SUPPLEMENT TO "NON-LOCAL PRIORS FOR HIGH-DIMENSIONAL ESTIMATION"

1. Walker's conditions

Conditions A1-A5 and B1-B4 are adapted from Walker (1969) and Johnson and Rossell (2010), retaining the original numbering system to facilitate comparison. In what follows $\mathbf{y}$ denotes the collection of random elements of interest, $f(\mathbf{y} \mid \boldsymbol{\theta})$ with $\boldsymbol{\theta} \in \Theta$ their Radon-Nikodym density with respect to a dominating measure μ.

A1: Θ is a closed set of points in $\mathbb{R}^s$, where s is finite.
A2: The sample space $\mathcal{Y} = \{\mathbf{y} : f(\mathbf{y} \mid \boldsymbol{\theta}) > 0\}$ is independent of $\boldsymbol{\theta}$.
A3: If $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ are distinct points of Θ,

$$\mu\{\mathbf{y} : f(\mathbf{y} \mid \boldsymbol{\theta}_1) \neq f(\mathbf{y} \mid \boldsymbol{\theta}_2)\} > 0.$$

A4: Let $\mathbf{y} \in \mathcal{Y}$, $\boldsymbol{\theta}' \in \Theta$. Then for all $\boldsymbol{\theta}$ such that $|\boldsymbol{\theta} - \boldsymbol{\theta}'| < \delta$, with δ sufficiently small,

$$|\log f(\mathbf{y} \mid \boldsymbol{\theta}) - \log f(\mathbf{y} \mid \boldsymbol{\theta}')| < H_\delta(\mathbf{y}, \boldsymbol{\theta}'),$$

where

$$\lim_{\delta \to 0} H_\delta(\mathbf{y}, \boldsymbol{\theta}') = 0,$$

and, for any $\boldsymbol{\theta}_0 \in \Theta$,

$$\lim_{\delta \to 0} \int_{\mathcal{Y}} H_\delta(\mathbf{y}, \boldsymbol{\theta}') f(\mathbf{y} \mid \boldsymbol{\theta}_0) d\mu = 0.$$

A5: If Θ is not bounded, then for any $\boldsymbol{\theta}_0 \in \Theta$, and sufficiently large Δ,

$$\log f(\mathbf{y} \mid \boldsymbol{\theta}) - \log f(\mathbf{y} \mid \boldsymbol{\theta}_0) < K_\Delta(\mathbf{y}, \boldsymbol{\theta}_0)$$

whenever $|\boldsymbol{\theta}| > \Delta$, where

$$\lim_{\delta \to 0} \int_{\mathcal{Y}} K_\Delta(\mathbf{y}, \boldsymbol{\theta}_0) f(\mathbf{y} \mid \boldsymbol{\theta}_0) d\mu < 0.$$

For the remaining conditions, let $\boldsymbol{\theta}_0$ be an interior point of Θ.
B1: $\log f(\mathbf{y} \mid \boldsymbol{\theta})$ is twice differentiable with respect to $\boldsymbol{\theta}$ in some neighborhood of $\boldsymbol{\theta}_0$.
B2: The matrix $\mathbf{J}(\boldsymbol{\theta}_0)$ with elements

$$J_{ij}(\boldsymbol{\theta}_0) = \int_{\mathcal{Y}} f_0 \left(\frac{\partial \log f_0}{\partial \boldsymbol{\theta}_{0,i}}\right) \left(\frac{\partial \log f_0}{\partial \boldsymbol{\theta}_{0,j}}\right) d\mu,$$

where f_0 denotes $f(\mathbf{y} \mid \boldsymbol{\theta}_0)$, is finite and positive definite. In the scalar case, this condition becomes $0 < J(\boldsymbol{\theta}_0) < \infty$, where

$$J(\boldsymbol{\theta}_0) = \int_{\mathcal{Y}} f_0 \left(\frac{\partial \log f_0}{\partial \boldsymbol{\theta}_0}\right)^2 d\mu.$$

B3:

$$\int_{\mathcal{Y}} \frac{\partial f_{0,i}}{\partial \boldsymbol{\theta}_{0,i}} d\mu = \int_{\mathcal{Y}} \frac{\partial^2 f_0}{\partial \boldsymbol{\theta}_{0,i} \partial \boldsymbol{\theta}_{0,j}} d\mu = 0$$

B4: If $|\boldsymbol{\theta} - \boldsymbol{\theta}_0| < \delta$, where δ is sufficiently small, then

$$\left| \frac{\partial^2 \log f(\mathbf{y} \mid \boldsymbol{\theta})}{\partial \boldsymbol{\theta}_i \partial \boldsymbol{\theta}_j} - \frac{\partial^2 \log f(\mathbf{y} \mid \boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_{0,i} \partial \boldsymbol{\theta}_{0,j}} \right| < M_\delta(\mathbf{y}, \boldsymbol{\theta}_0),$$

where

$$\lim_{\delta \to 0} \int_{\mathcal{Y}} M_\delta(\mathbf{y}, \boldsymbol{\theta}_0) f(\mathbf{y} \mid \boldsymbol{\theta}_0) d\mu = 0.$$

2. Proofs

2.1. **Proof of Proposition 1, Parts (i)-(ii).** We start by stating two useful lemmas. Lemma 1 states that the pMOM, peMOM and piMOM priors under a given M_k can be written as the product of a local prior times a bounded constant c_k, which will be useful later on to exchange limits and integration (*e.g.* via the Dominated Convergence Theorem). Lemma 2 establishes that $E^L(d(\boldsymbol{\theta}_k, \phi_k) \mid \mathbf{y}_n)$, which is the extra complexity penalty induced by NLPs, converges either to 0 or to a positive constant when M_k adds spurious parameters to a true model M_t or is missing parameters from M_t (respectively).

Lemma 1. *Let $\pi(\boldsymbol{\theta}_k, \phi_k) = d(\boldsymbol{\theta}_k, \phi_k)\pi^L(\boldsymbol{\theta}_k, \phi_k)$ be either the pMOM, peMOM or piMOM prior, where $d(\boldsymbol{\theta}_k, \phi_k) \to 0$ as $\theta_{ki} \to 0$ for any $i = 1, \ldots, dim(\boldsymbol{\theta}_k)$ and $\pi^L(\boldsymbol{\theta}_k, \phi_k)$ is a local prior. Then $\pi(\boldsymbol{\theta}_k, \phi_k) = \tilde{d}(\boldsymbol{\theta}_k, \phi_k)\tilde{\pi}^L(\boldsymbol{\theta}_k, \phi_k)$, where $\tilde{d}(\boldsymbol{\theta}_k, \phi_k) \leq c_k$ for some constant c_k and $\tilde{\pi}^L(\boldsymbol{\theta}_k, \phi_k)$ is a local prior.*

Proof. The result for the peMOM is direct with $\tilde{d}_k(\theta_{ki}, \phi_k) = \prod_{i=1}^{p_k} e^{\sqrt{2}} e^{-\tau\phi/\theta_{ki}^2} \leq e^{\sqrt{2}p_k}$ and $\tilde{\pi}^L(\boldsymbol{\theta}, \phi) = N(\boldsymbol{\theta}; \mathbf{0}, \tau\phi I)\pi(\phi)$. For the piMOM prior we multiply and divide the density by a Cauchy kernel, obtaining

$$\begin{aligned} \pi_k^I(\theta_{ki} \mid \phi_k) &= \frac{\sqrt{\tau\phi_k}}{\sqrt{\pi}\theta_{ki}^2} e^{-\tau\phi_k/\theta_{ki}^2} \pi \left(1 + \frac{\theta_{ki}^2}{\tau\phi_k}\right) \text{Cauchy}(\theta_{ki}; 0, \phi_k\tau) \\ &= \tilde{d}_k(\theta_{ki}, \phi_k) \text{Cauchy}(\theta_{ki}; 0, \phi_k\tau), \end{aligned} \tag{1}$$

where $\tilde{d}_k(\theta_{ki}, \phi_k) = \sqrt{\pi} \frac{\sqrt{\tau\phi_k}}{\theta_{ki}^2} e^{-\tau\phi_k/\theta_{ki}^2} \left(1 + \theta_{ki}^2/(\tau\phi_k)\right)$. By performing a change of variables $\eta_i = \theta_{ki}/\sqrt{\tau\phi_k}$ we obtain the implied prior $\pi_k^I(\eta_i \mid \phi_k) = \sqrt{\pi}\frac{1+\eta_i^2}{\eta_i^2} e^{-1/\eta_i^2} \pi^L(\eta_i \mid \phi_k)$. Now, $h(\eta_i) = \sqrt{\pi}\frac{1+\eta_i^2}{\eta_i^2} e^{-1/\eta_i^2}$ is continuous, has positive derivative for all $\eta_i > 0$ and negative for $\eta_i < 0$, $\lim_{\eta_i \to 0} h(\eta_i) = 0$ and $\lim_{\eta_i \to \pm\infty} h(\eta_i) = \sqrt{\pi}$, and hence $h(\eta_i) \leq \sqrt{\pi}$. In summary, $c_k = e^{\sqrt{2}p_k}$ for the product eMOM and $c = \pi^{p_k/2}$ for the product iMOM, where $p_k = \dim(\boldsymbol{\theta}_k)$.

The pMOM prior density has an unbounded term $\prod_{i\in M_k}\frac{\theta_{ki}^{2r}}{(2r-1)!!\phi^r\tau^r}$, but it can be rewritten as $\pi_k^M(\boldsymbol{\theta}_k \mid \phi_k) =$

$$\prod_{i\in M_k}\frac{\theta_{ki}^{2r}}{(2r-1)!!\phi_k^r\tau^r}\frac{N(\theta_{ki};0,\tau\phi_k I)}{N(\theta_{ki};0,(1+\epsilon)\tau\phi_k I)}N(\theta_{ki};0,(1+\epsilon)\tau\phi_k I) =$$

$$\prod_{i\in M_k}\frac{\theta_{ki}^{2r}}{(2r-1)!!\phi_k^r\tau^r}\exp\left\{-\frac{1}{2}\frac{\theta_{ki}^2}{\phi_k\tau(1+\epsilon^{-1})}\right\}N(\theta_{ki};0,(1+\epsilon)\tau\phi_k I) =$$

$$= \prod_{i\in M_k}\tilde{d}(\theta_{ki},\phi_k)N(\theta_{ki};0,(1+\epsilon)\tau\phi_k I) \tag{2}$$

for some $\epsilon \in (0,1)$, where it is straightforward to see that $\tilde{d}(\theta_{ki},\phi_k)$ is now bounded. $\square$

Lemma 2. *Let $d(\boldsymbol{\theta})$ be a continuous and differentiable function satisfying $0 \le d(\boldsymbol{\theta}) < c$ for all $\boldsymbol{\theta} \in \Theta$. Define*

$$g(\mathbf{y}_n) = \int d(\boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathbf{y}_n)d\boldsymbol{\theta},$$

where $\lim_{n\to\infty}\int_{\boldsymbol{\theta}\in N_\epsilon(A)}\pi(\boldsymbol{\theta} \mid \mathbf{y}_n) = 1$ almost surely for any fixed $\epsilon > 0$, some set A and a corresponding suitably defined ϵ-neighborhood $N_\epsilon(A)$. If $d(\boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in A$ then $g(\mathbf{y}_n) \longrightarrow 0$. Likewise, if $d(\boldsymbol{\theta}) > c'$ for all $\boldsymbol{\theta} \in A$ and some $c' > 0$ then $P(g(\mathbf{y}_n) \ge c') \longrightarrow 1$ almost surely as $n \longrightarrow \infty$. In particular, if $A = \{\boldsymbol{\theta}_0\}$ is a singleton, then $g(\mathbf{y}_n) \longrightarrow g(\boldsymbol{\theta}_0)$.

Proof. Consider

$$\begin{aligned} g(\mathbf{y}_n) &= \int_{\boldsymbol{\theta}\in N_\epsilon(A)} d(\boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathbf{y}_n)d\boldsymbol{\theta} + \int_{\boldsymbol{\theta}\notin N_\epsilon(A)} d(\boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathbf{y}_n)d\boldsymbol{\theta} = \\ &\le \delta_\epsilon P(\boldsymbol{\theta} \in N_\epsilon(A)) + cP(\boldsymbol{\theta} \notin N_\epsilon(A)) \le \delta_\epsilon + cP(\boldsymbol{\theta} \notin N_\epsilon(A)), \end{aligned} \tag{3}$$

where $\delta_\epsilon = \max_{\boldsymbol{\theta}\in N_\epsilon(A)} d(\boldsymbol{\theta})$ and the second term can be made arbitrarily small. Because $d(\boldsymbol{\theta})$ is continuous, if $d(\boldsymbol{\theta}) = 0$ for all $\boldsymbol{\theta} \in A$ then δ_ϵ can also be made arbitrarily small a.s. as $n \longrightarrow \infty$, and hence $g(\mathbf{y}_n) \longrightarrow 0$. Suppose now that $d(\boldsymbol{\theta}) > c'$ for all $\boldsymbol{\theta} \in A$, then from (3)

$$g(\mathbf{y}_n) > \int_{\boldsymbol{\theta}\in N_\epsilon(A)} d(\boldsymbol{\theta})\pi(\boldsymbol{\theta} \mid \mathbf{y}_n)d\boldsymbol{\theta} \ge \delta'_\epsilon \int_{\boldsymbol{\theta}\in N_\epsilon(A)} \pi(\boldsymbol{\theta} \mid \mathbf{y}_n)d\boldsymbol{\theta}, \tag{4}$$

where due to continuity $\delta'_\epsilon = \min_{\boldsymbol{\theta}\in N_\epsilon(A)} d(\boldsymbol{\theta})$ can be made arbitrarily close to c' for small enough ϵ and the integral on the right hand side of (4) can be made arbitrarily close to 1 as $n \longrightarrow \infty$. The proof for when $A = \{\boldsymbol{\theta}_0\}$ follows as an immediate implication. $\square$

Proof of Proposition 1, Part (i) The result follows from direct algebraic manipulation

$$m_k(\mathbf{y}_n) = \int\int f_k(\mathbf{y}_n \mid \boldsymbol{\theta}_k, \phi_k) d_k(\boldsymbol{\theta}_k, \phi_k) \pi^L(\boldsymbol{\theta}_k, \phi_k \mid M_k) d\boldsymbol{\theta}_k d\phi_k =$$
$$\int\int d_k(\boldsymbol{\theta}_k, \phi_k) \frac{f_k(\mathbf{y}_n \mid \boldsymbol{\theta}_k, \phi_k)\pi^L(\boldsymbol{\theta}_k, \phi_k \mid M_k)}{m_k^L(\mathbf{y}_n)} m_k^L(\mathbf{y}_n) d\boldsymbol{\theta}_k d\phi_k =$$
$$m_k^L(\mathbf{y}_n) \int\int d_k(\boldsymbol{\theta}_k, \phi_k) \pi^L(\boldsymbol{\theta}_k, \phi_k \mid \mathbf{y}_n, M_k) d\boldsymbol{\theta}_k, \phi_k = m_k^L(\mathbf{y}_n) g_k(\mathbf{y}_n),$$

as desired. In a slight abuse of notation, in the derivation above $d\boldsymbol{\theta}_k$ and $d\phi_k$ indicate integration with respect to the corresponding σ-finite dominating measures.

Proof of Proposition 1, Part (ii) We use Lemma 1, which states that piMOM and peMOM priors can be written as $d_k(\boldsymbol{\theta}_k, \phi_k)\pi^L(\boldsymbol{\theta}_k, \phi_k)$ with bounded $d_k(\boldsymbol{\theta}_k, \phi_k)$, so that

$$g_k(\mathbf{y}_n) = \int\int d_k(\boldsymbol{\theta}_k, \phi_k) \frac{f_k(\mathbf{y}_n \mid \boldsymbol{\theta}_k, \phi_k)\pi^L(\boldsymbol{\theta}_k, \phi_k)}{m_k^L(\mathbf{y}_n)} d\boldsymbol{\theta}_k d\phi_k \tag{5}$$

where by assumption $f_k(\mathbf{y}_n \mid \boldsymbol{\theta}_k^*, \phi_k^*)/f_k(\mathbf{y}_n \mid \tilde{\boldsymbol{\theta}}_k, \tilde{\phi}_k) \to \infty$ almost surely as $n \to \infty$ for any $(\boldsymbol{\theta}_k^*, \phi_k^*) \in A$ and $(\tilde{\boldsymbol{\theta}}_k, \tilde{\phi}_k) \notin A$. See *e.g.* Redner (1981) for such MLE consistency under general settings. We note that $\pi^L(\boldsymbol{\theta}_k, \phi_k)$ associated to either pMOM, piMOM or peMOM priors are products of independent Normal or Cauchy kernels assigning strictly positive density to any $\boldsymbol{\theta}_k \in \Theta_k$, which combined with MLE consistency guarantee that the limiting posterior concentrates arbitrarily large probability on any ϵ neighborhood of A as $n \to \infty$ (Ghosal, 2002). Part (ii) follows from Lemma 2. For the pMOM prior, from Lemma 1 $g_k(\mathbf{y}_n) =$

$$\int\int \tilde{d}_k(\boldsymbol{\theta}_k, \phi_k) \frac{f_k(\mathbf{y}_n \mid \boldsymbol{\theta}_k, \phi_k) N(\boldsymbol{\theta}_k; \mathbf{0}, \tau\phi_k(1+\epsilon)I)}{m_{k,\tau}^L(\mathbf{y}_n)} d\boldsymbol{\theta}_k d\phi_k =$$
$$\frac{m_{k,\tau(1+\epsilon)}^L(\mathbf{y}_n)}{m_{k,\tau}^L(\mathbf{y}_n)} \int\int \tilde{d}_k(\boldsymbol{\theta}_k, \phi_k) \pi_{\tau(1+\epsilon)}(\boldsymbol{\theta}_k, \phi_k \mid \mathbf{y}_n) d\boldsymbol{\theta}_k d\phi_k, \tag{6}$$

where $\epsilon \in (0,1)$, $\tilde{d}_k(\boldsymbol{\theta}_k, \phi_k)$ is bounded and $m_{k,\tau}(\mathbf{y}_n)$ is the integrated likelihood under a $N(\boldsymbol{\theta}_k; \mathbf{0}, (1+\epsilon)\tau\phi_k I)$ prior. Part (ii) follows from Lemma 2, which guarantees convergence for the integral in (6), and that by assumption $m_{k,\tau(1+\epsilon)}^L(\mathbf{y}_n)/m_{k,\tau}^L(\mathbf{y}_n) \to c \in (0,\infty)$ almost surely as $n \to \infty$. We note that from Bayes theorem

$$\frac{m_{k,\tau(1+\epsilon)}^L(\mathbf{y}_n)}{m_{k,\tau}^L(\mathbf{y}_n)} = \frac{\pi_{(1+\epsilon)\tau}^L(\boldsymbol{\theta}_k, \phi_k \mid \mathbf{y}_n)}{\pi_\tau^L(\boldsymbol{\theta}_k, \phi_k \mid \mathbf{y}_n)} \frac{\pi_\tau^L(\boldsymbol{\theta}_k, \phi_k)}{\pi_{(1+\epsilon)\tau}^L(\boldsymbol{\theta}_k, \phi_k)} \tag{7}$$

for any $(\boldsymbol{\theta}_k, \phi_k)$, where the second term in the right hand side is bounded (*e.g.* for $\boldsymbol{\theta}_k = \mathbf{0}$). The first term is the ratio of posterior densities under $N(\boldsymbol{\theta}; \mathbf{0}, (1+\epsilon)\tau\phi_k I)$ and $N(\boldsymbol{\theta}; \mathbf{0}, \tau\phi_k I)$, which for limiting normal posterior distributions with bounded covariance eigenvalues as in Condition D2 converges in probability to a bounded constant.

Now consider the particular case where the data-generating density $f^*(\mathbf{y}_n)$ belongs to the set of considered models, *i.e.* $f^*(\mathbf{y}_n) = f_t(\mathbf{y}_n \mid \boldsymbol{\theta}_t^*, \phi_t^*)$ for some $t \in \{1, \ldots, K\}$ of smallest $\dim(\boldsymbol{\theta}_t)$ amongst all such models. By definition of NLP if $M_t \subset M_k$ then the values $(\boldsymbol{\theta}_k^*, \phi_k^*)$ minimizing KL to f_t under M_k satisfy $d_k(\boldsymbol{\theta}_k^*, \phi_k^*) = 0$; further if $M_k \subset M_t$ then $(\boldsymbol{\theta}_k^*, \phi_k^*)$ satisfy $d_k(\boldsymbol{\theta}_k^*, \phi_k^*) > 0$. Therefore, $M_t \subset M_k$ implies that $g_k(\mathbf{y}_n) \xrightarrow{P} 0$ and $M_k \subseteq M_t$ implies $P(g_k(\mathbf{y}_n) \geq c) \longrightarrow 1$ for some constant $c > 0$. We note that for non-identifiable models the set A is no longer a singleton, but when $M_t \subset M_k$ by definition $d_k(\boldsymbol{\theta}_k, \phi_k) = 0$ for all $(\boldsymbol{\theta}_k, \phi_k) \in A$, hence we still obtain $g_k(\mathbf{y}_n) \xrightarrow{P} 0$. Also by NLP definition, when $M_k \subset M_t$ then $d_k(\boldsymbol{\theta}_k, \phi_k) > 0$ for all $(\boldsymbol{\theta}_k, \phi_k) \in A$, which implies $g_k(\mathbf{y}_n) \xrightarrow{P} c > 0$.

2.2. **Proof of Proposition 1, Part (iii).** We start by stating two lemmas. Lemma 3 shows that for linear models satisfying D1-D2 and NLPs where $d_k(\boldsymbol{\theta}_k, \phi_k)$ takes a product form the complexity penalty $g_k(\mathbf{y}_n)$ is a continuous function of $(\mathbf{m}_{k,n}, X_{k,n}' X_{k,n}, \hat{\phi}_{k,n})$. This allows us to use the Continuous Mapping Principle, which is useful given that under the proposition's assumptions $\mathbf{m}_{k,n}$ converges to $\hat{\boldsymbol{\theta}}_{k,n}$, and $(\hat{\boldsymbol{\theta}}_{k,n}, \phi_k)$ is strongly consistent to $(\boldsymbol{\theta}_k^*, \phi_k^*)$ minimizing KL to the data-generating truth. Lemma 4 shows that the mean $d_k(\boldsymbol{\theta}_k, \phi_k)$ under a posterior distribution centered on $(\boldsymbol{\theta}_k^*, \phi_k^*)$ converges to the penalty under the limiting posterior (essentially a point mass at $(\boldsymbol{\theta}_k^*, \phi_k^*)$), which is a partial result needed in the proof of Proposition 1, Part (iii).

Lemma 3. *Let $\mathbf{y}_n \sim N(X_{k,n}\boldsymbol{\theta}_k, \phi_k)$ be a linear model satisfying D1-D2, and consider a NLP $\pi(\boldsymbol{\theta}_k \mid \phi_k) = d_k(\boldsymbol{\theta}_k, \phi_k)\pi^L(\boldsymbol{\theta}_k \mid \phi_k)$. Assume that the NLP penalty takes the product form $d_k(\boldsymbol{\theta}_k, \phi_k) = \prod_{i \in M_k} d(\theta_{ki}, \phi_k)$, where $d_k(\theta_{ki}, \phi_k) = \frac{\theta_{ki}^{2r}}{(2r-1)!!(\tau\phi_k)^r}$ is either the MOM penalty or $d_k(\theta_{ki}, \phi_k) \leq c$ for all (θ_{ki}, ϕ_k) and some constant c, as in the eMOM or iMOM penalties. Then $g_k(\mathbf{y}_n)$ is a continuous function of $\mathbf{s}_{k,n} = (\mathbf{m}_{k,n}, X_{k,n}' X_{k,n}, \hat{\phi}_{k,n})$.*

Proof. To prove the result for bounded penalties $d(\theta_{ki}, \phi_k) \leq c$ recall that $\mathbf{s}_{k,n}$ is sufficient under M_k and hence we may write $g_k(\mathbf{y}_n) = g_k(\mathbf{s}_{k,n}) =$

$$\int\int \prod_{i \in M_k} d_k(\boldsymbol{\theta}_{ki}, \phi_k)\pi^L(\boldsymbol{\theta}_k \mid \mathbf{m}_{k,n}, X_{k,n}' X_{k,n}, \phi_k, M_k)$$
$$\pi^L(\phi_k \mid \mathbf{m}_{k,n}, X_{k,n}' X_{k,n}, \hat{\phi}_{k,n}, M_k) d\boldsymbol{\theta}_k d\phi_k \leq$$
$$\int\int c^{p_k} \pi^L(\boldsymbol{\theta}_k \mid \mathbf{m}_{k,n}, X_{k,n}' X_{k,n}, \phi_k, M_k)$$
$$\pi^L(\phi_k \mid \mathbf{m}_{k,n}, X_{k,n}' X_{k,n}, \hat{\phi}_{k,n}, M_k) d\boldsymbol{\theta}_k d\phi_k = c^{p_k}. \quad (8)$$

Now, letting $\mathbf{z} = (\mathbf{z}_1, X'_{k,n}X_{k,n}, z_2) \to (\mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \hat{\phi}_{k,n})$ and using the Dominated Convergence Theorem we obtain

$$\lim_{\mathbf{z}\to\mathbf{s}_{k,n}} g_k(z)c^{-p_k} = \int\int \prod_{i\in M_k} c^{-1}d_k(\boldsymbol{\theta}_{ki},\phi_k)\pi^L(\boldsymbol{\theta}_k \mid \mathbf{z}_1, X'_{k,n}X_{k,n}, \phi_k, M_k)$$
$$\pi^L(\phi_k \mid \mathbf{z}_1, X'_{k,n}X_{k,n}, z_2, M_k)d\boldsymbol{\theta}_k d\phi_k =$$
$$\int\int \prod_{i\in M_k} c^{-1}d_k(\boldsymbol{\theta}_{ki},\phi_k)\pi^L(\boldsymbol{\theta}_k \mid \mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \phi_k, M_k)$$
$$\pi^L(\phi_k \mid \mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \hat{\phi}_{k,n}, M_k)d\boldsymbol{\theta}_k d\phi_k, \tag{9}$$

and hence

$$\lim_{\mathbf{z}\to\mathbf{s}_{k,n}} g_k(z) = \int\int \prod_{i\in M_k} d_k(\boldsymbol{\theta}_{ki},\phi_k)\pi^L(\boldsymbol{\theta}_k \mid \mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \phi_k, M_k)$$
$$\pi^L(\phi_k \mid \mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \hat{\phi}_{k,n}, M_k)d\boldsymbol{\theta}_k d\phi_k, \tag{10}$$

showing that $g_k(\mathbf{s}_{k,n})$ is continuous.

Consider now the MOM prior case. For the particular prior choice $\phi_k \sim \mathrm{IG}(\alpha,\lambda)$, Johnson and Rossell (2012) showed that $g_k(\mathbf{s}_{k,n}) = E\left(\prod_{i\in M_k}\theta_{ki}^{2r}\right)$ where $\boldsymbol{\theta}_k \sim T_\nu(\mathbf{m}_{k,n}, V_{k,n})$, with $\nu = 2rp_k + n + \alpha$ and $V_{k,n} = S_{k,n}\nu/(\lambda + \mathbf{y}'_n\mathbf{y}_n - \mathbf{y}'_n X_{k,n}\mathbf{m}_{k,n}$. Kan (2008) gave explicit expressions for such products as a sum of continuous functions, and hence $g_k(\mathbf{s}_{k,n})$ is continuous. Lemma 1 ensures that the pMOM penalty is also bounded for more general priors $\pi_k(\phi_k)$. Therefore, $g_k(\mathbf{s}_{k,n}) =$

$$\int\int \prod_{i\in M_k} d(\theta_{ki},\phi_k)\frac{N(\mathbf{y}_n; X_{k,n}\boldsymbol{\theta}_k; \phi_k I)N(\boldsymbol{\theta}_k; \mathbf{0}, 2\tau\phi_k I)}{m^L_{k,\tau}(\mathbf{y}_n)}\pi_k(\phi_k)d\boldsymbol{\theta}_k d\phi_k =$$
$$\frac{m^L_{k,2\tau}(\mathbf{y}_n)}{m^L_{k,\tau}(\mathbf{y}_n)}\int\int \prod_{i\in M_k} d(\theta_{ki},\phi_k)\pi^L_{k,2\tau}(\boldsymbol{\theta}_k \mid \phi_k, \mathbf{s}_n)\pi_k(\phi_k)d\boldsymbol{\theta}_k d\phi_k \tag{11}$$

where $m^L_{k,\tau}(\mathbf{y}_n)$ is the integrated likelihood with respect to $N(\boldsymbol{\theta}_k; \mathbf{0}, \tau\phi_k I)$ and $\pi^L_{k,2\tau}(\boldsymbol{\theta}_k \mid \phi_k, \mathbf{s}_n)$ is the Normal posterior implied by the $N(\boldsymbol{\theta}_k; \mathbf{0}, 2\tau\phi I)$ prior. Because $d(\theta_{ki},\phi_k) \leq c$ for some constant c, the Dominated Convergence Theorem gives that

$$\lim_{\mathbf{z}\to\mathbf{s}_{k,n}} g_k(\mathbf{z})\frac{m^L_{k,\tau}(\mathbf{y}_n)}{m^L_{k,2\tau}(\mathbf{y}_n)} = \int\int \prod_{i\in M_k} d(\theta_{ki},\phi_k)\pi^L_{k,2\tau}(\boldsymbol{\theta}_k \mid \phi_k, \mathbf{s}_n)\pi_k(\phi_k)d\boldsymbol{\theta}_k d\phi_k, \tag{12}$$

so that direct algebraic manipulation after adding the integrated likelihood terms delivers

$$\lim_{\mathbf{z}\to\mathbf{s}_{k,n}} g_k(\mathbf{z}) = \int\int \prod_{i\in M_k} \frac{\theta_{ki}^{2r}}{(2r-1)!!\phi^r\tau^r}\pi^L_{k,\tau}(\boldsymbol{\theta}_k \mid \phi_k, \mathbf{s}_n)\pi_k(\phi_k)d\boldsymbol{\theta}_k d\phi_k, \tag{13}$$

which proves that $g_k(\mathbf{s}_{k,n})$ is continuous. $\square$

Lemma 4. *Let $d_k(\boldsymbol{\theta}_k, \phi_k)$ be as in Lemma 3 and $c_n =$*

$$\int\int d_k(\boldsymbol{\theta}_k, \phi_k)\pi_k^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X_{k,n}'X_{k,n}, \phi_k)\pi_k^L(\phi_k \mid \boldsymbol{\theta}_k^*, X_{k,n}'X_{k,n}, \phi_k^*)d\boldsymbol{\theta}_k d\phi_k \tag{14}$$

as in (24). Then

$$\lim_{n\to\infty} c_n = \int\int d_k(\boldsymbol{\theta}_k, \phi_k) \lim_{n\to\infty} \pi_k^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X_{k,n}'X_{k,n}, \phi_k)$$
$$\pi_k^L(\phi_k \mid \boldsymbol{\theta}_k^*, X_{k,n}'X_{k,n}, \phi_k^*)d\boldsymbol{\theta}_k d\phi_k \tag{15}$$

Proof. The proof runs analogous to that in Lemma 3, except that now the limit is taken with respect to n and the c^{-p_k} term may grow as $n \to \infty$. That is, for bounded $d(\theta_{ki}, \phi_k)$ the argument proceeds by using the Dominated Convergence Theorem to obtain $\lim_{n\to\infty} c_n c^{-p_k} =$

$$\int \prod_{i\in M_k} c^{-1} d_k(\theta_{ki}, \phi_k) \lim_{n\to\infty} \pi_k^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X_{k,n}'X_{k,n}, \phi_k)\pi_k^L(\phi_k)d\phi_k, \tag{16}$$

so that

$$\lim_{n\to\infty} c_n = \int \prod_{i\in M_k} d_k(\theta_{ki}, \phi_k) \lim_{n\to\infty} \pi_k^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X_{k,n}'X_{k,n}, \phi_k)\pi_k^L(\phi_k)d\phi_k. \tag{17}$$

For the MOM prior we adjust the argument slightly. From (26) we obtain $c_n =$

$$\int\int \prod_{i\in M_k} \frac{\theta_{ki}^{2r}}{(2r-1)!!\tau^2\phi^2} N(\boldsymbol{\theta}_k; \boldsymbol{\theta}_k^*, \phi_k(X_{k,n}'X_{k,n})^{-1})\mathrm{IG}\left(\phi_k; \frac{n}{2}, \frac{n\phi_k^*}{2}\right)$$
$$c_\phi^*(X_{k,n}, \tau)N(\boldsymbol{\theta}_k; 0, \tau\phi_k I)\pi_k(\phi_k)d\boldsymbol{\theta}_k^* d\phi_k, \tag{18}$$

where $c_\phi^*(X_{k,n}, \tau) = c_\phi(\mathbf{s}_{k,n})$ in (23) plugging in $\boldsymbol{\theta}_{k,n} = \boldsymbol{\theta}_k^*$, $\hat{\phi}_k = \phi_k^*$. Following the same argument as in the proof of Lemma 3, we divide and multiply by a $N(\boldsymbol{\theta}_k; \mathbf{0}, 2\tau)$ kernel to obtain

$$c_n = \int\int \prod_{i\in M_k} d(\theta_{ki}, \phi_k)N(\boldsymbol{\theta}_k; \boldsymbol{\theta}_k^*, \phi_k(X_{k,n}'X_{k,n})^{-1})N(\boldsymbol{\theta}_k; 0, 2\tau\phi_k I)$$
$$c_\phi^*(X_{k,n}, \tau)\mathrm{IG}\left(\phi_k; \frac{n}{2}, \frac{n\phi_k^*}{2}\right)\pi_k(\phi_k)d\boldsymbol{\theta}_k^* d\phi_k =$$
$$\frac{c_\phi^*(X_{k,n}, \tau)}{c_\phi^*(X_{k,n}, 2\tau)} \int\int \prod_{i\in M_k} d(\theta_{ki}, \phi_k)\frac{N(\boldsymbol{\theta}_k; \mathbf{m}_{2\tau}, S_{2\tau})}{c_\theta^*(\phi, X_{k,n}, 2\tau)}$$
$$c_\phi^*(X_{k,n}, 2\tau)\mathrm{IG}\left(\phi_k; \frac{n}{2}, \frac{n\phi_k^*}{2}\right)\pi_k(\phi_k)d\boldsymbol{\theta}_k^* d\phi_k, \tag{19}$$

where $d(\theta_{ki},\phi_k) = \frac{\theta_{ki}^{2r} N(\theta_{ki};0,\tau\phi_k)}{(2r-1)!!\tau^2\phi^2 N(\theta_{ki};0,2\tau\phi_k)} \leq c$ for some constant c, $S_{2\tau} = X'_{n,k}X_{n,k} + (2\tau)^{-1}I$, $\mathbf{m}_{2\tau} = S_{2\tau}^{-1}(X'_{k,n}X_{k,n})\boldsymbol{\theta}_k^*$, and $1/c_\theta^*(\phi, X_{k,n}, 2\tau) = \int N(\boldsymbol{\theta}_k; \boldsymbol{\theta}_k^*, \phi_k(X'_{k,n}X_{k,n})^{-1})N(\boldsymbol{\theta}_k; 0, 2\tau\phi_k I)d\boldsymbol{\theta}_k$. Now, because $d(\theta_{ki},\phi_k)$ is bounded and the remaining expression in (19) is a probability density function on $(\boldsymbol{\theta}_k, \phi_k)$, the Dominated Convergence Theorem gives

$$\lim_{n\to\infty} c_n \frac{c_\phi^*(X_{k,n}, 2\tau)}{c_\phi^*(X_{k,n},\tau)} c^{-p_k/2} = \int\int \lim_{n\to\infty} \prod_{i\in M_k} d(\theta_{ki},\phi_k) \frac{N(\boldsymbol{\theta}_k; \mathbf{m}_{2\tau}, S_{2\tau})}{c_\theta^*(\phi, X_{k,n}, 2\tau)}$$

$$c_\phi^*(X_{k,n}, 2\tau)\mathrm{IG}\left(\phi_k; \frac{n}{2}, \frac{n\phi_k^*}{2}\right)\pi_k(\phi_k)d\boldsymbol{\theta}_k^* d\phi_k, \tag{20}$$

which after rearranging terms gives

$$\lim_{n\to\infty} c_n = \int\int \lim_{n\to\infty} \prod_{i\in M_k} \frac{\theta_{ki}^{2r}}{\tau^2\phi_k^2(2r-1)!!} N(\boldsymbol{\theta}_k; \mathbf{m}_{k,n}, \phi_k S_{k,n}^{-1})$$

$$c_\phi(X_{k,n},\tau)\mathrm{IG}\left(\phi_k; \frac{n}{2}, \frac{n\phi_k^*}{2}\right)\pi_k(\phi_k)d\boldsymbol{\theta}_k^* d\phi_k, \tag{21}$$

concluding the proof.

□

We now proceed to prove Proposition 1(iii). The argument is slightly tedious but the main idea is to use MLE consistency and the previous lemmas to exchange limits and integration and obtain the desired result. We note that the eigenvalue conditions D2 $an < l_1(X'_{k,n}X_{k,n}) < l_k(X'_{k,n}X_{k,n}) < bn$ give $||(X'_{k,n}X_{k,n})^{-1}||_2^2 \leq 1/an \to 0$ for fixed a, which in turn guarantees $\hat{\boldsymbol{\theta}}_{k,n} \xrightarrow{a.s.} \boldsymbol{\theta}_k^*$ (Lai et al., 1979). This implies $\hat{\phi}_{k,n} = n^{-1}(\mathbf{y}_n - X_{k,n}\hat{\boldsymbol{\theta}}_{k,n})'(\mathbf{y}_n - X_{k,n}\hat{\boldsymbol{\theta}}_{k,n}) \xrightarrow{a.s.} n^{-1}(\mathbf{y}_n - X_{k,n}\boldsymbol{\theta}_k^*)'(\mathbf{y}_n - X_{k,n}\boldsymbol{\theta}_k^*) \xrightarrow{a.s.} \phi_k^*$, given that $V(Y - X_{k,n}\boldsymbol{\theta}_k^*) = \phi_k^* < \infty$ by assumption. Hence, $d_k(\hat{\boldsymbol{\theta}}_{k,n}, \hat{\phi}_{k,n}) \xrightarrow{a.s.} d_k(\boldsymbol{\theta}_k^*, \phi_k^*)$. Since $m_{k,n} \xrightarrow{P} \hat{\boldsymbol{\theta}}_{k,n}$ as $n \to \infty$ and $d_k(\boldsymbol{\theta}_k, \phi_k)$ is assumed continuous, the Continuous Mapping Principle gives $d_k(\mathbf{m}_{k,n}, \phi_k) \xrightarrow{a.s.} d_k(\boldsymbol{\theta}_k^*, \phi_k^*)$.

To show that $g_k(\mathbf{y}_n) \xrightarrow{P} d_k(\mathbf{m}_{k,n}, \phi_k)$, we note that $\mathbf{s}_{k,n} = (\mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \hat{\phi}_{k,n})$ is a one-to-one function with the sufficient statistic $(\hat{\boldsymbol{\theta}}_{k,n}, X'_{k,n}X_{k,n}, \hat{\phi}_{k,n})$ under M_k. Hence $\mathbf{s}_{k,n}$ is also sufficient and $g_k(\mathbf{y}_n)$ depends only on $\mathbf{s}_{k,n}$, so that we may write $g_k(\mathbf{s}_{k,n}) =$

$$\int\int d_k(\boldsymbol{\theta}_k, \phi_k)\pi_k^L(\boldsymbol{\theta}_k \mid \mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \phi_k)\pi_k^L(\phi_k \mid \mathbf{s}_{k,n})d\boldsymbol{\theta}_k d\phi_k, \tag{22}$$

where straightforward algebra shows that $\pi_k^L(\boldsymbol{\theta}_k \mid \mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \phi_k) =$

$$c_\theta(\phi_k, \mathbf{s}_{k,n})N(\boldsymbol{\theta}_k; \hat{\boldsymbol{\theta}}_{k,n}, \phi_k(X'_{k,n}X_{k,n})^{-1})\pi_k^L(\boldsymbol{\theta}_k \mid \phi_k)$$

$$\pi_k^L(\phi_k \mid \mathbf{s}_{k,n}) = \frac{c_\phi(\mathbf{s}_{k,n})}{c_\theta(\phi_k, \mathbf{s}_{k,n})}\phi_k^{-(n-k)/2} e^{-\frac{1}{2\phi_k}(\mathbf{y}'_n\mathbf{y}_n - \hat{\boldsymbol{\theta}}'_{k,n}X'_{k,n}X_{k,n}\hat{\boldsymbol{\theta}}_{k,n})}, \tag{23}$$

where $c_\theta(\phi_k, \mathbf{s}_{k,n})$ is the normalization constant for $\boldsymbol{\theta}_k$ (which may depend on ϕ_k) and $c_\phi(\mathbf{s}_{k,n})$ that for the marginal posterior of ϕ_k.

Lemma 3 gives that $g_k(\mathbf{s}_{k,n})$ is continuous in $\mathbf{s}_{k,n} = (\mathbf{m}_{k,n}, X'_{k,n}X_{k,n}, \hat{\phi}_{k,n})$, hence by the Continuous Mapping Principle

$$g_k(\mathbf{s}_{k,n}) \xrightarrow{P} \int\int d_k(\boldsymbol{\theta}_k, \phi_k)\pi_k^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n}, \phi_k) \pi_k^L(\phi_k \mid \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n}, \phi_k^*)d\boldsymbol{\theta}_k d\phi_k = c_n, \tag{24}$$

where $\pi_k^L(\phi_k \mid \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n}, \phi_k^*) \propto c_\theta(\phi_k, \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n})^{-1}\phi_k^{-(n-k)/2}e^{-\frac{n\phi_k^*}{2\phi_k}}$. For a fixed sequence of $X_{k,n}$ (24) is just a sequence in n. To complete the proof we just need to show that $\lim_{n\to\infty} c_n \to d_k(\boldsymbol{\theta}_k^*, \phi_k^*)$ for any sequence $X_{k,n}$ satisfying the theorem assumptions, which combined with $d_k(\mathbf{m}_{k,n}, \phi_k^*) \xrightarrow{P} d_k(\boldsymbol{\theta}_k^*, \phi_k^*)$ would give that $g_k(\mathbf{s}_{k,n}) \xrightarrow{P} d_k(\mathbf{m}_{k,n}, \phi_k^*)$. By Lemma 4,

$$\lim_{n\to\infty} c_n = \int\int d_k(\boldsymbol{\theta}_k, \phi_k) \lim_{n\to\infty} \pi_k^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n}, \phi_k) \pi_k^L(\phi_k \mid \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n}, \phi_k^*)d\boldsymbol{\theta}_k d\phi_k^* = \int\int d_k(\boldsymbol{\theta}_k, \phi_k) \lim_{n\to\infty} h_n(\boldsymbol{\theta}_k, \phi_k)d\boldsymbol{\theta}_k d\phi_k^*. \tag{25}$$

Now, from (22)-(24) we obtain $h_n(\boldsymbol{\theta}_k, \phi_k) \propto$

$$N(\boldsymbol{\theta}_k; \boldsymbol{\theta}_k^*, \phi_k(X'_{k,n}X_{k,n})^{-1})\mathrm{IG}(\phi_k; n/2, n\phi_k^*/2)\pi_k^L(\boldsymbol{\theta}_k \mid \phi_k)\pi_k^L(\phi_k), \tag{26}$$

where IG denotes the inverse gamma density function.

Informally, given the assumptions on $\pi_k^L(\boldsymbol{\theta}_k \mid \phi_k)$, for (26) to converge to a point mass at $(\boldsymbol{\theta}_k^*, \phi_k^*)$ we need the trace of $(X'_{k,n}X_{k,n})^{-1}$ to converge to 0. Note that $\mathrm{tr}((X'_{k,n}X_{k,n})^{-1}) \leq k/l_1$, which is satisfied as long as l_1 grows faster with n than k does, and that under our assumptions $k/l_1 < k/(an) \to 0$. Formally, Condition D2 on the eigenvalues of $X'_{k,n}X_{k,n}$ imply that for $n > n_0$,

$$h_n(\boldsymbol{\theta}_k) \leq \mathrm{IG}(\phi_k; n/2, n\phi_k^*/2)\pi^L(\phi_k)\times \times \left(\frac{b}{a}\right)^{p_k/2} \frac{(na)^{p_k/2}}{(2\pi)^{p_k/2}\phi_k^{p_k/2}} \exp\left\{-\frac{na}{2\phi_k}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)'(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)\right\}\pi_k^L(\boldsymbol{\theta}_k \mid \phi_k) \tag{27}$$

We first study the second line in (27). Given that $p_k = o(n)$, for bounded π^L we have $\pi^L(\boldsymbol{\theta}_k \mid \phi_k) < \infty$ for all $\boldsymbol{\theta}_k$ the second line in (27) converges to 0 as $n \to \infty$ for any given ϕ_k and all $\boldsymbol{\theta}_k \neq \boldsymbol{\theta}_k^*$, *i.e.* $\pi^L(\boldsymbol{\theta}_k \mid \boldsymbol{\theta}_k^*, X'_{k,n}X_{k,n}, \phi_k)$ converges to a point mass at $\boldsymbol{\theta}_k^*$.

Now suppose that $\pi^L(\boldsymbol{\theta}_k \mid \phi_k)$ is unbounded in a 0 Lebesgue measure set $\tilde{\Theta}_k$. In this case it also holds that

$$\lim_{\boldsymbol{\theta}_k\to\tilde{\boldsymbol{\theta}}_k} \exp\left\{-\frac{na}{2\phi_k}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)'(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)\right\}\pi_k^L(\boldsymbol{\theta}_k \mid \phi_k) = 0 \tag{28}$$

for any $\tilde{\boldsymbol{\theta}}_k \in \tilde{\Theta}_k$. This can be seen by contradiction, *i.e.* assume that for $||\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_k||^2 < \epsilon$ and an arbitrary small ϵ there exists some $\delta > 0$ such that

$\exp\left\{-\frac{na}{2\phi_k}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)'(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)\right\} \pi_k^L(\boldsymbol{\theta}_k \mid \phi_k) > \delta$ for some arbitrarily large values of n. Then the prior probability of $||\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_k||^2 < \epsilon$

$$\int_{||\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_k||^2 < \epsilon} \pi_k^L(\boldsymbol{\theta}_k \mid \phi_k) d\boldsymbol{\theta}_k > \delta \int_{||\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_k||^2 < \epsilon} \exp\left\{\frac{na}{2\phi_k}(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)'(\boldsymbol{\theta}_k - \boldsymbol{\theta}_k^*)\right\}, \tag{29}$$

but the integrand is positive and increasing with n and hence by the Monotone Convergence Theorem (29) converges to ∞ as $n \to \infty$, which would imply that $\pi_k^L(\boldsymbol{\theta}_k \mid \phi_k)$ is improper.

Finally, we note that given that $\pi_k^L(\phi_k)$ is bounded and continuous the first line in (27) converges to 0 as $n \to \infty$ for any $\phi_k \neq \phi_k^*$, hence (25) converges to $d(\boldsymbol{\theta}_k^*, \phi_k^*)$, which completes the proof.

The assertion that when $M_t \not\subset M_k$ we have $c = 0$ if and only if $(X_k' X_k)^{-1} X_k' X_t$ has a column converging to 0 arises from the fact that $(\boldsymbol{\theta}_k^*, \phi_k^*)$ minimizing KL to $N(\mathbf{y}_n \mid X_t \boldsymbol{\theta}_t^*, \phi_t^* I)$ has a zero element in $\boldsymbol{\theta}_k^*$ if and only if its corresponding linear projection of X_k onto the space generated by X_t has zero coefficients, *i.e.* all entries in the corresponding column of $(X_k' X_k)^{-1} X_k' X_t$ are 0.

2.3. **Proof of Proposition 2, Part (i).** The proof is analogous to that in (Johnson and Rossell, 2010) for MOM and iMOM priors with additive penalties (as opposed to the multiplicative penalties of the product MOM and product iMOM considered here, which induce a significantly different asymptotic behaviour). The basic idea is to use that under W69 the log-likelihood can be asymptotically approximated by a quadratic function, the MLE is consistent and $O_p(n^{-1/2})$ from the data-generating truth and the observed Fisher information matrix converges to a positive-definite matrix. These facts allow finding an asymptotic expression for the system of equations giving the posterior modes from which we can deduce their probabilistic order.

For ease of notation we drop the subindex k indicating the model and we denote $\dim(\boldsymbol{\theta}_k) = p_k$ as before. Consider first the pMOM prior and take $\tau = 1$ without loss of generality. The log-posterior density is

$$L_n(\boldsymbol{\theta}, \phi) + \sum_{i=1}^{p_k} \log(\theta_i^2) - p\log(\phi) - \frac{1}{2\phi} \sum_{i=1}^{p_k} \theta_i^2 + \log\pi(\phi), \tag{30}$$

where $L_n(\boldsymbol{\theta}, \phi)$ is the log-likelihood and $\pi(\phi)$ is the prior density on ϕ. Suppose that the sampling model satisfies the conditions in Walker (1969), then $L_n(\boldsymbol{\theta}, \phi)$ can be approximated by a second order Taylor expansion around an MLE $(\hat{\boldsymbol{\theta}}, \hat{\phi})$ maximizing $L_n(\boldsymbol{\theta}, \phi)$. Performing this expansion and setting the partial derivative with respect to θ_i to 0 delivers that all posterior modes $\tilde{\boldsymbol{\theta}}$ must satisfy for their i^{th} component

$$\sum_{j=1}^{p_k} h_{ij}(\tilde{\theta}_j - \hat{\theta}_j) + \frac{2}{\tilde{\theta}_i} - \frac{\tilde{\theta}_i}{\phi} = 0, \tag{31}$$

where h_{ij} is the (i,j) element of the Hessian of $L_n(\boldsymbol{\theta},\phi)$ evaluated at $(\boldsymbol{\theta},\phi) = (\hat{\boldsymbol{\theta}},\hat{\phi})$. Rearranging terms we obtain

$$\tilde{\theta}_i \left(\frac{\tilde{\theta}_i}{n\phi} - \frac{h_{ii}}{n}(\tilde{\theta}_i - \hat{\theta}_i) \right) - \tilde{\theta}_i \sum_{j \neq i} \frac{h_{ij}}{n}(\tilde{\theta}_j - \hat{\theta}_j) = \frac{2}{n} \tag{32}$$

We note that the Taylor approximation to (30) is a quadratic form in $\boldsymbol{\theta}$, which is convex in $\boldsymbol{\theta}$, plus $\sum_{i=1}^{p_k} \log(\theta_i^2)$ which is symmetric around the origin and convex in each quadrant of $\mathbb{R}^{p_k}$ (*i.e.* for fixed sign of $\theta_1,\ldots,\theta_{p_k}$) and converges to $-\infty$ as any $\theta_i \longrightarrow 0$. Therefore (30) has a global maxima at the quadrant where $\hat{\boldsymbol{\theta}}$ occurs and a local maxima in each other quadrant. Consider first the two modes for θ_i occurring when $\text{sign}(\tilde{\theta}_j) = \text{sign}(\hat{\theta}_j)$ for $j \neq i$. Under Walker's conditions $\tilde{\theta}_j - \hat{\theta}_j \xrightarrow{P} 0$, $h_{ij}/n \xrightarrow{P} J_{ij}$ with finite J_{ij} (condition B4), and $\tilde{\phi} \xrightarrow{P} \phi^*$, where $(\boldsymbol{\theta}^*,\phi^*)$ minimizes KL to the data-generating model and we assume that $\phi^* > 0$. Incorporating these facts into (32) gives that the left hand side in (32) converges in probability to $-J_{ii}\hat{\theta}_i(\tilde{\theta}_i - \hat{\theta}_i)$, and hence both posterior modes must satisfy

$$n\tilde{\theta}_i(\tilde{\theta}_i - \hat{\theta}_i) \xrightarrow{P} c \tag{33}$$

with $0 < c < \infty$. We note that (33) remains valid for linear models satisfying the eigenvalue conditions D2 given that then $L_n(\boldsymbol{\theta},\phi)$ is exactly quadratic and the condition ensures almost sure convergence of the MLE, which implies $\tilde{\theta}_j - \hat{\theta}_j \xrightarrow{P} 0$ and $\tilde{\phi} \xrightarrow{P} \phi^*$. The two roots of the quadratic equation given by (33) are given by $\tilde{\theta}_i^{(1)} = \hat{\theta}_i + 0.5\left(\sqrt{\hat{\theta}_i^2 + 4c/n} - \sqrt{\hat{\theta}_i}\right)$ $\tilde{\theta}_i^{(2)} = -0.5\left(\sqrt{\hat{\theta}_i^2 + 4c/n} - \sqrt{\hat{\theta}_i}\right)$, which using a first order Taylor expansion of $g(z) = \sqrt{z}$ around $z = \hat{\theta}_i$ gives that $\tilde{\theta}_i^{(1)} = \hat{\theta}_i + O(c/n)\hat{\theta}_i^{-1}$, $\tilde{\theta}_i^{(2)} = O(c/n)\hat{\theta}_i^{-1}$. Suppose first that $\theta_i^* = 0$, then the MLE $\hat{\theta}_i = O_p(n^{-1/2})$ and hence $\hat{\theta}_i - \tilde{\theta}_i^{(1)} = O(c/n)\hat{\theta}_i^{-1} = O_p(n^{-1/2})$, $\tilde{\theta}_i^{(2)} = O_p(n^{-1/2})$. Now suppose that $\theta_i^* = a_n$ where $a_n >> n^{-1/2}$, then $\hat{\theta}_i = O_p(a_n)$ and hence $\hat{\theta}_i - \tilde{\theta}_i^{(1)} = O_p(1/(na_n))$, $\tilde{\theta}_i^{(2)} = O_p(1/(na_n))$. In particular, if θ_i^* is fixed then $\hat{\theta}_i - \tilde{\theta}_i^{(1)} = O_p(1/n)$, $\tilde{\theta}_i^{(2)} = O_p(1/n)$. To summarize, the modes $\tilde{\theta}_i$ when $\text{sign}(\tilde{\theta}_j) = \text{sign}(\hat{\theta}_j)$ for $j \neq i$ are either $O_p(n^{-1})$ from the MLE (and are hence $O_p(a_n)$ when $\theta_i^* = a_n >> n^{-1/2}$ and $O_p(n^{-1/2})$ when $\theta_i^* = 0$) or $O_p(n^{-1/2})$ from 0. The modes in other quadrants are given by the intersection of the contours of an ellipse that is centered at $\hat{\boldsymbol{\theta}}$ and has all axis lengths shrinking at rate $O(n^{-1})$ (from boundedness of eigenvalues) and the contours $\sum_{i=1}^{p_k} \log(\theta_i) = c$ for some $0 < c < \infty$ (which do not depend on n). Hence modes $\tilde{\boldsymbol{\theta}}$ occurring at quadrants other than that of $\hat{\boldsymbol{\theta}}$ shrink towards 0 at the same rate than $\tilde{\theta}_i^{(1)}$ and $\tilde{\theta}_i^{(2)}$ above, and in particular $\tilde{\theta}_i = O_p(n^{-1/2})$ given that $a_n >> n^{-1/2}$.

The proof for the piMOM and peMOM priors follows in an analogous fashion. Performing a second order Taylor approximation to the piMOM posterior around an MLE $\hat{\boldsymbol{\theta}}$ and setting the partial derivative with respect to θ_i to 0 delivers that $\tilde{\theta}_i$ and $\tilde{\phi}_i$ must satisfy

$$\sum_{j=1}^{p_k} h_{ij}(\tilde{\theta}_i - \hat{\theta}_i) + h_{i,p_k+1}(\tilde{\phi} - \hat{\phi}) - \frac{2}{\tilde{\theta}_i} + \frac{2\tilde{\phi}}{\tilde{\theta}_i^3} = 0, \tag{34}$$

where as before h_{ij} indicates the Hessian of $L_n(\boldsymbol{\theta}, \phi)$ evaluated at $(\hat{\boldsymbol{\theta}}, \hat{\phi})$. Rearranging terms delivers

$$\frac{h_{ii}}{n}\tilde{\theta}_i^3(\tilde{\theta}_i - \hat{\theta}_i) + \tilde{\theta}_i^3\left(\sum_{j\neq i}\frac{h_{ij}}{n}(\tilde{\theta}_i - \hat{\theta}_i) + \frac{h_{i,p_k+1}}{n}(\tilde{\phi} - \hat{\phi})\right) - \frac{2}{n}\tilde{\theta}_i^2 + \frac{2\phi}{n} = 0. \tag{35}$$

We again consider the modes for $\tilde{\theta}_i$ when $\text{sign}(\tilde{\theta}_j) = \text{sign}(\hat{\theta}_j)$ for $j \neq i$. Either W69 or the eigenvalue conditions D2 for linear models guarantee that $h_{ij}/n \xrightarrow{P} J_{ij}$ for all i, j, whereas MLE consistency gives that $(\tilde{\theta}_i - \hat{\theta}_i) \xrightarrow{P} 0$, and $\tilde{\phi} - \hat{\phi} \xrightarrow{P} 0$. Therefore, $\tilde{\theta}_i$ must satisfy

$$n\tilde{\theta}_i^3(\tilde{\theta}_i - \hat{\theta}_i) \xrightarrow{P} c, \tag{36}$$

where $0 < c < \infty$. The two roots in (36) must satisfy either $(\tilde{\theta}_i - \hat{\theta}_i) \xrightarrow{P} 0$ (Mode 1) or $\tilde{\theta}_i^3 \xrightarrow{P} 0$ (Mode 2). Consider first the case of a fixed true parameter value $\theta_i^* \neq 0$, then $\hat{\theta}_i \xrightarrow{P} \theta_i^*$ and thus for Mode 1 $\tilde{\theta}_i^3 \xrightarrow{P} (\theta_i^*)^3 \neq 0$, so that $n(\tilde{\theta}_i - \hat{\theta}_i) \xrightarrow{P} c'$ with $0 < c' < \infty$. For Mode 2, which in particular satisfies $\text{sign}(\tilde{\theta}_i) \neq \text{sign}(\hat{\theta}_i)$ and hence $\tilde{\theta}_i - \hat{\theta}_i \xrightarrow{P} \theta_i^* \neq 0$, we obtain $n\tilde{\theta}_i^3 \xrightarrow{P} c'$.

Now consider the case when $\theta_i^* = 0$, here we shall first see that the posterior mode shrinks to 0 at a rate strictly slower than the MLE, *i.e.* $\tilde{\theta}_i = O_p(b_n)$ with $b_n >> n^{-1/2}$, and then show that this implies $n^{-1/4}\tilde{\theta}_i \xrightarrow{P} c$. To see that $b_n >> n^{-1/2}$, let us assume that $b_n = O(n^{-1/2})$ and see that this leads to a contradiction. Under this assumption $\tilde{\theta}_i - \hat{\theta}_i = o_p(n^{-1/2})$ so $n\tilde{\theta}_i^4(1 - \hat{\theta}_i/\tilde{\theta}_i) = O_p(1/n)(1 - \hat{\theta}_i/\tilde{\theta}_i)$ does not converge to a finite constant as required by (36), leading to a contradiction. Therefore $b_n >> n^{-1/2}$, which means that $1 - \hat{\theta}_i/\tilde{\theta}_i \xrightarrow{P} 0$ and from (36) we obtain $n^{-1/4}\tilde{\theta}_i \xrightarrow{P} c$.

Finally consider the case with vanishing $\theta_i^* = a_n \neq 0$. First focus on the mode with $\text{sign}(\tilde{\theta}_i) = \text{sign}(\hat{\theta}_i)$. If $n^{-1/2} << a_n << n^{-1/4}$, a similar argument to the $\theta_i^* = 0$ gives that $b_n >> a_n$ and hence $n^{-1/4}\tilde{\theta}_i \xrightarrow{P} c$. If $a_n >> n^{-1/4}$ we shall show that assuming $b_n >> a_n$ leads to a contradiction and that hence $b_n \asymp a_n$. If $b_n >> a_n$ then we would have $\tilde{\theta}_i - \hat{\theta}_i \asymp b_n$ (since $\hat{\theta}_i \asymp a_n$) and $n\tilde{\theta}_i^3(\tilde{\theta}_i - \hat{\theta}_i) \asymp nb_n^4$, which given that $b_n >> n^{-1/4}$ cannot converge to a finite constant as required by (36), leading to a contradiction. That is, the primary mode is of order a_n when $a_n >> n^{-1/4}$ and of order $n^{-1/4}$ otherwise, as desired. Now consider the mode with $\text{sign}(\tilde{\theta}_i) \neq \text{sign}(\hat{\theta}_i)$. If $a_n >> n^{-1/4}$ we have $\tilde{\theta}_i - \hat{\theta}_i = a_n + o_p(a_n)$ and hence $na_n\tilde{\theta}_i^3 \xrightarrow{P} c$. Note that by assumption $a_n >> n^{-1/4}$, hence $\tilde{\theta}_i = O_p(n^{-1/4})$ as desired.

Recall that these asymptotic rates apply to the modes with $\text{sign}(\tilde{\theta}_j) = \text{sign}(\hat{\theta}_j)$ for $j \neq i$. Similarly to the pMOM proof, all axes corresponding to the quadratic expansion contract exactly at rate n^{-1}, hence modes in all other quadrants are also $\tilde{\theta}_i = O_p(n^{-1/4})$. The proof for the peMOM case proceeds identically, with the only difference that term $-2\tilde{\theta}_i^2/n$ in (35) changes for $-\tilde{\theta}_i^4/(n\phi) \xrightarrow{P} 0$, hence one obtains the same convergence in probability for $\tilde{\theta}_i$.

2.4. **Proof of Proposition 2, Part (ii).** We first state a lemma regarding the derivatives of the univariate log-MOM, eMOM and iMOM prior densities with prior dispersion $\tau = 1$, which are useful in characterizing the asymptotic behaviour of Laplace approximations to integrated likelihoods and posterior means. We do not prove the lemma, as it follows from straightforward algebra.

Lemma 5. *Let $l(\theta_i, \phi) = log(\pi(\theta_i \mid \phi))$.*

(i) Let $\pi(\theta_i \mid \phi) \propto \phi^{-3/2}\theta_i^2 exp\{-\frac{1}{2}\theta_i^2/\phi\}$ be the MOM density, then

$$\frac{\partial^2 l}{\partial \theta_i^2} = -\frac{2}{\theta_i^2} - \frac{1}{\phi}; \frac{\partial^2 l}{\partial \theta_i \partial \phi} = \frac{\theta_i}{\phi^2}; \frac{\partial^2 l}{\partial \phi^2} = \frac{3}{2\phi^2} - \frac{\theta_i^2}{\phi^3}; \frac{\partial^3 l}{\partial \theta_i^3} = \frac{4}{\theta_i^3}.$$

(ii) Let $\pi(\theta_i \mid \phi) \propto exp\{-\phi/\theta_i^2\}\phi^{-1/2} exp\{-\frac{1}{2}\theta_i^2/\phi\}$ be the eMOM density, then

$$\frac{\partial^2 l}{\partial \theta_i^2} = -\frac{6\phi}{\theta_i^4} - \frac{1}{\phi}; \frac{\partial^2 l}{\partial \theta_i \partial \phi} = \frac{2}{\theta_i^3} + \frac{\theta_i}{\phi^2}; \frac{\partial^2 l}{\partial \phi^2} = \frac{1}{2\phi^2} - \frac{\theta_i^2}{\phi^3}; \frac{\partial^3 l}{\partial \theta_i^3} = \frac{24\phi}{\theta_i^5}.$$

(iii) Let $\pi(\theta_i \mid \phi) \propto \phi^{1/2}\theta_i^{-2} exp\{-\phi/\theta_i^2\}$ be the iMOM density, then

$$\frac{\partial^2 l}{\partial \theta_i^2} = \frac{2}{\theta_i^2} - \frac{6\phi}{\theta_i^4}; \frac{\partial^2 l}{\partial \theta_i \partial \phi} = \frac{2}{\theta_i^3}; \frac{\partial^2 l}{\partial \phi^2} = -\frac{1}{2\phi^2}; \frac{\partial^3 l}{\partial \theta_i^3} = -\frac{4}{\theta_i^3} + \frac{24\phi}{\theta_i^5}.$$

We now proceed to prove Proposition 2(ii) for models satisfying W69. For ease of notation we drop the subindex k and the conditioning on model M_k. The posterior mean of interest is $E(\theta_i \mid \mathbf{y}_n) =$

$$\frac{\int\int \theta_i \exp\left\{\log(\pi(\boldsymbol{\theta} \mid \phi)) + L_n(\boldsymbol{\theta}, \phi) + \log(\pi(\phi))\right\} d\boldsymbol{\theta} d\phi}{\int\int \exp\left\{\log(\pi(\boldsymbol{\theta} \mid \phi)) + L_n(\boldsymbol{\theta}, \phi) + \log(\pi(\phi))\right\} d\boldsymbol{\theta} d\phi} =$$

$$\frac{\int\int \theta_i e^{-nh_n(\boldsymbol{\theta},\phi)} d\boldsymbol{\theta} d\phi}{\int\int e^{-nh_n(\boldsymbol{\theta},\phi)} d\boldsymbol{\theta} d\phi}, \tag{37}$$

where $L_n(\boldsymbol{\theta}, \phi)$ is the log-likelihood function. We shall use Theorem 4 in Kass et al. (1990) to obtain a Laplace approximation to (37) by expanding $h_n(\boldsymbol{\theta}, \phi)$ around its main posterior mode $(\tilde{\boldsymbol{\theta}}, \tilde{\phi})$. We note that when the true parameter value $\theta_i^* = 0$ the posterior multi-modality does not vanish even as $n \to \infty$, but defer discussion of this point to later in the proof. We note that W69 ensure that the model is Laplace regular and hence Theorem 4 in Kass et al. (1990) can be used. To use the theorem we set $g(\boldsymbol{\theta}, \phi) = \theta_i$, $b(\boldsymbol{\theta}, \phi) = 1$ and $\gamma(\boldsymbol{\theta}, \phi) = \pi(\boldsymbol{\theta} \mid \phi)\pi(\phi)$ and note that $g(\boldsymbol{\theta}, \phi)$ are four times $\gamma(\boldsymbol{\theta}, \phi)$ and six times differentiable. We also note that when $\pi(\boldsymbol{\theta} \mid \phi)$ is either the eMOM or iMOM prior density, it is infinitely differentiable but not analytical at $\theta_i = 0$, but $\tilde{\theta}_i$ cannot occur at 0 (the prior density is 0) and hence we may ignore this set with 0 Lebesgue measure. Direct application of Theorem 4 in Kass et al. (1990) gives

$$E(\theta_i \mid \mathbf{y}_n) = \tilde{\theta}_i + \frac{1}{n}\sum_{j=1}^{p_k+1} h_{ij}\left(-\frac{1}{2}\sum_{r,s} h^{rs}h_{rsj}\right) + O\left(n^{-2}\right) \tag{38}$$

where $p_k = \dim(\boldsymbol{\theta})$, h_{ij} denotes the (i,j) element of the Hessian of $h_n(\boldsymbol{\theta}, \phi)$ evaluated at $(\tilde{\boldsymbol{\theta}}, \tilde{\phi})$, h^{ij} that of the inverse Hessian and h_{rsj} are third derivatives. That is,

$$\begin{aligned} h_{ii} &= \frac{1}{n}\frac{\partial^2}{\partial\theta_i^2}L_n(\boldsymbol{\theta},\phi) + \frac{1}{n}\frac{\partial^2}{\partial\theta_i^2}\log(\pi(\theta_i \mid \phi)), \\ h_{ij} &= \frac{1}{n}\frac{\partial^2}{\partial\theta_i\partial\theta_j}L_n(\boldsymbol{\theta},\phi), \\ h_{i,p_k+1} &= \frac{1}{n}\frac{\partial^2}{\partial\theta_i\partial\phi}L_n(\boldsymbol{\theta},\phi) + \frac{1}{n}\frac{\partial^2}{\partial\theta_i\partial\phi}\log(\pi(\theta_i \mid \phi)). \end{aligned} \tag{39}$$

From the Normal approximation to the likelihood we obtain that $h_{rsj} \xrightarrow{P} 0$ unless $r = s = j$, in which case $h_{jjj} \xrightarrow{P} \frac{\partial^3}{\partial\theta_j^3}\log(\pi(\tilde{\theta}_j \mid \phi))$. Hence,

$$E(\theta_i \mid \mathbf{y}_n) \xrightarrow{P} \tilde{\theta}_i - \frac{1}{2n}\left(\sum_{j=1}^{p_k+1} h_{ij}h^{jj}h_{jjj}\right) \tag{40}$$

W69 ensure that h_{ij} for $i \neq j$ converge in probability to a finite J_{ij}. Regarding h_{ii}, the first term converges to J_{ii} whereas the second term is $O_p(n^{-1})$ when $\theta_i^* \neq 0$ is fixed, and $O_p(1)$ when $\theta_i^* = 0$ for either the MOM, eMOM or iMOM prior (Proposition 2(i) and Lemma 5), hence $h_{ii} = O_p(1)$. When $\theta_i^* = a_n$ the second term in h_{ii} becomes $O_p(1/(na_n^2)) = o_p(1)$ (pMOM, $a_n >> n^{-1/2}$) or $O_p(1/(na_n^4)) = o_p(1)$ (peMOM and piMOM, $a_n >> n^{-1/4}$). This in turn implies that the Hessian converges in probability to J plus diagonal terms that either converge to 0 or are $O_p(1)$, and hence the elements in its inverse $h^{jj} = O_p(1)$. Finally consider h_{jjj}. From Lemma 5 when $\theta_j^* \neq 0$ we obtain $h_{jjj} = O_p(1/n)$ for either the MOM, eMOM or iMOM priors. When $\theta_j^* = 0$ for the MOM prior $h_{jjj} \xrightarrow{P} 4n^{-1}\tilde{\theta}_i^{-3} = n^{-1}O_p(n^{3/2}) = O_p(n^{1/2})$ for $j = 1, \ldots, p_k$ and $h_{jjj} \xrightarrow{P} O_p(1)$ for $j = p_k + 1$. For the eMOM and iMOM priors $h_{jjj} \xrightarrow{P} 24n^{-1}\tilde{\phi}\tilde{\theta}_i^{-5} = n^{-1}O_p(n^{5/4}) = O_p(n^{1/4})$ for $j = 1, \ldots, p_k$ and again $h_{jjj} \xrightarrow{P} O_p(1)$ for $j = p_k + 1$. Similarly when $\theta_i^* = a_n$ we get $h_{jjj} = o_p(\sqrt{n})$ (pMOM) and $h_{jjj} = o_p(n^{1/4})$ (peMOM, piMOM).

Putting these results together from (38) we obtain that if $\theta_j^* \neq 0$ is fixed for $j = 1, \ldots, p_k$ then $E(\theta_i \mid \mathbf{y}_n) \xrightarrow{P} \tilde{\theta}_i + O_p(n^{-2})$ if $\theta_j^* = a_n \neq 0$ for $j = 1, \ldots, p_k$ then $\tilde{\theta}_i + O_p(n^{-2}a_n^{-3}) = \tilde{\theta}_i + o_p(n^{-1/2})$ (pMOM) or $\tilde{\theta}_i + O_p(n^{-2}a_n^{-5}) = \tilde{\theta}_i + o_p(n^{-1/4})$ (peMOM, piMOM) , and finally if $\theta_j^* = 0$ for any $j = 1, \ldots, p_k$ then $E(\theta_i \mid \mathbf{y}_n) \xrightarrow{P} \tilde{\theta}_i + O_p(n^{-1/2})$ In particular, in cases of parameter orthogonality where $h_{ij} = 0$ for all $i \neq j$ then the difference between the posterior mean and posterior mode of θ_i is $O_p(n^{-1})$ whenever $\theta_i^* \neq 0$. To conclude the proof, we recall that the posterior is multi-modal and hence approximate $E(\theta_i \mid \mathbf{y}_n)$ by adding (38) across the 2^{p_k} modes. Proposition 2 gives that for such modes $\tilde{\theta}_i = O_p(n^{-1/2})$

for pMOM and $\tilde{\theta}_i = O_p(n^{-1/4})$ for peMOM and piMOM, hence $E(\theta_i \mid \mathbf{y}_n) = \hat{\theta}_i + O_p(n^{-1/2}) = \theta_i^* + O_p(n^{-1/2})$ for MOM and $E(\theta_i \mid \mathbf{y}_n) = \hat{\theta}_i + O_p(n^{-1/4}) = \theta_i^* + O_p(n^{-1/4})$ for eMOM or iMOM.

2.5. **Proof of Proposition 2, Part (iii).** The proof proceeds analogously to that of Part (ii). We consider linear models of growing dimensionality, again dropping the model subindex k for ease of notation. Although we assume that $X_n'X_n$ is a diagonal matrix, we state part of the argument for general $X_n'X_n$ (subject to the eigenvalue conditions in D2) and make explicit where the orthogonality assumption is needed. As argued during the proof of Proposition 2(i), the rates for posterior modes remain valid for linear models with such bounded eigenvalues. Regarding the posterior mean, the Condition D2 guarantees Laplace regularity (Kass et al., 1990) and hence the expansion (38) remains valid, where now $h_n(\boldsymbol{\theta}, \phi) =$

$$\frac{1}{2}\log(\phi) + \frac{1}{2\phi}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\frac{X_n'X_n}{n}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) - \frac{1}{n}\sum_{i=1}^{p_k}\log(\pi(\theta_i \mid \phi)) - \frac{1}{n}\log(\pi(\phi)) \tag{41}$$

Therefore h_{ij} is given by the (i,j) element in $\frac{X'X}{n\tilde{\phi}}$ for $i = 1, \ldots, p_k$, $i \neq j$, which is $O_p(1)$. For h_{ii} we add $\frac{1}{n}\frac{\partial^2}{\partial\theta_i^2}\log(\pi(\theta_i \mid \phi)$, which from Lemma 5 and Proposition 2 is $O_p(n^{-1})$ for the main mode $\tilde{\theta}_i$ when θ_i^* is fixed, $o_p(1)$ when $\theta_i^* = a_n$ (for pMOM with $a_n >> n^{-1/2}$ or peMOM/piMOM with $a_n >> n^{-1/4}$), and $O_p(1)$ when $\theta_i^* = 0$ or for any other mode (pMOM, peMOM and piMOM), hence in all cases $h_{ii} = O_p(1)$. The elements $h_{1,p_k+1}, \ldots, h_{p_k,p_k+1}$ are given by the vector

$$-\frac{1}{\tilde{\phi}^2}\frac{X_n'X_n}{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - \frac{1}{n}g(\tilde{\boldsymbol{\theta}}, \tilde{\phi}), \tag{42}$$

where $g(\tilde{\boldsymbol{\theta}}, \tilde{\phi})$ contains $\frac{\partial^2}{\partial\theta_i\partial\phi}\log(\pi(\theta_i \mid \phi))$ for $i = 1, \ldots, p_k$. Given that the eigenvalues of $X_n'X_n/n$ are bounded the first term in (42) converges in probability to 0 for the main mode where $\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}} \xrightarrow{P} 0$ and is $O_p(1)$ for all other modes. From Lemma 5 and Proposition 2(i) it is straightforward to see that $n^{-1}g(\tilde{\boldsymbol{\theta}}, \tilde{\phi}) \xrightarrow{P} \mathbf{0}$, hence $h_{i,p_k+1} = O_p(1)$ for $i = 1, \ldots, p_k$. Similarly, $h_{p_k+1,p_k+1} =$

$$\begin{aligned} -\frac{1}{2\phi^2} + \frac{1}{\phi^3}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}})'\frac{X_n'X_n}{n}(\tilde{\boldsymbol{\theta}} - \hat{\boldsymbol{\theta}}) - \frac{1}{n}\sum_{i=1}^{p_k}\frac{\partial^2}{\partial\phi^2}\log(\pi(\theta_i, \phi)) \\ + \frac{1}{n}\frac{\partial^2}{\partial\phi^2}\log(\pi(\phi)), \end{aligned} \tag{43}$$

which from Proposition 2(i) and Lemma 5 is $O_p(1)$.

Regarding the elements in the inverse Hessian h^{ij}, the Hessian is positive definite with $h_{ij} = O_p(1)$ and hence $h^{ij} = O_p(1)$ for $i, j = 1, \ldots, p_k + 1$.

Finally we obtain third derivatives h_{rsj}. Because h_{rs} is given by the corresponding element $X_n'X_n/(n\tilde{\phi})$, $h_{rsj} = 0$ for $r, s, j \in \{1, \ldots, p_k\}$. Consider now the term h_{jjj} corresponding to $r = s = j$. As in the proof of Proposition 2(i) (either with fixed

θ_i^* or $\theta_i^* = a_n$, $a_n >> n^{-1/2}$ for pMOM, $a_n >> n^{-1/4}$ for peMOM/piMOM), for the main mode $h_{jjj} = O_p(1)$ (Lemma 5) whereas for other modes $h_{jjj} = O_p(n^{1/2})$ under a pMOM or $O_p(n^{1/4})$ under a peMOM or piMOM priors (Proposition 2(i)). From (38), the contribution to $E(\theta_i \mid \mathbf{y}_n)$ from each mode is

$$\tilde{\theta}_i - \frac{1}{2n} \sum_{j=1}^{p_k+1} h_{ij} h^{jj} h_{jjj} \tag{44}$$

plus a lower order term.

Consider now that $X_n' X_n$ is orthogonal. In that case $h_{ij} = 0$ for $i \neq j$ and the two values $\tilde{\theta}_i^{(1)}, \tilde{\theta}_i^{(2)}$ maximizing the posterior are independent of θ_j for $j \neq i$. Therefore under a pMOM prior

$$E(\theta_i \mid \mathbf{y}) = \tilde{\theta}_i^{(1)} + \tilde{\theta}_i^{(2)} - \frac{1}{2n} O_p(n^{1/2}) = \theta_i^* + O_p(n^{-1/2}) \tag{45}$$

whereas

$$E(\theta_i \mid \mathbf{y}) = \tilde{\theta}_i^{(1)} + \tilde{\theta}_i^{(2)} - \frac{1}{2n} O_p(n^{1/4}) = \theta_i^* + O_p(n^{-1/4}) \tag{46}$$

under either a peMOM or piMOM prior, which concludes the proof.

2.6. **Proof of Proposition 3, Part (i).** The strategy is to show that Bayes factors between any M_k and the true model M_t can be approximated by ratios of Laplace approximations to the corresponding integrated likelihoods plus a negligible term, and then use the probabilistic order of the posterior modes obtained in Proposition 2 to obtain asymptotic Bayes factor rates.

Consider models M_k for $k = 1, \ldots, K$, all satisfying the W69 conditions. Let M_t be the true model and let k be such that $M_t \subset M_k$. Consider first the pMOM prior. We shall first characterize the asymptotic behaviour of the Bayes factor when $M_t \subset M_k$ for fixed $\boldsymbol{\theta}_t^*$, then consider $M_t \subset M_k$ with vanishing $\boldsymbol{\theta}_t^* = \boldsymbol{\theta}_0 a_n$ where throughout we assume that $\lim_{n\to\infty} a_n = 0$ with $a_n >> n^{-1/2}$ for pMOM and $a_n >> n^{-1/4}$ for peMOM and piMOM. The marginal likelihood $m_t(\mathbf{y}_n)$ under M_t can be approximated by a Laplace expansion around each posterior mode $(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)})$ for $m = 1, \ldots, 2^{p_t}$, obtaining for each mode

$$e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)})} \prod_i \frac{(\tilde{\theta}_{ti}^{(m)})^2}{\tau \tilde{\phi}^{(m)}} N(\tilde{\boldsymbol{\theta}}_t^{(m)}; \mathbf{0}, \tau \tilde{\phi}_t^{(m)} I) \pi(\tilde{\phi}_t^{(m)}) \left| H(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)}) \right|^{-1/2}, \tag{47}$$

where $L_n(\cdot)$ is the log-likelihood and $H(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)})$ the Hessian of the log-likelihood plus the log-prior density evaluated at $(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)})$. Expressions for the elements in $H(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)})$ are given in the proof of Proposition 2 for pMOM, peMOM and piMOM priors.

Without loss of generality denote by $(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})$ the mode located in the same quadrant as the MLE $(\hat{\boldsymbol{\theta}}, \hat{\phi})$. As seen in Proposition 2, under Walker's conditions

$(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)}) \xrightarrow{P} (\boldsymbol{\theta}_t^*, \phi_t^*)$ and $n^{-1}H(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)}) \xrightarrow{P} J$ for a positive-definite J, hence (47) converges in probability to

$$e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})} c_1 n^{-p_t/2} c_2, \tag{48}$$

where $c_1, c_2 > 0$. For modes in any other quadrant $e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)}) - L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})} \xrightarrow{P} e^{-nc_3}$, where $c_3 > 0$ is the KL between the data-generating density $f_t(\boldsymbol{\theta}_t^*, \phi_t^*)$ and that where some elements in $\boldsymbol{\theta}_t$ are set to 0 (which is positive by assumption). Further, in such quadrants $\prod_i \frac{(\tilde{\theta}_{ti}^{(m)})^2}{\tau \tilde{\phi}^{(m)}} = O_p(n^{-p_t})$ so that the sum of (47) across all modes $m = 1, \ldots, 2^{p_t}$ gives that the marginal likelihood $m_t(\mathbf{y}_n) \approx$

$$\begin{aligned} & e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})} \left(Z_1 Z_2 n^{-p_t/2} + \sum_m e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)}) - L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})} O_p(n^{-p_t}) Z_3 \right) \\ & \xrightarrow{P} n^{-p_t/2} e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})} Z_4, \end{aligned} \tag{49}$$

where $Z_j \xrightarrow{P} c_j > 0$ for $j = 1, \ldots, 4$.

Now consider M_k such that $M_t \subset M_k$ with $\boldsymbol{\theta}_t^* \neq 0$. Denote by $\boldsymbol{\theta}_{k1}$ the subset of $\boldsymbol{\theta}_k$ such that $\theta_{ki}^* = 0$ and $\boldsymbol{\theta}_{k2}^*$ that for $\theta_{ki}^* \neq 0$, where $\boldsymbol{\theta}_k^*$ minimizes KL to the data-generating $f_t(\mathbf{y}_n \mid \boldsymbol{\theta}_t, \phi_t)$ and $\dim(\boldsymbol{\theta}_{k1}) = p_k - p_t$. Following the same argument as for M_t, it suffices to focus on modes for which $\tilde{\boldsymbol{\theta}}_{k2}$ lies in the same quadrant as $\boldsymbol{\theta}_{k2}^*$. Adding up the Laplace approximations across all $2^{p_k - p_t}$ such modes delivers the Bayes factor
$\text{BF}_{kt} = \frac{m_k(\mathbf{y}_n)}{m_t(\mathbf{y}_n)} \xrightarrow{P}$

$$\sum_{m=1}^{2^{p_k - p_t}} \frac{e^{L_n(\tilde{\boldsymbol{\theta}}_k^{(m)}, \tilde{\phi}_k^{(m)})}}{e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})}} \times \frac{\prod_{i=1}^{p_k} \frac{(\tilde{\theta}_{ki}^{(m)})^2}{\tau \tilde{\phi}_k^{(m)}}}{\prod_{i=1}^{p_t} \frac{(\tilde{\theta}_{ti}^{(1)})^2}{\tau \tilde{\phi}_t^{(1)}}} \times \frac{\pi(\tilde{\phi}_k^{(m)})}{\pi(\tilde{\phi}_t^{(1)})} \times \frac{n^{-p_k/2}}{n^{-p_t/2}} \times \frac{\left| n^{-1} H(\tilde{\boldsymbol{\theta}}_k^{(m)}, \tilde{\phi}_k^{(m)}) \right|}{\left| n^{-1} H(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)}) \right|}, \tag{50}$$

where the first term is $O_p(1)$, the second term converges in probability to $n^{-(p_k - p_t)} Z_5$ for some random variable $Z_5 = O_p(1)$, the third and fourth terms converge in probability to a positive constant ($\pi(\phi)$ is bounded by assumption). Therefore each summand in (50) is $O_p(n^{-\frac{3}{2}(p_k - p_t)})$, and given that we are adding up a finite number of terms $\text{BF}_{kt} = O_p(n^{-\frac{3}{2}(p_k - p_t)})$.

Next consider that $M_t \subset M_k$ and $\boldsymbol{\theta}_t^* = \boldsymbol{\theta}_0 a_n$ with a_n vanishing a_n as above. As seen in Proposition 2 in this case $\tilde{\boldsymbol{\theta}}_t$ is of order a_n, from W69 $n^{-1}H(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)}) \xrightarrow{P} J$ as before and for the primary mode (47) converges in probability to $e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)})} c_1 a_n^{2p_t} n^{-p_t/2} c_2$. For other modes using that under W69 L_n can be asymptotically approximated by a quadratic expansion we obtain that KL $\left(f_t(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}_t^{(1)}), f_t(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}_t^{(m)}) \right)$ is of order $(\tilde{\boldsymbol{\theta}}_t^{(1)} - \tilde{\boldsymbol{\theta}}_t^{(m)})'(\tilde{\boldsymbol{\theta}}_t^{(1)} - \tilde{\boldsymbol{\theta}}_t^{(m)})$. Now, $\tilde{\boldsymbol{\theta}}_t^{(1)}$ is of order a_n and each element in $\tilde{\boldsymbol{\theta}}^{(m)}$ is $O_p(n^{-1/2})$ or $O_p(a_n)$, hence $L_n(\tilde{\boldsymbol{\theta}}_t^{(1)}, \tilde{\phi}^{(1)}) - L_n(\tilde{\boldsymbol{\theta}}_t^{(m)}, \tilde{\phi}^{(m)}) \xrightarrow{P} -na_n^2 c_3$. Adding across all modes we obtain $m_t(\mathbf{y}_n) = a_n^{2p_t} n^{-p_t/2} e^{L_n(\boldsymbol{\theta}_t^{(1)}, \phi_t^{(1)})} Z_4$. Similarly to the fixed $\boldsymbol{\theta}_t^*$ case we obtain the

Bayes factor

$$\sum_{m=1}^{2^{p_k-p_t}} \frac{e^{L_n(\tilde{\boldsymbol{\theta}}_k^{(m)},\tilde{\phi}_k^{(m)})}}{e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})}} \times \frac{\prod_{i=1}^{p_k} \frac{(\tilde{\theta}_{ki}^{(m)})^2}{\tau\tilde{\phi}_k^{(m)}}}{\prod_{i=1}^{p_t} \frac{(\tilde{\theta}_{ti}^{(1)})^2}{\tau\tilde{\phi}_t^{(1)}}} \times \frac{\pi(\tilde{\phi}_k^{(m)})}{\pi(\tilde{\phi}_t^{(1)})} \times \frac{n^{-p_k/2}}{n^{-p_t/2}} \times \frac{\left|n^{-1}H(\tilde{\boldsymbol{\theta}}_k^{(m)},\tilde{\phi}_k^{(m)})\right|}{\left|n^{-1}H(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})\right|}, \tag{51}$$

where now the first term is $o_p(1)$, the second term is of order $n^{-(p_k-p_t)}$ and the remaining terms are as before, hence $\text{BF}_{kt} = O_p(n^{-\frac{3}{2}(p_k-p_t)})$.

Let us now consider models M_k that do not contain M_t and that $\boldsymbol{\theta}_t^*$ is fixed. By assumption, the minimum Kullback-Leibler divergence $\text{KL}(M_t, M_k)$ between $f_t(\boldsymbol{\theta}_t^*, \phi_t^*)$ and any $f_k(\boldsymbol{\theta}_k^*, \phi_k^*)$ with $(\boldsymbol{\theta}_k, \phi_k) \in (\Theta_k, \Phi)$ is strictly positive. Hence by the law of large numbers $e^{L_n(\tilde{\boldsymbol{\theta}}_k^{(m)},\tilde{\phi}_k^{(m)})-L_n(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})} \xrightarrow{a.s.} e^{-n\text{KL}(f_t(\mathbf{y}_n|\boldsymbol{\theta}_t,\phi_t),f_k(\mathbf{y}_n|\boldsymbol{\theta}_k,\phi_k))}$ and $\text{BF}_{kt} = O_p(e^{-n})$. In the case where $\boldsymbol{\theta}_t^* = \boldsymbol{\theta}_0 a_n$ then the KL is of order $na_n^2(\boldsymbol{\theta}_{0t}^* - \boldsymbol{\theta}_{0k}^*)'(\boldsymbol{\theta}_{0t}^* - \boldsymbol{\theta}_{0k}^*)$ and hence $\text{BF}_{kt} = O_p(e^{-na_n^2})$.

The proof for the peMOM and piMOM are largely analogous. When $\boldsymbol{\theta}_t^*$ is fixed the marginal likelihood for M_t is $m_t(\mathbf{y}_n) \approx$

$$\begin{aligned} & e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})}\pi(\tilde{\phi}_t^{(1)}) \left|H(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})\right|^{-1/2} Z_1 \prod_i e^{-\tau\tilde{\phi}^{(1)}/(\tilde{\theta}_{ti}^{(1)})^2} \\ & \xrightarrow{P} n^{-p_t/2} e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})} Z_2 \end{aligned} \tag{52}$$

where $Z_1 = O_p(1)$ for the peMOM under any model, whereas for the piMOM $Z_1 = O_p(1)$ under M_t and $Z_1 = o_p(1)$ under any other M_k, and consequently $Z_2 = O_p(1)$. Consider k such that $M_t \subset M_k$, then from Proposition 2(i) for all modes with spurious $\tilde{\boldsymbol{\theta}}_{k2}$ in the same quadrant as $\boldsymbol{\theta}_{k2}^*$ we have $\prod_{i=1}^{p_k} \exp\{-\sqrt{n}\tau\tilde{\phi}^{(1)}/(n^{1/4}\tilde{\theta}_{ki}^{(1)})^2\} = \prod_i \exp\{-\sqrt{n}Z_{3i}\} = e^{-\sqrt{n}Z_4}$, where $Z_4 = O_p(1)$. Thus the Bayes factor

$$\text{BF}_{kt} \xrightarrow{P} \sum_m \frac{e^{L_n(\tilde{\boldsymbol{\theta}}_k^{(m)},\tilde{\phi}_k^{(m)})}}{e^{L_n(\tilde{\boldsymbol{\theta}}_t^{(1)},\tilde{\phi}_t^{(1)})}} \frac{e^{-\sqrt{n}Z_4} n^{-p_k/2}}{n^{-p_t/2} Z_2} = O_p(e^{-\sqrt{n}}). \tag{53}$$

When $\boldsymbol{\theta}_t^* = \boldsymbol{\theta}_0 a_n$ an analogous argument gives that $\text{BF}_{kt} \xrightarrow{P} O_p(e^{-\sqrt{n}})$.

The proof for the $M_t \not\subset M_k$ case proceeds in the same manner as for the pMOM, obtaining that if $\boldsymbol{\theta}_t^*$ is fixed then $\text{BF}_{kt} = O(e^{-n})$ whereas if $\boldsymbol{\theta}_t^* = \boldsymbol{\theta}_0 a_n$ then $\text{BF}_{kt} = O(e^{-na_n^2})$.

2.7. **Proof of Proposition 3, Part (ii).** The idea is to combine the rates for posterior model probabilities stemming from the Bayes factor rates from Part (i) with a Laplace approximation to the posterior mean conditional on a variable being in a model (which is asymptotically valid under W69).

We start by using the Bayes factor rates from Part (i) to derive rates for posterior model probabilities. Consider a model M_k such that $M_t \subset M_k$ and note that

$P(M_k \mid \mathbf{y}_n) < (1 + \mathrm{BF}_{tk} P(M_t)/P(M_k))^{-1}$. Under a pMOM prior

$$P(M_k \mid \mathbf{y}_n) < \frac{1}{1 + O_p(1) n^{\frac{3}{2}(p_k - p_t)} \frac{P(M_t)}{P(M_k)}} =$$

$$\frac{n^{-\frac{3}{2}(p_k-p_t)} \frac{P(M_t)}{P(M_k)}}{n^{-\frac{3}{2}(p_k-p_t)} \frac{P(M_t)}{P(M_k)} + O_p(1)} = n^{-\frac{3}{2}(p_k-p_t)} \frac{P(M_k)}{P(M_t)} O_p(1), \tag{54}$$

where the last equality follows from the assumption that $P(M_k)/P(M_t) = o(n^{\frac{3}{2}(p_k-p_t)})$ and hence the denominator is $O_p(1)$. The same argument applies under a peMOM or piMOM prior, where now $\mathrm{BF}_{kt} = e^{-\sqrt{n}}$ and hence $P(M_k \mid \mathbf{y}_n) < e^{-\sqrt{n}} \frac{P(M_k)}{P(M_t)} O_p(1)$. Finally, for models M_k such that $M_t \not\subset M_k$, from Proposition 3(i)
$P(M_k \mid \mathbf{y}_n) < \left(1 + e^{nO_p(1)} P(M_t)/P(M_k)\right)^{-1} = e^{-nO_p(1)} P(M_k)/P(M_t)$ if $\boldsymbol{\theta}^*$ is fixed, and $e^{-na_n^2 O_p(1)} P(M_k)/P(M_t)$ if $\boldsymbol{\theta}^* = \boldsymbol{\theta}_0 a_n$ with a_n as in C4.

The BMA posterior mean is $E(\theta_i \mid \mathbf{y}_n) =$

$$E(\theta_i \mid M_t, \mathbf{y}_n) P(M_t \mid \mathbf{y}_n) + \sum_{k: M_t \subset M_k} E(\theta_i \mid M_k, \mathbf{y}_n) P(M_k \mid \mathbf{y}_n) +$$

$$\sum_{k: M_t \not\subset M_k} E(\theta_i \mid M_k, \mathbf{y}_n) P(M_k \mid \mathbf{y}_n). \tag{55}$$

Suppose first that $\theta_i^* \neq 0$ is fixed. From Proposition 2(ii), $E(\theta_i \mid M_t, \mathbf{y}_n) = \hat{\theta}_i + O_p(n^{-1})$ for pMOM, peMOM and piMOM, where $\hat{\theta}_i$ is the MLE. Also, $E(\theta_i \mid M_k, \mathbf{y}_n)$ in the second term of (55) is $O_p(1)$ and $P(M_k \mid \mathbf{y}_n)$ is either $O_p(n^{-\frac{3}{2}(p_k-p_t)})$ (pMOM) or $O_p(e^{-\sqrt{n}})$ (peMOM, piMOM). Further, $P(M_k)/P(M_t) = o(n^{p_k-p_t})$ by assumption and hence the whole second term in (55) is $O_p(n^{-1})$. Regarding the third term in (55), $E(\theta_i \mid M_k, \mathbf{y}_n) = O_p(1)$ and $P(M_k \mid \mathbf{y}_n) = O_p(e^{-n})$. Summarizing, when $\theta_i^* \neq 0$ for the pMOM we have that $E(\theta_i \mid \mathbf{y}_n) =$

$$\left(\hat{\theta}_{ti} + O_p(n^{-1})\right) \left(1 + n^{-1} O_p(1)\right)^{-1} + O_p(n^{-1}) + O_p(e^{-n}) = \hat{\theta}_{ti} + O_p(n^{-1}) \tag{56}$$

$$= \theta_i^* + O_p(n^{-1/2}) \tag{57}$$

and for the peMOM or piMOM $E(\theta_i \mid \mathbf{y}_n) =$

$$\left(\hat{\theta}_{ti} + O_p(n^{-1})\right) \left(1 + e^{-\sqrt{n} O_p(1)}\right)^{-1} + O_p(e^{-\sqrt{n}}) + O_p(e^{-n}) = \hat{\theta}_{ti} + O_p(n^{-1}) \tag{58}$$

$$= \theta_i^* + O_p(n^{-1/2}). \tag{59}$$

Next consider that $\theta_i^* = \theta_{0i}^* a_n$ with fixed θ_{0i} and a_n as in C4. Then from Proposition 2(ii) $E(\theta_i \mid M_t, \mathbf{y}_n) = \theta_i^* + O_p(n^{-1/2})$ for pMOM and $\theta_i^* + O_p(n^{-1/4})$ for peMOM and piMOM. The second term in (55) has $E(\theta_i \mid M_k, \mathbf{y}_n) = O_p(a_n)$ and $P(M_k \mid \mathbf{y}_n)$ is either $O_p(n^{-3/2(p_k-p_t)})$ (pMOM) or $O_p(e^{-\sqrt{n}})$ (peMOM,piMOM), hence the whole term is $O_p(a_n n^{-1})$. Regarding the third term, $E(\theta_i \mid M_k, \mathbf{y}_n) = O_p(1)$ and

$P(M_k \mid \mathbf{y}_n) = e^{-na_n^2 O_p(1)}$. Summarizing for vanishing coefficients under a pMOM we obtain $E(\theta_i \mid \mathbf{y}_n) =$

$$(60) \qquad \left(\theta_i^* + O_p(n^{-1/2})\right)\left(1 + n^{-1}O_p(1)\right)^{-1} + O_p(n^{-1}) + O_p(e^{-na_n^2}) = \theta_i^* + O_p(n^{-1/2})$$

since $a_n << n$ by assumption, and for the peMOM or piMOM $E(\theta_i \mid \mathbf{y}_n) =$

$$(61) \qquad \left(\theta_i^* + O_p(n^{-1/4})\right)\left(1 + e^{-\sqrt{n}O_p(1)}\right)^{-1} + O_p(e^{-\sqrt{n}}) + O_p(e^{-na_n^2}) = \theta_i^* + O_p(n^{-1/4})$$

Finally consider the case $\theta_i^* = 0$. Obviously, M_t only includes non-zero coefficients and hence $E(\theta_i \mid M_t, \mathbf{y}_n) = 0$. In the second term of (55), from Proposition 2(ii) we have that $E(\theta_i \mid M_k, \mathbf{y}_n)$ is $O_p(n^{-1/2})$ for pMOM and $O_p(n^{-1/4})$ for peMOM and piMOM. Thus the whole second term is $O_p(n^{-2})\pi_{p_t+1}/P(M_t)$ for pMOM and $O_p(e^{-\sqrt{n}})\pi_{p_t+1}/P(M_t)$ for peMOM and piMOM, where $\pi_{p_t+1} = \max_{k:p_k=p_t+1} P(M_k)$ for $M_t \subset M_k$. As in the $\theta_i^* \neq 0$ case, the third term is $O_p(e^{-n})$. Summarizing, when $\theta_i^* = 0$ we obtain $E(\theta_i \mid \mathbf{y}_n) =$

$$(62) \qquad O_p(n^{-2})\frac{\pi_{p_t+1}}{P(M_t)}$$

and for the peMOM or piMOM $E(\theta_i \mid \mathbf{y}_n) =$

$$(63) \qquad O_p(e^{-\sqrt{n}})\frac{\pi_{p_t+1}}{P(M_t)},$$

as desired. The probabilistic orders of SSE_0 and SSE_1 are straightforward, since a random variable $Z_n = O_p(b_n)$ implies that $Z_n^2 = O_p(b_n^2)$ and $Z_{1n} = O_p(b_n), \ldots Z_{pn} = O_p(b_n)$ imply that $\sum_{i=1}^p Z_{in} = O_p(b_n)$ for finite p.

2.8. **Proof of Proposition 3, Part (iii).** We first determine the probabilistic order of the BMA posterior mean $\bar{\theta}_i = E(\theta_i \mid \mathbf{y}_n, \phi)$ and subsequently proceed to characterize the (asymptotic) frequentist expectation of $\text{SSE}_0 = \sum_{\theta_i^*=0}(\bar{\theta}_i - \theta_i^*)^2$ and $\text{SSE}_1 = \sum_{\theta_i^* \neq 0}(\bar{\theta}_i - \theta_i^*)^2$ under repeated sampling from $\mathbf{y}_n \sim N(X_n\boldsymbol{\theta}^*, \phi I)$. As stated in the proposition conditions we assume $X_n'X_n$ to be orthogonal and ϕ to be known.

2.8.1. *Proposition 3(iii). Probabilistic order of $E(\theta_i \mid \mathbf{y}_n, \phi)$.* The strategy is to find simple expressions for $E(\theta_i \mid \mathbf{y}_n, \phi)$ taking advantage of the fact that both the likelihood and prior factor across $\theta_1, \ldots, \theta_n$ and then use a Laplace expansion as in Part (ii) to derive its probabilistic order for peMOM and piMOM (for pMOM a closed-form expression is available).

We adjust the notation of the previous sections slightly to ease the exposition. Let θ_i for $i = 1, \ldots, p$ (where $p < n$) be the coefficient corresponding to variable i and $\delta_i = \mathrm{I}(\theta_i \neq 0)$ variable inclusion indicators. We aim to characterize $E(\theta_i \mid \mathbf{y}_n) = E(\theta_i \mid \mathbf{y}_n, \delta_i = 1)P(\delta_i = 1 \mid \mathbf{y}_n)$. We first derive $P(\delta_i \mid \mathbf{y}_n)$. Let $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)$ and $\boldsymbol{\delta}_{-i}$ be the result from removing δ_i from $\boldsymbol{\delta}$, and note that $P(\delta_i = 1 \mid \mathbf{y}_n) = \sum_{\boldsymbol{\delta}_{-i}} P(\delta_i = 1 \mid \boldsymbol{\delta}_{-i}, \mathbf{y}_n)P(\boldsymbol{\delta}_{-i} \mid \mathbf{y}_n)$. Because $X_n'X_n$ is orthogonal the likelihood factors across $i = 1, \ldots, p$, and given that the pMOM, peMOM and piMOM priors also factor straightforward algebra shows that

$P(\delta_i = 1 \mid \boldsymbol{\delta}_{-i}, \mathbf{y}_n) =$

$$\frac{\rho_1 m_{-i}(\mathbf{y}_n) \int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i}{\rho_1 m_{-i}(\mathbf{y}_n) \int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i + N(0; m_i, \phi v_i) m_{-i}(\mathbf{y}_n)(1-\rho_1)} =$$

$$\frac{\rho_1 \int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i}{\rho_1 \int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i + N(0; m_i, \phi v_i)(1-\rho_1)} \tag{64}$$

where $m_{-i}(\mathbf{y}_n) =$

$$\prod_{j \neq i, \delta_j = 1} \int d_j(\theta_j, \phi) N(\theta_j; m_j, \phi v_j) d\theta_j \prod_{j \neq i, \delta_j = 0} m_{j0}(\mathbf{y}_n), \tag{65}$$

and $\rho_1 = P(\delta_i = 1 \mid \boldsymbol{\delta}_{-i})$. For the pMOM prior $d_i(\theta_i, \phi) = \theta_i^2/\phi\tau$, $v_i = \tau/(1 + \tau \sum_{l=1}^n x_{il}^2)$ and $m_i = v_i \sum_{l=1}^n x_{il} y_l$. For the peMOM prior $d_i(\theta_i, \phi) = e^{-\tau\phi/\theta_i^2}$ and again $v_i = \tau/(1 + \tau \sum_{l=1}^n x_{il}^2)$ $m_i = v_i \sum_{l=1}^n x_{il} y_l$ and for piMOM $d_i(\theta_i, \phi) = \sqrt{\tau\phi}\theta_i^{-2} e^{-\tau\phi/\theta_i^2}$, $v_i = (\sum_{l=1}^n x_{il}^2)^{-1}$ $m_i = v_i \sum_{l=1}^n x_{il} y_l$. Assumption D2 gives that $an \leq \sum_{l=1}^n x_{il}^2 \leq bn$ for $n > n_0$ and some finite a, b, n_0, hence without loss of generality we assume that $\sum_{l=1}^n x_{il}^2 = n$ (*i.e.* covariates have mean 0 and variance 1), so that $v_i = \tau/(n\tau + 1)$ for pMOM and peMOM and $v_i = 1/n$ for piMOM. From de Finetti's theorem the assumption that $\delta_1, \ldots, \delta_p$ are exchangeable a priori gives

$$P(\delta_i = 1 \mid \boldsymbol{\delta}_{-i}) = \int P(\delta_i = 1 \mid \boldsymbol{\delta}_{-i}, w) P(w) dw = \int P(\delta_i = 1 \mid w) P(w) dw = P(\delta_i = 1),$$

and hence from (64) $P(\delta_i = 1 \mid \boldsymbol{\delta}_{-i}, \mathbf{y}_n) = P(\delta_i = 1 \mid \mathbf{y}_n)$. Thus $P(\delta_i = 1 \mid \mathbf{y}_n) =$

$$\frac{\int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i P(\delta_i = 1)}{\int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i P(\delta_i = 1) + N(0; m_i, \phi v_i) P(\delta_i = 0)}. \tag{66}$$

Following the same argument as in Proposition 3(ii), if $\theta_i^* \neq 0$ is fixed then $P(\delta_i = 1 \mid \mathbf{y}_n) = \left(1 - e^{-nO_p(1)} P(\delta_i = 0)/P(\delta_i = 1)\right)$ under either a pMOM, peMOM or piMOM prior. If $\theta_i^* = \theta_{0i}^* a_n$ with a_n as in C4 then $P(\delta_i = 1 \mid \mathbf{y}_n) = \left(1 - e^{-na_n^2} P(\delta_i = 0)/P(\delta_i = 1)\right)$ for pMOM, peMOM and piMOM. If $\theta_i^* = 0$ then $P(\delta_i = 1 \mid \mathbf{y}_n) = n^{-\frac{3}{2}(p_k - p_t)} P(\delta_i = 1)/P(\delta_i = 0)$ for pMOM and $P(\delta_i = 1 \mid \mathbf{y}_n) = e^{-\sqrt{n}O_p(1)} P(\delta_i = 1)/P(\delta_i = 0)$ for peMOM and piMOM.

We now characterize $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)$. Again, because of orthogonality this posterior mean is the same under any model with $\delta_i = 1$, giving

$$E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = \frac{\int \theta_i d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i}{\int d_i(\theta_i,\phi) N(\theta_i; m_i, \phi v_i) d\theta_i}, \tag{67}$$

As before for the pMOM prior $d_i(\theta_i, \phi) = \theta_i^2/(\phi\tau)$ and hence by using Normal moments of up to order 3 (67) becomes $m_i \left(1 + \frac{2\phi v_i}{m_i^2 + \phi v_i}\right)$, where $v_i = \tau/(n\tau + 1)$ and $m_i = v_i \sum_{i=1}^n x_{ji} y_i$. If $\theta_i^* \neq 0$ is fixed then $n^{-1} m_i^2/v_i \xrightarrow{P} \theta_i^*$ and hence $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = m_i + O_p(n^{-1}) = \theta_i^* + O_p(n^{-1/2})$. If $\theta_i^* = \theta_{0i}^* a_n$ then $n^{-1} a_n^{-2} m_i^2/v_i \xrightarrow{P} \theta_i^*$ and hence

$E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = m_i(1 + O_p(n^{-1}a_n^{-2})) = m_i(1 + o_p(1)) = \theta_i^* + O_p(n^{-1/2})$, since $a_n >> n^{-1/2}$ by assumption. If $\theta_i^* = 0$ then $m_i^2/v_i = O_p(1)$ and hence
$E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = m_i + m_i O_p(1) = O_p(n^{-1/2})$. For the peMOM and piMOM, using a Laplace approximation (Kass et al., 1990) around the two modes as in Proposition 2 gives that if $\theta_i^* \neq 0$ is fixed then $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = \hat{\theta}_i + O_p(n^{-1})$ (where $\hat{\theta}_i$ is the MLE), if $\theta_i^* = \theta_{0i}^* a_n$ then $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = \theta_i^* + O_p(n^{-1/4})$ and if $\theta_i^* = 0$ then $E(\theta_i \mid \delta_i, \mathbf{y}_n, \phi) = O_p(n^{-1/4})$.

Combining the rates derived above if $\theta_i^* \neq 0$ is fixed then
$E(\theta_i \mid \mathbf{y}_n, \phi) = E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)P(\delta_i = 1 \mid \mathbf{y}_n, \phi) = \hat{\theta}_i + O_p(n^{-1})$ for pMOM, peMOM and piMOM. If $\theta_i^* = \theta_{0i}^* a_n$ with a_n as in C4 then
$E(\theta_i \mid \mathbf{y}_n, \phi) = (\theta_i^* + O_p(n^{-1/2}))\left(1 - e^{-na_n^2}P(\delta_i = 1)/P(\delta_i = 0)\right) = \theta_i^* + O_p(n^{-1/2})$ for pMOM and $E(\theta_i \mid \mathbf{y}_n, \phi) = \theta_i^* + O_p(n^{-1/4})$ for peMOM and piMOM. If $\theta_i^* = 0$ then
$E(\theta_i \mid \mathbf{y}_n, \phi) = O_p(n^{-2})P(\delta_i = 1)/P(\delta_i = 0)$ for pMOM and
$E(\theta_i \mid \mathbf{y}_n, \phi) = e^{-\sqrt{n}O_p(1)}P(\delta_i = 1)/P(\delta_i = 0)$ for peMOM and piMOM.

2.8.2. *Proposition 3(iii). Sum of squared errors..* We now characterize $E_{\boldsymbol{\theta}^*}(SSE_0) = \sum_{\theta_i^*=0} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2)$ and $E_{\boldsymbol{\theta}^*}(SSE_1) = \sum_{\theta_i^*\neq 0} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2)$, where the expectation is with respect to the data-generating truth $N(\mathbf{y}_n; X\boldsymbol{\theta}^*, \phi I)$. Although we already characterized the probabilistic order of $(\bar{\theta}_i - \theta_i^*)^2$, unfortunately this does not imply that its expectation $E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2)$ is of the same order. Our strategy is to introduce Lemma 6, which under a stronger condition does bound the order of the expectation, and then show in Lemma 7 that the stronger condition is satisfied by pMOM, peMOM and piMOM. This will imply that $E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) = O(a_n)$ for some sequence a_n satisfying $\lim_{n\to\infty} a_n = 0$. Finally, Lemma 8 will guarantee that $\sum_{\theta_i^*=0} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) = o(p_0 b_n)$ for any b_n satisfying $a_n \ll b_n$, where $p_0 = \sum_{i=1}^p I(\theta_i^* = 0)$ is the number of truly inactive variables, and similarly for $\sum_{\theta_i^*=1} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2)$. We start by stating and proving Lemmas 6, 7 and 8.

Lemma 6. *Let $U_n \in [0,1]$ be a random variable such that $P(U_n > a_n) = O(a_n)$ where $\lim_{n\to\infty} a_n = 0$. Then $E(U_n) = O(a_n)$.*

Proof. of Lemma 6. Denote by P the probability measure of U_n. Then

$$E(U_n) = \int_0^{a_n} U_n dP(U_n) + \int_{a_n}^1 U_n dP(U_n) \leq a_n P(U_n < a_n) + P(U_n > a_n) \leq a_n + P(U_n > a_n) \tag{68}$$

and therefore

$$\limsup_{n\to\infty} \frac{E(U_n)}{a_n} \leq \limsup_{n\to\infty} 1 + \frac{P(U_n > a_n)}{a_n} \leq 1 + c$$

for some $c > 0$, where the last inequality follows from the assumption that
$P(U_n > a_n) = O(a_n)$. $\square$

Lemma 7. *Let $P(\delta_i = 1 \mid \mathbf{y}_n, \phi)$ where $\delta_i = I(\theta_i \neq 0)$ be the posterior inclusion probability under the sampling model $\mathbf{y}_n \sim N(X_n\boldsymbol{\theta}, \phi I)$ with diagonal $X_n'X_n$ and known ϕ satisfying the conditions in Proposition 3(iii). Suppose that conditional on $(\delta_1, \ldots, \delta_p)$ the prior on the non-zero elements in $\boldsymbol{\theta}$ has independent components.*

Denote by $P_{\boldsymbol{\theta}^*}(\cdot)$ the probability under the true data-generating distribution $\mathbf{y}_n \sim N(X_n\boldsymbol{\theta}^*, \phi I)$, and suppose that $\theta_i^* = 0$. Then if the prior on θ_i is a pMOM prior,

$$P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi) > \frac{r}{n^{3/2-\epsilon}}\right) = O\left(\exp\{-n^{\epsilon}\}n^{-\epsilon/2}\right) \tag{69}$$

for any fixed $\epsilon \in (0, 3/2)$. Further, if the prior on θ_i is a peMOM or piMOM prior then

$$P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi) > re^{-n^{1/2-\epsilon}}\right) = O\left(e^{-n^{1/2}}\right) \tag{70}$$

for any fixed $\epsilon \in (0, 1/2)$.

Now suppose that $\theta_i^* \neq 0$. If the prior on θ_i is a pMOM, peMOM or piMOM prior then

$$P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 0 \mid \mathbf{y}_n, \phi) > \frac{1}{r}e^{-n^{\epsilon}}\right) = O\left(e^{-n(\theta_i^*)^2}\right) \tag{71}$$

for any $\epsilon \in (0, 1)$ if θ_i^* is fixed. If $\theta_i^* = \theta_{0i}^* a_n$ where $\lim_{n\to\infty} a_n = 0$ with $a_n \gg n^{-1/2}$ (pMOM) or $a_n \gg n^{-1/4}$ (peMOM, piMOM) then (71) holds for any ϵ satisfying $n^{1/2}a_n \gg n^{\epsilon/2}$ (in particular, if $a_n = n^{-\alpha}$ then $\epsilon < 1 - 2\alpha$).

Proof. of Lemma 7. Denote by $\mathbf{x}_i = (x_{1i}, \ldots, x_{ni})'$ the i^{th} column in X_n, $m_i = v_i\mathbf{x}_i'\mathbf{y}_n$, $v_i = \tau/(1 + \tau\mathbf{x}_i'\mathbf{x}_i)$ and note that under the eigenvalue conditions D2 we have $v_i \asymp n^{-1}$. We start by proving the result when $\theta_i^* = 0$. Consider first the pMOM prior, under which $P(\delta_i = 1 \mid \mathbf{y}_n) = \left(1 + r^{-1}\mathrm{BF}_{01}(\mathbf{y}_n)\right)^{-1}$, where

$$\mathrm{BF}_{01}(\mathbf{y}_n) = \frac{\tau}{\sqrt{2\pi\phi}}\frac{1}{v_i^{3/2}}e^{-\frac{m_i^2}{2v_i\phi}}\frac{1}{(1 + \frac{m_i^2}{v_i\phi})}. \tag{72}$$

Trivially,

$$P_{\theta^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n) > \frac{r}{n^{3/2-\epsilon}}\right) = P_{\theta^*}\left(\frac{1}{r}\mathrm{BF}_{01}(\mathbf{y}_n) + 1 < \frac{n^{3/2-\epsilon}}{r}\right) \tag{73}$$

$$\leq P_{\theta^*}\left(\mathrm{BF}_{01}(\mathbf{y}_n) < n^{3/2-\epsilon}\right). \tag{74}$$

Now, we are looking for $P_{\theta^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n) > t\right)$ where $t \to 0$, which from monotonicity of $\mathrm{BF}_{01}(\mathbf{y}_n)$ as a function of m_i implies that $m_i \to 0$. For any fixed $c_1 \in (0, 1)$ there is a small enough m_i such that $e^{-m_i^2/(2v_i\phi)} > c_1$, implying that (74)

$$\leq P_{\theta^*}\left(\frac{\tau}{\sqrt{2\pi\phi}}\frac{1}{v_i^{3/2}}c_1\frac{1}{(1 + \frac{m_i^2}{v_i\phi})} < n^{3/2-\epsilon}\right) = P_{\theta^*}\left(\frac{m_i^2}{v_i\phi} > c_2 n^{\epsilon}\right) \tag{75}$$

where $c_1, c_2 > 0$ are constants and we used the fact that $v_i \asymp n^{-1}$. If truly $\theta_i^* = 0$ then $\frac{m_i^2}{v_i\phi} \sim N\left(0, \frac{\tau\mathbf{x}_i'\mathbf{x}_i}{1+\tau\mathbf{x}_i'\mathbf{x}_i}\right)$, from which (75) becomes $2\Phi\left(-\sqrt{c_3}n^{\epsilon/2}\right)$ with bounded $c_3 > 0$.

Using the bound that for any $z < 0$ the Normal cdf satisfies $\Phi(z) \leq e^{-z^2/2}\frac{1}{\sqrt{2\pi}|z|}$, we obtain that

$$(76) \qquad P_{\theta^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n) > \frac{r}{n^{3/2-\epsilon}}\right) \leq c_5 \frac{1}{n^{\epsilon/2}} \exp\left\{-n^{\epsilon} c_4/2\right\} = O\left(n^{-\epsilon/2} e^{-n^{\epsilon}}\right),$$

as desired.

Consider now the peMOM prior. Rossell et al. (2013) found that the integrated likelihood under $\delta_i = 1$ is

$$(77) \qquad \frac{e^{\sqrt{2}-\frac{1}{2\phi}\mathbf{y}_n'\mathbf{y}_n}(\phi v_i)^{\frac{1}{4}}}{(2\pi\phi)^{n/2}\tau^{\frac{1}{4}}} 2^{\frac{3}{4}} \sum_{l=0}^{\infty} \frac{\left(\sqrt{\tau} m_i^2/(2\phi v_i)^{\frac{3}{2}}\right)^l}{\Gamma(l+1/2)l!} K_{l+\frac{1}{2}}\left(\sqrt{\frac{2\tau}{\phi v_i}}\right),$$

where $K(\cdot)$ is the modified Bessel function of the second kind. Thus the Bayes factor is $\mathrm{BF}_{01}(\mathbf{y}_n) =$

$$(78) \qquad \frac{\tau^{\frac{1}{4}}}{e^{\sqrt{2}}(\phi v_i)^{\frac{1}{4}} 2^{\frac{3}{4}}} \left(\sum_{l=0}^{\infty} \frac{\left(\sqrt{\tau} m_i^2/(2\phi v_i)^{\frac{3}{2}}\right)^l}{\Gamma(l+1/2)l!} K_{l+\frac{1}{2}}\left(\sqrt{\frac{2\tau}{\phi v_i}}\right)\right)^{-1}.$$

Let $\epsilon \in (0, 1/2)$ be an arbitrary fixed constant, then

$$P_{\theta^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n) > r \exp\{-n^{1/2-\epsilon}\}\right) \leq P_{\theta^*}\left(\mathrm{BF}_{01}(\mathbf{y}_n) < \exp\{n^{1/2-\epsilon}\}\right) =$$

$$P_{\theta^*}\left(\exp\{n^{1/2-\epsilon}\} \sum_{l=0}^{\infty} \frac{\left(\sqrt{\tau} m_i^2/(2\phi v_i)^{\frac{3}{2}}\right)^l}{\Gamma(l+1/2)l!} K_{l+\frac{1}{2}}\left(\sqrt{\frac{2\tau}{\phi v_i}}\right) > \frac{\tau^{\frac{1}{4}} e^{\sqrt{2}}}{2^{\frac{3}{4}}(\phi v_i)^{\frac{1}{4}}}\right) \leq$$

$$(79) \qquad P_{\theta^*}\left(c_1 \exp\left\{n^{1/2-\epsilon} - \sqrt{\frac{2\tau}{v_i\phi}}\right\} \sum_{l=0}^{\infty} \frac{\left(m_i^2/(4\phi v_i)\right)^l}{l!} > \frac{\tau^{\frac{1}{4}} e^{\sqrt{2}}}{2^{\frac{3}{4}}(\phi v_i)^{\frac{1}{4}}}\right),$$

where the last inequality follows from the modified Bessel function of the second kind bound $K_{l+1/2}(z)/\Gamma(l+1/2) \leq c_1 e^{-z}(1/z)^l$ for all $l \geq 1$, $z \geq 5$ and some $c_1 > 0$. Now, the sum on the right hand side of (79) corresponds to the series expansion of the exponential function, giving

$$P_{\theta^*}\left(\exp\left\{\frac{m_i^2}{4\phi v_i}\right\} > \exp\left\{\sqrt{\frac{2\tau}{v_i\phi}} - n^{1/2-\epsilon}\right\} \frac{\tau^{\frac{1}{4}} e^{\sqrt{2}}}{c_1 2^{\frac{3}{4}}(\phi v_i)^{\frac{1}{4}}}\right)$$

$$(80) \qquad = P_{\theta^*}\left(\frac{m_i^2}{\phi v_i} > \sqrt{\frac{2\tau}{v_i\phi}} - n^{1/2-\epsilon} - \frac{1}{4}\log(\phi v_i) + \frac{1}{4}\log(\tau) + c_2\right)$$

where $c_2 = \sqrt{2} - \log(c_1) - \frac{3}{4}\log(2)$. Analogously to the MOM case when truly $\theta_i^* = 0$ then $\frac{m_i^2}{v_i\phi} \sim N(0, \frac{\tau\mathbf{x}_i'\mathbf{x}_i}{1+\tau\mathbf{x}_i'\mathbf{x}_i})$ and since for $z < 0$ the Normal cdf satisfies $\Phi(z) \leq e^{-z^2/2}\frac{1}{\sqrt{2\pi}|z|}$ we obtain that (80) is

$$\leq 2\exp\left\{-2\left(\sqrt{\frac{2\tau}{v_i\phi}} - n^{\frac{1}{2}-\epsilon}\right) - \frac{1}{2}\log(\phi v_i/\tau) - c_3\right\} \tag{81}$$

for fixed $c_3 \in \mathbb{R}$. Since $\sqrt{1/v_i} \asymp n^{1/2}$ we obtain that (81) is $O\left(\exp\{-n^{1/2}\}\right)$, as desired.

The proof for the piMOM is analogous. Intuitively, its penalty term behaves in the same exponential fashion as the eMOM's as $\theta_i \to 0$, hence the same limiting rates apply. More formally, from (66) we have that the iMOM's $P(\delta_i = 1 \mid \mathbf{y}_n, \phi)$ only differs from the eMOM's due to its marginal likelihood under $\delta_i = 1$ depending on $\int \frac{1}{\theta_i^2} e^{-\tau\phi/\theta_i^2} N(\theta_i; m_i, \phi v_i) d\theta_i$ rather than $\int e^{-\tau\phi/\theta_i^2} N(\theta_i; m_i, \phi v_i) d\theta_i$. Briefly, using integration by parts shows and the Normal cdf bound $\Phi(z) \leq e^{-z^2/2}\frac{1}{\sqrt{2\pi}|z|}$ it can be shown that

$$\int \frac{1}{\theta_i^2} e^{-\tau\phi/\theta_i^2} N(\theta_i; m_i, \phi v_i) d\theta_i \leq \frac{1}{2\tau\phi^2 v}\left(\int e^{-\tau\phi/\theta_i^2} N(\theta_i; m_i, \phi v_i) d\theta_i + g(m, v)\right) \tag{82}$$

for a certain function $g(m, v) = O(e^{-n})$ for any $m < 1/2$. From (82) it can then be shown that the term $g(m, v)$ can be ignored (as it has smaller order than $e^{-n^{1/2-\epsilon}}$), hence a derivation analogous to that for the eMOM shows that $P_{\theta^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n) > e^{-n^{1/2-\epsilon}}\right) = O\left(\exp\{-n^{1/2}\}\right)$.

We now prove the result for the case $\theta_i^* \neq 0$. Clearly,

$$P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 0 \mid \mathbf{y}_n, \phi) > \frac{1}{r}e^{-n^\epsilon}\right) = P_{\boldsymbol{\theta}^*}\left(r\mathrm{BF}_{10}(\mathbf{y}_n) + 1 < re^{n^\epsilon}\right) \leq P_{\boldsymbol{\theta}^*}\left(\mathrm{BF}_{10}(\mathbf{y}_n) < e^{n^\epsilon}\right). \tag{83}$$

For the pMOM prior from (72) we obtain that (83) is equal to

$$\begin{aligned} P_{\boldsymbol{\theta}^*}\left(\frac{\sqrt{2\pi\phi}}{\tau} v_i^{\frac{3}{2}} e^{\frac{m_i^2}{2v_i\phi}}\left(1 + \frac{m_i^2}{v_i\phi}\right) < e^{n^\epsilon}\right) &\leq P_{\boldsymbol{\theta}^*}\left(\frac{\sqrt{2\pi\phi}}{\tau} v_i^{\frac{3}{2}} e^{\frac{m_i^2}{2v_i\phi}} < e^{n^\epsilon}\right) \\ &\leq P_{\boldsymbol{\theta}^*}\left(\frac{m_i^2}{v_i\phi} < 2n^\epsilon - \frac{3}{2}\log(v_i) - \log\left(\frac{\tau}{\sqrt{2\pi\phi}}\right)\right), \end{aligned} \tag{84}$$

where for simplicity we may ignore the terms $-\frac{3}{2}\log(v_i) - \log\left(\frac{\tau}{\sqrt{2\pi\phi}}\right)$, as they are of smaller order than n^ϵ since $v_i \asymp 1/n$. From basic sampling theory we know that $m_i = v_i\mathbf{x}_i'\mathbf{x}_i\hat{\theta}_i \sim N(v_i\mathbf{x}_i'\mathbf{x}_i\theta_i^*, \phi\mathbf{x}_i'\mathbf{x}_i v_i^2)$ where $\hat{\theta}_i$ is the least squares estimator, implying that $m_i/(\sqrt{\phi\mathbf{x}_i'\mathbf{x}_i} v_i) \sim N\left(\theta_i^*\sqrt{\mathbf{x}_i'\mathbf{x}_i/\phi}, 1\right)$ and thus (84) is

$$\leq \Phi\left(\sqrt{\frac{2}{\mathbf{x}_i'\mathbf{x}_i v_i}n^\epsilon} - \sqrt{\frac{\mathbf{x}_i'\mathbf{x}_i}{\phi}}\theta_i^*\right) \leq \exp\left\{-\frac{1}{2}\left(\sqrt{c_1 n^\epsilon} - \sqrt{\frac{\mathbf{x}_i'\mathbf{x}_i}{\phi}}\theta_i^*\right)^2\right\}, \tag{85}$$

where the right hand side follows from the fact that $\mathbf{x}_i'\mathbf{x}_i v_i = \tau\mathbf{x}_i'\mathbf{x}_i/(1+\tau\mathbf{x}_i'\mathbf{x}_i)$ is bounded from below by eigenvalue assumption D2 and the Normal cdf inequality

$\Phi(z) \leq e^{-\frac{1}{2}z^2}/(\sqrt{2\pi}|z|)$. Given that $\mathbf{x}_i'\mathbf{x}_i \asymp n$ we obtain that if θ_i^* is fixed then $\frac{\mathbf{x}_i'\mathbf{x}_i}{\phi}\theta_i^* \gg n^\epsilon$ for any $\epsilon \in [0,1)$, so that (85) is $O(e^{-n(\theta_i^*)^2})$. If $\theta_i^* = \theta_{0i}^* a_n$ with fixed θ_{0i}^* and $a_n \gg n^{-\frac{1}{2}}$ then for any $\epsilon \geq 0$ satisfying $n^{\frac{1}{2}}a_n \gg n^{\epsilon/2}$ we again obtain that (85) is $O(e^{-n(\theta_i^*)^2}) = O(e^{-na_n^2})$, as we wished to prove.

Consider now the peMOM prior, again under the case $\theta_i^* \neq 0$. The modified Bessel function of the second kind satisfies $K_{l+\frac{1}{2}}(z)/\Gamma(l+1/2) \geq c_1 e^{-z^{1+\alpha}}/z^{\frac{l}{2}}$ for any $\alpha > 0$, all $l \geq 1$ and $z \in \mathbb{R}^+$, and some fixed finite $c_1 > 0$, hence from (78) we obtain that (83) is

$$\leq P_{\boldsymbol{\theta}^*}\left(\frac{e^{\sqrt{2}}(\phi v_i 2^3)^{\frac{1}{4}}}{\tau^{\frac{1}{4}}} e^{-\left(\frac{2\tau}{\phi v_i}\right)^{\frac{1+\alpha}{2}}} \sum_{l=0}^{\infty} \frac{c_1}{l!}\left(\frac{m_i^2}{4\phi v_i}\right)^l < e^{n^\epsilon}\right) = P_{\boldsymbol{\theta}^*}\left(c_2 e^{-\left(\frac{2\tau}{\phi v_i}\right)^{\frac{1+\alpha}{2}} + \frac{m_i^2}{4\phi v_i}} < e^{n^\epsilon}\right)$$

$$\leq P_{\boldsymbol{\theta}^*}\left(\frac{m_i^2}{\phi v_i} < n^\epsilon + c_3 n^{\frac{(1+\alpha)}{2}}\right) \leq \exp\left\{-\frac{1}{2}\left(\sqrt{n^{\tilde{\epsilon}}} - \sqrt{\frac{\mathbf{x}_i'\mathbf{x}_i}{\phi}}\theta_i^*\right)^2\right\} \tag{86}$$

where $c_2 = c_1(\phi v_i)^{\frac{1}{4}} e^{\sqrt{2}} 2^{\frac{3}{4}}/\tau^{\frac{1}{4}}$ is of smaller order than e^{n^ϵ} and for simplicity is dropped from the second line of (86), $c_3 = 2\tau/(\phi v_i n)$ is bounded since $v_i \asymp n$ by assumption, $\tilde{\epsilon} = \max\{\epsilon, \frac{1+\alpha}{2}\}$, and the last inequality was obtained from the sampling distribution of $m_i^2/(\phi v_i)$ and the Normal cdf bound as in (84). If $\theta_i^* \neq 0$ is fixed then (86) is $O(-n(\theta_i^*)^2)$ for any $\epsilon \in (0,1)$. If $\theta_i^* = \theta_{0i}^* a_n$ with fixed θ_{0i}^* and $a_n \gg n^{-\frac{1}{4}}$ then $n^{\frac{1}{2}}a_n \gg n^{\tilde{\epsilon}/2}$ for small enough α (recall that $\alpha > 0$ is an arbitrary positive constant used to bound the modified Bessel function), again giving that (86) is $O(-n(\theta_i^*)^2)$ for any $\epsilon \in (0,1)$, as desired.

To finalize the proof consider the piMOM prior when $\theta_i^* \neq 0$, for which we obtain

$$P_{\boldsymbol{\theta}^*}\left(\mathrm{BF}_{10}(\mathbf{y}_n) < e^{n^\epsilon}\right) = P_{\boldsymbol{\theta}^*}\left(\frac{\int \frac{\sqrt{\tau\phi}}{\theta_i^2} e^{-\tau\phi/\theta_i^2} N(\theta_i; m_i, v_i\phi) d\theta_i}{(\phi v_i)^{-\frac{1}{2}} e^{-\frac{m_i^2}{2\phi v_i}}} < e^{n^\epsilon}\right). \tag{87}$$

The strategy will be to lower bound the integral in (87). Analogously to the proof in Supplementary Section 4 for any fixed τ it is possible to find a fixed small enough $\tau' > 0$ such that $\tilde{d}(\theta_i, \phi) = e^{-\frac{\tau\phi}{\theta_i^2}}/(\theta_i^2 N(\theta_i; 0, \tau'\phi)$ is convex in θ_i. Hence (87) becomes

$$P_{\boldsymbol{\theta}^*}\left((\phi v_i)^{\frac{1}{2}} e^{\frac{m_i^2}{2\phi v_i}} \int \tilde{d}(\theta_i, \phi) N(\theta_i; 0, \tau'\phi) N(\theta_i; m_i, v_i\phi) d\theta_i < e^{n^\epsilon}\right)$$

$$\leq P_{\boldsymbol{\theta}^*}\left((\phi v_i)^{\frac{1}{2}} \frac{\sqrt{\tau\phi}}{m_i^2} e^{\frac{m_i^2}{2\phi}(\frac{1}{v_i} + \frac{1}{\tau'})} \tilde{d}(\tilde{m}_i, \phi) < e^{n^\epsilon}\right), \tag{88}$$

where the right hand side follows from Jensen's inequality, straightforward integration and setting $\frac{1}{\tilde{v}_i} = \frac{1}{v_i} + \frac{1}{\tau'}$, $\tilde{m}_i = m_i \frac{\tilde{v}_i}{v_i}$. Since $v_i \asymp 1/n$, for any fixed m_i as $n \to \infty$ we obtain that (88) converges to

$$P_{\boldsymbol{\theta}^*}\left((\phi v_i)^{\frac{1}{2}} \frac{\sqrt{\tau\phi}}{m_i^2} e^{\frac{m_i^2}{2\phi v_i}} < e^{n^\epsilon}\right), \tag{89}$$

$\square$

which following the same argument as in (86) is $O(e^{-n(\theta_i^*)^2})$ when θ_i^* is fixed for any $\epsilon \in (0,1)$. If $\theta_i^* = \theta_{0i}^* a_n$ with fixed θ_{0i}^* and $a_n \gg n^{-\frac{1}{4}}$ then (89) is $O(-n(\theta_i^*)^2)$ for any $\epsilon \in (0,1)$, as desired.

Lemma 8. *Let $b_{in} = o(a_n)$ for $i = 1, \ldots, q_n$ be a set of sequences, where q_n may grow with n, and $a_n > 0$. Then $\sum_{i=1}^{q_n} b_{in} = O(q_n a_n)$ as $n \to \infty$.*

Proof. of Lemma 8. We need to show that

$$\lim_{n\to\infty} \frac{\sum_{i=1}^{q_n} b_{in}}{q_n a_n} = \lim_{n\to\infty} \sum_{i=1}^{q_n} \frac{b_{in}}{a_n} \frac{1}{q_n} = \lim_{n\to\infty} \sum_{i=1}^{q_n} r_{in} < \infty,$$

where $r_{in} = b_{in}/(a_n q_n)$. To see that the series converges we shall use the ratio test. Redefining the series index to $m = q_n$ we want to prove that $\lim\limits_{m\to\infty} \sum_{i=1}^{m} r_{im} < \infty$, where we note that the terms of the series change with m. The increase in the series between m and $m+1$ is

$$\begin{aligned} &\sum_{i=1}^{m+1} r_{i,m+1} - \sum_{i=1}^{m} r_{im} = r_{m+1,m+1} + r_{m,m+1} - r_{mm} + \sum_{i=1}^{m-1} (r_{i,m+1} - r_{im}) \\ &< r_{m+1,m+1} + r_{m,m+1} - r_{mm} = \frac{b_{m+1,m+1}}{c_{m+1}(m+1)} + \frac{b_{m,m+1}}{c_{m+1}(m+1)} - \frac{b_{mm}}{c_m m}, \end{aligned} \tag{90}$$

the inequality following from the fact that r_{im} decrease with m. The assumption that $b_{im} = o(c_m)$ implies that the three terms on the right hand side of (90) go to 0 faster than $1/m$, *i.e.* the series increases at a rate strictly slower than $1/m$ and hence by the ratio test $\lim\limits_{m\to\infty} r_{im} < \infty$. $\square$

Now that we stated Lemmas 6-8 we proceed to characterize $E_{\theta^*}((\bar{\theta}_i - \theta_i^*)^2)$, which depends only on $\boldsymbol{\theta}^*$ through θ_i^* given that as seen earlier the posterior distribution of $\boldsymbol{\theta}$ under pMOM, peMOM or piMOM has independent components. Consider first the case $\theta_i^* = 0$. Clearly,

$$\begin{aligned} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) &= E_{\boldsymbol{\theta}^*}(\bar{\theta}_i^2) = E_{\boldsymbol{\theta}^*}\left(E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)^2 P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^2\right) \\ &\leq \mathrm{Var}_{\boldsymbol{\theta}^*}^{1/2}\left(E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)^2\right) \mathrm{Var}_{\boldsymbol{\theta}^*}^{1/2}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^2\right) \\ &\leq \mathrm{Var}_{\boldsymbol{\theta}^*}^{1/2}\left(E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)^2\right) E_{\boldsymbol{\theta}^*}^{1/2}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^4\right) \end{aligned} \tag{91}$$

the first inequality given by the Cauchy-Schwartz inequality. Under a pMOM prior $E(\theta_i \mid \mathbf{y}_n, \delta_i = 1, \phi) = m_i(1 + 2/(1 + m_i^2(\phi v_i)^{-1}) \leq 3m_i$ where recall that $m_i = v_i \mathbf{x}_i' \mathbf{y}_n$, hence from the eigenvalue conditions D2 we obtain $\mathrm{Var}_{\boldsymbol{\theta}^*}^{1/2}(E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)^2) \leq 3\mathrm{Var}_{\boldsymbol{\theta}^*}^{\frac{1}{2}}(m_i^2) \asymp 1/n$. Regarding the second term, Lemma 7 gives $P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^4 > r^4/n^{6-\epsilon}\right) = P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi) > r n^{-\frac{3}{2}+\tilde{\epsilon}}\right) = O\left(e^{-n^{\tilde{\epsilon}}} n^{-\tilde{\epsilon}/2}\right)$

for any $\tilde{\epsilon} = \epsilon/4 \in (0, \frac{3}{2})$. Thus from Lemma 6 $E_{\boldsymbol{\theta}^*}(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^4) = O\left(r^4/n^{6-\epsilon}\right)$ as long as $e^{-n^{\tilde{\epsilon}}} \ll r^4$, and hence (91) is $O\left(r^2/n^{4-\epsilon}\right)$. Noting that $\bar{\theta}_i^2$ is strictly decreasing with r, (91) is $O\left(r^2/n^{4-\epsilon}\right)$ also when $e^{-n^{\tilde{\epsilon}}} \succeq r^4$. Finally, by Lemma 8 we have $\sum_{\theta_i^*=0} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) = O(p_0 r^2/n^{4-\epsilon'})$ for any $\epsilon' > \epsilon$, as we wished to prove.

For the peMOM and piMOM we first consider the term $\mathrm{Var}_{\boldsymbol{\theta}^*}\left((E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi))^2\right)$ in (91). Lemma 9 (below) gives that $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) = m_i + \sqrt{v_i\phi}h(m_i, v_i)$ where $\mathrm{sign}(m_i) = \mathrm{sign}(h(m_i, v_i))$ and $E_{\boldsymbol{\theta}^*}(h^2(m_i, v_i)) < \infty$. Thus $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)^2 = m_i^2 + v_i\phi h^2(m_i, v_i) - 2m_i\sqrt{v_i\phi}h(m_i, v_i) \leq m_i^2 + v_i\phi h^2(m_i, v_i)$ and consequently $\mathrm{Var}_{\boldsymbol{\theta}^*}(E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)^2) \leq E_{\boldsymbol{\theta}^*}(m_i^2) + O(1/n)$, which is $O(1/n)$ when $\theta_i^* = 0$. We next consider the term $E_{\boldsymbol{\theta}^*}(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^4)$ in (91) and show that it is $O\left(r^4 e^{-n^{1/2-\epsilon}}\right)$. Lemma 7 gives that $P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^4 > r^4 e^{n^{-1/2-\epsilon}}\right) = P_{\boldsymbol{\theta}^*}\left(P(\delta_i = 1 \mid \mathbf{y}_n, \phi) > r e^{\frac{1}{4}n^{-1/2-\epsilon}}\right) = O\left(e^{-n^{1/2}}\right)$, which is $O(r^4 e^{-n^{1/2}})$ as long as $r \gg e^{-n^{1/2}}$. Hence from Lemma 6 we obtain $E_{\boldsymbol{\theta}^*}(P(\delta_i = 1 \mid \mathbf{y}_n, \phi)^4) = O\left(r^4 e^{-n^{1/2-\epsilon}}\right)$ as long as $r \gg e^{-n^{1/2}}$ and thus (91) is $O\left(r^2 e^{-n^{1/2-\epsilon}}\right)$. Given that $\bar{\theta}_i$ is strictly decreasing with r, (91) is $O\left(r^2 e^{-n^{1/2-\epsilon}}\right)$ also when $r \preceq e^{-n^{1/2}}$. Finally, by Lemma 8 we obtain $\sum_{\theta_i^*=0} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) = O(p_0 r^2 e^{-n^{1/2-\epsilon}})$ as we wished to prove.

Now consider the case $\theta_i^* \neq 0$. To show that $E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) = O(1/n)$ we shall decompose it in three terms and show that each of them is $O(1/n)$. By the triangle inequality $E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) \leq$

$$
\begin{aligned}
&E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi))^2 + (E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) - m_i)^2) + E_{\boldsymbol{\theta}^*}((m_i - \theta_i^*)^2) = \\
&- E_{\boldsymbol{\theta}^*}\left(E^2(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)P^2(\delta_i = 0 \mid \mathbf{y}_n, \phi)\right) + E_{\boldsymbol{\theta}^*}((E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) - m_i)^2) \\
&+ E_{\boldsymbol{\theta}^*}((m_i - \theta_i^*)^2)
\end{aligned}
$$

From standard theory and eigenvalue condition D2 the third term $E_{\boldsymbol{\theta}^*}((m_i - \theta_i^*)^2) \asymp 1/n$. Regarding the second term, for peMOM and piMOM Lemma 9 below gives that $E_{\boldsymbol{\theta}^*}((E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) - m_i)^2) = O(1/n)$. For the pMOM, easy algebra shows that

$$
(92) \qquad (E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) - m_i)^2 = \phi v_i \left(\frac{m_i/\sqrt{\phi v_i}}{1 + m_i^2/(\phi v_i)}\right)^2 \leq \frac{\phi v_i}{4} \asymp 1/n,
$$

given that $|z|/(1+z^2) \leq 1/2$ for all $z \in \mathbb{R}$. Therefore all that remains is showing that the first term on the right hand side of (92) is $O(1/n)$. For the pMOM this follows from Lemmas 6-7 and the Cachy-Schwarz inequality, since $E_{\boldsymbol{\theta}^*}^{\frac{1}{2}}\left(E^4(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)\right) \leq E_{\boldsymbol{\theta}^*}^{\frac{1}{2}}(3^4 m_i^4)$, which is a finite constant. For the peMOM and piMOM from Lemma 9 $E_{\boldsymbol{\theta}^*}\left(E^2(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)P^2(\delta_i = 0 \mid \mathbf{y}_n, \phi)\right)$
$< E_{\boldsymbol{\theta}^*}(m_i^2 P^2(\delta_i = 0 \mid \mathbf{y}_n, \phi)) + \phi v_i E_{\boldsymbol{\theta}^*}(h^2(m_i, v_i)P^2(\delta_i = 0 \mid \mathbf{y}_n, \phi))$
$< E_{\boldsymbol{\theta}^*}(m_i^2 P^2(\delta_i = 0 \mid \mathbf{y}_n, \phi)) + \phi v_i E_{\boldsymbol{\theta}^*}(h^2(m_i, v_i)) = o(1/n) + O(1/n)$. Lemma 6 coupled with Lemma 7 under the data-generating truth $\theta_i^* \neq 0$ gives that for pMOM, peMOM

and piMOM $E_{\boldsymbol{\theta}^*}\left(P^4(\delta_i=0 \mid \mathbf{y}_n,\phi)\right) = O(e^{-n(\theta_i^*)^2})$ as long as $1/r \ll e^{n^{\epsilon}}$ for some $\epsilon \in (0,1)$, which is $o(1/n)$ and hence (92) is $O(1/n)$. Finally, from Lemma 8 we obtain that $\sum_{\theta_i^*\neq 0} E_{\boldsymbol{\theta}^*}((\bar{\theta}_i - \theta_i^*)^2) = O(p_1/n^{1-\tilde{\epsilon}})$ for any fixed but arbitrarily small $\tilde{\epsilon}>0$, where $p_1 = \sum_{i=1}^{p} \mathrm{I}(\theta_i^* \neq 0)$ is the number of truly active variables.

Lemma 9. *Let $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi)$ be the posterior mean under the sampling model $\mathbf{y}_n \sim N(X_n\boldsymbol{\theta}, \phi I)$ with diagonal $X_n'X_n$, known ϕ as in Proposition 3(iii). Suppose that conditional on the variable inclusion indicators δ_i the prior on the non-zero elements in $\boldsymbol{\theta}$ has independent components and that an eMOM or iMOM prior is specified on θ_i.*

Let $\mathbf{x}_i = (x_{1i},\ldots,x_{ni})'$ be the i^{th} column in X_n, then $E(\theta_i \mid \delta_i=1,\mathbf{y}_n,\phi) = m_i + \sqrt{\phi v_i}h(m_i,v_i)$ where $sign(h(m_i,v_i)) = sign(m_i)$, $m_i = v_i\mathbf{x}_i'\mathbf{y}_n$ and $v_i = \tau/(1+\tau\mathbf{x}_i'\mathbf{x}_i)$ for the peMOM and $v_i = 1/\mathbf{x}_i'\mathbf{x}_i$ for the piMOM. Further, $\lim_{n\to\infty} h(m_i,v_i) = 0$ for any fixed m_i and $E_{\boldsymbol{\theta}^*}(h(m_i,v_i)) < \infty$, implying that $E_{\boldsymbol{\theta}^*}((E(\theta_i \mid \delta_i=1,\mathbf{y}_n,\phi) - m_i)^2) = O(1/n)$ by assumption D2.

Proof. of Lemma 9 From (67) we know that $E(\theta_i \mid \delta_i = 1, \mathbf{y}_n, \phi) =$

$$\frac{\int (z\sqrt{\phi v_i}+m_i)d(z\sqrt{\phi v_i}+m_i,\phi)N(z;0,1)dz}{\int d(z\sqrt{\phi v_i}+m_i,\phi)N(z;0,1)dz} = m_i + \sqrt{\phi v_i}h(m_i,v_i),$$

where

$$h(m_i,v_i) = \frac{\int z d(z\sqrt{\phi v_i}+m_i,\phi)N(z;0,1)dz}{\int d(z\sqrt{\phi v_i}+m_i,\phi)N(z;0,1)dz}. \tag{93}$$

This implies that

$$E_{\boldsymbol{\theta}^*}\left((E(\theta_i \mid \delta_i=1,\mathbf{y}_n,\phi)-m_i)^2\right) = \phi v_i E_{\boldsymbol{\theta}^*}\left(h(m_i,v_i)^2\right). \tag{94}$$

Note that the expectation in (94) is with respect to $m_i \sim N(\theta_i^*(\mathbf{x}_i'\mathbf{x}_i)v_i, \phi v_i)$, where for the piMOM $\theta_i^*(\mathbf{x}_i'\mathbf{x}_i)v_i = \theta_i^*$ and for the peMOM Assumption D2 implies $\lim_{n\to\infty} \theta_i^*(\mathbf{x}_i'\mathbf{x}_i)v_i = \theta_i^*$. For simplicity below we consider the peMOM case, but all arguments extend to the piMOM by noting that this latter prior can also be written in terms of a bounded penalty times a local prior (Lemma 1). Also for simplicity we denote $d(z) = d(z,\phi)$.

We first show that for any fixed m_i we have $\lim_{n\to\infty} h(m_i,v_i) = 0$. Symmetry in the integrand of the numerator of (93) gives that $h(m_i,v_i)=0$ for $m_i=0$, and similarly $h(m_i,v_i) > 0$ for $m_i > 0$ and $h(m_i,v_i) < 0$ for $m_i < 0$. Without loss of generality consider the case $m_i > 0$ (the case $m_i < 0$ is analogous). From (93) we obtain

$$h(m_i,v_i) < \frac{\int_0^\infty z d(z\sqrt{\phi v_i}+m_i)N(z;0,1)dz}{\int_0^\infty d(z\sqrt{\phi v_i}+m_i)N(z;0,1)dz} < \frac{\int_0^\infty zN(z;0,1)dz}{d(m_i)\int_0^\infty N(z;0,1)dz} = \frac{0.1995}{d(m_i)}, \tag{95}$$

where the second inequality is obtained by noting that $d(\cdot)$ is bounded and increasing in $\mathbb{R}^+$. Given that $h(m_i,v_i)$ is bounded by $0.1995/d(m_i)$ we may apply the Dominated

Convergence Theorem to obtain the limit as $n \to \infty$ (equivalently, as $v_i \to 0$)

(96)
$$\lim_{v_i \to 0} h(m_i, v_i) = \int z \lim_{v_i \to 0} \frac{d(z\sqrt{\phi v_i} + m_i) N(z; 0, 1) dz}{\int d(z\sqrt{\phi v_i} + m_i) N(z; 0, 1) dz} = \int z \frac{d(m_i) N(z; 0, 1) dz}{d(m_i)} = 0,$$

where the second equality follows from the fact that the denominator can be bounded by $1/d(m_i)$ and a second application of the Dominated Convergence Theorem. Expression (96) gives that the random variable $h(m_i, v_i)$ converges to 0 for any fixed m_i as $n \to \infty$, however it does not guarantee that $E(h(m_i, v_i)^2)$ is bounded and that therefore (94) is $O(1/n)$.

To characterize (94) we need a sharper upper bound on $h(m_i, v_i)$. Without loss of generality we shall show that $\int_0^\infty h^2(m_i, v_i) N(m_i; \theta_i^* \mathbf{x}_i' \mathbf{x}_i v_i, \phi v_i) dm_i < \infty$, which combined with an analogous argument for the integral over $m_i \in (-\infty, 0)$ gives $\int_{-\infty}^\infty h^2(m_i, v_i) N(m_i; \theta_i^* \mathbf{x}_i' \mathbf{x}_i v_i, \phi v_i) dm_i < \infty$. Briefly, let a be an arbitrary real number satisfying $a\sqrt{\phi v_i} + m_i < 0$, an argument akin to (95) gives

$$h(m_i, v_i) < \frac{\int_0^\infty z N(z; 0, 1) dz}{\int_{-\infty}^a d(z\sqrt{\phi v_i} + m_i) N(z; 0, 1) dz + \int_0^\infty d(z\sqrt{\phi v_i} + m_i) N(z; 0, 1) dz} \tag{97}$$

$$< \frac{0.3989}{d(a\sqrt{\phi v_i} + m_i)\Phi(a) + \frac{1}{2} d(m_i)}, \tag{98}$$

where $\Phi(\cdot)$ is the standard Normal cumulative distribution function. Let $b > 0$ be an arbitrary constant and consider the particular choice $a = -(m_i + b)/\sqrt{\phi v_i}$, which implies that $d(a\sqrt{\phi v_i} + m_i) = d(-b) = d(b)$ by symmetry of $d(\cdot)$. Therefore

$$\int_0^\infty h^2(m_i, v_i) N(\theta_i^* \mathbf{x}_i' \mathbf{x}_i v_i, \phi v_i) dm_i < \int_0^\infty \frac{0.3989^2 N(m_i; \theta_i^* \mathbf{x}_i' \mathbf{x}_i v_i, \phi v_i)}{\left(d(b)\Phi\left(-\frac{m_i + b}{\sqrt{\phi v}}\right) + \frac{1}{2} d(m_i) \right)^2} dm_i \tag{99}$$

$$= \int_0^\infty u(m_i, v_i) dm_i, \tag{100}$$

where $u(m_i, v_i)$ simply denotes the integrand. Now, $\int_0^\infty u(m_i, 1) dm_i$ is finite for any finite θ_i^* and $u(m_i, v_i)$ can be seen to be decreasing as $v_i \to 0$ for any $m_i > 0$, implying that $\int_0^\infty u(m_i, v_i) dm_i < \int_0^\infty u(m_i, 1) dm_i < \infty$ for any $v_i \leq 1$, as desired.

$\square$

2.9. **Proof of Proposition 4.** The goal is to show that for all $\epsilon > 0$ there exists $\eta > 0$ such that $d(\boldsymbol{\theta}) < \eta$ implies $\pi(\boldsymbol{\theta}) < \epsilon$. By construction, the conditional prior density is $\pi(\boldsymbol{\theta} \mid \lambda) = \pi^L(\boldsymbol{\theta}) \mathrm{I}(d(\boldsymbol{\theta}) > \lambda)/h(\lambda)$, where $h(\lambda) = P_u(d(\boldsymbol{\theta}) > \lambda) = \int \pi^L(\boldsymbol{\theta}) \mathrm{I}(d(\boldsymbol{\theta}) > \lambda) d\boldsymbol{\theta}$. Let $\boldsymbol{\theta}$ be a value such that $d(\boldsymbol{\theta}) < \eta$, and express the prior density as

$$\pi(\boldsymbol{\theta}) = \int \pi(\boldsymbol{\theta} \mid \lambda) \pi(\lambda) d\lambda =$$

$$\int_{\lambda \leq \eta} \frac{\pi^L(\boldsymbol{\theta}) \mathrm{I}(d(\boldsymbol{\theta}) > \lambda)}{h(\lambda)} \pi(\lambda) d\lambda + \int_{\lambda > \eta} \frac{\pi^L(\boldsymbol{\theta}) \mathrm{I}(d(\boldsymbol{\theta}) > \lambda)}{h(\lambda)} \pi(\lambda) d\lambda \tag{101}$$

The second term in (101) is 0, as by assumption $d(\boldsymbol{\theta}) < \eta$. Now, consider that for $\lambda \leq \eta$, $h(\lambda) = P_u(d(\boldsymbol{\theta}) > \lambda)$ is minimized at $\lambda = \eta$, and therefore (101) can be bounded by

$$\pi(\boldsymbol{\theta}) \leq \frac{\pi^L(\boldsymbol{\theta}) \int_{\lambda \leq \eta} \mathrm{I}(d(\boldsymbol{\theta}) > \lambda)\pi(\lambda)d\lambda}{h(\eta)} = \frac{\pi^L(\boldsymbol{\theta})P\left(\lambda < \min\{\eta, d(\boldsymbol{\theta})\}\right)}{h(\eta)} \tag{102}$$

Notice that the numerator can be made arbitrarily small by decreasing η, since $\pi^L(\boldsymbol{\theta})$ is bounded around $\boldsymbol{\theta}_0$, by assumption there is no prior mass at $\lambda = 0$ so that the cdf in the numerator converges to 0 as $\eta \to 0$, and that denominator converges to 1 as $\eta \to 0$. That is, it is possible to choose η such that $\pi(\boldsymbol{\theta}) \leq \epsilon$, which gives the result. $\square$

2.10. **Proof of Corollary 5.** Replace $\mathrm{I}(d(\boldsymbol{\theta}) > \lambda)$ by $\prod_{i=1}^p \mathrm{I}(d(\theta_i) > \lambda_i)$ in the proof of Proposition 4. Letting any λ_i go to 0 and applying the same argument delivers the result.

2.11. **Proof of Proposition 6.** We first note that in order for $\pi(\boldsymbol{\theta})$ to be proper the random variable $d(\boldsymbol{\theta})$ must have finite expectation with respect to $\pi^L(\boldsymbol{\theta})$. Now, the marginal prior for $\boldsymbol{\theta}$ is

$$\pi(\boldsymbol{\theta}) = \int \frac{\pi^L(\boldsymbol{\theta})\mathrm{I}(d(\boldsymbol{\theta}) > \lambda)}{P_u(d(\boldsymbol{\theta}) > \lambda)}\pi(\lambda)d\lambda = \pi^L(\boldsymbol{\theta}) \int_0^{d(\boldsymbol{\theta})} \frac{\pi(\lambda)}{h(\lambda)}d\lambda. \tag{103}$$

Suppose we set $\pi(\lambda) \propto h(\lambda)$, which we can do as long as $\int h(\lambda)d\lambda < \infty$. Then $\pi(\boldsymbol{\theta}) \propto \pi^L(\boldsymbol{\theta})d(\boldsymbol{\theta})$, which proves the result. The only step left is to show that indeed $\int h(\lambda)d\lambda < \infty$. In general

$$\int h(\lambda)d\lambda = \int P_u(d(\boldsymbol{\theta}) > \lambda)d\lambda = \int S_{d(\boldsymbol{\theta})}(\lambda)d\lambda, \tag{104}$$

where $S_{d(\boldsymbol{\theta})}(\lambda)$ is the survival function of the positive random variable $d(\boldsymbol{\theta})$ and therefore (104) is equal to its expectation $E_u\left(d(\boldsymbol{\theta})\right)$ with respect to $\pi^L(\boldsymbol{\theta})$, which is finite as discussed at the beginning of the proof. $\square$

2.12. **Proof of Corollary 7.** Analogously to the proof of Proposition 6 the marginal prior for $\boldsymbol{\theta}$ is $\pi(\boldsymbol{\theta}) =$

$$\int \cdots \int \frac{\pi^L(\boldsymbol{\theta}) \prod_{i=1}^p \mathrm{I}(d_i(\theta_i) > \lambda_i)}{P_u\left(d_1(\theta_1) > \lambda_1, \ldots, d_p(\theta_p) > \lambda_p\right)}\pi(\boldsymbol{\lambda})d\lambda_1, \ldots, d\lambda_p =$$

$$\pi^L(\boldsymbol{\theta}) \int_0^{d_1(\theta_1)} \cdots \int_0^{d_p(\theta_p)} \frac{\pi(\boldsymbol{\lambda})}{h(\boldsymbol{\lambda})}d\lambda_1, \ldots, d\lambda_p \propto \pi^L(\boldsymbol{\theta}) \prod_{i=1}^p d_i(\theta_i), \tag{105}$$

as by assumption $\pi(\boldsymbol{\lambda}) \propto h(\boldsymbol{\lambda})$. $\square$

2.13. **Proof of Proposition 8.** By definition, the marginal density $\pi(\boldsymbol{\theta}^{(m)}) =$

$$\pi^L(\boldsymbol{\theta}^{(m)}) \int \frac{\pi(\lambda)}{h(\lambda)} \prod_{i=1}^p \mathrm{I}(d(\theta_i) > \lambda)d\lambda = \pi^L(\boldsymbol{\theta}^{(m)}) \int \mathrm{I}(\lambda < d_{min}(\boldsymbol{\theta}^{(m)}))\frac{\pi(\lambda)}{h(\lambda)}d\lambda =$$

$$\pi^L(\boldsymbol{\theta}^{(m)})P_\lambda(d_{min}(\boldsymbol{\theta}^{(m)})) \int \frac{1}{h(\lambda)}\pi(\lambda \mid \lambda < d_{min}(\boldsymbol{\theta}^{(m)}))d\lambda, \tag{106}$$

where $P_\lambda(d_{min}(\boldsymbol{\theta}^{(m)})) = P(\lambda < d_{min}(\boldsymbol{\theta}^{(m)}))$ is the cdf of λ evaluated at $d_{min}(\boldsymbol{\theta}^{(m)})$. As $d_{min}(\boldsymbol{\theta}^{(m)}) \to 0$ we have that $\pi(\lambda \mid \lambda < d_{min}(\boldsymbol{\theta}^{(m)}))$ converges to a point mass at zero and hence the integral in the right hand side of (106) converges to $1/h(0) = 1$. To finish the proof of (i) notice that $P_\lambda(d_{min}(\boldsymbol{\theta}^{(m)})) = d_{min}(\boldsymbol{\theta}^{(m)})\pi(\lambda^{(m)})$ for some $\lambda^{(m)} \in (0, d_{min}(\boldsymbol{\theta}^{(m)}))$ by the Mean Value Theorem, as long as $P_\lambda(\cdot)$ is differentiable and continuous at 0^+, *i.e.* λ is a continuous random variable. In the particular case $\pi(\lambda) = ch(\lambda)$, note that $h(\cdot)$ is continuous and $h(0) = 1$.

To prove (ii) notice that $P_\lambda(d_{min}(\boldsymbol{\theta}^{(m)})) \to 1$ as $d_{min}(\boldsymbol{\theta}^{(m)}) \to \infty$ and that the integral in the right hand side of (106) is $m(d_{min}(\boldsymbol{\theta}^{(m)})) = E(1/h(\lambda) \mid \lambda < d_{min}(\boldsymbol{\theta}^{(m)}))$, which is increasing with $d_{min}(\boldsymbol{\theta}^{(m)})$ as $h(\lambda)$ is monotone decreasing in λ. Hence, $\lim_{m\to\infty} \pi(\boldsymbol{\theta}^{(m)})/\pi^L(\boldsymbol{\theta}^{(m)}) = \lim_{m\to\infty} m(d_{min}(\boldsymbol{\theta}^{(m)}))$ where $m(d_{min}(\boldsymbol{\theta}^{(m)}))$ increases as $m \to \infty$. Furthermore, if $\int \frac{\pi(\lambda)}{h(\lambda)} < \infty$ the Monotone Converge Theorem applies and $m(d_{min}(\boldsymbol{\theta}^{(m)}))$ converges to a finite constant. $\square$

2.14. **Proof of Corollary 9.** Because $\lambda_1, \ldots, \lambda_p$ have independent marginals, $\pi(\boldsymbol{\theta}^{(m)}) = \pi^L(\boldsymbol{\theta}^{(m)}) \prod_{i=1}^p P_{\lambda_i}(d_i(\theta_i)) \times$

$$\int \cdots \int \frac{1}{h(\boldsymbol{\lambda})} \pi\left(\boldsymbol{\lambda} \mid \lambda_1 < d_1(\theta_1^{(m)}), \ldots, \lambda_p < d_p(\theta_p^{(m)})\right) d\lambda_1 \ldots d\lambda_p =$$

$$\pi^L(\boldsymbol{\theta}^{(m)}) E\left(h(\boldsymbol{\lambda})^{-1} \mid \forall \lambda_i < d_i(\theta_i^{(m)})\right) \prod_{i=1}^p P_{\lambda_i}(d_i(\theta_i^{(m)})), \tag{107}$$

where $h(\boldsymbol{\lambda})$ is a multivariate survival function and decreases as $d_i(\theta_i^{(m)}) \to 0$. Hence as $d_i(\theta_i^{(m)}) \to 0$ $E\left(h(\boldsymbol{\lambda})^{-1} \mid \forall \lambda_i < d_i(\theta_i^{(m)})\right)$ decreases. To find the limit as $d_i(\theta_i^{(m)}) \to 0$ we note that the integral is bounded by the finite integral obtained plugging $d_i(\theta_i^{(m)}) = 1$ into the integrand. Hence, the Dominated Convergence Theorem applies and $\lim_{m\to\infty} E\left(h(\boldsymbol{\lambda})^{-1} \mid \forall \lambda_i < d_i(\theta_i^{(m)})\right) = E\left(h(\mathbf{0})\right) = 1$ and from (107) $\lim_{m\to\infty} \pi(\boldsymbol{\theta}^{(m)}) / \left(\pi^L(\boldsymbol{\theta}^{(m)}) \prod_{i=1}^p P_{\lambda_i}(d_i(\theta_i^{(m)}))\right) = 1$. Since $\lambda_1, \ldots, \lambda_p$ are continuous the Mean Value Theorem applies, so that $P_{\lambda_i}\left(d_i(\theta_i^{(m)})\right) = d_i(\theta_i^{(m)})\pi(\lambda_i^{(m)})$ for some $\lambda_i^{(m)} \in (0, d_i(\theta_i^{(m)}))$. To prove (ii), notice that $P_{\lambda_i}(d_i(\theta_i^{(m)})) \to 1$ as $d_i(\theta_i^{(m)}) \to \infty$ and that $m(\boldsymbol{\theta}^{(m)}) = E\left(h(\boldsymbol{\lambda})^{-1} \mid \forall \lambda_i < d_i(\theta_i^{(m)})\right)$ increases as $d_i(\theta_i^{(m)}) \to \infty$. Hence, $\lim_{m\to\infty} \pi(\boldsymbol{\theta}^{(m)}) / \left(\pi^L(\boldsymbol{\theta}^{(m)}) m(\boldsymbol{\theta}^{(m)})\right) = 1$ where $m(\boldsymbol{\theta}^{(m)})$ increases with $d_i(\theta_i^{(m)})$, which proves (ii). Further, if $E(h(\boldsymbol{\lambda})^{-1}) < \infty$ the Monotone Convergence Theorem applies and $\lim_{m\to\infty} m(\boldsymbol{\theta}^{(m)}) = c$ for finite $c > 0$. $\square$

3. Multivariate Normal sampling under outer rectangular truncation

The goal is to sample $\boldsymbol{\theta} \sim N(\boldsymbol{\mu}, \Sigma)\mathrm{I}(\boldsymbol{\theta} \in T)$ with truncation region $T = \{\boldsymbol{\theta} : \theta_i < l_i \text{ or } \theta_i > u_i, i = 1, \ldots, p\}$. We generalize the Gibbs sampling of Rodriguez-Yam et al. (2004) and importance sampling of Hajivassiliou (1993) and Keane (1993) to the non-convex region T.

Let $D = \text{chol}(\Sigma)$ be the Cholesky decomposition of Σ and $K = D^{-1}$ its inverse, so that $K\Sigma K' = KDD'K' = I$ is the identity matrix, and define $\boldsymbol{\alpha} = K\boldsymbol{\mu}$. The random variable $\mathbf{Z} = K\boldsymbol{\theta}$ follows a $N(\boldsymbol{\alpha}, I)\mathrm{I}(\mathbf{Z} \in S)$ distribution with truncation region S. Since $\boldsymbol{\theta} = K^{-1}\mathbf{Z} = D\mathbf{Z}$, denoting $\mathbf{d}_{i.}$ as the i^{th} row in D we obtain the truncation region $S = \{\mathbf{Z} : \mathbf{d}_{i.}Z \leq l_i \text{ or } \mathbf{d}_{i.}Z \geq u_i, i = 1, \ldots, p\}$.

The full conditionals for Z_i given $Z_{(-i)} = (Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_p)$ needed for Gibbs sampling follow from straightforward algebra. Denote by d_{jk} the (j, k) element in D, then $Z_i \mid Z_{(-i)} \sim N(\alpha_i, 1)$ truncated so that either $d_{ji}Z_i \leq l_j - \sum_{k \neq i} d_{jk}Z_k$ or $d_{ji}Z_i \geq u_j - \sum_{k \neq i} d_{jk}Z_k$ hold simultaneously for $j = 1, \ldots, p$. We now adapt the algorithm to address the fact that this truncation region is non-convex.

The region excluded from sampling can be written as $S_i^c = \bigcup_{j=1}^{p}(a_j, b_j)$, $a_j = (l_j - \sum_{k \neq i} d_{jk}Z_k)/d_{ji}$ when $d_{ji} > 0$ and $a_j = (u_j - \sum_{k \neq i} d_{jk}Z_k)/d_{ji}$ when $d_{ji} < 0$ (analogously for b_j). S_i^c as given is the union of possibly non-disjoint intervals, which complicates sampling. Fortunately, it can be expressed as a union of disjoint intervals $S_i = \bigcup_{j=1}^{K}(\tilde{a}_j, \tilde{b}_j)$ with the following algorithm. Suppose that l_i are sorted increasingly, set $\tilde{l}_1 = l_1$, $\tilde{u}_1 = u_1$ and $K = 1$. For $j = 2, \ldots, p$ repeat the following two steps.

(1) If $l_j > \tilde{u}_K$ set $K = K + 1$, $\tilde{l}_K = l_j$ and $\tilde{u}_K = u_j$, else if $l_j \leq \tilde{u}_K$ and $u_j \geq \tilde{u}_K$ set $\tilde{u}_K = u_j$.
(2) Set $j = j + 1$.

Finally, because $(\tilde{l}_1, \tilde{u}_1), \ldots, (\tilde{l}_K, \tilde{u}_K)$ are disjoint and increasing, we may draw a uniform number u in $(0, 1)$ excluding intervals $(\Phi(\tilde{l}_j), \Phi(\tilde{u}_j))$ and set $Z_i = \Phi^{-1}(u)$, where $\Phi(\cdot)$ is the inverse Normal$(\alpha_i, 1)$ cdf.

4. Monotonicity and inverse of iMOM prior penalty

Consider the product iMOM prior as given in (9). We first study the monotonicity of the penalty $d(\theta_i, \lambda)$, which for simplicity here we denote as $d(\theta)$, and then provide an algorithm to evaluate its inverse function. Equivalently, it is convenient to consider the log-penalty $\log(d(\theta)) =$

$$\frac{1}{2}\left(\log(\tau\tau_N) + 2\log(\phi) + \log(2)\right) - \log\left((\theta - \theta_0)^2\right) - \frac{\tau\phi}{(\theta - \theta_0)^2} + \frac{1}{2\tau_N\phi}(\theta - \theta_0)^2, \tag{108}$$

as its inverse uniquely determines the inverse of $d(\theta)$. Denoting $z = (\theta - \theta_0)^2$, (108) can be written as

$$g(z) = \frac{1}{2}\left(\log(\tau\tau_N) + 2\log(\phi) + \log(2)\right) - \log(z) - \frac{\tau\phi}{z} + \frac{1}{2\tau_N\phi}z. \tag{109}$$

To show the monotonicity of (109) we compute its derivative $g'(z) = -\frac{1}{z} + \frac{\tau\phi}{z^2} + \frac{1}{2\tau_N\phi}$ and show that it is positive for all z. Clearly, both when $z \to 0$ and $z \to \infty$ we have positive $g'(z)$. Hence we just need to see that there is some τ_N for which all roots of $g'(z)$ are imaginary, so that $g'(z) > 0$ for all z. Simple algebra shows that the roots of $g'(z)$ are $z = \tau_N\phi \pm \tau_N\phi\sqrt{1 - \frac{2\tau}{\tau_N}}$, so that for $\tau_N \leq 2\tau$ there are no real roots. Hence, for $\tau_N \leq 2\tau$ $g(z)$ is monotone increasing.

We now provide an algorithm to evaluate the inverse. That is, given a threshold t we seek z_0 such that $g(z_0) = t$. Our strategy is to obtain an initial guess from an approximation to $g(z)$ and then use continuity and monotonicity to bound the desired z_0 and conduct a linear interpolation based search. Inspecting the expression for $g(z)$ in (109) we see that the term $\log(z)$ is dominated by $\tau\phi/z$ when z approaches 0 and by $\frac{z}{2\tau_N\phi}$ when z is large. Hence, we approximate $g(z)$ by dropping the $\log(z)$ term, obtaining

$$g(z) \approx \frac{1}{2}(\log(\tau\tau_N) + 2\log(\phi) + \log(2)) - \frac{\tau\phi}{z} + \frac{1}{2\tau_N\phi}z. \tag{110}$$

Setting (110) equal to t and solving for z gives $z_0 = \tau_N\phi\left(-b + \sqrt{b^2 - 2\frac{\tau}{\tau_N}}\right)$ as an initial guess, where $b = \log(\tau\tau_N) + 2\log(\phi) + \log(2) - t$.

If $g(z_0) < t$ we set a lower bound $z_l = z_0$ and an upper bound z_u obtained by increasing z_0 by a factor of 2 until $g(z_0) > t$. Similarly, if $g(z_0) > t$ we set the upper bound $z_u = z_0$ and find a lower bound by successively dividing z_0 by a factor of 0.5. Once (z_l, z_u) are determined, we use a linear interpolation to update z_0, evaluate $g(z_0)$ and update either z_l or z_u. The process continues until $|g(z_0) - t|$ is below some tolerance (we used 10^{-5}). In our experience the initial guess is often quite good and the algorithm converges in very few iterations.

5. Model search and posterior parameter sampling algorithm

For our examples in Sections 5.2 and 5.3 we first obtained posterior samples from the model space using a modification of the Gibbs sampling algorithm in Johnson and Rossell (2012) and subsequently used Algorithm 2 to obtain the corresponding posterior samples of $\boldsymbol{\theta}_k$ given $\mathbf{y}_n$ and the sampled M_k. Specifically, to obtain a total of B samples we used Algorithm 3.

Algorithm 3. Joint posterior sampling for models and parameters

(1) Draw model realizations $m^{(1)}, \ldots, m^{(B)}$ from the target posterior with probabilities $P(M_k \mid \mathbf{y}_n)$ for $k = 1, \ldots, K$ using the Gibbs sampling algorithm in Johnson and Rossell (2012), Section 3, and starting with the null model with no covariates.

(2) Let $m_1^*, \ldots, m_{\tilde{B}}^*$ be the distinct models visited in Step 1 and $v_1, \ldots, v_{\tilde{B}}$ the corresponding number of visits. For $b = 1, \ldots, \tilde{B}$, obtain v_b samples from $P(\boldsymbol{\theta}_k, \phi_k \mid M_k = m_b^*, \mathbf{y}_n)$ using Algorithm 2 (Section 4.2) with a burnin of $\max\{0.1v_b, 100\}$ samples.

Steps 1 and 2 are implemented in functions `modelSelection` and `rnlp` in R package `mombf`, respectively. In Step 1 for pMOM priors we evaluated $m_k(\mathbf{y}_n)$ exactly using

Expression (6) in Johnson and Rossell (2012) and the recursive formula in Kan (2008) for $g_k(\mathbf{s}_{k,n}) = E\left(\prod_{i \in M_k} \theta_{ki}^{2r}\right)$ where $\boldsymbol{\theta}_k \sim T_\nu(\mathbf{m}_{k,n}, V_{k,n})$, with $\nu = 2rp_k + n + \alpha$ and $V_{k,n} = S_{k,n}\nu/(\lambda + \mathbf{y}_n'\mathbf{y}_n - \mathbf{y}_n' X_{k,n}\mathbf{m}_{k,n}$. For piMOM priors we used Laplace approximations as described in Johnson and Rossell (2012), Section 3.

The Gibbs sampling algorithm in Step 1 was initialized at the null model. Although we set the simulation so that the first $\theta_1 = \ldots = \theta_{p_0} = 0$, which implies that spurious covariates are given the chance to enter the model before truly active covariates, it is important to assess that the algorithm did not get stuck during the exploration. First we note that (Johnson, 2004) used a similar strategy and provided a formal discussion of numerical convergence through a coupling argument. His findings suggest that rapid convergence and high accuracy of Bayesian model selection procedures is often achieved, with potential concerns arising in cases where only a small number of observations is available for inference. To assess convergence in our simulation study of Section 5.2 we compared the marginal inclusion probabilities $P(\theta_i \neq 0 \mid \mathbf{y}_n)$ obtained from the proportion of MCMC draws where $\theta_i \neq 0$ or from renormalizing $m_k(\mathbf{y}_n \mid M_k)P(M_k)$ across all visited models, a discrepancy between these two estimates would signal lack of convergence. Figure 6 shows a very strong agreement between the two, supporting that convergence has been achieved by B =5,000. To assess the robustness of the results to initialization in our example of Section 5.3 we considered five alternative initial models: the null model, a model initialized after a greedy search for large marginal associations and three models initialized at random with 10, 20 and 30 predictors respectively. For both pMOM and piMOM priors all initializations produced the same MAP model, the same median model (that obtained by selecting $P(\theta_i \neq 0 \mid \mathbf{y}_n) > 0.5$) and estimates of marginal inclusion probabilities are highly correlated. Further, all initializations produced the same ranking in the top 5 models, with very similar estimates of their posterior probabilities based on 25,000 MCMC samples. These diagnostics suggest that our estimates are not appreciably sensitive to alternative initializations. Detailed results are reported in Supplementary Figure 7 and Supplementary Table 3.

Although Laplace approximations greatly facilitate the required computations the integrand is multi-modal, thus the approximation may under-estimate $m_k(\mathbf{y}_n)$ for models containing spurious parameters (hence being conservative). This stems from the fact that for spurious parameters ($\theta_i^* = 0$) the two modes vanish at the same rate, whereas for non-spurious parameters ($\theta_i^* \neq 0$) one of the modes dominates the other asymptotically (Proposition 2 and proof of Proposition 3). A practical solution is to use function `pimomMarginalU` in `mombf` to estimate $m_k(\mathbf{y}_n)$ via Importance Sampling for the models visited in Step 1 of Algorithm 3. Nevertheless our examples in Sections 5.2 and 5.3 suggest that Laplace approximations often provide reasonably good results. Naturally Step 2 avoids any multi-modality issues by using our MCMC-exact latent truncation approach.

Further, we assessed the accuracy of Laplace approximations empirically in several simulation scenarios and found they were asymptotically accurate. For instance, when we consider model selection with pMOM priors, we observed that Laplace approximations closely resembled Importance Sampling estimates of the marginal likelihood, with

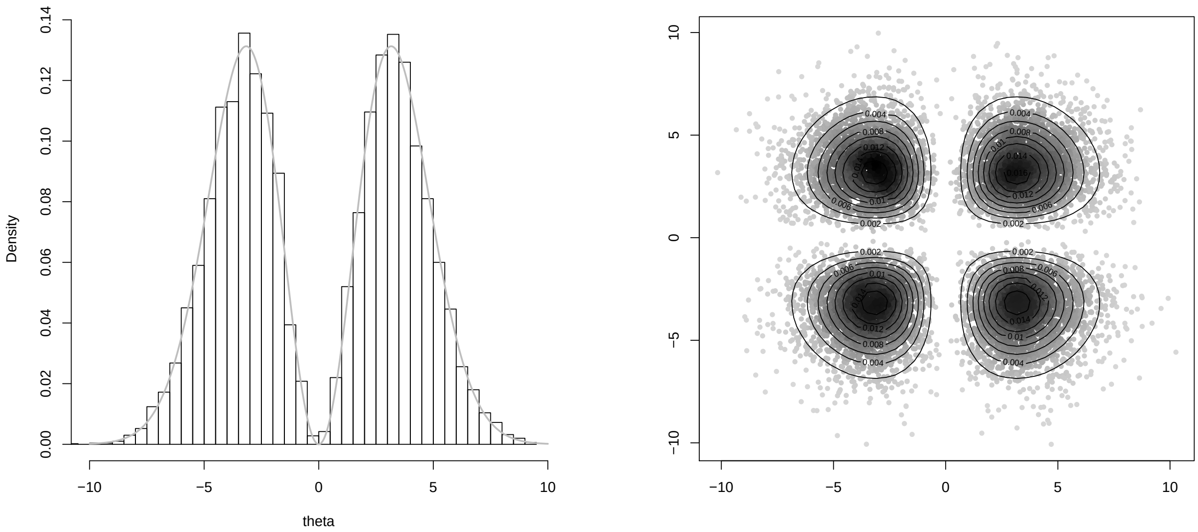

FIGURE 1. 10,000 independent univariate (left) and bivariate (right) pMOM prior draws ($\tau = 5$). Lines indicate true density.

$\theta_1 = 0.5$, $\theta_2 = 1$			
	MOM	iMOM	eMOM
$\theta_1 = 0$, $\theta_2 = 0$	0	0	0
$\theta_1 = 0$, $\theta_2 \neq 0$	2.8e-78	2.72e-78	6.86e-79
$\theta_1 \neq 0$, $\theta_2 = 0$	1.95e-191	3.82-e191	5.90e-191
$\theta_1 \neq 0$, $\theta_2 = 0$	1	1	1
$\theta_1 = 0$, $\theta_2 = 1$			
$\theta_1 = 0$, $\theta_2 = 0$	1.69e-225	4.39e-225	1.08e-224
$\theta_1 = 0$, $\theta_2 \neq 0$	0.999	1	1
$\theta_1 \neq 0$, $\theta_2 = 0$	1.82e-193	1.64e-192	6.80e-192
$\theta_1 \neq 0$, $\theta_2 = 0$	8.83e-05	3.30e-09	3.17e-09

TABLE 1. Posterior model probabilities with 2 predictors ($\theta_1 = 0.5$ or 0, $\theta_2 = \phi = 1$, $n = 1000$)

Laplace approximations growing more conservative as the number of spurious parameters increases. Specifically, we report results on data simulated from a scheme similar to that of Section 5.2. Considering $p = 50$ potential predictors, we observed that Laplace approximations were slightly conservative for small sample sizes ($n = 50$), although the approximations improved as the sample size grew to $n = 100$ and $n = 200$. Detailed results are in Supplementary Figure 8.

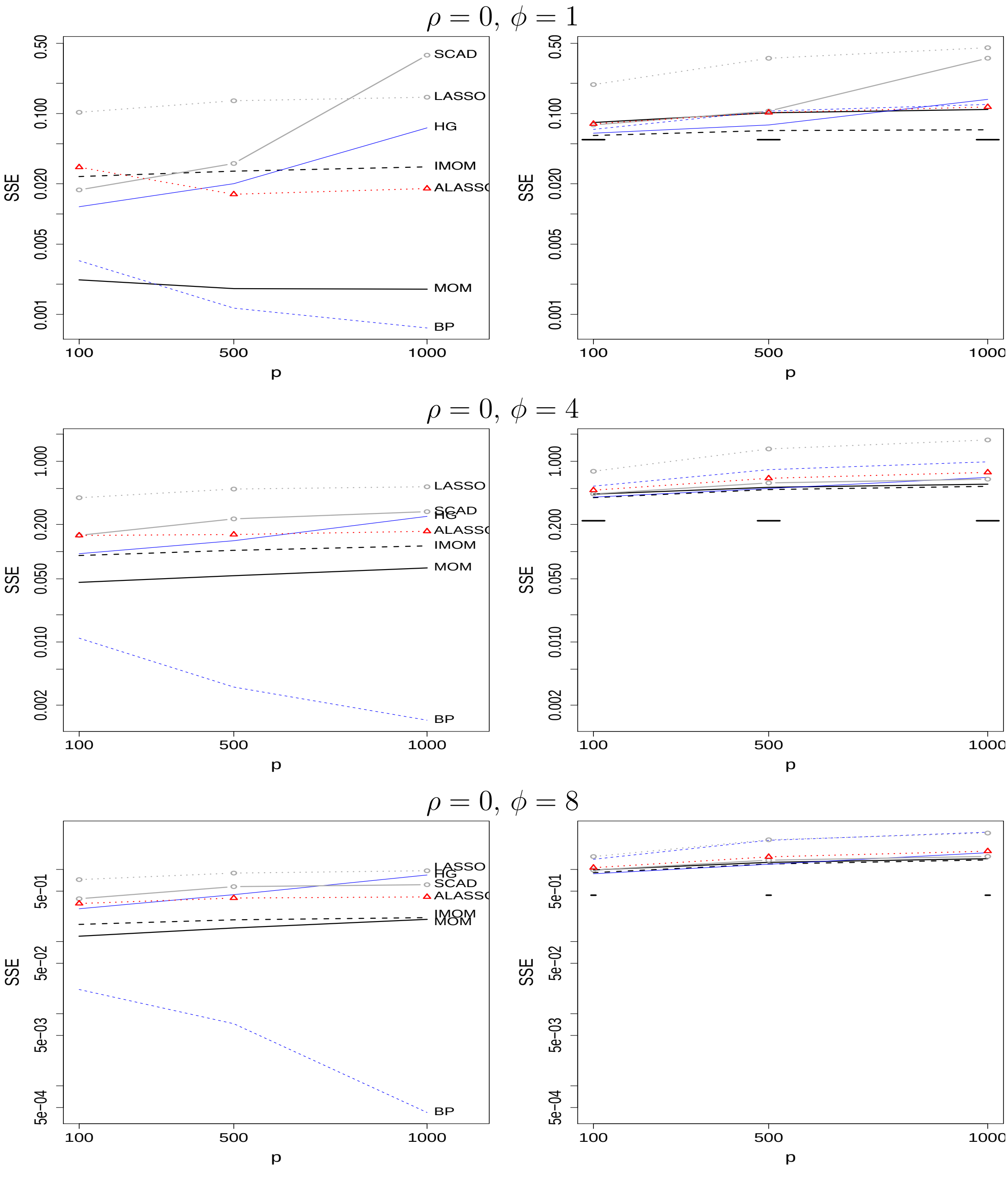

FIGURE 2. Mean SSE for $\theta_i = 0$ (left) and $\theta_i \neq 0$ (right) when $\phi = 1, 4, 8$. Simulation settings: $\rho = 0$, $n = 100$, $p = 100, 500, 1000$ and 5 non-zero coefficients $0.6, 1.2, 1.8, 2.4, 3.0$.

References

S. Ghosal. A review of consistency and convergence of posterior distribution. Technical report, Indian Statistical Institute, 2002.

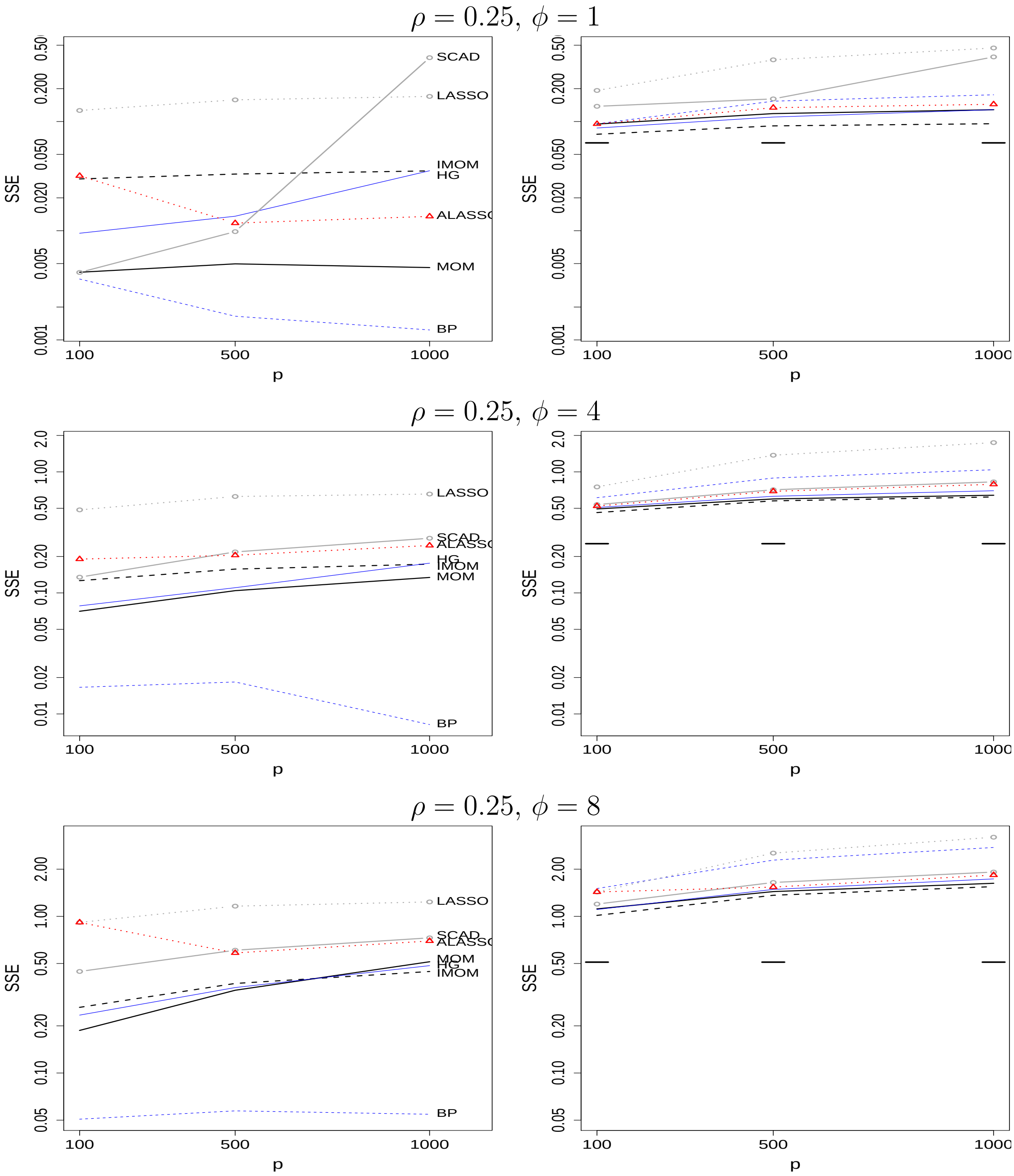

FIGURE 3. Mean SSE for $\theta_i = 0$ (left) and $\theta_i \neq 0$ (right) when $\phi = 1, 4, 8$. Simulation settings: $\rho = 0.25$, $n = 100$, $p = 100, 500, 1000$ and 5 non-zero coefficients $0.6, 1.2, 1.8, 2.4, 3.0$.

V.A. Hajivassiliou. Simulating normal rectangle probabilities and their derivatives: the effects of vectorization. *Econometrics*, 11:519–543, 1993.

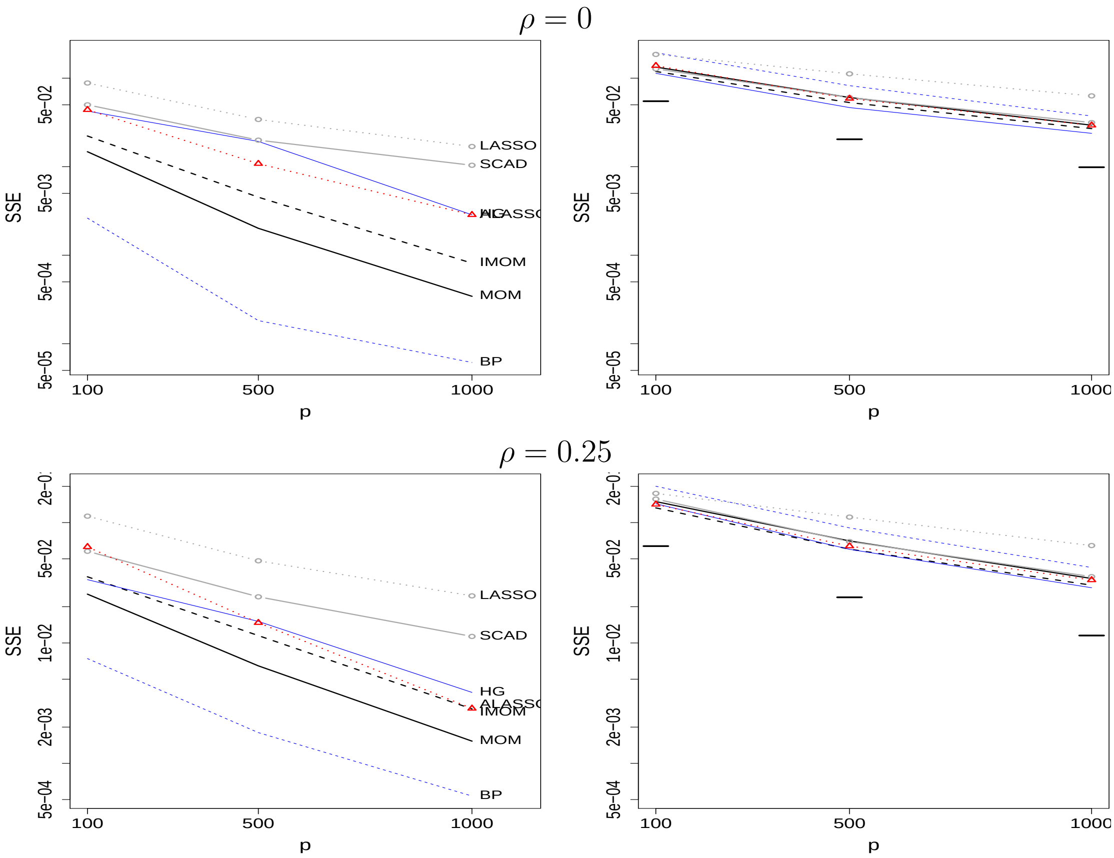

FIGURE 4. Mean SSE for $\theta_i = 0$ (left) and $\theta_i = n^{-1/4}(0.6, 1.2, 1.8, 2.4, 3.0)$ (right) and $\rho = 0, 0.25$ ($\phi = 1$). Simulation settings: ($n = 100, p = 100$), ($n = 250, p = 500$), ($n = 500, p = 1000$)

$\theta_1 = 0.5,\ \theta_2 = 1$			
	MOM	iMOM	eMOM
θ_1	0.096	0.110	0.018
θ_2	0.034	0.134	0.019
ϕ	-0.016	0.069	0.027
$\theta_1 = 0,\ \theta_2 = 1$			
θ_1	0.115	0.032	0.049
θ_2	0.134	0.122	0.042
ϕ	-0.040	0.327	0.353

TABLE 2. Serial correlation with 2 predictors ($\theta_1 = 0.5$ or 0, $\theta_2 = \phi = 1$, $n = 1000$)

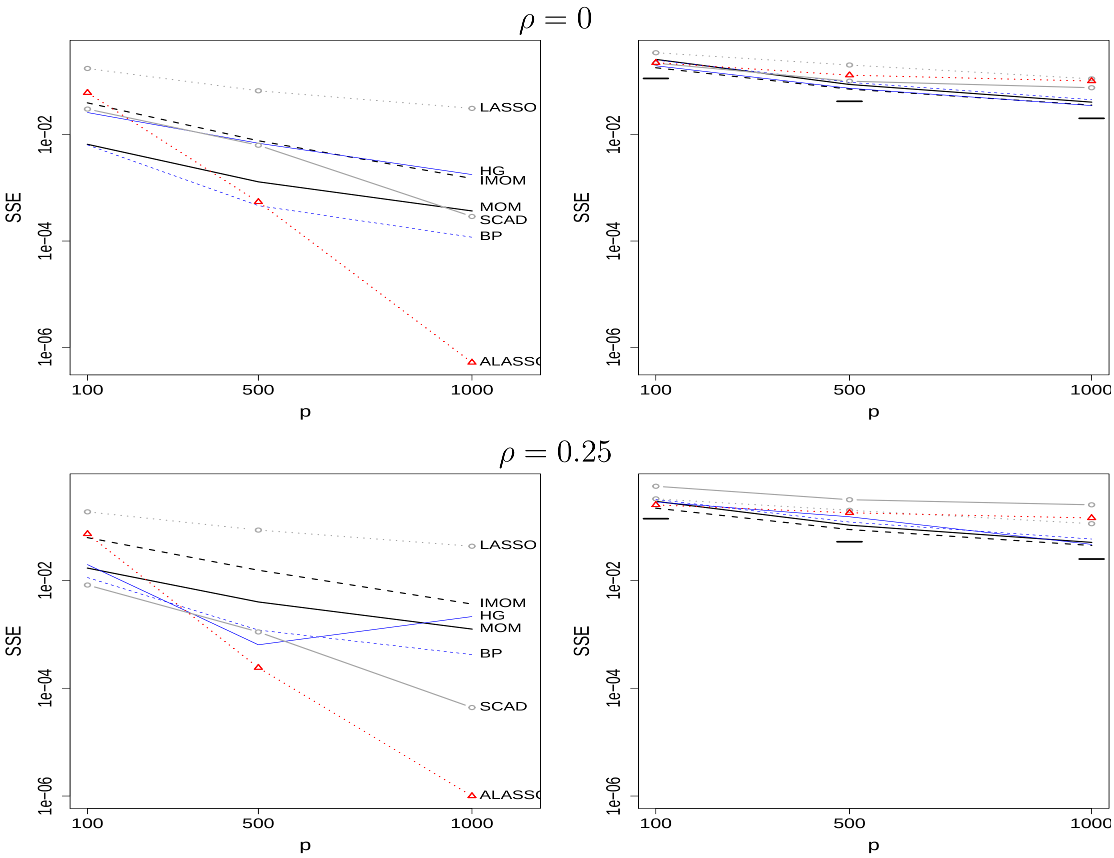

FIGURE 5. Mean SSE for $\theta_i = 0$ (left) and $(\theta_{p-10}, \ldots, \theta_{p-6}) = n^{-1/4}(0.6, 1.2, 1.8, 2.4, 3)$, $(\theta_{p-5}, \ldots, \theta_p) = (0.6, 1.2, 1.8.2.4, 3)$ (right) and $\rho = 0, 0.25$ ($\phi = 1$). Simulation settings: $(n = 100, p = 100)$, $(n = 250, p = 500)$, $(n = 500, p = 1000)$

V.E. Johnson. On numerical aspects of Bayesian model selection in high dimensional and ultrahigh-dimensional settings. *Bayesian Analysis*, 1(1):1–17, 2004.

V.E. Johnson and D. Rossell. Prior densities for default Bayesian hypothesis tests. *Journal of the Royal Statistical Society B*, 72:143–170, 2010.

V.E. Johnson and D. Rossell. Bayesian model selection in high-dimensional settings. *Journal of the American Statistical Association*, 24(498):649–660, 2012.

R. Kan. From moments of sum to moments of product. *Journal of Multivariate Analalysis*, 99:542–554, 2008.

R.E. Kass, L. Tierney, and J.B. Kadane. *The validity of posterior expansions based on Laplace's method*, pages 473–488. Elsever Science, 1990.

M.P. Keane. Simulation estimation for panel data models with limited dependent variables. *Econometrics*, 11:545–571, 1993.

T.L. Lai, H. Robbins, and C.Z. Wei. Strong consistency of least squares in multiple regression. *Journal of multivariate analysis*, 9:343–361, 1979.

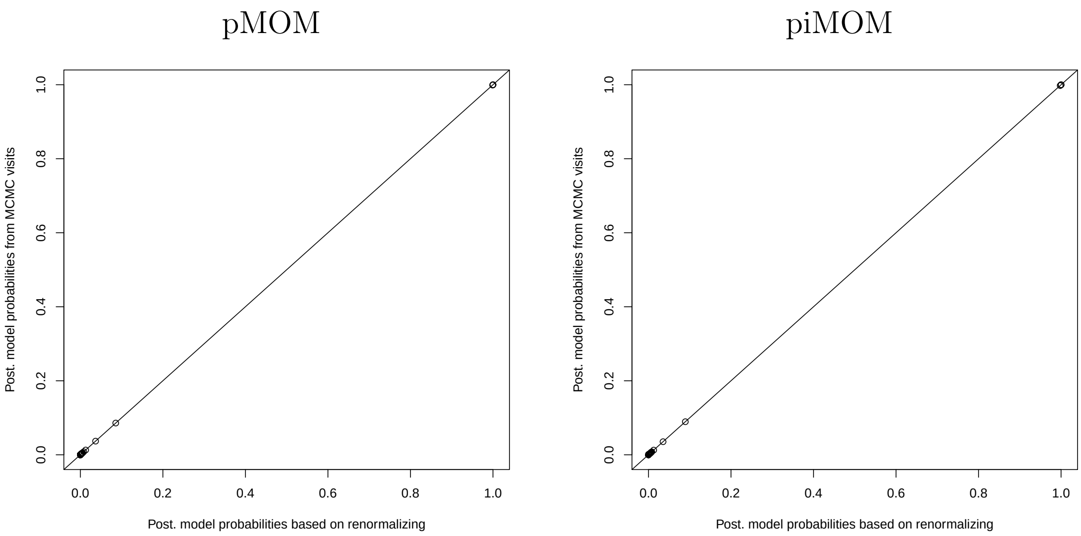

FIGURE 6. Marginal posterior inclusion probabilities in a simulations from Section 5.2 with $p = 1,000$, $\rho = 0$, $\phi = 4$ as estimated from MCMC draws or re-normalizing $m_k(\mathbf{y}_n)P(M_k)$ across visited models. Left: pMOM. Right: piMOM

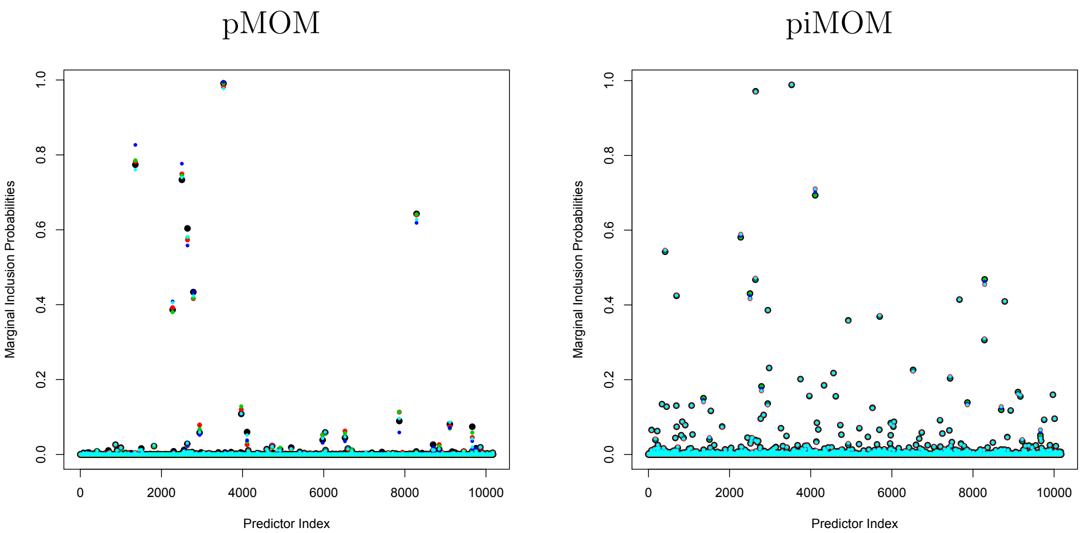

FIGURE 7. Marginal posterior inclusion probabilities for all predictors in the TGFB example of Section 5.3. Different colours represent different initialization strategies

R. Redner. Note on the consistency of the maximum likelihood estimator for nonidentifiable distributions. *Annals of Statistics*, 9(1):225–228, 1981.

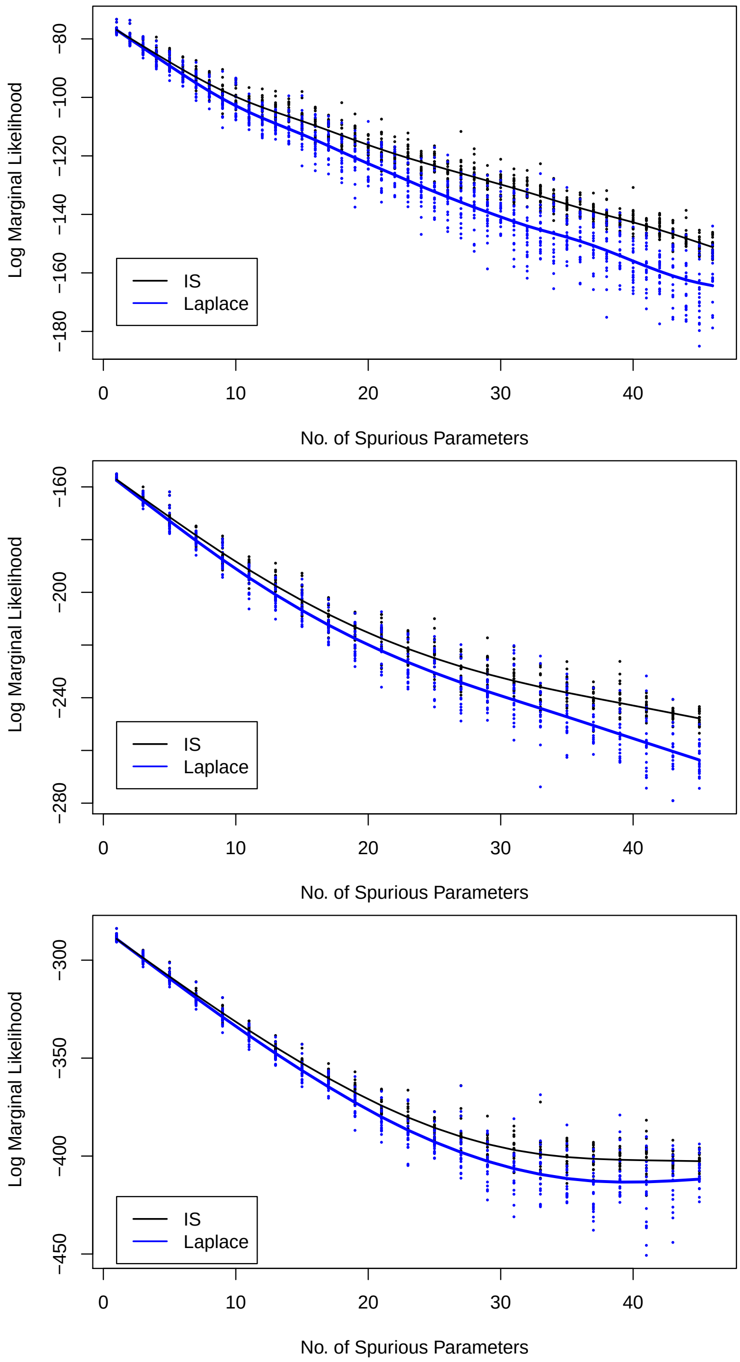

FIGURE 8. Importance Sampling (IS) estimates of pMOM log-marginal likelihoods and Laplace approximations for 900 random models fitted to simulated data and $n = 50, 100, 200$ (top, middle, bottom)

Predictor index	**Posterior probabilities**				
	Initialization				
pMOM	Null	10	20	30	Greedy
1357,2504,2642,2786,3530,8288	0.2357	0.2347	0.2363	0.2368	0.2350
1357,2275,2504,3530	0.1295	0.1290	0.1299	0.1302	0.1292
1357,2275,3530	0.1269	0.1263	0.1272	0.1275	0.1265
1357,2275,2504,3530,3967	0.0355	0.0354	0.0356	0.0357	0.0354
1357,2504,3530,8288	0.0218	0.0217	0.0218	0.0219	0.0217
	Initialization				
piMOM	Null	10	20	30	Greedy
1357,2504,2642,2786,3530,8288	0.1821	0.1833	0.1821	0.1822	0.1833
2504,2642,2940,3530,3967,5973,6524,7863,8288,8846	0.0757	0.0762	0.0757	0.0758	0.0761
1357,2504,2642,2786,3530,6037,8288	0.0292	0.0294	0.0292	0.0291	0.0294
867,1357,2504,2642,2786,3530,8288	0.0191	0.0192	0.0191	0.0192	0.0192
1357,2504,2642,2786,3530,8288,9862	0.0185	0.0186	0.0185	0.0186	0.0186

TABLE 3. Top 5 most likely models with associated posterior probabilities estimated using alternative initialization strategies.

G. Rodriguez-Yam, R.A. Davis, and L.L. Scharf. *Efficient Gibbs sampling of truncated multivariate normal with application to constrained linear regression*. PhD thesis, Department of Statistics, Colorado State University, 2004.

D. Rossell, D. Telesca, and V.E. Johnson. High-dimensional Bayesian classifiers using non-local priors. In *Statistical Models for Data Analysis XV*, pages 305–314. Springer, 2013.

A.M. Walker. On the asymptotic behaviour of posterior distributions. *Jornal of the Royal Statistical Society B*, 31(1):80–88, 1969.