Supplementary Material: Proof of Theorem 1 and Corollaries

The lemmas given here are used in the proof of Theorem 1. Lemmas 1 and 2 follow directly from Boole's inequality and are stated for later reference. The first lemma is used to obtain bounds on the Bayes factors between arbitrary models $\mathbf{k} \in \mathcal{J}$ and $\mathbf{t}$, which take the form of products of functions of random variables.

Lemma 1 *Let $Z_1, \ldots, Z_n$ denote non-negative random variables defined with respect to an underlying probability space and let $z_1, \ldots, z_n$ denote arbitrary constants. Then*

$$\mathbf{P}\left[\prod_{i=1}^{n} Z_i > \prod_{i=1}^{n} z_i\right] \leq \sum_{i=1}^{n} \mathbf{P}\left[Z_i > z_i\right]. \tag{19}$$

Proof of Lemma 1. Suppose $n = 2$ and define $B = \{\omega \in \Omega : Z_1(\omega)Z_2(\omega) > z_1 z_2\}$, $A_1 = \{\omega \in \Omega : Z_1(\omega) > z_1\}$, and $A_2 = \{\omega \in \Omega : Z_2(\omega) > Z_2\}$. Because $B \subset A_1 \cup A_2$, equation (19) follows by Boole's inequality. Cases for which $n > 2$ follow by induction.

The next lemma is used to show that sums of Bayes factors that determine the posterior probability of the true model converge in probability to 0.

Lemma 2 *For $d = 1, \ldots, p$, $p \leq n$, define $h = \binom{p}{d}$ and let $Z_{d,i}$, $i = 1, \ldots, h$ denote a sequence of non-negative random variables. Suppose $a, b, c > 0$, and there exist $s > 0$*

and N such that for all $n > N$

$$\mathbf{P}\left[Z_{d,i} \geq \frac{ab^d}{n^{d(1+s)}}\right] < \frac{c}{n^{d(1+s)}}.$$

Then

$$\sum_{d=1}^{p}\sum_{i=1}^{h} Z_{d,i} \xrightarrow{p} 0$$

as $n \to \infty$.

Proof of Lemma 2. *Letting $S_d = \sum_{i=1}^{h} Z_{d,i}$ and noting $h < n^d$,*

$$\begin{aligned}\mathbf{P}\left[S_d > \frac{ab^d}{n^{ds}}\right] &\leq \sum_{i=1}^{h}\mathbf{P}\left[Z_{d,i} > \frac{ab^d}{n^{d+ds}}\right] \\ &\leq \frac{c}{n^{ds}}.\end{aligned}$$

Also, because the $S_d > 0$ for $d = 1, \ldots, p$, it follows that

$$\mathbf{P}\left[\sum_{d=1}^{p} S_d > K\right] \leq \frac{c(1-n^{-p})}{n^s - 1} \to 0,$$

where

$$K = \sum_{d=1}^{p} \frac{ab^d}{n^{ds}} = \frac{a(1-\frac{b^p}{n^{ps}})}{(\frac{n^s}{b}-1)} \to 0.$$

The next lemma provides a loose (but convenient) bound on the probability that a non-central chi-squared random variable exceeds a specified threshold.

Lemma 3 *Suppose the random variable T has a chi-squared distribution on s degrees of freedom and non-centrality parameter λ. Let $a > 0$ and define*

$$x = t - \lambda - \frac{a}{2\sqrt{\lambda}}.$$

If

$$t > 2s - 4 + \lambda + \frac{a}{2\sqrt{\lambda}},$$

then

$$\mathbf{P}(T > t) < \frac{\sqrt{2\lambda}}{a\sqrt{\pi}} \exp\left(-\frac{a^2}{8\lambda}\right) + \frac{2}{\sqrt{\pi s}} \left(\frac{x}{s}\right)^{s/2-1} \exp\left[-\frac{1}{2}(x-s)\right]. \tag{20}$$

Furthermore, if $\lambda = 0$ *and* $t > s$*, then*

$$\mathbf{P}(T > t) < \frac{2}{\sqrt{\pi s}} \left(\frac{t}{s}\right)^{s/2-1} \exp\left[-\frac{1}{2}(t-s)\right]. \tag{21}$$

Proof of Lemma 3. Letting Z_i, $i = 1, \ldots, s$ denote independent standard normal random variables, it follows that T can be expressed as

$$T = (Z_s + \sqrt{\lambda})^2 + \sum_{i=1}^{s-1} Z_i^2 = \lambda + 2\sqrt{\lambda} Z_s + \sum_{i=1}^{s} Z_i^2.$$

It follows that

$$\mathbf{P}(T > t) \leq \mathbf{P}\left[Z_s > \frac{a}{2\sqrt{\lambda}}\right] + \mathbf{P}\left[\sum_{i=1}^{s} Z_i^2 > t - \lambda - \frac{a}{2\sqrt{\lambda}}\right].$$

The result follows by applying standard bounds on the survival function for standard normal and chi-squared random variables [3].

The remaining lemmas follow from relationships between eigenvalues of a matrix and quadratic forms as found in, for example, Schott (1997; Theorems 3.15–20) and Arnold (1981; Theorem 7.4). In particular, we note that the eigenvalue bounds $\mathbf{X}_n'\mathbf{X}_n$ assumed in the Theorem imply the same conditions for all submatrices of $\mathbf{X}_n$, i.e., to

$\mathbf{X}_\mathbf{k}'\mathbf{X}_\mathbf{k}$. For convenience and without loss of generality, we also assume that c and M are chosen so that the eigenvalue bounds for $\lambda_1(\mathbf{X}_n'\mathbf{X}_n) < nM$ and $\lambda_p(\mathbf{X}_n'\mathbf{X}_n) > nc$, stated in Theorem 1, apply also to the matrices $\mathbf{X}_\mathbf{k}'\mathbf{X}_\mathbf{k} + \mathbf{A}_\mathbf{k}/\tau$ for all model $\mathbf{k}$.

The conditions of Theorem 1 are assumed to hold throughout the remainder of this section.

To begin, note that

$$c^k < \left| \frac{\mathbf{X_k}'\mathbf{X_k}}{n} \right| < M^k,$$

and

$$c\boldsymbol{\beta}_\mathbf{k}'\boldsymbol{\beta}_\mathbf{k} < \boldsymbol{\beta}_\mathbf{k}' \frac{\mathbf{X_k}'\mathbf{X_k}}{n} \boldsymbol{\beta}_\mathbf{k} < M\boldsymbol{\beta}_\mathbf{k}'\boldsymbol{\beta}_\mathbf{k}.$$

The prior density in equation (2) can thus be bounded by the product of central moments of circular normal densities. In particular, the prior constant $d_\mathbf{k}$ is bounded by

$$\left(\frac{c}{M}\right)^{k/2} \left[\frac{c^r}{(2r-1)!!}\right]^k < d_\mathbf{k} < \left(\frac{M}{c}\right)^{k/2} \left[\frac{M^r}{(2r-1)!!}\right]^k.$$

Similarly, the Bayes factor between models $\mathbf{k}$ and $\mathbf{t}$ is bounded by

$$\begin{aligned} \frac{m_\mathbf{k}(\mathbf{y}_n)}{m_\mathbf{t}(\mathbf{y}_n)} &< \left[\frac{M^{r+1}}{c\sigma^{2r}\tau^{r+\frac{1}{2}}(2r-1)!!}\right]^k \left[\frac{M\sigma^{2r}\tau^{r+\frac{1}{2}}(2r-1)!!}{c^{r+1}}\right]^t n^{\frac{t-k}{2}} \frac{Q_\mathbf{k}}{Q_\mathbf{t}} \exp\left[-\frac{R_\mathbf{k}-R_\mathbf{t}}{2\sigma^2}\right] \\ &\equiv \alpha^k \gamma^t \, n^{\frac{t-k}{2}} \frac{Q_\mathbf{k}}{Q_\mathbf{t}} \exp\left[-\frac{R_\mathbf{k}-R_\mathbf{t}}{2\sigma^2}\right], \end{aligned} \tag{22}$$

where for each $\mathbf{k} \in \mathcal{J}$

$$Q_\mathbf{k} = \mathbf{E}_\mathbf{k}\left(\prod_i^k \beta_{k_i}^{2r}\right),$$

where $\mathbf{E_k}$ is the expectation defined in equation (5).

The next three lemmas provide bounds on the expectations $Q_\mathbf{k}$ under various model assumptions.

Lemma 4 *Suppose that the conditions of Theorem 1 hold. Then*

$$Q_\mathbf{t} \xrightarrow{a.s.} \prod_{i=1}^{t} (\beta^0_{t_i})^{2r} > \delta^{2rt} > 0.$$

The proof of this lemma follows simply by noting that the conditions of Theorem 1 are sufficient to guarantee almost sure convergence of the least squares estimates to $\boldsymbol{\beta}^0_\mathbf{t}$ [2], that $\tilde{\boldsymbol{\beta}}$ differs from the least squares estimate by a term that is $O(n^{-1})$, and application of Slutsky's theorem.

Lemma 5 *Suppose that the conditions of Theorem 1 apply and that* $t \subset k$*. Then for* $0 < \epsilon < \frac{1}{2}$,

$$\mathbf{P}\left[Q_\mathbf{k} > \left(\frac{M}{c}\right)^{k/2} H^t G^{k-t}\right] < \frac{\sqrt{2}\, t\sigma}{\Delta\sqrt{cn\pi}} \exp\left[-\frac{cn\Delta^2}{2\sigma^2}\right] + \frac{\sqrt{2}\,(k-t)\sigma}{n^\epsilon\sqrt{c\pi}} \exp\left[-\frac{cn^{2\epsilon}}{2\sigma^2}\right] \tag{23}$$

where

$$G = 4^{2r} n^{-r+2r\epsilon} + (2r-1)!! \left(\frac{4\sigma^2}{cn}\right)^r \qquad \textit{and} \qquad H = (4\Delta)^{2r} + (2r-1)!! \left(\frac{4\sigma^2}{cn}\right)^r.$$

Note that $H^t G^{k-t} = O\left[n^{-r(k-t)(1-2\epsilon)}\right]$.

Proof of Lemma 5. From the conditions of the theorem, it follows from assumed bounds on the eigenvalues of $\mathbf{X}_n'\mathbf{X}_n$ that

$$Q_{\mathbf{k}} < \int |\mathbf{C}_{\mathbf{k}}|^{1/2}(2\pi\sigma^2)^{-k/2}\prod_{i=1}^{k}\beta_{k_i}^{2r}\exp\left[-\frac{cn}{2\sigma^2}(\boldsymbol{\beta}_{\mathbf{k}}-\tilde{\boldsymbol{\beta}}_{\mathbf{k}})'(\boldsymbol{\beta}_{\mathbf{k}}-\tilde{\boldsymbol{\beta}}_{\mathbf{k}})\right]d\boldsymbol{\beta}_{\mathbf{k}}.$$

Because $(\boldsymbol{\beta}_{k_i}-\tilde{\boldsymbol{\beta}}_{k_i})^2 < (2\boldsymbol{\beta}_{k_i})^2+(2\tilde{\boldsymbol{\beta}}_{k_i})^2$, moment properties of circular normal distributions imply that

$$\begin{aligned} Q_{\mathbf{k}} &< \left(\frac{M}{c}\right)^{k/2}\prod_{i=1}^{k}\left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r}+(2r-1)!!\left(\frac{4\sigma^2}{cn}\right)^{r}\right] && (24)\\ &= \left(\frac{M}{c}\right)^{k/2}\prod_{k_i\in\mathbf{t}}\left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r}+(2r-1)!!\left(\frac{4\sigma^2}{cn}\right)^{r}\right]\\ &\quad\times\prod_{k_i\notin\mathbf{t},k_i\in\mathbf{k}}\left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r}+(2r-1)!!\left(\frac{4\sigma^2}{cn}\right)^{r}\right]. && (25)\end{aligned}$$

Considering first the components of $\tilde{\boldsymbol{\beta}}_{\mathbf{k}}$ belonging to $\mathbf{t}$ and referring to Lemma 1,

$$\begin{aligned}\mathbf{P}\left[\prod_{k_i\in\mathbf{t}}\left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r}+(2r-1)!!\left(\frac{4\sigma^2}{cn}\right)^{r}\right]>H^t\right] &\le \sum_{k_i\in\mathbf{t}}\mathbf{P}\left[|\tilde{\boldsymbol{\beta}}_{k_i}|>2\Delta\right]\\ &\le \sum_{k_i\in\mathbf{t}}\mathbf{P}\left[\frac{|\tilde{\boldsymbol{\beta}}_{k_i}|-\Delta}{\sigma/\sqrt{nc}}>\frac{\Delta\sqrt{nc}}{\sigma}\right]. \quad (26)\end{aligned}$$

Noting that

$$\frac{1}{\sqrt{2\pi}}\int_u^\infty \exp(-x^2/2)dx < \frac{1}{\sqrt{2\pi}\,u}\exp(-u^2/2), \qquad (27)$$

and that each component of $\tilde{\boldsymbol{\beta}}_{k_i}$, $k_i\in\mathbf{t}$, is normally distributed with a mean that is smaller in magnitude than Δ and marginal variance smaller than $\sigma^2/(cn)$, it follows that

$$\mathbf{P}\left[\prod_{k_i \in \mathbf{t}} \left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r} + (2r-1)!!\left(\frac{4\sigma^2}{cn}\right)^r\right] > H^t\right] < \frac{\sqrt{2}\,t\sigma}{\Delta\sqrt{cn\pi}} \exp\left[-\frac{cn\Delta^2}{2\sigma^2}\right]. \tag{28}$$

A result similar to equation (28) can be obtained for terms in the second product of equation (25) by replacing Δ with $n^{-1/2+\epsilon}$ and noting that the sampling mean of each $\tilde{\boldsymbol{\beta}}_{k_i}$, $k_i \notin \mathbf{t}, k_i \in \mathbf{k}$, is bounded in magnitude by v/n, where $v < \sqrt{a_2 t\Delta/c}$. Because the bias term v/n is bounded by $2n^{-1/2+\epsilon}$ for sufficiently large n, it follows that

$$\mathbf{P}\left[\prod_{k_i \notin \mathbf{t}, k_i \in \mathbf{k}} \left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r} + (2r-1)!!\left(\frac{4\sigma^2}{cn}\right)^r\right] > G^{k-t}\right] < \frac{\sqrt{2}\,(k-t)\sigma}{n^{\epsilon}\sqrt{c\pi}} \exp\left[-\frac{cn^{2\epsilon}}{2\sigma^2}\right], \tag{29}$$

which completes the proof of the lemma.

Lemma 6 *Suppose that the conditions of Theorem 1 apply and that $t \not\subset k$. Define $\mathbf{s} = \mathbf{k} \cup \mathbf{t}$ and*

$$\lambda = \frac{1}{\sigma^2}\,\boldsymbol{\beta}'_{\mathbf{t}}\mathbf{X}'_t\mathbf{X}_t\boldsymbol{\beta}_{\mathbf{t}}.$$

Then

$$\mathbf{P}\left[Q_{\mathbf{k}} > q^{rk}\right] < \mathbf{P}\left[T > \left(\frac{c}{M}\right)^{0.5/r} \frac{nc^3kq}{4M^2\sigma^2} - \frac{c^2k[(2r-1)!!]^{1/r}}{M^2}\right], \tag{30}$$

where T follows a chi-squared distribution on $|\mathbf{s}|$ degrees of freedom and non-centrality parameter λ.

Proof of Lemma 6. From equation (24),

$$Q_{\mathbf{k}} < \left(\frac{M}{c}\right)^{k/2} \prod_{i=1}^{k} \left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r} + (2r-1)!! \left(\frac{4\sigma^2}{cn}\right)^{r} \right].$$

From the conditions of Theorem 1, it follows that

$$\begin{aligned} \tilde{\boldsymbol{\beta}}_{\mathbf{k}}' \tilde{\boldsymbol{\beta}}_{\mathbf{k}} &= \mathbf{y}_n' \mathbf{X}_{\mathbf{k}} \mathbf{C}_{\mathbf{k}}{}^{-2} \mathbf{X}_{\mathbf{k}}' \mathbf{y}_n \leq \frac{1}{(nc)^2} \mathbf{y}_n' \mathbf{X}_{\mathbf{k}} \mathbf{X}_{\mathbf{k}}' \mathbf{y}_n \\ &\leq \frac{1}{(nc)^2} \mathbf{y}_n' \mathbf{X}_{\mathbf{s}} \mathbf{X}_{\mathbf{s}}' \mathbf{y}_n \leq \frac{M^2}{c^2} \mathbf{y}_n' \mathbf{X}_{\mathbf{s}} (\mathbf{X}_{\mathbf{s}}' \mathbf{X}_{\mathbf{s}})^{-2} \mathbf{X}_{\mathbf{s}}' \mathbf{y}_n \leq \frac{M^2}{c^2} \hat{\boldsymbol{\beta}}_{\mathbf{s}}' \hat{\boldsymbol{\beta}}_{\mathbf{s}} \equiv V, \end{aligned} \tag{31}$$

where $\hat{\boldsymbol{\beta}}_{\mathbf{s}}$ is the least squares estimate of $\boldsymbol{\beta}_{\mathbf{s}}$. This implies that

$$\sum_{i=1}^{k} \left[4\tilde{\boldsymbol{\beta}}_{\mathbf{k}_i}^2 + [(2r-1)!!]^{0.5/r} \left(\frac{4\sigma^2}{cn}\right) \right] \leq 4V + k[(2r-1)!!]^{0.5/r} \left(\frac{4\sigma^2}{cn}\right).$$

If $\sum_1^k z_i \leq S$ and $z_i > 0$ for $i = 1, \ldots, k$, then it follows that $\prod_1^k z_i$ is maximized when $z_i = S/k$ for $i = 1, \ldots, k$. It follows that

$$\begin{aligned} Q_{\mathbf{k}} &< \left(\frac{M}{c}\right)^{k/2} \prod_{i=1}^{k} \left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2r} + (2r-1)!! \left(\frac{4\sigma^2}{cn}\right)^{r} \right] < \\ &\left(\frac{M}{c}\right)^{k/2} \prod_{i=1}^{k} \left[(2\tilde{\boldsymbol{\beta}}_{k_i})^{2} + [(2r-1)!!]^{1/r} \left(\frac{4\sigma^2}{cn}\right) \right]^{r} < \\ &\left(\frac{M}{c}\right)^{k/2} \left(\frac{4V}{k} + [(2r-1)!!]^{1/r} \frac{4\sigma^2}{cn} \right)^{rk}. \end{aligned} \tag{32}$$

Also,

$$V \leq \frac{M^2}{nc^3} \hat{\boldsymbol{\beta}}_s' \mathbf{X}_{\mathbf{s}}' \mathbf{X}_{\mathbf{s}} \hat{\boldsymbol{\beta}}_s \equiv \frac{M^2 \sigma^2}{nc^3} T,$$

where T is distributed as a non-central chi-squared distribution on s degrees of freedom and non-centrality parameter λ. It follows that

$$\begin{aligned}
\mathbf{P}\left[Q_{\mathbf{k}} > q^{rk}\right] &< P\left\{\left(\frac{M}{c}\right)^{k/2}\left[\frac{4V}{k} + [(2r-1)!!]^{1/r}\left(\frac{4\sigma^2}{cn}\right)\right]^{rk} > q^{rk}\right\} \\
&< \mathbf{P}\left[T > \left(\frac{c}{M}\right)^{0.5/r}\frac{nc^3kq}{4M^2\sigma^2} - \frac{c^2k[(2r-1)!!]^{1/r}}{M^2}\right]. \qquad (33)
\end{aligned}$$

With these lemmas in place, the proof of Theorem 1 can now be presented.

Proof of Theorem 1. The posterior probability $p(\mathbf{t}\,|\,\mathbf{y}_n)$ can be expressed as

$$\begin{aligned}
p(\mathbf{t}\,|\,\mathbf{y}_n) &= \frac{p(\mathbf{t})m_{\mathbf{t}}(\mathbf{y}_n)}{\sum_k p(\mathbf{k})m_{\mathbf{k}}(\mathbf{y}_n)} \\
&= \left[1 + \sum_{\mathbf{t}\subset\mathbf{k}}\frac{p(\mathbf{k})m_{\mathbf{k}}(\mathbf{y}_n)}{p(\mathbf{t})m_{\mathbf{t}}(\mathbf{y}_n)} + \sum_{\mathbf{t}\not\subset\mathbf{k}}\frac{p(\mathbf{k})m_{\mathbf{k}}(\mathbf{y}_n)}{p(\mathbf{t})m_{\mathbf{t}}(\mathbf{y}_n)}\right]^{-1}. \qquad (34)
\end{aligned}$$

Referring to equation (22), consider first the sum

$$\sum_{\mathbf{t}\subset\mathbf{k}}\frac{p(\mathbf{k})m_{\mathbf{k}}(\mathbf{y}_n)}{p(\mathbf{t})m_{\mathbf{t}}(\mathbf{y}_n)} < \frac{1}{Q_{\mathbf{t}}}\sum_{\mathbf{t}\subset\mathbf{k}}\alpha^k\gamma^t n^{-\frac{k-t}{2}}Q_{\mathbf{k}}\exp\left[-\frac{1}{2\sigma^2}(R_{\mathbf{k}} - R_{\mathbf{t}})\right]. \qquad (35)$$

Let $d = k - t$ and $\epsilon = s$ in Lemma 2. From Lemma 1 it follows that

$$\begin{aligned}
&\mathbf{P}\left(\alpha^k\gamma^t n^{-\frac{d}{2}}Q_{\mathbf{k}}\exp\left[-\frac{1}{2\sigma^2}(R_{\mathbf{k}} - R_{\mathbf{t}})\right] > \left[\frac{M}{c}\right]^{k/2}\frac{\alpha^d[\gamma\alpha]^t}{n^{d(1+\epsilon)}}\right]\right) < \\
&\mathbf{P}\left[n^{-6\epsilon d}Q_{\mathbf{k}} > \left[\frac{M}{c}\right]^{k/2}\frac{1}{n^{d(1+\epsilon)+d}}\right] + \mathbf{P}\left(n^{-(1/2-6\epsilon)d}\exp\left[-\frac{1}{2\sigma^2}(R_{\mathbf{k}} - R_{\mathbf{t}})\right] > n^d\right), \qquad (36)
\end{aligned}$$

for $d = 1, \ldots, p - t$.

For $r \geq 2$ in Lemma 5, $H^t G^d$ is of smaller order than $O(n^{-2d+4d\epsilon})$. Thus, it follows from Lemma 5 that

$$\mathbf{P}\left[Q_{\mathbf{k}} > \left[\frac{M}{c}\right]^{k/2} \frac{1}{n^{2d-5d\epsilon}}\right]$$

decreases exponentially fast with n as $n \to \infty$. This in turn implies that

$$\mathbf{P}\left[\alpha^k \gamma^t n^{-6\epsilon d} Q_{\mathbf{k}} > \left[\frac{M}{c}\right]^{k/2} \frac{\alpha^d (\gamma/\alpha)^t}{n^{d(2+\epsilon)}}\right] < \frac{1}{n^{d(1+\epsilon)}}$$

for large n.

The term $(R_{\mathbf{t}} - R_{\mathbf{k}})/\sigma^2$ can be bounded by using properties of the classical least squares estimates. For $\mathbf{k} \in \mathcal{J}$, let $R^*_{\mathbf{k}}$ denote the usual sum-of-squares obtained by taking $1/\tau = 0$. Then $R_{\mathbf{k}} > R^*_{\mathbf{k}}$, and $R_{\mathbf{t}} < R^*_{\mathbf{t}} + \omega$, where $\omega = a_1 \hat{\boldsymbol{\beta}}_{\mathbf{t}}{}' \hat{\boldsymbol{\beta}}_{\mathbf{t}} / \tau$. Under the assumptions of Theorem 1, the quantity $(R^*_{\mathbf{t}} - R^*_{\mathbf{k}})/\sigma^2$ has a $\chi^2_d(0)$ distribution. It follows from Lemma 3 that

$$\mathbf{P}\left(n^{-(1/2-6\epsilon)d} \exp\left[-\frac{(R_{\mathbf{k}} - R_{\mathbf{t}})}{2\sigma^2}\right] > n^d\right) < \mathbf{P}\left(n^{-(1/2-6\epsilon)d} \exp\left[-\frac{(R^*_{\mathbf{k}} - R^*_{\mathbf{t}} - \omega)}{2\sigma^2}\right] > n^d\right). \quad (37)$$

From Lemma 1,

$$\begin{aligned}
\mathbf{P}\left(n^{-(1/2-6\epsilon)d}\exp\left[-\frac{(R^*_{\mathbf{k}}-R^*_{\mathbf{t}}-\omega)}{2\sigma^2}\right]>n^d\right) &< \mathbf{P}\left(n^{-(1/2-7\epsilon)d}\exp\left[\frac{(R^*_{\mathbf{t}}-R^*_{\mathbf{k}})}{2\sigma^2}\right]>n^d\right)\\
&\quad+\mathbf{P}\left(n^{-\epsilon d}\exp\left[\frac{\omega}{2\sigma^2}\right]>1\right)\\
&= \mathbf{P}\left[\frac{R^*_{\mathbf{t}}-R^*_{\mathbf{k}}}{\sigma^2}>(3-14\epsilon)\ d\log(n)\right] \quad (38)\\
&\quad+\mathbf{P}\left[\frac{\omega}{\sigma^2}>2\epsilon d\log(n)\right]. \quad (39)
\end{aligned}$$

The quantity in equation (38) represents the probability that a central χ^2_d random variable exceeds $(3-14\epsilon)d\log(n)$. From Lemma 3, this probability is less than

$$\frac{2}{\sqrt{\pi}}d^{-1/2}[(3-14\epsilon)\log(n)]^{d/2-1}\exp^{d/2} n^{(-3/2+7\epsilon)d},$$

which for small ϵ is less than $O(n^{5d/4})$ for $d \leq n$. The quantity ω converges in probability to $a_1\boldsymbol{\beta}'_{\mathbf{t}}\boldsymbol{\beta}_{\mathbf{t}}/\tau$ and satisfies the inequality

$$\frac{nc\,\omega}{\sigma^2}<\frac{a_1}{\tau\sigma^2}\hat{\boldsymbol{\beta}}_{\mathbf{t}}{}'(\mathbf{X}'_{\mathbf{t}}\mathbf{X}_{\mathbf{t}})\hat{\boldsymbol{\beta}}_{\mathbf{t}}=\frac{a_1}{\tau}L, \qquad L\sim\chi^2_t(\boldsymbol{\beta}'_{\mathbf{t}}\mathbf{X}'_t\mathbf{X}_t\boldsymbol{\beta}_{\mathbf{t}}). \tag{40}$$

Applying Lemma 3 and re-expressing the quantity in equation (39) as

$$\mathbf{P}\left[\frac{nc\omega}{\sigma^2}>2\epsilon c\ dn\log(n)\right] \tag{41}$$

shows that this quantity also decreases exponentially with n as n becomes large. Thus,

$$\mathbf{P}\left(n^{-(1/2-6\epsilon)(k-t)}\exp\left[-\frac{1}{2\sigma^2}(R_{\mathbf{k}}-R_{\mathbf{t}})\right]>n^d\right)<\frac{1}{n^{d(1+\epsilon)}}$$

as $n \to \infty$.

Because $1/Q_{\mathbf{t}}$ converges almost surely to a positive constant (Lemma 4), the application of Lemma 2 and Slutsky's theorem implies that

$$\sum_{\mathbf{t} \subset \mathbf{k}} \frac{p(\mathbf{k}) m_{\mathbf{k}}(\mathbf{y}_n)}{p(\mathbf{t}) m_{\mathbf{t}}(\mathbf{y}_n)} \xrightarrow{p} 0.$$

Again referring to equation (22), consider next the sum

$$\sum_{\mathbf{t} \not\subset \mathbf{k}} \frac{p(\mathbf{k}) m_{\mathbf{k}}(\mathbf{y}_n)}{p(\mathbf{t}) m_{\mathbf{t}}(\mathbf{y}_n)} < \frac{1}{Q_{\mathbf{t}}} \sum_{\mathbf{t} \not\subset \mathbf{k}} \alpha^k \gamma^t n^{-\frac{k-t}{2}} Q_{\mathbf{k}} \exp\left[-\frac{1}{2\sigma^2} (R_{\mathbf{k}} - R_{\mathbf{t}}) \right]. \qquad (42)$$

Terms in this sum will be split into two groups. The first group includes terms for which $k < n^{1-1/(8r)}$; in this group the non-central chi-squared contribution from $\exp[-0.5(R_{\mathbf{k}} - R_{\mathbf{t}})/\sigma^2]$ dominates the Bayes factor so that Lemma 1 can be applied. For terms in the second group with larger k, the loss of degrees of freedom in the exponentiated chi-squared term is offset by the value of $E(\prod \beta_{k_i}^{2r})$, which becomes small even when $\mathbf{k} \not\subset \mathbf{t}$.

Define models $\mathbf{u}$ and $\mathbf{s}$ according to $\mathbf{u} = \mathbf{k} \cup \mathbf{t}$ and $\mathbf{s} = \mathbf{k} \cap \mathbf{t}$, and consider first the sum over terms for which $k \leq n^{1-1/(8r)}$.

From Lemma 1, terms in this series obey the following inequality:

$$\mathbf{P}\left(\alpha^k\gamma^t n^{-\frac{k-t}{2}} Q_{\mathbf{k}} \exp\left[-\frac{1}{2\sigma^2}(R_{\mathbf{k}} - R_{\mathbf{t}})\right] > \frac{\alpha^k\gamma^t}{n^{k(1+p)}}\right)$$
$$\leq \mathbf{P}\left(n^t \exp\left[-\frac{1}{2\sigma^2}(R^*_{\mathbf{k}} - R^*_{\mathbf{u}})\right] > \frac{1}{n^{3k(1+p)}}\right) + \mathbf{P}\left(n^{-\frac{t}{2}} \exp\left[\frac{\omega}{2\sigma^2}\right] > 1\right)$$
$$+ \mathbf{P}\left(n^{-\frac{k}{2}} \exp\left[\frac{1}{2\sigma^2}(R^*_{\mathbf{t}} - R^*_{\mathbf{u}})\right] > n^{k(1+p)}\right) + \mathbf{P}\left[Q_{\mathbf{k}} > n^{k(1+p)}\right]. \quad (43)$$

Define $\mathbf{W}$ to be the $(t-s) \times t$ submatrix of $\mathbf{I}_t$ obtained by selecting the rows i of $\mathbf{I}_t$ for which $i \in \mathbf{t}$ and $i \notin \mathbf{s}$. From the assumptions of the theorem, the quantity

$$\frac{1}{\sigma^2}(R^*_{\mathbf{k}} - R^*_{\mathbf{u}})$$

has a $\chi^2_{u-k}(\lambda)$ distribution with non-centrality parameter

$$\lambda = \boldsymbol{\beta}'_{\mathbf{t}}\mathbf{W}'\left(\mathbf{W}(\mathbf{X}'_{\mathbf{t}}\mathbf{X}_{\mathbf{t}})^{-1}\mathbf{W}'\right)^{-1}\mathbf{W}\boldsymbol{\beta}_{\mathbf{t}} > nc\delta^2. \quad (44)$$

If Z has a standard normal distribution and $6k(1+p) - 2t > 0$, it follows that

$$\mathbf{P}\left(\exp\left[-\frac{1}{2\sigma^2}(R^*_{\mathbf{k}} - R^*_{\mathbf{u}})\right] > \frac{1}{n^{3k(1+p)-t}}\right)$$
$$= \mathbf{P}\left(\frac{R^*_{\mathbf{k}} - R^*_{\mathbf{u}}}{\sigma^2} < [6k(1+p) - 2t]\log(n)\right)$$
$$< \mathbf{P}\left\{(Z - \sqrt{\lambda})^2 < [6k(1+p) - 2t]\log(n)\right\}$$
$$< \mathbf{P}\left[Z > \sqrt{\lambda} - \sqrt{[6k(1+p) - 2t]}\right]$$
$$< \mathbf{P}\left[Z > \sqrt{nc\delta^2} - \sqrt{[6n^{1-1/(8r)}(1+p) - 2t]\log(n)}\right], \quad (45)$$

which is $O(e^{-cn})$ for some $c > 0$ as $n \to \infty$. (If $6k(1+p) - 2t < 0$, then $k < t$ and $R_k - R_t$ is $O_p(n)$, so that the term on the left-hand side of equation (43) decreases exponentially fast with n. Because there is a finite number of such terms, they may be ignored.)

As noted in equation (41), the second term on the right-hand side of equation (43) decreases exponentially fast with n as $n \to \infty$.

In the third term on the right-hand side of equation (43), $(R^*_{\mathbf{t}} - R^*_{\mathbf{u}})/\sigma^2$ follows a $\chi^2_{u-t}(0)$ distribution, so from Lemma (3) it follows that

$$\mathbf{P}\left(n^{-\frac{k}{2}} \exp\left[\frac{1}{2\sigma^2}(R^*_{\mathbf{t}} - R^*_{\mathbf{u}})\right] > n^{k(1+p)}\right) < \\ 2\sqrt{\frac{u-t}{\pi}}\left(\frac{2k(1+p)+k}{u-t}\log(n)\right)^{\frac{u-t}{2}-1} n^{-k(1+p)-k/2}\exp[(u-t)/2], \quad (46)$$

which is smaller than $n^{-k(1+p)}$ for large n and $k < n^{1-1/(8r)}$.

The probability of the final term in equation (43) can be bounded by $n^{-k(1+p)}$ by taking $q = n^{(1+p)/r}$ in Lemma 6 and $a = n^{3/4}$ in Lemma 3. The leading term in the value of t in Lemma 3 is then $O(kn^{(1+(1+p)/r})$, which guarantees that the final probability in equation (43) is $O(e^{-cn})$ for some $c > 0$ as $n \to \infty$.

For models $\mathbf{k}$ with $k > n^{1-1/(8r)}$ and $\mathbf{t} \not\subset \mathbf{k}$,

$$\begin{aligned}
\mathbf{P}\left(\alpha^k\gamma^t n^{-\frac{k-t}{2}} Q_{\mathbf{k}} \exp\left[-\frac{1}{2\sigma^2}(R_{\mathbf{k}} - R_{\mathbf{t}})\right] > \frac{\alpha^k\gamma^t}{n^{k(1+p)}}\right) \leq \\
\mathbf{P}\left(\exp\left[-\frac{1}{2\sigma^2}(R^*_{\mathbf{k}} - R^*_{\mathbf{u}} - \omega)\right] > 1\right) \\
+ \mathbf{P}\left[Q_{\mathbf{k}} > n^{-3k(1+p)/4}\right] \\
+ \mathbf{P}\left(n^{-\frac{k-t}{2}} \exp\left[\frac{1}{2\sigma^2}(R^*_{\mathbf{t}} - R^*_{\mathbf{u}})\right] > n^{-k(1+p)/4}\right). \quad (47)
\end{aligned}$$

The probability of the term in the second line of equation (47) is $O(e^{-cn})$ for some $c > 0$.

To obtain a bound for the term in the third line of equation (47), we apply Lemma 6 with $p < 1/6$. In that lemma, take $q = n^{-3(1+p)/(4r)}$. Then the probability in the third line is bounded by the probability that a non-central chi-squared random variable is greater than

$$\left(\frac{c}{M}\right)^{0.5/r} \frac{nc^3kq}{4M^2\sigma^2} - \frac{c^2k[(2r-1)!!]^{1/r}}{M^2} > O(n^{3/2}) \quad (48)$$

since $k > n^{1-1/(8r)}$ and $r \geq 2$. The degree of freedom of the non-central chi-squared random variable is less than n, and its non-centrality parameter is bounded by $t\Delta Mn/\sigma^2$. By Lemma 3, the probability in the third line thus decreases exponentially with n as n grows.

The probability in the fourth line represents the probability that a central chi-squared random variable on $u-t$ degrees of freedom exceeds $(k/2 - pk/2 - t)\log(n)$.

Because $u - t < k$ and $p < 1/6$, the bound in Lemma 3 implies that this probability also decreases exponentially with n for large n.

The proof of the theorem follows by application of Lemma 2.

Proof of Corollary 1: The proof of this corollary follows along the same lines as that of Theorem 1. For brevity, only an outline and correspondence to the proof of Theorem 1 are provided. To simplify the notation and without loss of generality, we consider the limiting case of $\alpha, \psi \to 0$.

Using results reported in [5] (Section 3), it follows that

$$E_{\mathbf{k}}^T \left(\prod_{i \in \mathbf{k}} \beta_i^{2r} \right) \leq \left(\frac{M}{c} \right)^{\frac{k}{2}} \left(\frac{s_{\mathbf{k}}}{\sqrt{nc}} \right)^{2rk} \left[\frac{(4e)^r \Gamma(r + \frac{1}{2})}{\sqrt{\pi}} \right]^k + \left(\frac{M}{c} \right)^{\frac{k}{2}} 2^{2rk} \prod_{i \in \mathbf{k}} \mathbf{b}_{k_i}^{2r}.$$

From this inequality and Lemma 3, bounds similar to those presented in Lemmas 5-6 for $E_{\mathbf{k}} \left(\prod_{i \in \mathbf{k}} \beta_i^{2r} \right)$ under a Gaussian expectation can be obtained. Note that the degrees of freedom associated with $E_{\mathbf{k}}^T$ are $n + 2rk$, which guarantees the existence of required moments.

The determinant in equation (6) plays similar roles in the proofs of the corollary and in Theorem 1.

The remaining terms in equation (6) that must be evaluated to establish the convergence of the posterior probability of the true model obey the following inequality

[3]:

$$\left(\frac{\nu_{\mathbf{k}}{s_{\mathbf{k}}}^2}{2}\right)^{-\nu_{\mathbf{k}}/2}\Gamma\left(\frac{\nu_{\mathbf{k}}}{2}\right) \leq \sqrt{2\pi}\left(\frac{\nu_{\mathbf{k}}}{2}-\frac{1}{6}\right)^{-\frac{1}{2}}\left(\frac{eR_{\mathbf{k}}}{\nu_{\mathbf{k}}}\right)^{-\frac{\nu_{\mathbf{k}}}{2}}. \tag{49}$$

For $\mathbf{t} \subset \mathbf{k}$, the ratio of corresponding terms in $m_{\mathbf{k}}(\mathbf{y}_n)$ and $m_{\mathbf{t}}(\mathbf{y}_n)$ thus satisfies

$$\frac{\left(\nu_{\mathbf{k}}{s_{\mathbf{k}}}^2/2\right)^{-\nu_{\mathbf{k}}/2}\Gamma\left(\nu_{\mathbf{k}}/2\right)}{\left(\nu_{\mathbf{t}}{s_{\mathbf{t}}}^2/2\right)^{-\nu_{\mathbf{t}}/2}\Gamma\left(\nu_{\mathbf{t}}/2\right)} \leq A \times B \times C \times D, \tag{50}$$

where

$$A = \left(1+\frac{R_{\mathbf{t}}-R_{\mathbf{k}}}{R_{\mathbf{k}}}\right)^{\frac{\nu_{\mathbf{k}}}{2}} \quad B = \left(\frac{R_{\mathbf{t}}}{\nu_{\mathbf{t}}}\right)^{\frac{\nu_{\mathbf{t}}-\nu_{\mathbf{k}}}{2}} \quad C = e^{r(t-k)} \quad D = \left(1+\frac{r(t-k)}{(n+2rk)/2}\right)^{-\frac{n+2rk}{2}}. \tag{51}$$

The factor A can be bounded according to

$$\begin{aligned} \mathbf{P}\left[A > n^{k-t}\right] &= \mathbf{P}\left[\frac{R_{\mathbf{t}}-R_{\mathbf{k}}}{R_{\mathbf{k}}} > n^{\frac{2(k-t)}{\nu_{\mathbf{k}}}}-1\right] \\ &\leq \mathbf{P}\left[\frac{n-k}{k-t}\frac{R_{\mathbf{t}}-R_{\mathbf{k}}}{R_{\mathbf{k}}} > 2\frac{(n-k)}{\nu_{\mathbf{k}}}\log(n)\right] \\ &= \mathbf{P}\left[F_1 > 2\frac{(n-k)}{(n+2rk)}\log(n)\right] \end{aligned} \tag{52}$$

where F_1 denotes a random variable distributed according to a central F-distribution on $(k-t, n-k)$ degrees of freedom. Because $F_1 \to \chi^2_{k-t}(0)/(n-k)$ as $(n-k) \to \infty$, it is possible to obtain bounds on A that are similar to the bounds obtained for $\exp[-0.5(R_{\mathbf{k}}-R_{\mathbf{t}})/\sigma^2]$ in equation (35).

The factor B in equation (50) converges to σ^{-rk}, and corresponds to the same factor appearing in the Bayes factor resulting from the marginal densities for the known variance case in equation (5). Finally, $C \times D \to 1$ as $n \to \infty$.

For $\mathbf{t} \not\subset \mathbf{k}$, if $\mathbf{s} = \mathbf{k} \cup \mathbf{t}$ it follows that

$$\frac{(eR_{\mathbf{k}}/\nu_{\mathbf{k}})^{-\nu_{\mathbf{k}}/2}}{(eR_{\mathbf{t}}/\nu_{\mathbf{t}})^{-\nu_{\mathbf{t}}/2}} = G \times H \times J \times C \times D, \tag{53}$$

where

$$G = \left(\frac{R_{\mathbf{s}}}{R_{\mathbf{k}}}\right)^{\frac{\nu_{\mathbf{k}}}{2}} = \left(1 + \frac{r-k}{n-r} F_2\right)^{-\frac{\nu_{\mathbf{k}}}{2}}, \tag{54}$$

$$H = \left(\frac{R_{\mathbf{t}}}{R_{\mathbf{s}}}\right)^{\frac{\nu_{\mathbf{k}}}{2}} = \left(1 + \frac{r-t}{n-r} F_3\right)^{\frac{\nu_{\mathbf{k}}}{2}}, \tag{55}$$

and

$$J = \left(\frac{R_{\mathbf{t}}}{\nu_{\mathbf{t}}}\right)^{r(t-k)} = \left(\frac{\sigma^2}{\nu_{\mathbf{k}}} X\right)^{r(t-k)}. \tag{56}$$

$$\frac{\left(\frac{eR_{\mathbf{k}}}{\nu_{\mathbf{k}}}\right)^{-\frac{\nu_{\mathbf{k}}}{2}}}{\left(\frac{eR_{\mathbf{t}}}{\nu_{\mathbf{t}}}\right)^{-\frac{\nu_{\mathbf{t}}}{2}}} = \left(\frac{R_{\mathbf{s}}}{R_{\mathbf{k}}}\right)^{\frac{\nu_{\mathbf{k}}}{2}} \left(\frac{R_{\mathbf{t}}}{R_{\mathbf{s}}}\right)^{\frac{\nu_{\mathbf{k}}}{2}} \left(\frac{R_{\mathbf{t}}}{\nu_{\mathbf{t}}}\right)^{r(t-k)} \times C \times D \tag{57}$$

$$= \left(1 + \frac{r-k}{n-r} F_2\right)^{-\frac{\nu_{\mathbf{k}}}{2}} \left(1 + \frac{r-t}{n-r} F_3\right)^{\frac{\nu_{\mathbf{k}}}{2}} \left(\frac{\sigma^2}{\nu_{\mathbf{k}}} X\right)^{r(t-k)} \times C \times D \tag{58}$$

In equation (54), F_2 is distributed as a non-central F random variable on $(r-k, n-r)$ degrees of freedom and noncentrality parameter λ, equation (44); whereas in equation (55) the quantity F_3 is distributed as a central F random variable on $(r-t, n-r)$ degrees of freedom. The factor J is similar to B in equation (50); X is distributed as a chi-squared random variable on $n-t$ degrees of freedom. Noting that $F_2 \longrightarrow \chi^2_{r-k}(\lambda)/(r-k)$ and $F_3 \longrightarrow \chi^2_{r-t}(0)/(r-t)$, as $(n-r) \to \infty$ when $k < bn$, inequalities like those displayed for the case $\mathbf{t} \subset \mathbf{k}$ (i.e., equation 52) can be applied to the product in equation (57) in order to obtain bounds similar to those obtained for equations (43) and (47) in the known variance case.

Proof of Corollary 2: Consistency for the piMOM priors can be established from the consistency of pMOM priors under similar conditions. Taking $\mathbf{A}_\mathbf{k} = \mathbf{I}_k$ in equation (2) and factoring the densities in equations (2) and (3) componentwise for a given value of (τ, σ^2), it follows from the inequality

$$\begin{aligned}\left(\frac{\boldsymbol{\beta}_{\mathbf{k}_i}^2}{\tau\sigma^2}\right)^{-(r+1)/2} \exp\left(-\frac{\tau\sigma^2}{\boldsymbol{\beta}_{\mathbf{k}_i}^2}\right) &= \left(\frac{\boldsymbol{\beta}_{\mathbf{k}_i}^2}{\tau\sigma^2}\right)^{-(r+1)/2} \frac{1}{\sum_{j=0}^{\infty} \frac{1}{j!}\left(\frac{\tau\sigma^2}{\boldsymbol{\beta}_{\mathbf{k}_j}^2}\right)^j} \\ &< r! \left(\frac{\boldsymbol{\beta}_{\mathbf{k}_i}^2}{\tau\sigma^2}\right)^{m-(r+1)/2}, \qquad m \in \mathcal{Z}^+, \qquad (59)\end{aligned}$$

that factors of the piMOM density for small values of $\boldsymbol{\beta}_{\mathbf{k}_i}$ are bounded by the corresponding factors of a pMOM density of order 4 or greater, provided that $m > 4 + (r+1)/2$. Conversely,

$$\left(\frac{\boldsymbol{\beta}_{\mathbf{k}_i}^2}{\tau\sigma^2}\right)^{-(r+1)/2} \exp\left(-\frac{\tau\sigma^2}{\boldsymbol{\beta}_{\mathbf{k}_i}^2}\right) > c\left(\frac{\boldsymbol{\beta}_{\mathbf{k}_i}^2}{\tau\sigma^2}\right)^{m-(r+1)/2} \exp\left(-\frac{1}{2\tau\sigma^2}\boldsymbol{\beta}_{\mathbf{k}_i}^2\right) \qquad (60)$$

for some constant c and $|\boldsymbol{\beta}_{\mathbf{k}_i}| > \epsilon$ for an arbitrary $\epsilon > 0$. Thus, the penalty that results from small values of $\boldsymbol{\beta}_{\mathbf{k}_i}$ on the marginal densities obtained with piMOM prior densities imposed on regression coefficients is greater than it is under a comparable pMOM prior; whereas the penalty incurred for values of the components bounded away from 0 is of the same asymptotic order. For $\mathbf{k} \neq \mathbf{t}$, if we assume that the prior density on σ^2 is the same under both the piMOM and pMOM models, it follows that the ratio of the marginal density of the data under a piMOM prior density on $\boldsymbol{\beta}_{\mathbf{k}}$ to the corresponding marginal density obtained under a pMOM prior density converges to a constant less than 1. At the true model, the ratio of the marginal densities obtained under the corresponding pMOM and piMOM priors converges in probability to the ratio of their prior densities evaluated at $\boldsymbol{\beta}_{\mathbf{t}}^0$, which is bounded away from 0. Thus, $p(\mathbf{t} \,|\, \mathbf{y}_n) \xrightarrow{p} 1$.

Proof of Theorem 2: For $\mathbf{k} \in \mathcal{J}_1$, it follows that

$$\begin{aligned} \frac{m_k(\mathbf{y}_n)}{m_t(\mathbf{y}_n)} &= \frac{(2\pi\sigma^2)^{k/2}\,|\mathbf{X}_{\mathbf{k}}'\mathbf{X}_{\mathbf{k}}|^{-1/2}}{(2\pi\sigma^2)^{t/2}\,|\mathbf{X}_{\mathbf{t}}'\mathbf{X}_{\mathbf{t}}|^{-1/2}} \exp\left(-\frac{R_k - R_t}{2\sigma^2}\right) \\ &\quad \times \frac{\int_{\mathcal{R}^k} \pi_{\mathbf{k}}^L(\boldsymbol{\beta}_{\mathbf{k}})\ \phi_k\left[\boldsymbol{\beta}_{\mathbf{k}}; \hat{\boldsymbol{\beta}}_k, \sigma^2(\mathbf{X}_{\mathbf{k}}'\mathbf{X}_{\mathbf{k}})^{-1}\right] d\boldsymbol{\beta}_{\mathbf{k}}}{\int_{\mathcal{R}^t} \pi_{\mathbf{t}}^L(\boldsymbol{\beta}_{\mathbf{t}})\ \phi_t\left[\boldsymbol{\beta}_{\mathbf{t}}; \hat{\boldsymbol{\beta}}_t, \sigma^2(\mathbf{X}_{\mathbf{t}}'\mathbf{X}_{\mathbf{t}})^{-1}\right] d\boldsymbol{\beta}_{\mathbf{t}}} \\ &\geq n^{-1/2}\left(\frac{c}{M}\right)^{t/2}\sqrt{\frac{2\pi\sigma^2}{M}}\, \frac{\int_{\mathcal{R}^k} \pi_{\mathbf{k}}^L(\boldsymbol{\beta}_{\mathbf{k}})\ \phi_k\left[\boldsymbol{\beta}_{\mathbf{k}}; \hat{\boldsymbol{\beta}}_k, \sigma^2(\mathbf{X}_{\mathbf{k}}'\mathbf{X}_{\mathbf{k}})^{-1}\right] d\boldsymbol{\beta}_{\mathbf{k}}}{\int_{\mathcal{R}^t} \pi_{\mathbf{t}}^L(\boldsymbol{\beta}_{\mathbf{t}})\ \phi_t\left[\boldsymbol{\beta}_{\mathbf{t}}; \hat{\boldsymbol{\beta}}_t, \sigma^2(\mathbf{X}_{\mathbf{t}}'\mathbf{X}_{\mathbf{t}})^{-1}\right] d\boldsymbol{\beta}_{\mathbf{t}}}, \end{aligned} \tag{61}$$

where $\phi_k(\cdot; \boldsymbol{\mu}, \mathbf{C})$ denotes a k-variate normal distribution with mean equal to $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$, and $\hat{\boldsymbol{\beta}}$ denotes the least squares estimate under the referenced model.

For sufficiently large n, the almost sure convergence of the least squares estimates under each model [2] guarantees that

$$\frac{\int_{\mathcal{R}^k} \pi_{\mathbf{k}}^L(\boldsymbol{\beta}_{\mathbf{k}})\ \phi_k\left[\boldsymbol{\beta}_{\mathbf{k}}; \hat{\boldsymbol{\beta}}_k, \sigma^2(\mathbf{X}_{\mathbf{k}}'\mathbf{X}_{\mathbf{k}})^{-1}\right] d\boldsymbol{\beta}_{\mathbf{k}}}{\int_{\mathcal{R}^t} \pi_{\mathbf{t}}^L(\boldsymbol{\beta}_{\mathbf{t}})\ \phi_t\left[\boldsymbol{\beta}_{\mathbf{t}}; \hat{\boldsymbol{\beta}}_t, \sigma^2(\mathbf{X}_{\mathbf{t}}'\mathbf{X}_{\mathbf{t}})^{-1}\right] d\boldsymbol{\beta}_{\mathbf{t}}} \xrightarrow{a.s.} \frac{\pi_{\mathbf{k}}^L(\boldsymbol{\gamma}_{\mathbf{k}})}{\pi_{\mathbf{t}}^L\boldsymbol{\beta}_{\mathbf{t}})} > c_L. \tag{62}$$

Hence,

$$\mathbf{P}\left[\frac{m_k(\mathbf{y}_n)}{m_t(\mathbf{y}_n)} > n^{-1/2}\left(\frac{c}{M}\right)^{t/2}\sqrt{\frac{2\pi\sigma^2}{M}}c_L\right] \xrightarrow{a.s.} 1.$$

Noting that there are $(p-t)$ models in $\mathcal{J}_1$, the proof of the theorem follows from the fact that

$$\begin{aligned} p(\mathbf{t} \mid \mathbf{y}_n) &= \frac{p(\mathbf{t})m_{\mathbf{t}}(\mathbf{y}_n)}{\sum_{k\in\mathcal{J}} p(\mathbf{k})m_{\mathbf{k}}(\mathbf{y}_n)} \\ &< \frac{p(\mathbf{t})m_{\mathbf{t}}(\mathbf{y}_n)}{\sum_{k\in\mathcal{J}_1} p(\mathbf{k})m_{\mathbf{k}}(\mathbf{y}_n)}. \end{aligned}$$

Extended simulation results

The following tables provide numerical values for Figures 3 and 4 in the main article, as well as supplementary data. In the tables, pp_0 and pp_1 refer to the marginal posterior probability that a zero and a non-zero coefficient were included in a sampled model, respectively. FDR refers to the false discovery rate observed in each simulation, and MSE refers to the mean squared error of prediction within the simulation setting. For the IBF and BIC summaries, the MSE was evaluated at the least squares estimates of the regression coefficient calculated under the MAP model. The least squares estimates were used because of the difficulty in obtaining the MAP estimate for the regression coefficients under the intrinsic prior and the lack of a formal prior (and thus a MAP estimate) for the BIC.

$\sigma^2 = 1,\ \rho = 0$								
	$n = 100$		$n = 200$		$n = 300$		$n = 500$	
	pp_0	pp_1	pp_0	pp_1	pp_0	pp_1	pp_0	pp_1
pMOM $r = 1$	0.000	0.951	0.000	0.999	0.000	1.000	0.000	1.000
pMOM $r = 2$	0.000	0.846	0.000	0.988	0.000	1.000	0.000	1.000
piMOM	0.001	0.982	0.000	1.000	0.000	1.000	0.000	1.000
IBF	0.049	0.997	0.028	1.000	0.021	1.000	0.014	1.000
BIC	0.043	0.993	0.330	0.991	0.709	0.987	0.828	0.990
$\sigma^2 = 1,\ \rho = 0$								
pMOM $r = 1$	0.000	0.943	0.000	0.999	0.000	1.000	0.000	1.000
pMOM $r = 2$	0.000	0.893	0.000	0.995	0.000	1.000	0.000	1.000
piMOM	0.001	0.973	0.000	1.000	0.000	1.000	0.000	1.000
IBF	0.034	0.994	0.019	1.000	0.014	1.000	0.009	1.000
BIC	0.017	0.979	0.096	0.995	0.369	0.988	0.744	0.983
$\sigma^2 = 1.5,\ \rho = 0$								
pMOM $r = 1$	0.000	0.904	0.000	0.992	0.000	0.999	0.000	1.000
pMOM $r = 2$	0.000	0.835	0.000	0.960	0.000	0.997	0.000	1.000
piMOM	0.001	0.937	0.000	0.996	0.000	1.000	0.000	1.000
IBF	0.066	0.990	0.041	1.000	0.029	1.000	0.020	1.000
BIC	0.126	0.971	0.596	0.981	0.802	0.982	0.844	0.988
$\sigma^2 = 1.5,\ \rho = 0$								
pMOM $r = 1$	0.000	0.897	0.000	0.981	0.000	0.998	0.000	1.000
pMOM $r = 2$	0.000	0.865	0.000	0.976	0.000	0.998	0.000	1.000
piMOM	0.001	0.917	0.000	0.990	0.000	0.999	0.000	1.000
IBF	0.043	0.969	0.025	0.998	0.018	1.000	0.012	1.000
BIC	0.031	0.948	0.233	0.987	0.612	0.979	0.824	0.980
$\sigma^2 = 2,\ \rho = 0$								
pMOM $r = 1$	0.000	0.868	0.000	0.961	0.000	0.996	0.000	1.000
pMOM $r = 2$	0.000	0.820	0.000	0.929	0.000	0.986	0.000	1.000
piMOM	0.001	0.900	0.000	0.978	0.000	0.998	0.000	1.000
IBF	0.080	0.975	0.051	0.999	0.038	1.000	0.027	1.000
BIC	0.232	0.948	0.729	0.974	0.848	0.981	0.882	0.985
$\sigma^2 = 2,\ \rho = 0.25$								
pMOM $r = 1$	0.000	0.866	0.000	0.946	0.000	0.990	0.000	1.000
pMOM $r = 2$	0.001	0.850	0.000	0.943	0.000	0.989	0.000	1.000
piMOM	0.001	0.882	0.000	0.960	0.000	0.993	0.000	1.000
IBF	0.053	0.953	0.032	0.994	0.023	1.000	0.015	1.000
BIC	0.056	0.920	0.418	0.969	0.752	0.973	0.848	0.981

Table 1: Marginal inclusion probabilities for zero (pp_0) and non-zero (pp_1) coefficients.

	$p(\mathbf{t}\|\mathbf{y}_n)$	pp_0	pp_1	$\mathbf{P}(\hat{\mathbf{t}} = \mathbf{t})$	FDR	MSE
$n = 100$						
pMOM $r = 1$	0.746	0.000	0.951	0.767	0.000	1.146
pMOM $r = 2$	0.229	0.000	0.846	0.190	0.000	1.614
piMOM	0.852	0.001	0.982	0.922	0.001	1.077
IBF	0.032	0.049	0.997	0.746	0.051	1.093
BIC	0.224	0.043	0.993	0.929	0.017	5.107
SCAD				0.135	0.377	1.110
LASSO				0.001	0.714	1.289
$n = 200$						
pMOM $r = 1$	0.746	0.000	0.951	0.767	0.000	1.146
pMOM $r = 2$	0.229	0.000	0.846	0.190	0.000	1.614
piMOM	0.852	0.001	0.982	0.922	0.001	1.077
IBF	0.032	0.049	0.997	0.746	0.051	1.093
BIC	0.224	0.043	0.993	0.929	0.017	5.107
SCAD				0.135	0.377	1.110
LASSO				0.001	0.714	1.289
$n = 300$						
pMOM $r = 1$	0.993	0.000	0.999	0.997	0.000	1.026
pMOM $r = 2$	0.936	0.000	0.988	0.953	0.000	1.109
piMOM	0.980	0.000	1.000	0.997	0.001	1.005
IBF	0.021	0.028	1.000	0.778	0.044	1.011
BIC	0.127	0.330	0.991	0.717	0.192	22.666
SCAD				0.314	0.261	1.035
LASSO				0.001	0.748	1.165
$n = 500$						
pMOM $r = 1$	0.999	0.000	1.000	1.000	0.000	1.002
pMOM $r = 2$	1.000	0.000	1.000	1.000	0.000	1.021
piMOM	0.999	0.000	1.000	1.000	0.000	0.995
IBF	0.011	0.014	1.000	0.776	0.044	1.024
BIC	0.011	0.828	0.990	0.071	0.895	158.161
SCAD				0.823	0.038	1.018
LASSO				0.001	0.778	1.065

Table 2: $\sigma^2 = 1$, $\rho = 0$, $t = 5$. $\boldsymbol{\beta}'_{\mathbf{t}} = c(0.6, 1.2, 1.8, 2.4, 3.0)$

$n = 100$						
	$p(\mathbf{t}\|\mathbf{y}_n)$	pp_0	pp_1	$\mathbf{P}(\hat{\mathbf{t}} = \mathbf{t})$	FDR	MSE
pMOM $r = 1$	0.703	0.000	0.943	0.728	0.000	1.155
pMOM $r = 2$	0.454	0.000	0.893	0.456	0.000	1.378
piMOM	0.798	0.001	0.973	0.880	0.001	1.090
IBF	0.078	0.034	0.994	0.800	0.040	1.089
BIC	0.258	0.017	0.979	0.912	0.005	1.089
SCAD				0.395	0.157	1.074
LASSO				0.000	0.712	1.235
$n = 200$						
pMOM $r = 1$	0.988	0.000	0.999	0.996	0.000	1.043
pMOM $r = 2$	0.968	0.000	0.995	0.983	0.000	1.083
piMOM	0.977	0.000	1.000	0.998	0.000	1.026
IBF	0.061	0.019	1.000	0.806	0.039	1.031
BIC	0.250	0.096	0.995	0.954	0.018	13.812
SCAD				0.847	0.030	1.093
LASSO				0.000	0.765	1.137
$n = 300$						
pMOM $r = 1$	0.997	0.000	1.000	0.999	0.000	0.999
pMOM $r = 2$	0.998	0.000	1.000	1.000	0.000	1.037
piMOM	0.992	0.000	1.000	0.999	0.000	1.024
IBF	0.050	0.014	1.000	0.818	0.035	1.023
BIC	0.108	0.369	0.988	0.633	0.227	25.801
SCAD				0.983	0.003	1.064
LASSO				0.000	0.788	1.112
$n = 500$						
pMOM $r = 1$	0.999	0.000	1.000	1.000	0.000	1.033
pMOM $r = 2$	0.999	0.000	1.000	1.000	0.000	1.018
piMOM	0.996	0.000	1.000	0.999	0.000	1.049
IBF	0.040	0.009	1.000	0.806	0.038	1.020
BIC	0.016	0.744	0.983	0.100	0.841	162.107
SCAD				1.000	0.000	1.042
LASSO				0.000	0.814	1.088

Table 3: $\sigma^2 = 1$, $\rho = 0.25$, $t = 5$. $\boldsymbol{\beta}'_{\mathbf{t}} = c(0.6, 1.2, 1.8, 2.4, 3.0)$

$n = 100$						
	$p(\mathbf{t}\|\mathbf{y}_n)$	pp_0	pp_1	$\mathbf{P}(\hat{\mathbf{t}} = \mathbf{t})$	FDR	MSE
pMOM $r = 1$	0.510	0.000	0.904	0.515	0.000	1.772
pMOM $r = 2$	0.181	0.000	0.835	0.142	0.000	2.065
piMOM	0.638	0.001	0.937	0.687	0.002	1.687
IBF	0.014	0.066	0.990	0.630	0.075	1.697
BIC	0.145	0.126	0.971	0.753	0.065	14.305
SCAD				0.116	0.410	1.681
LASSO				0.001	0.715	1.937
$n = 200$						
pMOM $r = 1$	0.954	0.000	0.992	0.966	0.000	1.579
pMOM $r = 2$	0.798	0.000	0.960	0.814	0.000	1.702
piMOM	0.961	0.000	0.996	0.984	0.000	1.566
IBF	0.006	0.041	1.000	0.696	0.062	1.588
BIC	0.048	0.596	0.981	0.339	0.556	263.843
SCAD				0.136	0.407	1.546
LASSO				0.001	0.748	1.721
$n = 300$						
pMOM $r = 1$	0.995	0.000	0.999	0.999	0.000	1.507
pMOM $r = 2$	0.984	0.000	0.997	0.987	0.000	1.567
piMOM	0.990	0.000	1.000	0.997	0.000	1.531
IBF	0.005	0.029	1.000	0.732	0.057	1.521
BIC	0.016	0.802	0.982	0.086	0.876	201.630
SCAD				0.249	0.315	1.550
LASSO				0.001	0.759	1.669
$n = 500$						
pMOM $r = 1$	0.999	0.000	1.000	1.000	0.000	1.544
pMOM $r = 2$	1.000	0.000	1.000	1.000	0.000	1.522
piMOM	0.998	0.000	1.000	1.000	0.000	1.532
IBF	0.003	0.020	1.000	0.692	0.060	1.557
BIC	0.009	0.844	0.988	0.068	0.896	310.988
SCAD				0.560	0.147	1.517
LASSO				0.001	0.777	1.599

Table 4: $\sigma^2 = 1.5,\ \rho = 0,\ t = 5.\ \boldsymbol{\beta}'_{\mathbf{t}} = c(0.6, 1.2, 1.8, 2.4, 3.0)$

	$p(\mathbf{t}\|\mathbf{y}_n)$	pp_0	pp_1	$\mathbf{P}(\hat{\mathbf{t}}=\mathbf{t})$	FDR	MSE
$n = 100$						
pMOM $r = 1$	0.474	0.000	0.897	0.475	0.001	1.748
pMOM $r = 2$	0.315	0.000	0.865	0.291	0.000	1.936
piMOM	0.539	0.001	0.917	0.575	0.003	1.705
IBF	0.041	0.043	0.969	0.686	0.045	1.696
BIC	0.174	0.031	0.948	0.712	0.015	3.196
SCAD				0.170	0.268	1.679
LASSO				0.000	0.713	1.911
$n = 200$						
pMOM $r = 1$	0.899	0.000	0.981	0.917	0.000	1.548
pMOM $r = 2$	0.875	0.000	0.976	0.896	0.000	1.627
piMOM	0.929	0.000	0.990	0.966	0.000	1.563
IBF	0.031	0.025	0.998	0.784	0.043	1.549
BIC	0.168	0.233	0.987	0.834	0.099	31.000
SCAD				0.468	0.136	1.582
LASSO				0.000	0.764	1.704
$n = 300$						
pMOM $r = 1$	0.988	0.000	0.998	0.992	0.000	1.501
pMOM $r = 2$	0.986	0.000	0.998	0.992	0.000	1.530
piMOM	0.989	0.000	0.999	0.996	0.000	1.540
IBF	0.024	0.018	1.000	0.778	0.045	1.531
BIC	0.044	0.612	0.979	0.262	0.614	94.263
SCAD				0.810	0.038	1.551
LASSO				0.000	0.787	1.657
$n = 500$						
pMOM $r = 1$	0.999	0.000	1.000	1.000	0.000	1.504
pMOM $r = 2$	0.999	0.000	1.000	1.000	0.000	1.511
piMOM	0.996	0.000	1.000	1.000	0.000	1.518
IBF	0.018	0.012	1.000	0.784	0.043	1.533
BIC	0.011	0.824	0.980	0.064	0.899	281.106
SCAD				0.990	0.002	1.542
LASSO				0.000	0.814	1.616

Table 5: $\sigma^2 = 1.5$, $\rho = 0.25$, $t = 5$. $\boldsymbol{\beta}'_{\mathbf{t}} = c(0.6, 1.2, 1.8, 2.4, 3.0)$

	$p(\mathbf{t}\|\mathbf{y}_n)$	pp_0	pp_1	$\mathbf{P}(\hat{\mathbf{t}}=\mathbf{t})$	FDR	MSE
			$n=100$			
pMOM $r=1$	0.335	0.000	0.868	0.313	0.000	2.339
pMOM $r=2$	0.134	0.000	0.820	0.101	0.000	2.635
piMOM	0.462	0.001	0.900	0.468	0.002	2.346
IBF	0.006	0.080	0.975	0.580	0.079	2.212
BIC	0.098	0.232	0.948	0.562	0.149	39.303
SCAD				0.103	0.414	2.226
LASSO				0.001	0.712	2.555
			$n=200$			
pMOM $r=1$	0.800	0.000	0.961	0.808	0.000	2.095
pMOM $r=2$	0.643	0.000	0.929	0.643	0.000	2.237
piMOM	0.871	0.000	0.978	0.903	0.001	2.063
IBF	0.003	0.051	0.999	0.678	0.070	2.128
BIC	0.023	0.729	0.974	0.159	0.766	165.308
SCAD				0.111	0.457	2.080
LASSO				0.001	0.749	2.339
			$n=300$			
pMOM $r=1$	0.976	0.000	0.996	0.982	0.000	2.054
pMOM $r=2$	0.928	0.000	0.986	0.935	0.000	2.091
piMOM	0.985	0.000	0.998	0.994	0.000	2.020
IBF	0.002	0.038	1.000	0.648	0.073	2.047
BIC	0.011	0.848	0.981	0.062	0.911	309.592
SCAD				0.136	0.427	2.079
LASSO				0.001	0.763	2.274
			$n=500$			
pMOM $r=1$	0.999	0.000	1.000	1.000	0.000	2.044
pMOM $r=2$	0.999	0.000	1.000	0.999	0.000	2.036
piMOM	0.997	0.000	1.000	1.000	0.000	2.000
IBF	0.001	0.027	1.000	0.654	0.073	2.071
BIC	0.008	0.882	0.985	0.049	0.926	521.929
SCAD				0.332	0.274	2.046
LASSO				0.000	0.779	2.129

Table 6: $\sigma^2=2$, $\rho=0$, $t=5$. $\boldsymbol{\beta}'_{\mathbf{t}}=c(0.6,1.2,1.8,2.4,3.0)$

$n = 100$						
	$p(\mathbf{t}\|\mathbf{y}_n)$	pp_0	pp_1	$\mathbf{P}(\hat{\mathbf{t}} = \mathbf{t})$	FDR	MSE
pMOM $r = 1$	0.326	0.000	0.866	0.313	0.002	2.325
pMOM $r = 2$	0.246	0.001	0.850	0.228	0.001	2.539
piMOM	0.383	0.001	0.882	0.390	0.003	2.311
IBF	0.022	0.053	0.953	0.596	0.056	2.239
BIC	0.128	0.056	0.920	0.546	0.029	9.828
SCAD				0.111	0.316	2.219
LASSO				0.000	0.713	2.509
$n = 200$						
pMOM $r = 1$	0.723	0.000	0.946	0.735	0.001	2.146
pMOM $r = 2$	0.708	0.000	0.943	0.713	0.000	2.177
piMOM	0.780	0.000	0.960	0.795	0.001	2.190
IBF	0.015	0.032	0.994	0.732	0.050	2.127
BIC	0.093	0.418	0.969	0.575	0.289	60.616
SCAD				0.214	0.254	2.118
LASSO				0.000	0.764	2.281
$n = 300$						
pMOM $r = 1$	0.945	0.000	0.990	0.956	0.000	2.084
pMOM $r = 2$	0.942	0.000	0.989	0.951	0.000	2.064
piMOM	0.954	0.000	0.993	0.967	0.001	2.004
IBF	0.013	0.023	1.000	0.756	0.049	2.120
BIC	0.021	0.752	0.973	0.112	0.829	175.327
SCAD				0.492	0.127	2.084
LASSO				0.000	0.789	2.147
$n = 500$						
pMOM $r = 1$	0.998	0.000	1.000	1.000	0.000	2.002
pMOM $r = 2$	0.999	0.000	1.000	1.000	0.000	2.003
piMOM	0.997	0.000	1.000	1.000	0.000	1.999
IBF	0.009	0.015	1.000	0.742	0.056	2.068
BIC	0.010	0.848	0.981	0.059	0.910	418.415
SCAD				0.917	0.016	2.044
LASSO				0.000	0.813	2.117

Table 7: $\sigma^2 = 2$, $\rho = 0.25$, $t = 5$. $\boldsymbol{\beta}'_{\mathbf{t}} = c(0.6, 1.2, 1.8, 2.4, 3.0)$

R code for Section 4 simulation results

The results reported in Section 4 were obtained with the R package `mombf`. This package includes the full MCMC scheme described in Section 3 for pMOM and piMOM priors, approximate and Monte Carlo evaluations of marginal likelihoods for pMOM and piMOM priors, functions that evaluate non-local prior densities, and the ability to specify several choices for the prior on the model space. Examples can be accessed in the manual by typing `vignette("mombf")` in the R command line (http://cran.r-project.org/web/packages/mombf/index.html).

The R code used to produce the simulation results follows.

```
#SIMULATION SETTINGS
# n: sample size
# p: number of predictors
# p0: number of predictors with non-zero coefficients
# phi: residual variance
# rho: correlation  between covariates
# theta: coefficient values
n <- 100; p <- 100; p0 <- 5; phi <- 1; rho <- 0; theta <- c(rep(0,p-p0),
seq(0,3,length=p0+1)[-1]); nsim <- 1000

#Set number of processors, used to run jobs in parallel with multicore package.
# If multicore package is not available for your system, the code still works.
#Simply replace the 'mclapply' calls for regular 'lapply' (delete the
# arguments mc.cores=mc.cores,mc.preschedule=FALSE)
mc.cores <- 12

#Load packages
library(mombf)
library(multicore)
library(ncvreg)
library(lars)
```

```
library(parcor)
library(BMS)

##########################################
#1. DEFINE FUNCTIONS
##########################################
## Intrinsic BF

bfIntrin <- function(y, x, logscale=TRUE, includeIntercept=TRUE) {
#Intrinsic BF comparing the alternative model lm(y~x) with null model
#lm(y~1) (large values favor the alternative)
#Note: for includeIntercept==FALSE models lm(y~-1+x) vs lm(y~-1) are compared
  if (is.matrix(y)) y <- as.vector(y)
  if (is.vector(x)) x <- matrix(x,ncol=1)
  if ((includeIntercept & (ncol(x)>1)) | (!includeIntercept & (ncol(x)>0))) {
    n <- nrow(x)
    if (includeIntercept) {
      lm0 <- lm(y~1); lm1 <- lm(y~x); k0 <- 1; k <- ncol(x)+1
    } else {
      lm0 <- lm(y~ -1); lm1 <- lm(y~ -1+x); k0 <- 0; k <- ncol(x)
    }
    b <- sum(residuals(lm1)^2)/sum(residuals(lm0)^2)
    f <- function(psi,ct=0,logscale=FALSE) {
      sinsq <- (k+1)*sin(psi)^2
      num <- k0*log(sin(psi)) + (.5*(n-k))*log(n+sinsq)
      den <- (.5*(n-k+k0))*log(n*b+sinsq)
      ans <- num-den-ct; ans[psi==0] <- -ct
      if (!logscale) ans <- exp(ans)
      return(ans)
    }
    ct <- max(f(seq(0,.5*pi,length=1000),ct=0,logscale=TRUE))
    ans <- log(integrate(f, lower=0, upper=.5*pi, ct=ct)$value) +
    (.5*k0)*log(k+1) +log(2/pi) + ct
  } else {
    ans <- 0
  }
  if (!logscale) ans <- exp(ans)
  return(ans)
}

###Priors on model space
```

```
unifPrior <- function(sel, logscale=TRUE) { ifelse(logscale,-length(sel)*log(2),
    2^(-length(sel)))  }
bbPrior <- function(sel, logscale=TRUE) {
  ans <- lbeta(sum(sel) + 1, sum(!sel) + 1) - lbeta(1,1)
  ifelse(logscale,ans,exp(ans))
}

### Simulate data and perform model selection based on non-local priors
simNonLocal <- function(simid,theta,n,rho=0,niter=10^3,tauseq,priorCoef,priorDelta,
priorVar,phi,alpha=.001,lambda=.001,phiknown=TRUE) {
  if (missing(priorVar)) priorVar <- new("msPriorSpec",priorType='nuisancePars',
  priorDistr='invgamma',
  priorPars=c(alpha=alpha,lambda=lambda))
  p <- length(theta)
  S <- diag(p); S[upper.tri(S)] <- rho; S[lower.tri(S)] <- rho
  x <- rmvnorm(n,sigma=S)
  y <- x %*% matrix(theta,ncol=1) + rnorm(n,0,sd=sqrt(phi))
  #
  selection <- matrix(NA,nrow=p,ncol=length(tauseq))
  postprob <- matrix(NA,nrow=p+2,ncol=length(tauseq))
  colnames(selection) <- paste(priorCoef@priorDistr,tauseq,sep='_')
  rownames(selection) <- paste('theta',1:p,sep='')
  colnames(postprob) <- colnames(selection)[1:ncol(postprob)]
  rownames(postprob) <- c(paste('margpp',1:p,sep='_'),'propCorrectVars',
  'propTrueModel')
  #
  if (missing(priorDelta)) priorDelta <- new("msPriorSpec",
  priorType='modelIndicator',priorDistr='uniform',
  priorPars=double(0))
  deltaini <- rep(FALSE,p)
  for (i in 1:length(tauseq)) {
    priorCoef@priorPars['tau'] <- tauseq[i]
    if (phiknown) {
      ms <- modelSelection(y=y, x=x, center=FALSE, scale=FALSE, niter=niter,
      priorCoef=priorCoef, priorDelta=priorDelta, priorVar=priorVar, phi=phi,
      method='Laplace', deltaini=deltaini, initSearch='greedy', verbose=FALSE)
    } else {
      ms <- modelSelection(y=y, x=x, center=FALSE, scale=FALSE, niter=niter,
      priorCoef=priorCoef, priorDelta=priorDelta, priorVar=priorVar,
      method='Laplace', deltaini=deltaini, initSearch='greedy',
      verbose=FALSE)
```

```
    }
    pCorrect <- colMeans(t(ms$postSample)==(theta>0))
    postprob[,i] <- c(ms$margpp,mean(pCorrect),mean(pCorrect==1))
    selection[,i] <- ms$postMode
    deltaini <- as.logical(ms$postMode)
  }
  ans <- list(theta=theta,coef=ms$coef,selection=selection,postprob=postprob)
  cat('.')
  return(ans)
}

### Simulate data and perform model selection based on non-local priors
simLocal <- function(simid,theta,n,rho=0,niter=10^3,prior='Zellner',
tauseq,phi,alpha=.001,lambda=.001,mprior) {
  if (mprior=='uniform') priorFunction <-
  unifPrior else if (mprior=='betabin') priorFunction <-
  bbPrior else stop("Incorrect mprior")
  #
  p <- length(theta)
  S <- diag(p); S[upper.tri(S)] <- rho; S[lower.tri(S)] <- rho
  x <- rmvnorm(n,sigma=S)
  y <- x %*% matrix(theta,ncol=1) + rnorm(n,0,sd=sqrt(phi))
  #
  deltaini <- rep(FALSE,p)
  if (prior=='Intrinsic') {
    selection <- matrix(NA,nrow=p,ncol=1); coef <- matrix(0,nrow=p,ncol=1)
    postprob <- matrix(NA,nrow=p+2,ncol=1)
    colnames(selection) <- colnames(coef) <- colnames(postprob) <- 'Intrinsic'
    rownames(selection) <- rownames(coef) <- paste('theta',1:p,sep='')
    rownames(postprob) <- c(paste('margpp',1:p,sep='_'),
    'propCorrectVars','propTrueModel')
    ms <- mombf:::modelSelectionR(y=y,x=x,niter=niter,marginalFunction=bfIntrin,
    priorFunction=priorFunction,includeIntercept=FALSE,
    deltaini=deltaini,verbose=FALSE)
    pCorrect <- colMeans(t(ms$postSample)==(theta>0))
    postprob[,1] <- c(ms$margpp,mean(pCorrect),mean(pCorrect==1))
    selection[,1] <- ms$postMode
    coef[selection[,1],1] <- coef(lm(y ~ -1 + x[,selection[,1]]))
  } else if (prior=='BIC') {
    if (mprior=='betabin') mprior <- 'random'
    selection <- matrix(NA,nrow=p,ncol=1); coef <- matrix(0,nrow=p,ncol=1)
```

```
    postprob <- matrix(NA,nrow=p+2,ncol=1)
    colnames(selection) <- colnames(coef) <- 'BIC'
    rownames(selection) <- rownames(coef) <- paste('theta',1:p,sep='')
    colnames(postprob) <- colnames(selection)[1:ncol(postprob)]
    rownames(postprob) <- c(paste('margpp',1:p,sep='_'),
    'propCorrectVars','propTrueModel')
    ms <- try(bms(data.frame(y,x), burn=.1*niter, iter=niter,
    mcmc='bd', g='UIP', mprior=mprior, user.int=FALSE, nmodel=1000))
    if (class(ms)!='try-error') {
      topmod <- topmodels.bma(ms)
      selection[,1] <- as.logical(topmod[1:ncol(x),1]==1)
      coef[selection[,1],1] <- coef(lm(y ~ -1 + x[,selection[,1]]))
      whichcorrect <- which(colSums((topmod[1:ncol(x),]!=0)==(theta !=0))==ncol(x))
      pCorrect <- ifelse(length(whichcorrect)==0,0,topmod['PMP (MCMC)',
      whichcorrect])
      margpp <- coef(ms)[paste('X',1:ncol(x),sep=''),'PIP']
      postprob[,1] <- c(margpp,NA,pCorrect)
    }
  }
  ans <- list(theta=theta,coef=coef,selection=selection,
  postprob=postprob)
  cat('.')
  return(ans)
}


### Simulate data and perform model selection via SCAD/LASSO
simFreq <- function(simid,theta,n,rho=0,phi=1,method='SCAD') {
  require(mvtnorm)
  p <- length(theta)
  S <- diag(p); S[upper.tri(S)] <- rho; S[lower.tri(S)] <- rho
  x <- rmvnorm(n,sigma=S)
  y <- x %*% matrix(theta,ncol=1) + rnorm(n,0,sd=sqrt(phi))
  #
  selection <- coef <- matrix(NA,nrow=p,ncol=1)
  colnames(selection) <- colnames(coef) <- method
  rownames(selection) <- rownames(coef) <- paste('theta',1:p,sep='')
  if (method=='SCAD') {
    cvscad <- cv.ncvreg(X=x,y=y,family="gaussian",penalty="SCAD",nfolds=10,
    dfmax=1000,max.iter=10^4)
    coef[,1] <- ncvreg(X=x,y=y,penalty='SCAD',dfmax=1000,
```

```
    lambda=rep(cvscad$lambda[cvscad$cv],2))$beta[-1,1]
    selection[,1] <- coef[,1]!=0
  } else if (method=='LASSO') {
    lars1 <- mylars(X=x,y=y,k=10,use.Gram=TRUE,normalize=TRUE)
    coef[,1] <- lars1$coef
    selection[,1] <- lars1$coef!=0
  }
  ans <- list(theta=theta,coef=coef,selection=selection)
  cat('.')
  return(ans)
}


################################
# 2. PERFORM SIMULATIONS
################################
#Set random seed
d <- date(); d <- substr(d,nchar(d)-12,nchar(d)-5)
set.seed(prod(as.numeric(strsplit(d,':')[[1]])))

priorDelta <- new("msPriorSpec",priorType='modelIndicator',
priorDistr='binomial',priorPars=c(alpha.p=1,beta.p=1))

priorCoef <- new("msPriorSpec",priorType='coefficients',
priorDistr='pMOM',priorPars=c(tau=1,r=1))
simmom1 <- mclapply(1:nsim,simNonLocal,theta=theta,
n=n,rho=rho,niter=5000,tauseq=.348,priorCoef=priorCoef,
priorDelta=priorDelta,phi=phi,
phiknown=FALSE,mc.cores=mc.cores,mc.preschedule=FALSE)

priorCoef <- new("msPriorSpec",priorType='coefficients',
priorDistr='pMOM',priorPars=c(tau=1,r=2))
simmom2 <- mclapply(1:nsim,simNonLocal,theta=theta,n=n,rho=rho,
niter=5000,tauseq=.07216,priorCoef=priorCoef,priorDelta=priorDelta,
phi=phi,phiknown=FALSE,
mc.cores=mc.cores,mc.preschedule=FALSE)

priorCoef <- new("msPriorSpec",priorType='coefficients',
priorDistr='piMOM',priorPars=c(tau=1))
simimom <- mclapply(1:nsim,simNonLocal,theta=theta,
n=n,rho=rho,niter=5000,tauseq=.133,priorCoef=priorCoef,
```

```
priorDelta=priorDelta,phi=phi,
phiknown=FALSE,mc.cores=mc.cores,mc.preschedule=FALSE)

simIntrin <- mclapply(1:nsim,simLocal,theta=theta,n=n,rho=rho,niter=5000,
prior='Intrinsic',phi=phi,mprior='betabin',mc.cores=mc.cores,mc.preschedule=FALSE)
simBIC <- mclapply(1:nsim,simLocal,theta=theta,n=n,rho=rho,niter=5000,
prior='BIC',phi=phi,mprior='betabin',mc.cores=mc.cores,mc.preschedule=FALSE)
simSCAD <- mclapply(1:nsim,simFreq,theta=theta,n=n,rho=rho,phi=phi,
method='SCAD',mc.cores=mc.cores,mc.preschedule=FALSE)
simLASSO <- mclapply(1:nsim,simFreq,theta=theta,n=n,rho=rho,phi=phi,
method='LASSO',mc.cores=mc.cores,mc.preschedule=FALSE)


#INTERPRETING THE OUTPUT
#
# Results are stored as lists, one element for each simulation,
# e.g. simmom1[[i]] contains the MOM fit (r=1) for the ith
#simulated dataset
# - simmom1[[i]]$theta: true coefficient values (simulation truth)
# - simmom1[[i]]$coef: estimated coefficients (posterior mode)
# - simmom1[[i]]$selection: matrix indicating which variables are
# included in the model at each MCMC iteration
# - simmom1[[i]]$postprob: elements 1:p contain the marginal
# posterior probabilities of inclusion for each covariate
# (via Rao-Blackwellization, i.e. more precise than
# colMeans(simmom1[[i]])$selection. Element p+1
# (propCorrectVars) indicates the average proportion of
# correctly included/excluded variables. Element p+2
# (propTrueModel) the proportion of MCMC visits to the true model.
#
# Output for simmom2 (MOM with r=2) and simimom
# (iMOM) are identical.
# simIntrin and simBIC are identical except that
# only the posterior mode is returned in the slot 'selection',
# and the coefficients are estimated via least-squares
# for the selected model.
# simSCAD and simLASSO return only the selected
# model in the 'selection' slot, and NAs for 'postprob'.
```

References

[1] Kan, R. (2008), "From Moments of Sum to Moments of Product," *Journal of Multivariate Analysis,* 99 542–554.

[2] Lai, T.L., Robbins, H. and Wei, C.Z. (1978), "Strong Consistency of Least Squares Estimates in Multiple Regression," *Proceedings of the National Academy of Science,* 75, 3034–3036.

[3] Natalini, P. and Palumbo, B. (2000), "Inequalities for the Incomplete Gamma Function," *Mathematical Inequalities and Applications,* 3, 69–77.

[4] Schott, J.R. (1997), *Matrix Analysis for Statistics*, Wiley, New York.

[5] Sutradhar, B. (1986), "On the Characteristic Function of the Multivariate Student T-distribution," *Canadian Journal of Statistics,* 14, 329-337.