \right]^{-1/2},
\]

for all \( \delta \in (0, e^{-1}(\kappa/(\kappa + 1))^{1/2}/A(\theta_0, \tau)] \) and any statistic \( \tilde{\theta} \), where \( \bar{C}_3 = \bar{C}_1 + 2 \bar{C}_2 \). Since \( r_\tau^2(\theta_0) \leq r_\tau^2(\theta_0) \) for \( \tau \geq 1 \), the last display yields that for any \( \theta_0 \in \mathbb{R}^p \), any statistic \( \tilde{\theta} \) and any \( \delta \in (0, e^{-1}(\kappa/(\kappa + 1))^{1/2}/A(\theta_0, \tau)] \),

(S11) \( E_{\theta_0} \tilde{p}(\|X \theta - X \hat{\theta}\| \leq \delta r(\theta_0) |Y|) \leq \bar{C}_3 \delta A(\theta_0, \tau) \left[ \log \left( \delta A(\theta_0, \tau) \right) \right]^{-1/2}. \)

where \( \bar{C}_3 = \bar{C}_1 + 1 + 2 \bar{C}_2 \).

To establish coverage, note that by the definition of \( \bar{p} \) given in (2.19), Markov’s inequality and the estimates in (S10) and (S11), we have

\[
P_{\theta_0}(X \theta_0 \notin B(X \hat{\theta}, M \log(ep)^{1/2} \hat{\rho})) \leq P_{\theta_0}(\|X \theta_0 - X \hat{\theta}\| \geq \log(ep)^{1/2} M \hat{\rho}, \hat{\rho} \geq \delta r(\theta_0)) + P_{\theta_0}(\hat{\rho} < \delta r(\theta_0)) \leq P_{\theta_0}(\|X \theta_0 - X \hat{\theta}\| \geq \log(ep)^{1/2} M \delta r(\theta_0)) + P_{\theta_0}(\tilde{p}(\|X \theta - X \hat{\theta}\| \leq \delta r(\theta_0) |Y|) \geq 1 - \gamma) \leq \frac{C}{M^2 \delta^2 \log(ep)} (1 - \gamma)^{-1} \bar{C}_3 \delta A(\theta_0, \tau) \left[ \log \left( 1/(\delta A(\theta_0, \tau)) \right) \right]^{1/2},
\]
for any \( \delta \in (0, e^{-1}(\kappa/(\kappa + 1))^{1/2}/A(\theta_0, \tau)] \). Now for \( M \geq e^{3/2}(\kappa/(\kappa + 1))^{-3/4} \), with the choice \( \delta = M^{-2/3}/A(\theta_0, \tau) \), the bound reduces to
\[
M^{-2/3} \left( \frac{C A^2(\theta_0, \tau)}{\log(ep)} + (2/3)^{1/2}(1 - \gamma)^{-1}C_3(\log M)^{1/2} \right).
\]

For each \( \theta_0 \in \Theta_{eb}(t, \tau) \), \( A^2(\theta_0, \tau) \leq (t + \tau) \log(ep) \), and hence for any given \( t > 0 \), the expression can be bounded by \( \epsilon_1 \) simultaneously for all \( \theta_0 \in \Theta_{eb}(t, \tau) \) by choosing \( M \) sufficiently large (depending on \( t \) and \( \epsilon_1 \) only).

Finally, to verify the size property, observe that for all \( \theta_0 \in \mathbb{R}^p \),
\[
P_{\theta_0}(\tilde{\rho} \geq Lr(\theta_0))
\leq P_{\theta_0}(\tilde{\pi}(\|X\theta - \hat{X}\hat{\theta}\| \geq Lr(\theta_0)|Y) > \gamma)
\leq \gamma^{-1}[E_{\theta_0}\tilde{\pi}(\|X\theta - \hat{X}\theta\| \geq Lr(\theta_0)/2|Y) + P_{\theta_0}(\|X\theta - \hat{X}\theta\| \geq Lr(\theta_0)/2)]
\leq 8\gamma^{-1}C/L^2.
\]

Clearly, a sufficiently large \( L > 0 \) makes the expression smaller than any given \( \epsilon_2 > 0 \). \( \square \)

4. Proof of Theorem 4. To prove (4.4), we establish the conditional bound
\[
E_{\theta_0}[\tilde{\mathbb{E}}(\|Z\theta - Z\theta_0\|^2|X,Y)|X] \leq Cr^2(\theta_0|X).
\]

Then as remarked in the part 3 of the discussions section, the unconditional bound \( Cr^2(\theta_0) \) follows by integration.

We can proceed as in the proof of Theorem 1 with appropriate modifications of the estimates. Below we briefly point out the modifications of the expressions appearing in the course of the proof. The rest of the proof can be completed as before.

Recall that in the additive regression framework, two indexes \( I \leq [p] \) and \( J \in \mathbb{N} \) are used, \( I \) is the class of all index-pairs \( (I, J) \) such that \( Z_{I',J}Z_{I,J} \) is invertible, and the projection \( H_{I,J} \) is given by \( H_{I,J} = Z_{I,J}(Z'_{I,J}Z_{I,J})^{-1}Z_{I,J} \) for \( (I, J) \in I \).

As \( Z\theta(Y, X, I, J) \sim N_{n}(H_{I,J}Y, \frac{\sigma^2}{\kappa+1}H_{I,J}) \), we obtain
\[
E_{\theta_0}[\tilde{\mathbb{E}}(\|Z\theta - X\theta_0\| |Y, X)|X] \leq \sum_{(I,J) \in I} (r^2(I, J), \theta_0|X) + \sigma^2\|H_{I,J}\|^2 \hat{\pi}(I, J)|Y, X)
\]
and
\[
E_{\theta_0}(\|Z\theta - Z\theta_0\|) \geq Mr(\theta_0|X)|Y, X)
\leq \sum_{(I,J) \in I} r^2(I, J, \theta_0|X)E_{\theta_0}(\hat{\pi}(I, J)|Y, X)|X]
\leq \frac{M^2r^2(\theta_0|X)}{M^2r^2(\theta_0|X)}
\]
where bound $E$.

Since $\lambda_{ep}/r$ \( (S12) \)

Now proceeding as before, for any two pairs \((I, J), (I_0, J_0)\) \( \in \mathcal{I} \), we obtain the bound (S12)

\[
E_{\theta_0} \left[ \hat{\pi}((I, J)|Y, X)|X] \right] \leq \left( \frac{\lambda_{I,J}}{\lambda_{I_0,J_0}} \right)^h \frac{\exp \left\{ \frac{h}{2\sigma^2} \theta_0'Z'(I-hH_{(I,J),(I_0,J_0)})^{-1}H_{I,J}Z\theta_0 \right\}}{(1+\kappa)^{h(|I|+J)-|I_0|}/\sqrt{\det(I-hH_{(I,J),(I_0,J_0)})}}.
\]

Since $\lambda_{I,J}/\lambda_{I_0,J_0} = \exp \{-\kappa(|I|\log(ep/|I|) - \tau_0) - \beta(J - J_0)\}$ and $\det(I - h(H_{I,J} - H_{I_0,J_0})) \geq (1-h)^{|I|+|J|}$, it follows that the bound for $E_{\theta_0}(\hat{\pi}(I, J)$ is given by

\[
(ep/|I|)^{-c_1} |I| e^{-c_1 J} \exp \left\{ \frac{h(1+h)}{2\sigma^2(1-h)} \theta_0'Z'(I-H_{I_0})Z\theta_0 + \kappa h|I| \log \frac{ep}{|I_0|} + \beta h J_0 \right\}
\]

\[
+ \frac{h}{2}(|I_0| + J_0) \log(1+\kappa) - \frac{h(1-3h)}{2\sigma^2(1-h)} \theta_0'Z'(I-H_{I,J})Z\theta_0 - (h\kappa - c_1) |I| \log \frac{ep}{|I_0|} \left( 1+h(1-h)^{1/2} \right)
\]

\[-(h\beta - c_1^J) |I| - (|I| + J) \log \left( 1+\kappa \right) \left( 1+h(1-h)^{1/2} \right) \],

where $c_1$ is as before and $c_1^J = h\beta/2$. Now choose $(I_0, J_0)$ to be the standard oracle $(I_0, J_0)$. Then observing that the square of the mimicable oracle rate $r^2((I, J), \theta_0|X) = r^2((I, J), \theta_0|X)$ given by (4.3) contains the terms $\sigma^2(1+|I|+|J| \log(ep/|I|)$, we can bound $E_{\theta_0}[\hat{\pi}((I, J)|Y, X)|X]$ by

\[(S13) \quad (ep/|I|)^{-c_1} |I| e^{-c_1 J} \exp \left\{ -c_2 \sigma^{-2} (r^2((I, J), \theta_0|X) - c_3 \sigma^2(\theta_0|X)) \right\},
\]

where the positive constants $c_2, c_3$ can be chosen the same as in Theorem 1.

Using bounds as in Theorem 1, for $\mathcal{O}(\tau_0, \theta_0) = \{(I, J) : r^2((I, J), \theta_0|X) \leq \tau_0 \sigma^2(\theta_0|X)\}$, the expression $\sum_{(I,J) \in \mathcal{O}(\tau_0, \theta_0)} r^2((I, J), \theta_0|X) [E_{\theta_0}[\hat{\pi}((I, J)|Y, X)]^{1/2}$ is bounded by a constant multiple of $\sigma^2$.

In the present context, $(\|H_{I,J} \epsilon\|^2|X) \sim \chi^2_{|I|+J}$, and so

\[
E[\|H_{I,J} \epsilon\|^2|X] = (1 + |I|J)^2 + 2(1 + |I|J) \leq 4r^4((I, J), \theta_0|X)/\sigma^4.
\]

Combining the last relation with (S13) and arguing as in Theorem 1, we obtain that for some constant $C_1 > 0$,

\[
\sigma^2 E_{\theta_0} \left[ \sum_{(I,J) \in \mathcal{O}(\tau_0, \theta_0)} \|H_{I,J} \epsilon\|^2 \hat{\pi}((I, J)|Y, X)|X \right] \leq C_1 \sigma^2.
\]

Observe that by the norm-decreasing property of projection operators, and since $J$ is linearly ordered, the maximum value of $\|H_{I,J} \epsilon\|^2$ over $(I, J) \in \mathcal{O}(\tau_0, \theta_0)$ is bounded by $\max \{\|H_{I,J} \epsilon\|^2 : |I| = m\}$, where $m$ is the largest integer such that $\sigma^2[1 + m^2 + m \log(ep/m)] \leq \tau_0 \sigma^2(\theta_0|X)$. Each $\|H_{I,J} \epsilon\|^2$ given $X$ has $\chi^2_{1+m^2}$ distribution, which has
finite moment generating function at \( t \) for \( t \leq (1 - e^{-1})/2 \approx 0.31 \), so by Lemma 1, the expectation of the maximum is bounded by

\[
\frac{1}{0.31} \left\{ \log(1 + m^2) + \frac{1 + m^2}{2} \log \frac{1}{1 - 0.6} \right\} \leq C_2 \sigma^{-2} r^2(\theta_0|X)
\]

for some constant \( C_2 > 0 \). As in Theorem 1, by using the last display and (S13), we obtain that

\[
\sigma^2 E_{\theta_0} \left[ \sum_{(I,J) \in \Omega(\tau,\theta_0)} \|H_{I,J}\|_2^2 \hat{\pi}((I,J)|Y,X)|X| \right] \leq C_3 r^2(\theta_0|X)
\]

for some constant \( C_3 > 0 \). As argued before in Theorem 1, this implies that

\[
\sum_{(I,J) \in \mathcal{I}} r^2((I,J),\theta_0|X) E_{\theta_0} \hat{\pi}((I,J)|Y,X) \leq C' r^2(\theta_0|X)
\]

for some constant \( C' > 0 \), since \( \sigma^2 \) is a part of the oracle risk. As argued before, this leads to the conclusion that \( E\{\hat{\pi}|Z\theta - Z\theta_0|^2|Y,X|X\} \) is bounded by a constant multiple of \( r^2(\theta_0|X) \), which leads to the conclusions of the theorem.

5. Proof of Theorem 5. First we need to establish an analog of Lemma 2 to assure that \((I, J)\) chosen from the posterior can rarely be much smaller than a \( \tau \)-oracle \((I^*_\tau, J^*_\tau) = (I^*_\tau(\theta_0), J^*_\tau(\theta_0))\) if \( \tau \) is chosen sufficiently large. We shall not try to identify the constants explicitly, as that is not essential for the proof, but can be done if desired, in linear regression.

Lemma 4. For \( \varrho \in [0,1) \) we can choose positive constants \( \tau, \alpha \) and \( C \) such that for any true \( \theta_0 \),

\[
E_{\theta_0} [\hat{\pi}(|I| \leq \varrho|I^*_\tau|, J \leq \varrho J^*_\tau|Y,X)|X| \leq C \exp \{- \alpha|I^*_\tau| \log(ep/|I^*_\tau|) - \alpha J^*_\tau \}
\]

Moreover, the above bound can be reduced to \( \varrho^\alpha \) for some \( \alpha_0 > 0 \) if \( \varrho \) is sufficiently small.

Proof. Fix a true parameter sequence \( \theta_0 \). For \((I, J) \in \mathcal{I}\) such that \(|I| \leq \varrho|I^*_\tau|, J \leq \varrho J^*_\tau\), let \( I_0 = I \cup I^*_\tau \) for some \( I^*_\tau \subseteq I^*_\tau \), \( J_0 = J^*_\tau \), such that \((I_0, J_0) \in \mathcal{I}\) and \( \text{col}(X_{I_0,J_0}) = \text{col}(X_{I^*_\tau,J^*_\tau}) \). From the estimate \(|I_0| \leq |I| + |I^*_\tau| \leq (1 + \varrho)|I^*_\tau|\), as before, we have \(|I| \log(ep/|I|) \leq \varrho(1 + \log(1/\varrho))|I^*_\tau| \log(ep/|I^*_\tau|)\). Moreover the difference \( H_{I_0,J_0} - H_{I^*_\tau,J^*_\tau} \) is also a projection. This leads to

\[
\sigma^{-2} \theta_0^T Z'(H_{I_0,J_0} - H_{I,J}) \theta_0 \\
\geq \sigma^{-2} \theta_0^T Z'(I - H_{I,J}) \theta_0 - \theta_0^T Z'(1 - H_{I^*_\tau,J^*_\tau}) \theta_0 \\
\geq \tau \left( |I^*_\tau| \log \frac{ep}{|I^*_\tau|} - |I| \log \frac{ep}{|I|} + J^*_\tau - J \right),
\]
in view of the fact that $r_2^2((I_o^r, J_o^r), \theta_0|X) \leq r_2^2((I, J), \theta_0|X)$. This leads to, with $a(\varrho) = 1 - \varrho(1 + \log(1/\varrho))$, the bound $\sigma^{-2}\theta'_0 Z'(H_{I_o, J_o} - H_{I, J})Z\theta_0 \geq \tau a(\varrho)|I_o^r| \log(\varrho|/|I_o^r|))$.

Using arguments analogous to those used in linear regression, for $|I| \leq \varrho|I_o^r|$, $J \leq \varrho J_o^r$, $E_{\theta_0}\hat{\pi}((I, J)|Y, X)$ is bounded by

$$\lambda_{I,J}E_{\theta_0} \exp \left\{ -\frac{1}{2\pi^2} Y'(H_{I_o, J_o} - H_{I, J})Y \right\}$$

$$= \frac{\lambda_{I,J} \exp \left\{ -\frac{1}{2\pi^2} \theta'_0 Z'(H_{I_o, J_o} - H_{I, J})Z\theta_0 \right\}}{(1 + \kappa)^{|I|+|J|+J_0}/2}$$

$$\leq \lambda_{c, I,J} \exp \left\{ -\frac{1}{4} \tau a(\varrho)|I_o^r| \log \frac{ep}{|I_o^r|} - \tau(1 - \varrho)J_o^r + \kappa|I_0| \log \frac{ep}{|I_o^r|} \right\}$$

$$+ \beta J_o^r + \frac{1}{2}(|I_o| - |I| + J_o - J) \log(1 + \kappa)$$

$$\leq C\lambda_{I,J} \exp \left\{ -\alpha \left( |I_o^r| \log(\varrho|/|I_o^r|) + J_o^r \right) \right\}$$

for suitable positive constants $C$ and $\alpha$. Now summing over $(I, J)$ gives the first part. The second part follows readily from the first part by separating into cases the combinations of $|I_o^r| \leq \varrho^{-1}$ and $|I_o^r| > \varrho^{-1}$ with $J_o^r \leq \log \varrho^{-1}$ and $J_o^r > \log \varrho^{-1}$.

**Proof of Theorem 5.** We prove the result for the EBMS posterior only. For the EBMA posterior, a proof can be constructed following analogous arguments to modify the proof of Theorem 3 after an analog of Lemma 3 is established.

To establish the coverage of the credible ball $B(Z\hat{\theta}, M\hat{\rho})$, as before, we intersect with the events $\{\hat{\rho} > \delta r(\theta_0)\}$ and $\{\hat{\rho} \leq \delta r(\theta_0)\}$ separately. The first intersection contains $B(Z\hat{\theta}, M\delta r(\theta_0))$, and by the proof of Theorem 4, $P_{\theta_0}(Z\theta_0 \notin B(Z\hat{\theta}, M\delta \hat{\rho})|X) \leq CM^{-2}\delta^{-2}$. To bound $P_{\theta_0}(\hat{\rho} < \delta r(\theta_0)|X)$ for $X$ such that

(S14) \[ r_2^2(\theta_0|X) \leq (1 + t)[1 + |I_o^r| J_o^r + |I_o^r| \log \frac{ep}{|I_o^r|}] \]

for some $t > 0$, we have by Lemma 4, $P_{\theta_0}\{\hat{\rho} \leq \delta r(\theta_0|X)\} \leq (\delta(1 + t))^{\alpha_0}$ for the positive constant $\alpha_0$ appearing there, provided that $\delta > 0$ is chosen sufficiently small. Thus the noncoverage probability for any given $X$ satisfying (S14) is bounded by $CM^{-2}\delta^{-2} + (\delta(1 + t))^{\alpha_0}$, which can be made smaller than any given $\epsilon_1$ by first choosing $\delta$ sufficiently small, and then $M$ sufficiently large. Now if $\theta_0 \in \Theta_{\epsilon_1-\alpha_0}(t, \tau)$, then the probability that (S14) fails is at most $\epsilon$, leading to the conclusion about coverage.

To verify the size property with high probability as claimed, define $G(L) = G(L, X) = \{(I, J) \in I : \sigma^2[(I, J) + 1 + |I| \log(ep/|I|)] \geq L^2 r(\theta_0|X)\}$. Note that $\hat{\rho}$ used to construct the credible ball $B(Z\hat{\theta}, M\hat{\rho})$ is smaller than $L r(\theta_0|X)$ if $(I, J) \notin G(L)$, so it suffices to bound the probability $P_{\theta_0}((I, J) \notin G(L)|X) = E_{\theta_0}\hat{\pi}((I, J) \notin G(L)|X)$.

Note that for any $\theta_0 \in \mathbb{R}^p$ and all $(I, J) \in G(L)$,

$$\sigma^{-2}(r^2((I, J), \theta_0|X) - c_3 r^2(\theta_0|X)) \geq 1 + |I| + |J| \log(ep/|I|) - c_3 \sigma^{-2} r^2(\theta_0|X)$$
\[ \geq (L^2 - c_3)\sigma^{-2}r^2(\theta_0|X). \]

From (5.7) and the last relation it follows that for any \((I, J) \in \mathcal{G}(L)\)
\[
P_{\theta_0}(\hat{I} = I, \hat{J} = J|X) \leq (ep/|I|)^{-c_1|I|}e^{-c_1J} \exp\{-c_2\sigma^{-2}(r^2((I, J), \theta_0|X) - c_2\sigma^{-2}r^2(\theta_0|X))\}
\leq (ep/|I|)^{-c_1|I|}e^{-c_1J} \exp\{-c_2(L^2 - c_3)\sigma^{-2}r^2(\theta_0|X)\},
\]
where \(c_3\) is the constant that appeared in (S13).

Therefore, using the bound (S13), it follows that
\[
E_{\theta_0}\hat{\pi}((I, J) \in \mathcal{G}(L)|Y, X) \leq \exp\{-c_2(L^2 - c_3)\sigma^{-2}r^2(\theta_0|X)\} \sum_{(I, J) \in I} (ep/|I|)^{-c_1|I|}e^{-c_1J},
\]
which is bounded by a constant multiple of \(e^{-c'_3L^2}\) for all sufficiently large \(L\) and for some constant \(c'_3 > 0\), since the sum above is convergent and \(\sigma^{-2}r^2(\theta_0|X) \geq 1\) by the definition of an oracle. The unconditional probability bound for the radius follows from the argument given at the end of part Random predictors of Section 3. \(\square\)

6. **Proof of Corollary 5.** By the definition of the functional class \(\mathcal{F}(s, \alpha)\), it follows that for any \(J \geq 1\)
\[
(S15) \quad E\left|f_{i,0}(X_i) - \sum_{j=1}^{J} \theta_{ij,0}B_j(X_i)\right|^2 = \sum_{j > J} q_{ij}^2 \leq QJ^{-2\alpha},
\]
uniformly in \(F_0 \in \mathcal{F}(s, \alpha)\).

Let \(I_0 = I_0(\theta_0)\) be the true set of active predictors and \(J_0 = J_0(\theta_0)\) be the integer part of \(n^{1/(1+2\alpha)}\). Clearly the oracle rate \(r^2(\theta_0)\) is bounded by the rate at \((I_0, J_0)\) and the latter is bounded as follows:
\[
r_1^2((I_0, J_0), \theta_0) \leq nE\left|\sum_{i \in I_0} \sum_{j > J_0} \theta_{ij,0}B_j(X_i)\right|^2 + \sigma^2(sJ_0 + 1) + \sigma^2s \log \frac{ep}{s}
\leq \sum_{i \in I_0} \sum_{j > J_0} \theta^2_{ij,0} + \sigma^2(sJ_0 + 1) + \sigma^2s \log \frac{ep}{s}
\leq ns\left((K + \sigma^2)n^{-2\alpha/(1+2\alpha)} + \frac{\sigma^2}{\pi} \log \frac{ep}{s}\right),
\]
where we used the independence of the components of the predictor, orthonormality of the basis functions, the norm-decreasing property of a projection operator, (S15) and the choice of \(J_0\). Hence the posterior contraction and estimation rates for the normalized distances respectively \(\|Z\theta - F_0(X)\|^2_n\) and \(\|Z\hat{\theta} - F_0(X)\|^2_n\) are both bounded by \(\epsilon_n^2 = \frac{\sigma^2}{\pi} \log \frac{ep}{s}\).
max\((sn^{-2\alpha/(1+2\alpha)}, sn^{-1} \log(p/s)})\), uniformly in \(F_0 \in \mathcal{F}(s, \alpha)\) for the sparsity regime \(s\) and the smoothness \(\alpha\).

The second assertion follows directly from Theorem 5 under the \(\epsilon\)-EBR condition, since \(r(\theta_0)\) under the stated sparsity assumption is bounded by a constant multiple of the rate sequence \(\epsilon_n\).

REFERENCES