Appendix to "Power-Expected-Posterior Priors for Variable Selection in Gaussian Linear Models"

D. Fouskakis*, I. Ntzoufras† and D. Draper‡

5 December 2012

1 Prior mean and covariance matrix for model parameters under the Z-PEP prior

The Z-PEP marginal prior distributions can be expressed as

$$\begin{aligned}
\pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell|\mathrm{X}_\ell^*,\delta) &= \int \pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell,\sigma_\ell^2|\mathrm{X}_\ell^*,\delta)\,d\sigma_\ell^2 \\
&= \int \left\{ \int f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,w\,\widehat{\boldsymbol{\beta}}_\ell^*,\delta\,w\,(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right] f_{IG}\left(\sigma_\ell^2\,;\,a_\ell+\frac{n^*}{2},b_\ell+\frac{SS_\ell^*}{2}\right)d\sigma_\ell^2\right\}\times \\
&\quad\; m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^* \\
&= \int f_{St_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,2\,a_\ell+n^*,\,w\,\widehat{\boldsymbol{\beta}}_\ell^*,\,\delta\,w\,(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\frac{b_\ell+\frac{SS_\ell^*}{2}}{a_\ell+\frac{n^*}{2}}\right] m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^* \qquad (1)
\end{aligned}$$

and

$$\begin{aligned}
\pi_\ell^{Z\text{-}PEP}(\sigma_\ell^2|\mathrm{X}_\ell^*,\delta) &= \int \pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell,\sigma_\ell^2|\mathrm{X}_\ell^*,\delta)\,d\boldsymbol{\beta}_\ell \\
&= \int \left\{ \int f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\,w\,\widehat{\boldsymbol{\beta}}_\ell^*,\delta\,w\,(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\sigma_\ell^2\right] d\boldsymbol{\beta}_\ell \right\} \\
&\qquad \times f_{IG}\left(\sigma_\ell^2\,;\,a_\ell+\frac{n^*}{2},b_\ell+\frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^* \\
&= \int f_{IG}\left(\sigma_\ell^2\,;\,a_\ell+\frac{n^*}{2},b_\ell+\frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^*\,. \qquad (2)
\end{aligned}$$

THEOREM 1: *Under the baseline prior setup given by equation (15) in the main paper, when* $(a_\ell>1, a_0>1)$ *the Z-PEP prior mean of* $\boldsymbol{\beta}_\ell$ *is* $\mathsf{E}(\boldsymbol{\beta}_\ell)=\mathbf{0}$, *and when* $(a_\ell>2, a_0>1)$ *the Z-PEP prior covariance matrix is*

$$\mathsf{V}(\boldsymbol{\beta}_\ell)=\left\{\frac{\delta\,w}{a_\ell-1+\frac{n^*}{2}}\left[b_\ell+\frac{1}{2}\frac{b_0}{a_0-1}tr(\Lambda_\ell^*\Lambda_0^{*-1})\right]\mathrm{I}_{d_\ell}+\frac{w^2b_0}{a_0-1}(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\mathrm{X}_\ell^{*T}\Lambda_0^{*-1}\mathrm{X}_\ell^*\right\}(\mathrm{X}_\ell^{*T}\mathrm{X}_\ell^*)^{-1}\,, \quad (3)$$

where $tr(\mathrm{A})$ *is the trace of the matrix* A.

*D. Fouskakis is with the Department of Mathematics, National Technical University of Athens, Zografou Campus, Athens 15780 Greece; email fouskakis@math.ntua.gr

†I. Ntzoufras is with the Department of Statistics, Athens University of Economics and Business, 76 Patision Street, Athens 10434 Greece; email ntzoufras@aueb.gr

‡D. Draper is with the Department of Applied Mathematics and Statistics, Baskin School of Engineering, University of California, 1156 High Street, Santa Cruz CA 95064 USA; email draper@ams.ucsc.edu

Proof of Theorem 1. From (1), the prior mean is

$$\begin{aligned}
\mathsf{E}(\boldsymbol{\beta}_\ell) &= \int \boldsymbol{\beta}_\ell \, \pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell \,|\, \mathrm{X}_\ell^*, \delta) \, d\boldsymbol{\beta}_\ell \\
&= \int \left\{ \int \boldsymbol{\beta}_\ell \, f_{St_{d_\ell}} \left[\boldsymbol{\beta}_\ell \, ; \, 2\, a_\ell + n^*, \, w \, \widehat{\boldsymbol{\beta}}_\ell^*, \delta \, w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \frac{b_\ell + \frac{SS_\ell^*}{2}}{a + \frac{n^*}{2}} \right] d\boldsymbol{\beta}_\ell \right\} m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \\
&= \int \mathsf{E}_{St_{d_\ell}} \left[\boldsymbol{\beta}_\ell \, ; \, 2\, a_\ell + n^*, \, w \, \widehat{\boldsymbol{\beta}}_\ell^*, \delta \, w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \frac{b_\ell + \frac{SS_\ell^*}{2}}{a_\ell + \frac{n^*}{2}} \right] m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \, ; \qquad (4)
\end{aligned}$$

here SS_ℓ^* is defined in Section 2.2 of the main paper and $\mathsf{E}_{St_d}\left[\xi(\boldsymbol{z}) \, ; \, df, \, \boldsymbol{\mu}, \, \Sigma\right]$ is the expectation of a function $\xi(\boldsymbol{z})$ of $\boldsymbol{z}$, where z follows a d-dimensional Student distribution with density $f_{St_d}(\boldsymbol{z} \, ; \, df, \boldsymbol{\mu}, \Sigma)$ given by

$$f_{St_d}(\boldsymbol{y} \, ; \, df, \boldsymbol{\mu}, \Sigma) = \frac{\Gamma\left(\frac{df+d}{2}\right)}{\Gamma\left(\frac{df}{2}\right)} (df \, \pi)^{-\frac{d}{2}} |\Sigma|^{-\frac{1}{2}} \left[1 + \frac{1}{df} (\boldsymbol{y} - \boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) \right]^{-\frac{df+d}{2}} . \qquad (5)$$

For $\xi(\boldsymbol{z}) = \boldsymbol{z}$, the expectation is $\boldsymbol{\mu}$, yielding

$$\begin{aligned}
\mathsf{E}(\boldsymbol{\beta}_\ell) &= \int w \, \widehat{\boldsymbol{\beta}}_\ell^* \, m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* = w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^{*T} \int \boldsymbol{y}^* \, m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \\
&= w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^{*T} \int \boldsymbol{y}^* f_{St_{n^*}} \left(\boldsymbol{y}^* \, ; \, 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda_0^{*-1} \right) d\boldsymbol{y}^* \\
&= w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^{*T} \, \mathsf{E}_{St_{n^*}} \left(\boldsymbol{y}^* \, ; \, 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda_0^{*-1} \right) \\
&= w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \mathrm{X}_\ell^{*T} \mathbf{0} = \mathbf{0} \, . \qquad (6)
\end{aligned}$$

Since the mean is zero, the covariance matrix is

$$\begin{aligned}
\mathsf{V}(\boldsymbol{\beta}_\ell) &= \mathsf{E}(\boldsymbol{\beta}_\ell \boldsymbol{\beta}_\ell^T) = \int \boldsymbol{\beta}_\ell \boldsymbol{\beta}_\ell^T \, \pi_\ell^{Z\text{-}PEP}(\boldsymbol{\beta}_\ell \,|\, \mathrm{X}_\ell^*, \delta) \, d\boldsymbol{\beta}_\ell \\
&= \int \mathsf{E}_{St_{d_\ell}} \left[\boldsymbol{\beta}_\ell \boldsymbol{\beta}_\ell^T \, ; \, 2\, a_\ell + n^*, \, w \, \widehat{\boldsymbol{\beta}}_\ell^*, \delta \, w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \frac{b_\ell + \frac{SS_\ell^*}{2}}{a_\ell + \frac{n^*}{2}} \right] m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta) \, d\boldsymbol{y}^* \, . \qquad (7)
\end{aligned}$$

For $\boldsymbol{z} \sim St(df, \boldsymbol{\mu}, \Sigma)$, $\mathsf{E}(\boldsymbol{z}\boldsymbol{z}^T)$ is given by

$$\mathsf{E}(\boldsymbol{z}\boldsymbol{z}^T) = \mathsf{V}(\boldsymbol{z}) + \mathsf{E}(\boldsymbol{z}) \left[\mathsf{E}(\boldsymbol{z})\right]^T = \frac{df}{df - 2} \Sigma + \boldsymbol{\mu}\boldsymbol{\mu}^T \, , \qquad (8)$$

from which

$$\begin{aligned}
\mathsf{E}_{St_{d_\ell}} \left[\boldsymbol{\beta}_\ell \boldsymbol{\beta}_\ell^T \, ; \, 2\, a_\ell + n^*, \, w \, \widehat{\boldsymbol{\beta}}_\ell^*, \, \delta \, w \, (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} \frac{b_\ell + \frac{SS_\ell^*}{2}}{a_\ell + \frac{n^*}{2}} \right] &= \delta \, w \left(\frac{b_\ell + \frac{SS_\ell^*}{2}}{a_\ell - 1 + \frac{n^*}{2}} \right) (\mathrm{X}_\ell^{*T} \mathrm{X}_\ell^*)^{-1} + \\
& \quad w^2 \widehat{\boldsymbol{\beta}}_\ell^* \widehat{\boldsymbol{\beta}}_\ell^{*T} \, . \qquad (9)
\end{aligned}$$

Substitution into (7) yields

$$
\begin{aligned}
\mathsf{V}(\boldsymbol{\beta}_\ell) &= \int \left[\delta\, w \left(\frac{b_\ell + \frac{SS^*_\ell}{2}}{a_\ell - 1 + \frac{n^*}{2}} \right) (\mathrm{X}^{*T}_\ell \mathrm{X}^*_\ell)^{-1} + w^2 \widehat{\boldsymbol{\beta}}^*_\ell \widehat{\boldsymbol{\beta}}^{*T}_\ell \right] m^N_0(\boldsymbol{y}^* | \mathrm{X}^*_0, \delta)\, d\boldsymbol{y}^* \\
&= \delta\, w \frac{1}{a_\ell - 1 + \frac{n^*}{2}} \left[b_\ell + \frac{1}{2} \int SS^*_\ell\, m^N_0(\boldsymbol{y}^* | \mathrm{X}^*_0, \delta)\, d\boldsymbol{y}^* \right] (\mathrm{X}^{*T}_\ell \mathrm{X}^*_\ell)^{-1} + \\
&\quad w^2 \int \widehat{\boldsymbol{\beta}}^*_\ell \widehat{\boldsymbol{\beta}}^{*T}_\ell\, m^N_0(\boldsymbol{y}^* | \mathrm{X}^*_0, \delta)\, d\boldsymbol{y}^* \\
&= \delta\, w \frac{1}{a_\ell - 1 + \frac{n^*}{2}} \left[b_\ell + \frac{1}{2} \mathsf{E}_{St_{n^*}} \left(\boldsymbol{y}^{*T} \Lambda^*_\ell \boldsymbol{y}^*; 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda^{*-1}_0 \right) \right] (\mathrm{X}^{*T}_\ell \mathrm{X}^*_\ell)^{-1} + \\
&\quad w^2 (\mathrm{X}^{*T}_\ell \mathrm{X}^*_\ell)^{-1} \mathrm{X}^{*T}_\ell \mathsf{E}_{St_{n^*}} \left(\boldsymbol{y}^* \boldsymbol{y}^{*T}; 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda^{*-1}_0 \right) \mathrm{X}^*_\ell (\mathrm{X}^{*T}_\ell \mathrm{X}^*_\ell)^{-1}\,. \qquad (10)
\end{aligned}
$$

Now (see, e.g., Scott, 1997, Theorem 9.18) for any symmetric matrix A and any random vector $\boldsymbol{z}$ with mean $\boldsymbol{\mu}$ and covariance matrix $\mathsf{V}(\boldsymbol{z})$,

$$
E(\boldsymbol{z}^T \mathrm{A} \boldsymbol{z}) = tr\Big[\mathrm{A}\, \mathsf{V}(\boldsymbol{z}) \Big] + \boldsymbol{\mu}^T \mathrm{A} \boldsymbol{\mu}\,, \qquad (11)
$$

if $E(\boldsymbol{z}\boldsymbol{z}^T)$ exists. Therefore for $\boldsymbol{z} \sim St(df, \boldsymbol{\mu}, \Sigma)$,

$$
E(\boldsymbol{z}^T \mathrm{A} \boldsymbol{z}) = \frac{df}{df - 2} tr(A\Sigma) + \boldsymbol{\mu}^T \mathrm{A} \boldsymbol{\mu}\,, \qquad (12)
$$

from which, in our case,

$$
\mathsf{E}_{St_{n^*}} \left(\boldsymbol{y}^{*T} \Lambda^*_\ell\, \boldsymbol{y}^*; 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda^{*-1}_0 \right) = \frac{b_0}{a_0 - 1} tr(\Lambda^*_\ell \Lambda^{*-1}_0)\,. \qquad (13)
$$

Moreover,

$$
\mathsf{E}_{St_{n^*}} \left(\boldsymbol{y}^* \boldsymbol{y}^{*T}; 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda^{*-1}_0 \right) = \mathsf{V}_{St_{n^*}} \left(\boldsymbol{y}^*; 2\, a_0, \mathbf{0}, \frac{b_0}{a_0} \Lambda^{*-1}_0 \right) + \mathbf{0} = \frac{b_0}{a_0 - 1} \Lambda^{*-1}_0\,. \qquad (14)
$$

Substituting (13) and (14) into (10), we obtain (3) as desired.

□

THEOREM 2: *Under the baseline prior setup given by equation (15) in the main paper, for* $(a_\ell > 1, a_0 > 1)$ *the Z-PEP prior mean of* σ^2_ℓ *is*

$$
\mathsf{E}\left(\sigma^2_\ell\right) = \frac{b_0}{a_0 - 1} \frac{\frac{1}{2} tr\left(\Lambda^*_\ell \Lambda^{*-1}_0\right) + \frac{(a_0 - 1) b_\ell}{b_0}}{\frac{n^*}{2} + a_\ell - 1}\,, \qquad (15)
$$

and for $(a_\ell > 2, a_0 > 2)$ *the Z-PEP prior variance is*

$$
\begin{aligned}
\mathsf{V}\left(\sigma^2_\ell\right) &= \left[\left(\frac{n^*}{2} + a_\ell - 1 \right) \left(\frac{n^*}{2} + a_\ell - 2 \right) \right]^{-1} \left\{ b^2_\ell + \frac{b_\ell\, b_0\, tr(\Lambda^*_\ell \Lambda^{*-1}_0)}{a_0 - 1} + \right. \\
&\quad \left. \frac{b^2_0 \left[2\, tr(\Lambda^*_\ell \Lambda^{*-1}_0 \Lambda^*_\ell \Lambda^{*-1}_0) + tr(\Lambda^*_\ell \Lambda^{*-1}_0)^2 \right]}{4(a_0 - 1)(a_0 - 2)} \right\} - \\
&\quad \left(\frac{b_0}{a_0 - 1} \right)^2 \left[\frac{\frac{1}{2} tr(\Lambda^*_\ell \Lambda^{*-1}_0) + \frac{(a_0 - 1) b_\ell}{b_0}}{\frac{n^*}{2} + a_\ell - 1} \right]^2. \qquad (16)
\end{aligned}
$$

Proof of Theorem 2. The prior mean of the variance parameter is

$$\begin{aligned}
\mathsf{E}\left(\sigma_\ell^2\right) &= \int \sigma_\ell^2\, \pi_\ell^{Z\text{-}PEP}(\sigma_\ell^2|\mathrm{X}_\ell^*, \delta)\, d\sigma_\ell^2 \\
&= \int \left[\int \sigma_\ell^2 f_{IG}\left(\sigma_\ell^2\,;\, a_\ell + \frac{n^*}{2}, b_\ell + \frac{SS_\ell^*}{2}\right) d\sigma_\ell^2\right] m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \\
&= \int \mathsf{E}_{IG}\left(\sigma_\ell^2;\, a_\ell + \frac{n^*}{2}, b_\ell + \frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \qquad (17)
\end{aligned}$$

Here $\mathsf{E}_{IG}\left[\xi(Z);\, a, b\right]$ is the expectation of a function $\xi(Z)$ of Z, where Z follows an Inverse-Gamma distribution with parameters a and b; for $\xi(Z) = Z$ this expectation is $\frac{b}{a-1}$. Thus the prior mean of σ_ℓ^2 is

$$\begin{aligned}
\mathsf{E}\left(\sigma_\ell^2\right) &= \int \frac{b_\ell + \frac{SS_\ell^*}{2}}{\frac{n^*}{2} + a_\ell - 1} m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \\
&= \left(\frac{n^*}{2} + a_\ell - 1\right)^{-1} \left[b_\ell + \frac{1}{2}\mathsf{E}_{St_{n^*}}\left(SS_\ell^*; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right)\right] \\
&= \left(\frac{n^*}{2} + a_\ell - 1\right)^{-1} \left[b_\ell + \frac{1}{2}\mathsf{E}_{St_{n^*}}\left(\boldsymbol{y}^*\Lambda_\ell^*\,\boldsymbol{y}^*; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right)\right] . \qquad (18)
\end{aligned}$$

From (13),

$$\begin{aligned}
\mathsf{E}\left(\sigma_\ell^2\right) &= \left(\frac{n^*}{2} + a_\ell - 1\right)^{-1} \left[b_\ell + \frac{1}{2}\frac{b_0}{a_0 - 1} tr(\Lambda_\ell^*\Lambda_0^{*-1})\right] \\
&= \frac{b_0}{a_0 - 1}\, \frac{\frac{1}{2}tr(\Lambda_\ell^*\Lambda_0^{*-1}) + \frac{(a_0-1)b_\ell}{b_0}}{\frac{n^*}{2} + a_\ell - 1} . \qquad (19)
\end{aligned}$$

The prior variance of σ_ℓ^2 can be written as

$$\begin{aligned}
\mathsf{V}\left(\sigma_\ell^2\right) &= \mathsf{E}\left(\sigma_\ell^4\right) - \left[\mathsf{E}\left(\sigma_\ell^2\right)\right]^2 \\
&= \mathsf{E}\left(\sigma_\ell^4\right) - \left(\frac{b_0}{a_0 - 1}\right)^2 \left[\frac{\frac{1}{2}tr(\Lambda_\ell^*\Lambda_0^{*-1}) + \frac{(a_0-1)b_\ell}{b_0}}{\frac{n^*}{2} + a_\ell - 1}\right]^2 , \qquad (20)
\end{aligned}$$

where

$$\begin{aligned}
\mathsf{E}\left(\sigma_\ell^4\right) &= \int \sigma_\ell^4\, \pi_\ell^{Z\text{-}PEP}(\sigma_\ell^2|\mathrm{X}_\ell^*, \delta)\, d\sigma_\ell^2 \\
&= \int \left[\int \sigma_\ell^4 f_{IG}\left(\sigma_\ell^2\,;\, a_\ell + \frac{n^*}{2}, b_\ell + \frac{SS_\ell^*}{2}\right) m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\sigma_\ell^2\right] d\boldsymbol{y}^* \\
&= \int \mathsf{E}_{IG}\left[(\sigma_\ell^2)^2\,;\, a_\ell + \frac{n^*}{2}, b_\ell + \frac{SS_\ell^*}{2}\right] m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\, d\boldsymbol{y}^* \,. \qquad (21)
\end{aligned}$$

For a random variate Z that follows an Inverse-Gamma distribution with parameters a and b,

$$\mathsf{E}(Z^2) = \mathsf{V}(Z) + \mathsf{E}(Z)^2 = \frac{b^2}{(a-1)^2(a-2)} + \left(\frac{b}{a-1}\right)^2 = \frac{b^2}{(a-1)(a-2)} \qquad (22)$$

for $a > 2$. Hence

$$\begin{aligned}
\mathsf{E}\left(\sigma_\ell^4\right) &= \int \frac{\left(b_\ell + \frac{SS_\ell^*}{2}\right)^2}{\left(\frac{n^*}{2}+a_\ell-1\right)\left(\frac{n^*}{2}+a_\ell-2\right)}\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^* \\
&= \int \frac{\left(b_\ell^2 + b_\ell SS_\ell^* + \frac{1}{4}SS_\ell^{*2}\right)}{\left(\frac{n^*}{2}+a_\ell-1\right)\left(\frac{n^*}{2}+a_\ell-2\right)}\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^* \\
&= \left[\left(\frac{n^*}{2}+a_\ell-1\right)\left(\frac{n^*}{2}+a_\ell-2\right)\right]^{-1}\left[b_\ell^2 + b_\ell \int SS_\ell^*\, m_0^N(\boldsymbol{y}^*;\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^* + \right. \\
&\qquad \left. \frac{1}{4}\int SS_\ell^{*2}\, m_0^N(\boldsymbol{y}^*;\mathrm{X}_0^*,\delta)\,d\boldsymbol{y}^*\right] \\
&= \left[\left(\frac{n^*}{2}+a_\ell-1\right)\left(\frac{n^*}{2}+a_\ell-2\right)\right]^{-1}\left\{b_\ell^2 + b_\ell\, \mathsf{E}_{St_{n^*}}\left(\boldsymbol{y}^{*^T}\Lambda_\ell^*\,\boldsymbol{y}^*; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right) + \right. \\
&\qquad \left. \frac{1}{4}\,\mathsf{E}_{St_{n^*}}\left[(\boldsymbol{y}^{*^T}\Lambda_\ell^*\,\boldsymbol{y}^*)^2; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right]\right\}. \qquad (23)
\end{aligned}$$

The expectation $\mathsf{E}_{St_{n^*}}\left(\boldsymbol{y}^{*^T}\Lambda_\ell^*\boldsymbol{y}^*; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right)$ is given by (13). Moreover, if A is a symmetric matrix and $\boldsymbol{z} \sim N_d(\mathbf{0},\Sigma)$ then (see Scott, 1997, Theorem 9.21)

$$\mathsf{E}\left[(\boldsymbol{z}^T\mathrm{A}\boldsymbol{z})^2\right] = 2\,tr(\mathrm{A}\Sigma\mathrm{A}\Sigma) + tr(\mathrm{A}\Sigma)^2\,. \qquad (24)$$

By rewriting the multivariate Student distribution with density $f_{St_d}(\boldsymbol{y}\,;\,df,\boldsymbol{\mu},\Sigma)$ as a Normal-Inverse-Gamma scale mixture, we can calculate $\mathsf{E}_{St_{n^*}}\left[(\boldsymbol{y}^{*^T}\Lambda_\ell^*\,\boldsymbol{y}^*)^2; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right]$, as follows: if $\boldsymbol{z} \sim St_d(df,\mathbf{0},\Sigma)$ then

$$\begin{aligned}
\mathsf{E}_{St_d}\left[(\boldsymbol{z}^T\mathrm{A}\boldsymbol{z})^2\,;\,df,\mathbf{0},\Sigma\right] &= \int (\boldsymbol{z}^T\mathrm{A}\boldsymbol{z})^2 f_{St_d}(\boldsymbol{z}\,;\,df,\mathbf{0},\Sigma)\,d\boldsymbol{z} \\
&= \int (\boldsymbol{z}^T\mathrm{A}\boldsymbol{z})^2\left[\int f_{N_d}(\boldsymbol{z}\,;\,\mathbf{0},\Sigma\psi)\,f_{IG}(\psi\,;\,\frac{df}{2},\frac{df}{2})\,d\psi\right]d\boldsymbol{z} \\
&= \int\left[\int (\boldsymbol{z}^T\mathrm{A}\boldsymbol{z})^2 f_{N_d}(\boldsymbol{z}\,;\,\mathbf{0},\Sigma\psi)\,d\boldsymbol{z}\right] f_{IG}(\psi\,;\,\frac{df}{2},\frac{df}{2})\,d\psi \qquad (25) \\
&= \left[2\,tr(\mathrm{A}\Sigma\mathrm{A}\Sigma) + tr(\mathrm{A}\Sigma)^2\right]\int \psi^2 f_{IG}(\psi\,;\,\frac{df}{2},\frac{df}{2})d\psi \\
&= \left[2\,tr(\mathrm{A}\Sigma\mathrm{A}\Sigma) + tr(\mathrm{A}\Sigma)^2\right]\frac{\frac{df^2}{4}}{\left(\frac{df}{2}-1\right)\left(\frac{df}{2}-2\right)}\,. \qquad (26)
\end{aligned}$$

It now follows that

$$\mathsf{E}_{St_{n^*}}\left[(\boldsymbol{y}^{*^T}\Lambda_\ell^*\,\boldsymbol{y}^*)^2; 2\,a_0, \mathbf{0}, \frac{b_0}{a_0}\Lambda_0^{*-1}\right] = \frac{b_0^2}{(a_0-1)(a_0-2)}\left[2\,tr(\Lambda_\ell^*\Lambda_0^{*-1}\Lambda_\ell^*\Lambda_0^{*-1}) + tr(\Lambda_\ell^*\Lambda_0^{*-1})^2\right]. \qquad (27)$$

By substituting (13) and (27) in (23), we obtain

$$\begin{aligned}
\mathsf{E}\left(\sigma_\ell^4\right) &= \left[\left(\frac{n^*}{2}+a_\ell-1\right)\left(\frac{n^*}{2}+a_\ell-2\right)\right]^{-1}\left\{b_\ell^2 + \frac{b_\ell\,b_0\,tr(\Lambda_\ell^*\Lambda_0^{*-1})}{a_0-1} + \right. \qquad (28) \\
&\qquad \left.\frac{b_0^2\left[2\,tr(\Lambda_\ell^*\Lambda_0^{*-1}\Lambda_\ell^*\Lambda_0^{*-1}) + tr(\Lambda_\ell^*\Lambda_0^{*-1})^2\right]}{4(a_0-1)(a_0-2)}\right\}. \qquad (29)
\end{aligned}$$

Finally, the prior variance is then

$$
\begin{aligned}
\mathsf{V}\left(\sigma_\ell^2\right) &= \mathsf{E}\left(\sigma_\ell^4\right) - \left[\mathsf{E}\left(\sigma_\ell^2\right)\right]^2 \qquad (30)\\
&= \left[\left(\frac{n^*}{2} + a_\ell - 1\right)\left(\frac{n^*}{2} + a_\ell - 2\right)\right]^{-1} \left\{ b_\ell^2 + \frac{b_\ell\, b_0\, tr(\Lambda_\ell^* \Lambda_0^{*-1})}{a_0 - 1} + \right.\\
&\left. \frac{b_0^2\left[2\, tr(\Lambda_\ell^* \Lambda_0^{*-1} \Lambda_\ell^* \Lambda_0^{*-1}) + tr(\Lambda_\ell^* \Lambda_0^{*-1})^2\right]}{4(a_0-1)(a_0-2)} \right\} - \left(\frac{b_0}{a_0 - 1}\right)^2 \left[\frac{\frac{1}{2} tr(\Lambda_\ell^* \Lambda_0^{*-1}) + \frac{(a_0-1)b_\ell}{b_0}}{\frac{n^*}{2} + a_\ell - 1}\right]^2 .
\end{aligned}
$$

□

2 MCMC for sampling from the posterior

To generate MCMC samples from the posterior distributions defined by equation (14) or equation (19) in the main paper (under the two baseline prior choices), we consider the following conditional distribution:

$$
\begin{aligned}
f(\boldsymbol{\beta}_\ell, \sigma_\ell^2, \boldsymbol{y}^* | \boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \;\propto\; & f_{N_{d_\ell}}\big(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N \sigma_\ell^2\big) f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N) \times \\
& m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)\,. \qquad (31)
\end{aligned}
$$

The parameters in the above Normal-Inverse-Gamma distribution are given in equations (11) and (20) in the main paper for the baseline Jeffreys and g-priors, respectively. From the above we have that $f(\sigma_\ell^2|\boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N)$, $f(\boldsymbol{\beta}_\ell|\sigma_\ell^2, \boldsymbol{y}, \boldsymbol{y}^*; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{N_{d_\ell}}\big(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N, \widetilde{\Sigma}^N\sigma_\ell^2\big)$ and

$$
\begin{aligned}
f(\boldsymbol{y}^*|\boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) &\propto \iint f(\boldsymbol{\beta}_\ell, \sigma_\ell^2, \boldsymbol{y}^*|\boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, d\boldsymbol{\beta}_\ell\; d\sigma_\ell^2 \\
&\propto \left[\iint f_{N_{d_\ell}}\big(\boldsymbol{\beta}_\ell\,;\, \widetilde{\boldsymbol{\beta}}^N,\, \widetilde{\Sigma}^N\sigma_\ell^2\big)\, f_{IG}(\sigma_\ell^2\,;\, \widetilde{a}_\ell^N, \widetilde{b}_\ell^N)\, d\boldsymbol{\beta}_\ell\, d\sigma_\ell^2\right] \times \\
&\qquad m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta) \\
&\propto m_\ell^N(\boldsymbol{y}|\boldsymbol{y}^*;\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta) \\
&\propto m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y};\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)\, \frac{m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)}{m_\ell^N(\boldsymbol{y}^*|\mathrm{X}_\ell^*, \delta)}\,, \qquad (32)
\end{aligned}
$$

with

$$
m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y},\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \;=\; \iint f(\boldsymbol{y}^*|\boldsymbol{\beta}_\ell, \sigma_\ell^2, M_\ell\,; \mathrm{X}_\ell^*, \delta)\, f(\boldsymbol{\beta}_\ell, \sigma_\ell^2|\boldsymbol{y}, M_\ell\,;\, \mathrm{X}_\ell)\, d\boldsymbol{\beta}_\ell\, d\sigma_\ell^2\,. \qquad (33)
$$

For the baseline g-prior, (33) becomes

$$
\begin{aligned}
m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y},\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) \;=\; & \iint f_{N_{n^*}}(\boldsymbol{y}^*\,;\, \mathrm{X}_\ell^*\boldsymbol{\beta}_\ell, \delta\,\sigma_\ell^2\, \mathrm{I}_{n^*})\, f_{N_{d_\ell}}\left[\boldsymbol{\beta}_\ell\,;\, \frac{g}{g+1}\widehat{\boldsymbol{\beta}}_\ell, \frac{g}{g+1}(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}\sigma_\ell^2\right] \times \\
& f_{IG}\left(\sigma_\ell^2\,;\, a_\ell + \frac{n}{2}, b_\ell + \frac{SS_\ell}{2}\right) d\boldsymbol{\beta}_\ell\, d\sigma_\ell^2 \\
= & f_{St_{n^*}}\left\{\boldsymbol{y}^*\,;\, 2\,a_\ell + n,\, \frac{g}{g+1}\mathrm{X}_\ell^*\widehat{\boldsymbol{\beta}}_\ell\,,\, \frac{2b_\ell + SS_\ell}{2\,a_\ell + n}\left[\delta\, \mathrm{I}_{n^*} + \frac{g}{g+1}\mathrm{X}_\ell^*(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}\mathrm{X}_\ell^{*T}\right]\right\}, \qquad (34)
\end{aligned}
$$

where $SS_\ell = \boldsymbol{y}^T\left[\mathrm{I}_n - \frac{g}{g+1}\mathrm{X}_\ell\,(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}\mathrm{X}_\ell^T\right]\boldsymbol{y}$; for the Jeffreys baseline prior the expression is the same with $\frac{g}{g+1} = 1$, $a_\ell = -\frac{d_\ell}{2}$ and $b_\ell = 0$. Therefore for the Jeffreys baseline prior, equation (33) becomes

$$
m_\ell^N(\boldsymbol{y}^*|\boldsymbol{y},\, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = f_{St_{n^*}}\left\{\boldsymbol{y}^*\,;\, n - d_\ell,\, \mathrm{X}_\ell^*\widehat{\boldsymbol{\beta}}_\ell\,,\, \frac{SS_\ell}{n - d_\ell}\left[\delta\, \mathrm{I}_{n^*} + \mathrm{X}_\ell^*(\mathrm{X}_\ell^T\mathrm{X}_\ell)^{-1}\mathrm{X}_\ell^{*T}\right]\right\}, \qquad (35)
$$

with the posterior sum of squares now given by $SS_\ell = \boldsymbol{y}^T \left[\mathrm{I}_n - \mathrm{X}_\ell \, (\mathrm{X}_\ell^T \mathrm{X}_\ell)^{-1} \mathrm{X}_\ell^T \right] \boldsymbol{y}$.

Using the above expressions, we can specify the following MCMC scheme, in which the Inverse-Gamma distribution $IG(a, b)$ was previously defined in Section 2.1.1 of the main paper:

(1) Generate $\boldsymbol{y}^*$ from (32);

(2) Generate σ_ℓ^2 from $IG(\widetilde{a}_\ell^N, \widetilde{b}_\ell^N)$; and

(3) Generate $\boldsymbol{\beta}_\ell$ from $N_{d_\ell}(\widetilde{\boldsymbol{\beta}}^N, \widetilde{\Sigma}^N \sigma_\ell^2)$.

In Step 1, we can generate the imaginary data $\boldsymbol{y}^*$ by using a Metropolis-Hastings algorithm with proposal $q(\boldsymbol{y}^{*'}) = m_\ell^N(\boldsymbol{y}^{*'} | \boldsymbol{y}, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)$ given in (34) or (35) (for the baseline g-prior or Jeffreys prior choices, respectively) and acceptance probability

$$\alpha = \min \left[1, \frac{m_0^N(\boldsymbol{y}^{*'} | \mathrm{X}_0^*, \delta) \, m_\ell^N(\boldsymbol{y}^* | \mathrm{X}_\ell^*, \delta)}{m_\ell^N(\boldsymbol{y}^{*'} | \mathrm{X}_\ell^*, \delta) \, m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta)} \right], \tag{36}$$

where $m_\ell^N(\boldsymbol{y}^* | \mathrm{X}_\ell^*, \delta)$ is given in equations (9) and (16) of the main paper for the baseline Jeffreys and g-prior choices, respectively.

3 Marginal-likelihood computation

In the main paper we noted that four different Monte-Carlo estimates of the marginal likelihood of any model $M_\ell \in \mathcal{M}$ are possible, in settings in which the marginal likelihood is not analytically tractable. Here we present details on the two least successful of these approaches, denoted by schemes (3) and (4).

(3) Generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_0^N(\boldsymbol{y}^* | \mathrm{X}_0^*, \delta)$ and estimate the marginal likelihood by

$$\hat{m}_\ell^{PEP}(\boldsymbol{y} | \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = \frac{1}{T} m_\ell^N(\boldsymbol{y} | \mathrm{X}_\ell, \mathrm{X}_\ell^*) \sum_{t=1}^{T} \frac{m_\ell^N(\boldsymbol{y}^{*(t)} | \boldsymbol{y}, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)}{m_\ell^N(\boldsymbol{y}^{*(t)} | \mathrm{X}_\ell^*, \delta)}. \tag{37}$$

This approach was also proposed by Perez and Berger (2002).

(4) Generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_0^N(\boldsymbol{y}^* | \boldsymbol{y}, \mathrm{X}_0, \mathrm{X}_0^*, \delta)$ and estimate the marginal likelihood by

$$\hat{m}_\ell^{PEP}(\boldsymbol{y} | \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta) = m_0^N(\boldsymbol{y} | \mathrm{X}_0, \mathrm{X}_0^*) \left[\frac{1}{T} \sum_{t=1}^{T} \frac{m_\ell^N(\boldsymbol{y} | \boldsymbol{y}^{*(t)}, \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)}{m_0^N(\boldsymbol{y} | \boldsymbol{y}^{*(t)}, \mathrm{X}_0, \mathrm{X}_0^*, \delta)} \right]. \tag{38}$$

The third and fourth Monte-Carlo schemes are straightforward, since we only need to obtain a single sample of $\boldsymbol{y}^*$ from the prior and the posterior predictive distribution from model M_0, respectively; then we estimate the marginal likelihood of every model using those simulated values. Nevertheless we expect those estimates for the marginal likelihood of model M_ℓ to have large Monte-Carlo error, since the imaginary data are generated from importance functions that do not make full use of the data $\boldsymbol{y}$ (in the third Monte-Carlo scheme) or the stochastic structure of model M_ℓ (in the third and fourth schemes). See Section 5.1.1 of the main paper for numerical comparisons between these two schemes and the other two approaches.

4 Implementation of the EIBF

The results in this paper based on the Expected Intrinsic Bayes Factor (EIBF) were obtained using the following procedure:

(1) Generate S random splits.

(2) For every random split $s = 1, 2, \ldots, S$ we denote by $\boldsymbol{y}_{[s]}$ and $\mathrm{X}_{[s]}$ the data used for the evaluation of the posterior distribution and by $\boldsymbol{y}_{[\setminus s]}$ and $\mathrm{X}_{[\setminus s]}$ the remaining data used for the calculation of the Bayes factor.

(3) For every split we calculate the marginal likelihood of model M_ℓ by

$$\psi(\boldsymbol{y}|M_\ell, s) = \int f(\boldsymbol{y}_{[\setminus s]}|\boldsymbol{\beta}_\ell\,, \sigma^2_\ell, M_\ell; \mathrm{X}_{\ell[\setminus s]})\, f(\boldsymbol{\beta}_\ell\,, \sigma^2_\ell|\boldsymbol{y}_{[s]}, M_\ell; \mathrm{X}_{\ell[s]})\, d\boldsymbol{\beta}_\ell\, d\sigma^2_\ell\,, \tag{39}$$

where $\mathrm{X}_{\ell[s]}$ and $\mathrm{X}_{\ell[\setminus s]}$ are the submatrices of $\mathrm{X}_{[s]}$ and $\mathrm{X}_{[\setminus s]}$ corresponding to model M_ℓ, respectively.

(4) We then calculate the Bayes factor of model M_ℓ versus the reference model M_0 under the split s by

$$IBF^{[s]}_{\ell 0} = \frac{\psi(\boldsymbol{y}|M_\ell, s)}{\psi(\boldsymbol{y}|M_0, s)}\,. \tag{40}$$

(5) We calculate the EIBF by computing the arithmetic mean of $IBF^{[s]}_{\ell 0}$ over all splits $s = 1, \ldots, S$:

$$EIBF_{\ell 0} = \frac{1}{S}\sum_{s=1}^{S} IBF^{[s]}_{\ell 0}\,. \tag{41}$$

(6) All weights based on the EIBF are calculated by

$$W^{EIBF}_\ell = \frac{EIBF_{\ell 0}}{\sum_{M_\ell \in \mathcal{M}} EIBF_{\ell 0}}\,. \tag{42}$$

For the IBF approach we generate one random split and we calculate the Bayes factor as in step (4).

5 Model-search algorithm

For any number of variables p under consideration in our model-uncertainty problem, the number of models for which we need to evaluate the marginal likelihood is 2^p, which is enormous even when p is only moderately large. As a result, full enumeration (across all models) of the marginal likelihoods and the corresponding posterior model probabilities needed in Bayesian variable-selection problems becomes infeasible. For this reason, in such problems, advanced MCMC methods are typically used as model-search algorithms to identify the most important models and variables. Estimation of posterior model odds can then be performed efficiently within reduced model spaces in which unimportant variables have been excluded, according to the model search algorithm; see Fouskakis, Ntzoufras and Draper (2009) for an example of this approach in practice.

When the marginal likelihood is given in closed form, we may use the MCMC model composition (MC^3: Madigan and York, 1995) method, which is a simple Metropolis algorithm that can be employed to explore large model spaces. The MC^3 algorithm can be summarized as follows:

(1) For the current model $m \in \mathcal{M}$, propose a move to model $m' \in \mathcal{M}$ with probability $j(m, m')$.

(2) Calculate and store the marginal likelihood $f(\boldsymbol{y}|m')$ of model m'.

(3) Set $m = m'$ (i.e., accept the proposed model m') with probability $\alpha = \min\left[1, \frac{f(m'|\boldsymbol{y})}{f(m|\boldsymbol{y})}\frac{j(m',m)}{j(m,m')}\right]$.

(4) Store m as the current model.

(5) Repeat steps (1)–(4) until a target number of models is visited or a pre-specified CPU budget is exhausted.

Posterior model probabilities can be estimated in two ways: by considering the marginal likelihoods of the visited and proposed models stored in step (2), or by a frequency tabulation of the visited models in the output of the MCMC sampler.

When the marginal likelihood is not analytically tractable under the PEP prior, it can be obtained by one of the Monte-Carlo schemes described in Section 3 of the main paper and in Section 3 of this Appendix. We can exploit the fact that, in order to estimate the marginal likelihood of any model $M_\ell \in \mathcal{M}$, in the third and fourth Monte-Carlo schemes we need only to sample $\boldsymbol{y}^*$ from $m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)$ and $m_0^N(\boldsymbol{y}^*|\boldsymbol{y}; \mathrm{X}_0, \mathrm{X}_0^*, \delta)$, respectively.

Here (when marginal likelihoods are not available analytically) we propose to modify the standard MC^3 method by sampling a binary vector $\boldsymbol{\gamma}$ indicating the variables included in the model (see, e.g., George and McCulloch, 1993), using a Metropolis-within-Gibbs approach as follows.

(1) Generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_0^N(\boldsymbol{y}^*|\mathrm{X}_0^*, \delta)$ (this is the third Monte-Carlo marginal likelihood scheme) or $m_0^N(\boldsymbol{y}^*|\boldsymbol{y}; \mathrm{X}_0, \mathrm{X}_0^*, \delta)$ (this is the fourth scheme).

(2) For the current model M_ℓ, corresponding to the set of variable-inclusion indicators $\boldsymbol{\gamma}_\ell$, repeat the following:

   For $j = 1, \ldots, p$ (selected in random order), repeat the following steps:

   (a) Propose $\gamma_j' = 1 - \gamma_j$ with probability one.
   (b) Keep the other covariates the same: $\gamma_l' = \gamma_l$ for all $l \neq j$.
   (c) Identify $M_{\ell'}$ corresponding to the vector $\boldsymbol{\gamma}_{\ell'}$ with elements γ_k', $k = 1, \ldots, p$.
   (d) If $M_{\ell'}$ is not previously visited, calculate and store its estimated marginal likelihood $f(\boldsymbol{y}|M_{\ell'}) = \hat{m}_{\ell'}^{PEP}(\boldsymbol{y}|\mathrm{X}_{\ell'}, \mathrm{X}_{\ell'}^*, \delta)$ given by equation (37) or (38) in this Appendix for the third or fourth schemes, respectively.
   (e) Set $M_\ell = M_{\ell'}$ (i.e., accept the proposed model $M_{\ell'}$) with probability

$$\alpha = \min\left[1, \frac{f(M_{\ell'}|\boldsymbol{y})}{f(M_\ell|\boldsymbol{y})}\right] = \min\left[1, \frac{\hat{m}_{\ell'}^{PEP}(\boldsymbol{y}|\mathrm{X}_{\ell'}, \mathrm{X}_{\ell'}^*, \delta)}{\hat{m}_{\ell}^{PEP}(\boldsymbol{y}|\mathrm{X}_{\ell}, \mathrm{X}_{\ell}^*, \delta)}\frac{f(M_{\ell'})}{f(M_\ell)}\right], \quad (43)$$

   where $f(M_\ell)$ is the prior probability for model ℓ.

(3) Store M_ℓ as the current model.

(4) Repeat steps (2)–(3) until a target number of models is visited or a pre-specified CPU budget is exhausted.

For the first and second Monte-Carlo marginal-likelihood estimates, we start the above MC^3 algorithm from step (2) and in step (2)(d) we generate $\boldsymbol{y}^{*(t)}$ $(t = 1, \ldots, T)$ from $m_{\ell'}^N(\boldsymbol{y}^*|\,\boldsymbol{y}; \mathrm{X}_\ell, \mathrm{X}_\ell^*, \delta)$, which now depends on the proposed model, and estimate the marginal likelihood of that model using expressions (25) and (26) from the main paper, respectively.

6 Ozone data set details

Here we present a brief description of the data and the transformations used in the analyses of Section 5.2 of the main paper, based on the original ozone data.

- The **response variable** ozone (daily maximum of 24 hourly averages: midnight–1am, 1–2am, ..., 11pm–midnight) had substantial positive skew; within the Box-Cox power transformation family the optimal transformation was $ozone^{0.17}$. This is not far from the log transform, which is easier to interpret, so we standardized log(ozone) (subtracting off its mean and dividing by its standard deviation: all of the standardizations here and below are to stabilize the numerical work and reduce the correlations between main effects and their squares); the standardized variable is called `s_log_ozone`.

- **Month, day and weekday predictors**: Weekday had no effect on ozone and was omitted. We combined month and day to create a variable called `day_of_year` that ran through the consecutive integers 1–366 from 1 Jan to 31 Dec, and we then standardized this variable; the standardized variable is called `s_day_of_year` . We then created a variable called `s_day_of_year_2` by squaring `s_day_of_year`.

- **Temperature predictors**: Temperature at Sandburg and temperature at El Monte were very highly correlated, and the El Monte temperature variable had 139 missing values (versus only 2 missing values for the Sandburg temperature), so we omitted the El Monte temperature variable.

- The **remaining 8 covariates** were `pressure_500`, `humidity`, `temp_sandburg`, `inversion_height`, `wind`, `pressure_gradient`, `inversion_temp`, and `visibility`; these were standardized as above (the variable names of the resulting standardized versions are the same as those of the original variables preceded by `s_`). We also calculated squared versions of all standardized variables with a naming convention similar to that above (for example, the squared standardized wind variable was called `s_wind_2`).

- Omitting all rows of data for which one or more of the predictors were missing, our regression modeling was based on 330 days of data.

- Local-regression (`loess`) descriptive analyses of the relationships between the outcome `s_log_ozone` and each of our 9 predictor variables revealed cubic relationships between the outcome and the predictors `temp_sandburg` and `inversion_temp`, so we raised each of `s_temp_sandburg` and `s_inversion_temp` to the third power and included these two cubic terms among the total set of predictor variables.

- With 9 main effects there are $\frac{9 \cdot 8}{2} = 36$ pairwise interactions among the main effects; we also created these 36 variables by multiplying the standardized versions of the predictors in a pairwise manner. The resulting variables had names of the form `humidityXwind`.

- Our total set of predictor variables therefore had 9 main effects, 9 quadratic terms, 2 cubic terms, and 36 two-way interactions, for a total of 56 predictors.

- The final data set contains 330 rows and 57 columns; the first column is `s_log_ozone`, and the other 56 columns are the predictors.

Additional details on the ozone data analysis are available in a supplemental document provided on request; our version of the data set is also available from us.

7 Reduced model space in the ozone example

Index	Name
1	Day of year
2	Wind speed at LAX
5	Temperature at Sandburg
7	PG from LAX to Daggett
8	Inversion base temperature at LAX
9	Visibility at LAX
10	(Day of year)^2
12	(500 mb pressure height at VAFB)^2
13	(Humidity at LAX)^2
15	(Inversion base height at LAX)^2
16	(PG from LAX to Daggett)^2
18	(Visibility at LAX)^2
20	(Inversion base temperature at LAX)^3
23	(Day of year) × (Humidity at LAX)
26	(Day of year) × (PG from LAX to Daggett)
30	(Wind speed at LAX) × (Humidity at LAX)
36	(500 mb pressure height at VAFB) × (Humidity at LAX)
39	(500 mb pressure height at VAFB) × (PG from LAX to Daggett)
42	(Humidity at LAX) × (Temperature at Sandburg)
43	(Humidity at LAX) × (Inversion base height at LAX)
48	(Temperature at Sandburg) × (PG from LAX to Daggett)
51	(Inversion base height at LAX) × (PG from LAX to Daggett)

Notes: (1) Abbreviations used in this table: LAX = Los Angeles International Airport, mb = millibar, VAFB = Vandenberg Air Force Base, PG = pressure gradient (mm Hg). (2) Wind speed is measured in mph, temperature and inversion base temperature in °F, visibility in miles, inversion base height in feet, humidity in % and pressure height in m. (3) As mentioned in the text, all variables were standardized (mean 0, standard deviation 1) before exploring quadratic/cubic terms and interactions, to minimize collinearity.

References in the Appendix

Fouskakis, D., Ntzoufras, I. and Draper, D. (2009). Bayesian variable selection using cost-adjusted BIC, with application to cost-effective measurement of quality of health care. *Annals of Applied Statistics*, **3**, 663–690.

George, E. and McCulloch, R. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, **88**, 881–889.

Madigan, D. and York, J. (1995). Bayesian graphical models for discrete data. *International Statistical Review*, **63**, 215–232.

Perez, J. M. and Berger, J. O. (2002). Expected-posterior prior distributions for model selection. *Biometrika*, **89**, 491–511.

Scott, J. (1997). *Matrix Analysis for Statistics*. Wiley Series in Probability and Statistics, New York, USA.