Supplement to "Incorporating Historical Models with Adaptive Bayesian Updates"

Philip S. Boonstra   Ryan P. Barbaro

S1 $Z_{\mathbf{NAB}}(\phi, \eta, \tau, \lambda)$

$Z_{\text{NAB}}(\phi, \eta, \tau, \lambda) = \left(\int_{\beta^o} N(\beta^o | m_\alpha, \eta \boldsymbol{S}_\alpha/\phi) \prod_{j=1}^p N(\beta_j | 0, \tilde\theta_j^2) d\beta^o\right)^{-1}$ is the normalizing constant from Equation (4.13) in the manuscript, where $\tilde\theta_j \equiv \tilde\theta_j(\phi, \tau, \lambda) = \left(\frac{1}{d^2} + \frac{1-\phi}{c^2\tau^2\lambda_j^2}\right)^{-1/2}$ since we only need to consider the elements corresponding to β^o. This must be included in the posterior calculation when any of ϕ, η, τ, or λ are random. For ease of notation, let $\boldsymbol{\Theta}_o = \text{diag}\{\tilde\theta_1^2, \tilde\theta_2^2, \ldots, \tilde\theta_p^2\}$ be the $p \times p$ diagonal matrix comprised of the prior variances on β^o from the regularized horseshoe prior. We can calculate its inverse, $Z_{\text{NAB}}(\phi, \eta, \tau, \lambda)^{-1}$, as follows:

$$
\begin{aligned}
&Z_{\text{NAB}}(\phi, \eta, \tau, \lambda)^{-1} \\
&= \int_{\beta^o} N(\beta^o | m_\alpha, \eta \boldsymbol{S}_\alpha/\phi) N(\beta^o | 0, \boldsymbol{\Theta}_o) d\beta^o \\
&\quad= \int_{\beta^o} |\eta\phi^{-1}\boldsymbol{S}_\alpha|^{-1/2} \exp\{-(1/2)(\beta^o - m_\alpha)^\top (\eta\boldsymbol{S}_\alpha/\phi)^{-1} (\beta^o - m_\alpha)\} \\
&\qquad\quad \times |\boldsymbol{\Theta}_o|^{-1/2} \exp\{-(1/2)\beta^{o\top}\boldsymbol{\Theta}_o^{-1}\beta^o\} d\beta^o \\
&\quad= |\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} \int_{\beta^o} \exp\{-(1/2)(\beta^o - m_\alpha)^\top (\eta\boldsymbol{S}_\alpha/\phi)^{-1} (\beta^o - m_\alpha) - (1/2)\beta^{o\top}\boldsymbol{\Theta}_o^{-1}\beta^o\} d\beta^o \\
&\quad= |\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} \\
&\qquad \times \int_{\beta^o} \exp\left\{-(1/2)\left[\beta^{o\top}\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}\right)\beta^o - 2\beta^{o\top}\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta\right){m_\alpha}^\top + m_\alpha^\top (\eta\boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha\right]\right\} d\beta^o \\
&\quad= |\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} \exp\{-(1/2)m_\alpha^\top (\eta\boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha\} \\
&\qquad \times \int_{\beta^o} \exp\left\{-(1/2)\left[\beta^{o\top}\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}\right)\beta^o - 2\beta^{o\top}\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta\right)m_\alpha\right]\right\} d\beta^o
\end{aligned}
$$

Define $\boldsymbol{S}_\alpha^* = (\phi \boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1})^{-1}$ and $m_\alpha^* = \boldsymbol{S}_\alpha^*(\phi \boldsymbol{S}_\alpha^{-1}/\eta) m_\alpha$. Then the expression evaluates to

$$\begin{aligned}
&|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} \exp\{-(1/2) m_\alpha^\top (\eta \boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha\} \\
&\quad \times \int_{\beta^o} \exp\left\{-(1/2)\left[\beta^{o\top}\boldsymbol{S}_\alpha^{*-1}\beta^o - 2\beta^{o\top}\boldsymbol{S}_\alpha^{*-1} m_\alpha^*\right]\right\} d\beta^o \\
&= |\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} \exp\{-(1/2) m_\alpha^\top (\eta \boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha\} \\
&\quad \times \exp\{(1/2) m_\alpha^{*\top}\boldsymbol{S}_\alpha^{*-1} m_\alpha^*\} \int_{\beta^o} \exp\{-(1/2)(\beta^o - m_\alpha^*)^\top \boldsymbol{S}_\alpha^{*-1}(\beta^o - m_\alpha^*)\} d\beta^o \\
&= |\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} \exp\{-(1/2) m_\alpha^\top (\eta \boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha\} \\
&\quad \times \exp\{(1/2) m_\alpha^{*\top}\boldsymbol{S}_\alpha^{*-1} m_\alpha^*\} \times |\boldsymbol{S}_\alpha^*|^{1/2} \\
&= |\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2} |\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}|^{-1/2} \\
&\quad \times \exp\{-(1/2) m_\alpha^\top (\eta \boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha\} \exp\{(1/2) m_\alpha^\top (\phi\boldsymbol{S}_\alpha^{-1}/\eta)\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}\right)^{-1}(\phi\boldsymbol{S}_\alpha^{-1}/\eta) m_\alpha\}.
\end{aligned}$$

Combining the expressions for the determinants, we have $|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o|^{-1/2}|\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}|^{-1/2} = |\eta\phi^{-1}\boldsymbol{S}_\alpha + \boldsymbol{\Theta}_o|^{-1/2}$. We can combine the quadratic form in the exponential term as follows:

$$\begin{aligned}
&\exp\left\{-(1/2)\left[m_\alpha^\top (\eta\boldsymbol{S}_\alpha/\phi)^{-1} m_\alpha - (\phi/\eta)^2 m_\alpha^\top \boldsymbol{S}_\alpha^{-1}\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}\right)^{-1}\boldsymbol{S}_\alpha^{-1} m_\alpha\right]\right\} \\
&\quad = \exp\left\{-(1/2) m_\alpha^\top \left[\phi\boldsymbol{S}_\alpha^{-1}/\eta - (\phi/\eta)^2\boldsymbol{S}_\alpha^{-1}\left(\phi\boldsymbol{S}_\alpha^{-1}/\eta + \boldsymbol{\Theta}_o^{-1}\right)^{-1}\boldsymbol{S}_\alpha^{-1}\right] m_\alpha\right\} \\
&\quad = \exp\left\{-(1/2) m_\alpha^\top \left[\eta\phi^{-1}\boldsymbol{S}_\alpha + \boldsymbol{\Theta}_o\right]^{-1} m_\alpha\right\}.
\end{aligned}$$

The last expression comes from the Woodbury matrix identity and gives that

$$Z_{\text{NAB}}(\phi,\eta,\tau,\lambda) = |\eta\phi^{-1}\boldsymbol{S}_\alpha + \boldsymbol{\Theta}_o|^{1/2} \exp\left\{(1/2) m_\alpha^\top \left[\eta\phi^{-1}\boldsymbol{S}_\alpha + \boldsymbol{\Theta}_o\right]^{-1} m_\alpha\right\} \qquad \text{(S1)}$$

S2 $Z_{\mathbf{SAB}}(\phi,\eta,\tau,\lambda)$

$Z_{\text{SAB}}(\phi,\eta,\tau,\lambda) = \left(\iint_{\beta^o,\beta^a} N(\{\beta^o + \boldsymbol{P}\beta^a\}|m_\alpha, \eta\boldsymbol{S}_\alpha/\phi)\prod_{j=1}^{p+q} N(\beta_j|0,\tilde{\theta}_j^2) d\beta^o d\beta^a\right)^{-1}$ is the normalizing constant from Equation (4.14) in the manuscript, where

$$\tilde{\theta}_j(\phi,\tau,\lambda) = \begin{cases} \left(\dfrac{1}{d^2} + \dfrac{1-\phi}{c^2\tau^2\lambda_j^2}\right)^{-1/2}, & j = 1,\ldots,p \\ \left(\dfrac{1}{d^2} + \dfrac{1}{c^2\tau^2\lambda_j^2}\right)^{-1/2}, & j = p+1,\ldots,p+q \end{cases}$$

As with NAB, the value of $Z_{\text{SAB}}(\phi,\eta,\tau,\lambda)$ must be included in the posterior calculation when any of ϕ, η, τ, or λ are random. As before, let $\boldsymbol{\Theta}_o = \text{diag}\{\tilde{\theta}_1^2, \tilde{\theta}_2^2, \ldots, \tilde{\theta}_p^2\}$ be the $p \times p$ diagonal matrix comprised of the prior variances on β^p. from the regularized horseshoe prior, and let $\mathbf{\Theta}_a = \text{diag}\{\tilde{\theta}^2_{p+1}, \tilde{\theta}^2_{p+2}, \ldots, \tilde{\theta}^2_{p+q}\}$ be the analogous $q \times q$ diagonal matrix corresponding to β^a. To simplify the double integral, we first break up the historical prior as follows:

$$
\begin{aligned}
&N(\{\beta^o + \boldsymbol{P}\beta^a\} | m_\alpha, \eta \boldsymbol{S}_\alpha/\phi) \\
&\quad = |\eta\phi^{-1}\boldsymbol{S}_\alpha|^{-1/2} \exp\left\{-(1/2)\left(\beta^o + \boldsymbol{P}\beta^a - m_\alpha\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\beta^o + \boldsymbol{P}\beta^a - m_\alpha\right)\right\} \\
&\quad = |\eta\phi^{-1}\boldsymbol{S}_\alpha|^{-1/2} \exp\left\{-(1/2)\left(\beta^o - m_\alpha\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\beta^o - m_\alpha\right)\right\} \\
&\qquad \times \exp\left\{-(1/2)\left(\boldsymbol{P}\beta^a\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\boldsymbol{P}\beta^a\right)\right\} \\
&\qquad \times \exp\left\{\left(\beta^o - m_\alpha\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\boldsymbol{P}\beta^a\right)\right\} \\
&\quad = |\eta\phi^{-1}\boldsymbol{S}_\alpha|^{-1/2} \exp\left\{-(1/2)\left(\beta^o - m_\alpha\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\beta^o - m_\alpha\right)\right\} \\
&\qquad \times \exp\left\{-(1/2)\beta^{a\top}\boldsymbol{P}^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\} \\
&\qquad \times \exp\left\{\beta^{o\top} \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\} \\
&\qquad \times \exp\left\{-m_\alpha^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}
\end{aligned}
$$

The integral in the calculation of $Z_{\text{SAB}}(\phi, \eta, \tau, \lambda)$ is then given as follows:

$$
\begin{aligned}
&\iint_{\beta^o,\beta^a} N(\{\beta^o + \boldsymbol{P}\beta^a\} | m_\alpha, \eta \boldsymbol{S}_\alpha/\phi) N(\beta^o | 0, \mathbf{\Theta}^o) N(\beta^a | 0, \mathbf{\Theta}^a) d\beta^o d\beta^a \qquad \text{(S2)} \\
&\quad = \iint_{\beta^o,\beta^a} |\eta\phi^{-1}\boldsymbol{S}_\alpha|^{-1/2} \exp\left\{-(1/2)\left(\beta^o - m_\alpha\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\beta^o - m_\alpha\right)\right\} \\
&\qquad \times \exp\left\{-(1/2)\beta^{a\top}\boldsymbol{P}^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\} \\
&\qquad \times \exp\left\{\beta^{o\top} \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\} \\
&\qquad \times \exp\left\{-m_\alpha^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\} \\
&\qquad \times |\mathbf{\Theta}_o|^{-1/2} \exp\{-(1/2)\beta^{o\top}\mathbf{\Theta}_o^{-1}\beta^o\} \\
&\qquad \times |\mathbf{\Theta}_a|^{-1/2} \exp\{-(1/2)\beta^{a\top}\mathbf{\Theta}_a^{-1}\beta^a\} d\beta^o d\beta^a
\end{aligned}
$$

Using the same identity used in calculating the normalizing constant in the Adaptive Bayesian Updating scheme, let $\boldsymbol{S}^*_\alpha = (\phi\boldsymbol{S}_\alpha^{-1}/\eta + \mathbf{\Theta}_o^{-1})^{-1}$ and $m^*_\alpha = \boldsymbol{S}^*_\alpha(\phi\boldsymbol{S}_\alpha^{-1}/\eta)m_\alpha$. Then, we have that

$$
\begin{aligned}
&\exp\left\{-(1/2)\left(\beta^o - m_\alpha\right)^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\beta^o - m_\alpha\right)\right\} \exp\{-(1/2)\beta^{o\top}\mathbf{\Theta}_o^{-1}\beta^o\} \\
&= \exp\{-(1/2)m_\alpha^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1} m_\alpha\} \exp\left\{-(1/2)\left[\beta^{o\top}\boldsymbol{S}^{*-1}_\alpha\beta^o - 2\beta^{o\top}\boldsymbol{S}^{*-1}_\alpha m^*_\alpha\right]\right\} \\
&= \exp\{-(1/2)m_\alpha^\top \left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1} m_\alpha\} \exp\{(1/2)m^{*\top}_\alpha\boldsymbol{S}^{*-1}_\alpha m^*_\alpha\} \\
&\quad \times \exp\left\{-(1/2)\left(\beta^o - m^*_\alpha\right)^\top \boldsymbol{S}^{*-1}_\alpha\left(\beta^o - m^*_\alpha\right)\right\}
\end{aligned}
$$

Then, (S2) simplifies to

$$
\begin{aligned}
&|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{-1/2}|\boldsymbol{S}^*_\alpha|^{1/2}\exp\{-(1/2)m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\}\exp\{(1/2)m^{*\top}_\alpha\boldsymbol{S}^{*-1}_\alpha m^*_\alpha\}\\
&\quad\times\int_{\beta^a}\exp\left\{-(1/2)\beta^{a\top}\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\left\{-m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\{-(1/2)\beta^{a\top}\boldsymbol{\Theta}_a^{-1}\beta^a\}\\
&\qquad\int_{\beta^o}|\boldsymbol{S}^*_\alpha|^{-1/2}\exp\left\{\beta^{o\top}\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\left\{-(1/2)\left(\beta^o-m^*_\alpha\right)^\top\boldsymbol{S}^{*-1}_\alpha\left(\beta^o-m^*_\alpha\right)\right\}d\beta^o d\beta^a\\
&=|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{-1/2}|\boldsymbol{S}^*_\alpha|^{1/2}\exp\{-(1/2)m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\}\exp\{(1/2)m^{*\top}_\alpha\boldsymbol{S}^{*-1}_\alpha m^*_\alpha\}\\
&\quad\times\int_{\beta^a}\exp\left\{-(1/2)\beta^{a\top}\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\left\{-m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\{-(1/2)\beta^{a\top}\boldsymbol{\Theta}_a^{-1}\beta^a\}\\
&\qquad\exp\left\{m^{*\top}_\alpha\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\left\{(1/2)\beta^{a\top}\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{S}^*_\alpha\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}d\beta^a
\end{aligned}
\tag{S3}
$$

The equality immediately above comes from an application of the multivariate normal moment-generating function. Define

$$
\begin{aligned}
\boldsymbol{S}^{**}_\alpha&=\left(\boldsymbol{\Theta}_a^{-1}+\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}-\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{S}^*_\alpha\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\right)^{-1}\\
&=\left(\boldsymbol{\Theta}_a^{-1}+\boldsymbol{P}^\top\left[\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}-\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{S}^*_\alpha\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\right]\boldsymbol{P}\right)^{-1}\\
&=\left(\boldsymbol{\Theta}_a^{-1}+\boldsymbol{P}^\top\left[\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}-\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}+\boldsymbol{\Theta}_o^{-1}\right)^{-1}\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\right]\boldsymbol{P}\right)^{-1}\\
&=\left(\boldsymbol{\Theta}_a^{-1}+\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi+\boldsymbol{\Theta}_o\right)^{-1}\boldsymbol{P}\right)^{-1}
\end{aligned}
$$

and

$$
\begin{aligned}
m^{**}_\alpha&=\boldsymbol{S}^{**}_\alpha\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(m_\alpha-m^*_\alpha\right)\\
&=\boldsymbol{S}^{**}_\alpha\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(m_\alpha-\boldsymbol{S}^*_\alpha\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\right)\\
&=\boldsymbol{S}^{**}_\alpha\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(m_\alpha-\left(\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}+\boldsymbol{\Theta}_o^{-1}\right)^{-1}\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\right)\\
&=\boldsymbol{S}^{**}_\alpha\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\left(\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}+\boldsymbol{\Theta}_o^{-1}\right)^{-1}\boldsymbol{\Theta}_o^{-1}m_\alpha
\end{aligned}
$$

Then, continuing from Equation (S3),

$$
\begin{aligned}
&|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{-1/2}|\boldsymbol{S}^*_\alpha|^{1/2}\exp\{-(1/2)m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\}\exp\{(1/2)m_\alpha^{*\top}\boldsymbol{S}_\alpha^{*-1}m^*_\alpha\}\\
&\quad\times\int_{\beta^a}\exp\left\{-(1/2)\beta^{a\top}\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\left\{-m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\{-(1/2)\beta^{a\top}\boldsymbol{\Theta}_a^{-1}\beta^a\}\\
&\qquad\exp\left\{m_\alpha^{*\top}\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}\exp\left\{(1/2)\beta^{a\top}\boldsymbol{P}^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{S}^*_\alpha\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}\boldsymbol{P}\beta^a\right\}d\beta^a\\
&=|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{-1/2}|\boldsymbol{S}^*_\alpha|^{1/2}\exp\{-(1/2)m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\}\exp\{(1/2)m_\alpha^{*\top}\boldsymbol{S}_\alpha^{*-1}m^*_\alpha\}\\
&\quad\times\int_{\beta^a}\exp\left\{-1/2\left[\beta^{a\top}\boldsymbol{S}_\alpha^{**-1}\beta^a-2\beta^{a\top}\boldsymbol{S}_\alpha^{**-1}m_\alpha^{**}\right]\right\}d\beta^a\\
&=|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{-1/2}|\boldsymbol{S}^*_\alpha|^{1/2}\exp\{-(1/2)m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\}\exp\{(1/2)m_\alpha^{*\top}\boldsymbol{S}_\alpha^{*-1}m^*_\alpha\}\\
&\quad\times\exp\{(1/2)m_\alpha^{**\top}\boldsymbol{S}_\alpha^{**-1}m_\alpha^{**}\}|\boldsymbol{S}_\alpha^{**}|^{1/2}\\
&\quad\times\int_{\beta^a}|\boldsymbol{S}_\alpha^{**}|^{-1/2}\exp\left\{-1/2(\beta^a-m_\alpha^{**})^\top\boldsymbol{S}_\alpha^{**-1}(\beta^a-m_\alpha^{**})\right\}d\beta^a
\end{aligned}
$$

The integrand is proportional to a density, and therefore the integral is proportional to a constant not depending on any parameters. Thus, the normalizing constant $Z_{\text{SAB}}(\phi,\eta,\tau,\lambda)$ is equal to the reciprocal of the above expression after dropping the integral:

$$
\begin{aligned}
&Z_{\text{SAB}}(\phi,\eta,\tau,\lambda)\\
&\quad=|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{1/2}|\boldsymbol{S}^*_\alpha|^{-1/2}|\boldsymbol{S}^{**}_\alpha|^{-1/2}\\
&\qquad\times\exp\{(1/2)m_\alpha^\top\left(\eta\boldsymbol{S}_\alpha/\phi\right)^{-1}m_\alpha\}\\
&\qquad\times\exp\{-(1/2)m_\alpha^{*\top}\boldsymbol{S}_\alpha^{*-1}m^*_\alpha\}\\
&\qquad\times\exp\{-(1/2)m_\alpha^{**\top}\boldsymbol{S}_\alpha^{**-1}m^{**}_\alpha\}\\
&\quad=|\eta\phi^{-1}\boldsymbol{S}_\alpha\boldsymbol{\Theta}_o\boldsymbol{\Theta}_a|^{1/2}|\eta^{-1}\phi\boldsymbol{S}_\alpha^{-1}+\boldsymbol{\Theta}_o^{-1}|^{1/2}|\boldsymbol{S}^{**}_\alpha|^{-1/2}\\
&\qquad\times\exp\left\{(1/2)m_\alpha^\top\left[\eta\phi^{-1}\boldsymbol{S}_\alpha+\boldsymbol{\Theta}_o\right]^{-1}m_\alpha\right\}\\
&\qquad\times\exp\{-(1/2)m_\alpha^{**\top}\boldsymbol{S}_\alpha^{**-1}m^{**}_\alpha\}\\
&\quad=|\eta\phi^{-1}\boldsymbol{S}_\alpha+\boldsymbol{\Theta}_o+\boldsymbol{P}\boldsymbol{\Theta}_a\boldsymbol{P}^\top|^{1/2}\\
&\qquad\times\exp\left\{(1/2)m_\alpha^\top\left(\eta\phi^{-1}\boldsymbol{S}_\alpha+\boldsymbol{\Theta}_o+\boldsymbol{P}\boldsymbol{\Theta}_a\boldsymbol{P}^\top\right)^{-1}m_\alpha\right\}
\end{aligned}
\tag{S4}
$$

S3 STAN Code

Regularized Horseshoe

```
// Regularized Horseshoe Prior
data {
  int<lower = 1> n_stan; // n_curr
  int<lower = 1> p_stan; // number of original covariates
  int<lower = 0> q_stan; // number of augmented (or added) covariates
  int<lower = 0,upper = 1> y_stan[n_stan]; // outcome
  matrix[n_stan, p_stan + q_stan] x_scaled_stan; //covariates (no intercept)
```

```
  real<lower = 0> local_dof_stan; // dof of pi(lambda), = 1
  real<lower = 0> global_dof_stan; // dof of pi(tau), = 1
  real<lower = 0> beta_orig_scale_stan; // c, Section 2
  real<lower = 0> beta_aug_scale_stan; // c, Section 2
  real<lower = 0> slab_precision_stan; // 1/d^2, Section 2
  vector<lower = 0,upper = 1>[n_stan] intercept_offset_stan;//allow for up to two intercepts
  int<lower = 0,upper = 1> only_prior;//if 1, ignore the model and data and generate from the prior only
}
parameters {
  real mu;
  real mu_offset;
  vector[p_stan + q_stan] beta_raw;
  // tau is decomposed into chi-square and inverse-gamma portions
  real<lower = 0> tau_glob_base_sq;
  real<lower = 0> tau_glob_scale_sq;
  // lambdas are decomposed into chi-square and inverse-gamma portions
  vector<lower = 0>[p_stan] lambda_orig_base_sq;
  vector<lower = 0>[p_stan] lambda_orig_scale_sq;
  vector<lower = 0>[q_stan] lambda_aug_base_sq;
  vector<lower = 0>[q_stan] lambda_aug_scale_sq;
}
transformed parameters {
  vector[p_stan + q_stan] beta;
  vector<lower = 0,upper = sqrt(1/slab_precision_stan)>[p_stan] beta_orig_scale;//theta
  vector<lower = 0,upper = sqrt(1/slab_precision_stan)>[q_stan] beta_aug_scale;//theta
  real<lower = 0> tau_glob_sq;//tau^2
  vector<lower = 0>[p_stan] lambda_orig_sq;//lambda^2
  vector<lower = 0>[q_stan] lambda_aug_sq;//lambda^2
  tau_glob_sq = tau_glob_base_sq * tau_glob_scale_sq;
  lambda_orig_sq = lambda_orig_base_sq .* lambda_orig_scale_sq;
  lambda_aug_sq = lambda_aug_base_sq .* lambda_aug_scale_sq;
  beta_orig_scale = 1 ./ sqrt(slab_precision_stan + (1 ./ (beta_orig_scale_stan^2 * tau_glob_sq * lambda_orig_sq)));
  beta_aug_scale = 1 ./ sqrt(slab_precision_stan + (1 ./ (beta_aug_scale_stan^2 * tau_glob_sq * lambda_aug_sq)));
  beta = append_row(beta_orig_scale, beta_aug_scale) .* beta_raw;
}
model {
  beta_raw ~ normal(0.0, 1.0);
  tau_glob_base_sq ~ chi_square(1.0);
  tau_glob_scale_sq ~ inv_gamma(global_dof_stan/2.0, global_dof_stan/2.0);
  lambda_orig_base_sq ~ chi_square(1.0);
  lambda_orig_scale_sq ~ inv_gamma(local_dof_stan/2.0, local_dof_stan/2.0);
  lambda_aug_base_sq ~ chi_square(1.0);
  lambda_aug_scale_sq ~ inv_gamma(local_dof_stan/2.0, local_dof_stan/2.0);
  mu ~ logistic(0.0, 5.0);
  mu_offset ~ logistic(0.0, 2.5);
  if(only_prior == 0)
    y_stan ~ bernoulli_logit(mu + intercept_offset_stan * mu_offset + x_scaled_stan * beta);
}
```

Naive Adaptive Bayes

For computational expediency, we re-characterize the historical part of the prior. Let $\boldsymbol{S}_\alpha \equiv \boldsymbol{QDQ}^\top$ denote the eigendecomposition of the posterior covariance matrix of α, where $\boldsymbol{Q}$ is the matrix of eigenvectors and $\boldsymbol{D}$ is a diagonal matrix of the eigenvalues.

$$\begin{aligned}
&\beta^o \sim N(m_\alpha, \eta \boldsymbol{S}_\alpha/\phi) \\
\Rightarrow &\beta^o \sim N\left(m_\alpha, \eta\left(\boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^\top\right)/\phi\right) \\
\Rightarrow &\frac{\boldsymbol{Q}^\top(\beta^o - m_\alpha)}{(\phi/\eta)^{-1/2}} \sim N(0, \boldsymbol{D}) \\
\Rightarrow &\boldsymbol{\Gamma}\boldsymbol{Q}^\top(\beta^o - m_\alpha) \dot{\sim} N(0, \boldsymbol{D}),
\end{aligned} \tag{S5}$$

where $\boldsymbol{\Gamma}$ is a $p\times p$ diagonal matrix with the ith diagonal element given by $\boldsymbol{\Gamma}_{[ii]} = \left([1-\phi]\boldsymbol{S}_{\alpha[ii]}/225 + \phi/\eta\right)^{1/2}$. The term $(1-\phi)\boldsymbol{S}_{\alpha[ii]}/225$ requires some explanation; it is introduced for computational purposes. When $\phi \approx 0$, the left-hand side of (S5) is effectively zero but does not vanish. In other words, the marginal prior variances from the coefficients corresponding to each original predictor increase to 225 (on a standardized scale) rather than infinity. Numerically, this is nearly equivalent to using $(\phi/\eta)^{1/2}$ alone, but we found that using $\boldsymbol{\Gamma}$ as defined above improves the computational stability of our approach when draws of ϕ are small.

```
//Naive Adaptive Bayes
data {
  int<lower = 1> n_stan; // n_curr
  int<lower = 1> p_stan; // number of original covariates
  int<lower = 0> q_stan; // number of augmented (or added) covariates
  int<lower = 0,upper = 1> y_stan[n_stan]; // outcome
  matrix[n_stan, p_stan + q_stan] x_scaled_stan; // covariates (no intercept)
  vector[p_stan] beta_orig_prior_mean_stan; // prior mean of alpha, m_alpha
  matrix[p_stan, p_stan] beta_orig_prior_cov_stan; // prior covariance of alpha, S_alpha
  vector[p_stan] sqrt_eigenval_hist_var_stan; // sqrt of eigenvalues of S_alpha; D^1/2 in Equation (S5)
  matrix[p_stan, p_stan] eigenvec_hist_var_stan; // eigenvectors of S_alpha; Q^T in Equation (S5)
  real<lower = 0> local_dof_stan; // dof of pi(lambda), = 1
  real<lower = 0> global_dof_stan; // dof of pi(tau), = 1
  real<lower = 0> beta_orig_scale_stan; // c, Section 2
  real<lower = 0> beta_aug_scale_stan; // c, Section 2
  real<lower = 0> beta_aug_relaxed_scale_stan; // tilde~c = 0.05, Section 4.1
  real<lower = 0> slab_precision_stan; // 1/d^2, Section 2
  vector<lower = 0>[p_stan] scale_to_variance225; //Equation (S5); equal to diag(S_alpha) / 225;
  real<lower = 0, upper = 1> phi_mean; // mean of phi in (0,1) using normal distribution truncated to [0,1].
  real<lower = 0> phi_sd; // sd of phi using normal distribution truncated to [0,1]
  int<lower = 0,upper = 1> only_prior;//if 1, ignore the model and data and generate from the prior only
}
transformed data {
  vector[p_stan] zero_vec;
  zero_vec = rep_vector(0.0, p_stan);
}
parameters {
  real mu;
  vector[p_stan] beta_raw_orig; // unscaled
  vector[q_stan] beta_raw_aug; // unscaled
  real<lower = 0> eta;
  vector<lower=0>[q_stan] lambda_aug_sq_relaxed; // tilde lambda^2
  real<lower = 0> tau_glob;// tau
  vector<lower = 0>[p_stan] lambda_orig;// lambda^o
  vector<lower = 0>[q_stan] lambda_aug;// lambda^a
  real<lower = 0, upper = 1> phi;// phi
}
transformed parameters {
  vector[p_stan] normalized_beta; //entire LHS of Equation (S5)
  vector[p_stan] beta_orig; // scaled
  vector[q_stan] beta_aug; // scaled
  vector[p_stan + q_stan] beta;
  vector<lower = 0,upper = sqrt(1/slab_precision_stan)>[p_stan] beta_orig_scale; // theta
```

```
  vector<lower = 0,upper = sqrt(1/slab_precision_stan)>[q_stan] beta_aug_scale; // theta
  vector<lower = 0,upper = sqrt(1/min(scale_to_variance225))>[p_stan] hist_orig_scale; // Gamma in LHS of Eqn (S5)
  matrix[p_stan,p_stan] normalizing_cov;// S_alpha * hist_orig_scale  + Theta^o
  real<lower = 0, upper = 1> phi_copy; // copy of phi
  if(phi_sd > 0) {
    phi_copy = phi;
  } else {
    phi_copy = phi_mean;
  }
  beta_orig_scale = 1 ./ sqrt(slab_precision_stan +
   ((1 - phi_copy) ./ (tau_glob^2 * square(lambda_orig))));
  beta_aug_scale = 1 ./ sqrt(slab_precision_stan +
   ((1 - phi_copy) ./ (tau_glob^2 * square(lambda_aug))) +
(phi_copy ./ (beta_aug_relaxed_scale_stan^2 * lambda_aug_sq_relaxed)));
  beta_orig = beta_orig_scale .* beta_raw_orig;
  beta_aug = beta_aug_scale .* beta_raw_aug;
  beta = append_row(beta_orig, beta_aug);
  hist_orig_scale = 1 ./ sqrt(scale_to_variance225 * (1 - phi_copy) + phi_copy / eta);
  normalizing_cov = quad_form_diag(beta_orig_prior_cov_stan,hist_orig_scale);
  for(i in 1:p_stan) {
    normalizing_cov[i,i] = normalizing_cov[i,i] + beta_orig_scale[i]^2;
  }
  normalized_beta = eigenvec_hist_var_stan * (beta_orig - beta_orig_prior_mean_stan) ./ hist_orig_scale;
}
model {
  beta_raw_orig ~ normal(0.0, 1.0);
  beta_raw_aug ~ normal(0.0, 1.0);
  lambda_aug_sq_relaxed ~ inv_gamma(0.5, 0.5);
  eta ~ inv_gamma(2.5, 2.5);
  tau_glob ~ student_t(global_dof_stan, 0.0, 1.0);
  lambda_orig ~ student_t(local_dof_stan, 0.0, beta_orig_scale_stan);
  lambda_aug ~ student_t(local_dof_stan, 0.0, beta_aug_scale_stan);
  mu ~ logistic(0.0, 5.0);
  if(phi_sd > 0) {
    phi ~ normal(phi_mean, phi_sd);
  }
  // Equation (S5) The next two lines together comprise the naive adaptive prior contribution
  normalized_beta ~ normal(0.0, sqrt_eigenval_hist_var_stan);
  // Scaling normalized_beta to be independent ends up dropping a necessary determinant calculation: we add that back in here:
  target += -(1.0 * sum(log(hist_orig_scale)));
  // Z_NAB (Normalizing constant)
  target += -(1.0 * multi_normal_lpdf(beta_orig_prior_mean_stan|zero_vec, normalizing_cov));
  if(only_prior == 0)
    y_stan ~ bernoulli_logit(mu + x_scaled_stan * beta);
}
```

Sensible Adaptive Bayes

We apply the analogous decomposition to the historical prior as given in Equation (S5):

$$\begin{aligned}
&\beta^o + \boldsymbol{P}\beta^a \sim N(m_\alpha, \eta \boldsymbol{S}_\alpha/\phi) \\
\Rightarrow &\beta^o + \boldsymbol{P}\beta^a \sim N\left(m_\alpha, \eta\left(\boldsymbol{QDQ}^\top\right)/\phi\right) \\
\Rightarrow &\frac{\boldsymbol{Q}^\top(\beta^o + \boldsymbol{P}\beta^a - m_\alpha)}{(\phi/\eta)^{-1/2}} \sim N\left(0, \boldsymbol{D}\right) \\
\Rightarrow &\boldsymbol{\Gamma}\boldsymbol{Q}^\top(\beta^o + \boldsymbol{P}\beta^a - m_\alpha) \dot{\sim} N\left(0, \boldsymbol{D}\right),
\end{aligned} \tag{S6}$$

```
//Sensible Adaptive Bayes
data {
  int<lower = 1> n_stan; // n_curr
  int<lower = 1> p_stan; // number of original covariates
```

```
  int<lower = 0> q_stan; // number of augmented (or added) covariates
  int<lower = 0,upper = 1> y_stan[n_stan]; // outcome
  matrix[n_stan, p_stan + q_stan] x_scaled_stan; // covariates (no intercept)
  matrix[p_stan, q_stan] aug_projection_stan; // {v^o}^(-1) %*% (E[V^a|v^o]- E[V^a|v^o = 0]), Eqn (9)
  vector[p_stan] beta_orig_prior_mean_stan; // prior mean of alpha, m_alpha
  matrix[p_stan, p_stan] beta_orig_prior_cov_stan; // prior covariance of alpha, S_alpha
  vector[p_stan] sqrt_eigenval_hist_var_stan; // sqrt of eigenvalues of S_alpha; D^1/2 in Equation (S6)
  matrix[p_stan, p_stan] eigenvec_hist_var_stan; // eigenvectors of S_alpha; Q^T in Equation (S6)
  real<lower = 0> local_dof_stan; // dof of pi(lambda), = 1
  real<lower = 0> global_dof_stan; // dof of pi(tau), = 1
  real<lower = 0> beta_orig_scale_stan; // c, Section 2
  real<lower = 0> beta_aug_scale_stan; // c, Section 2
  real<lower = 0> slab_precision_stan; // 1/d^2, Section 2
  vector<lower = 0>[p_stan] scale_to_variance225; //Equation (S6); equal to diag(S_alpha) / 225;
  real<lower = 0, upper = 1> phi_mean; // mean of phi in (0,1) using normal distribution truncated to [0,1].
  real<lower = 0> phi_sd; // sd of phi using normal distribution truncated to [0,1]
  int<lower = 0,upper = 1> only_prior;//if 1, ignore the model and data and generate from the prior only
}
transformed data {
  vector[p_stan] zero_vec;
  zero_vec = rep_vector(0.0, p_stan);
}
parameters {
  real mu;
  vector[p_stan] beta_raw_orig; // unscaled
  vector[q_stan] beta_raw_aug; // unscaled
  real<lower = 0> eta;
  real<lower = 0> tau_glob; // tau
  vector<lower = 0>[p_stan] lambda_orig;// lambda^o
  vector<lower = 0>[q_stan] lambda_aug; // lambda^a
  real<lower = 0, upper = 1> phi;// phi
}
transformed parameters {
  vector[p_stan] normalized_beta; // entire LHS of Equation (S6)
  vector[p_stan] beta_orig; // scaled
  vector[q_stan] beta_aug; // scaled
  vector[p_stan + q_stan] beta;
  vector<lower = 0,upper = sqrt(1/slab_precision_stan)>[p_stan] beta_orig_scale;// theta
  vector<lower = 0,upper = sqrt(1/slab_precision_stan)>[q_stan] beta_aug_scale;// theta
  vector<lower = 0,upper = sqrt(1/min(scale_to_variance225))>[p_stan] hist_orig_scale;// Gamma in LHS of Eqn (S6)
  matrix[p_stan,p_stan] normalizing_cov;// S_alpha * hist_orig_scale  + Theta^o
  real<lower = 0, upper = 1> phi_copy;// copy of phi
  if(phi_sd > 0) {
    phi_copy = phi;
  } else {
    phi_copy = phi_mean;
  }
  beta_orig_scale = 1 ./ sqrt(slab_precision_stan + ((1 - phi_copy) ./ (tau_glob^2 * square(lambda_orig))));
  beta_aug_scale = 1 ./ sqrt(slab_precision_stan + (1 ./ (tau_glob^2 * square(lambda_aug))));
  beta_orig = beta_orig_scale .* beta_raw_orig;
  beta_aug = beta_aug_scale .* beta_raw_aug;
  beta = append_row(beta_orig, beta_aug);
  hist_orig_scale = 1 ./ sqrt(scale_to_variance225 * (1 - phi_copy) + phi_copy / eta);
  normalizing_cov = (quad_form_diag(beta_orig_prior_cov_stan,hist_orig_scale)) + tcrossprod(diag_post_multiply(aug_projection_st
  for(i in 1:p_stan) {
    normalizing_cov[i,i] = normalizing_cov[i,i] + beta_orig_scale[i]^2;
  }
  normalized_beta = eigenvec_hist_var_stan * (beta_orig + aug_projection_stan * beta_aug - beta_orig_prior_mean_stan) ./ hist_or
}
model {
  beta_raw_orig ~ normal(0.0, 1.0);
  beta_raw_aug ~ normal(0.0, 1.0);
  eta ~ inv_gamma(2.5, 2.5);
  tau_glob ~ student_t(global_dof_stan, 0.0, 1.0);
  lambda_orig ~ student_t(local_dof_stan, 0.0, beta_orig_scale_stan);
  lambda_aug ~ student_t(local_dof_stan, 0.0, beta_aug_scale_stan);
  mu ~ logistic(0.0, 5.0);
  if(phi_sd > 0) {
```

```
    phi ~ normal(phi_mean, phi_sd);
  }
  // Equation (S6) The next two lines together comprise the sensible adaptive prior contribution
  normalized_beta ~ normal(0.0, sqrt_eigenval_hist_var_stan);
  // Scaling normalized_beta to be independent ends up dropping a necessary determinant calculation: we add that back in here:
  target += -(1.0 * sum(log(hist_orig_scale)));
  // Z_SAB (Normalizing constant)
  target += -(1.0 * multi_normal_lpdf(beta_orig_prior_mean_stan|zero_vec, normalizing_cov));
  if(only_prior == 0)
    y_stan ~ bernoulli_logit(mu + x_scaled_stan * beta);
}
```

S4 Supplemental Figures

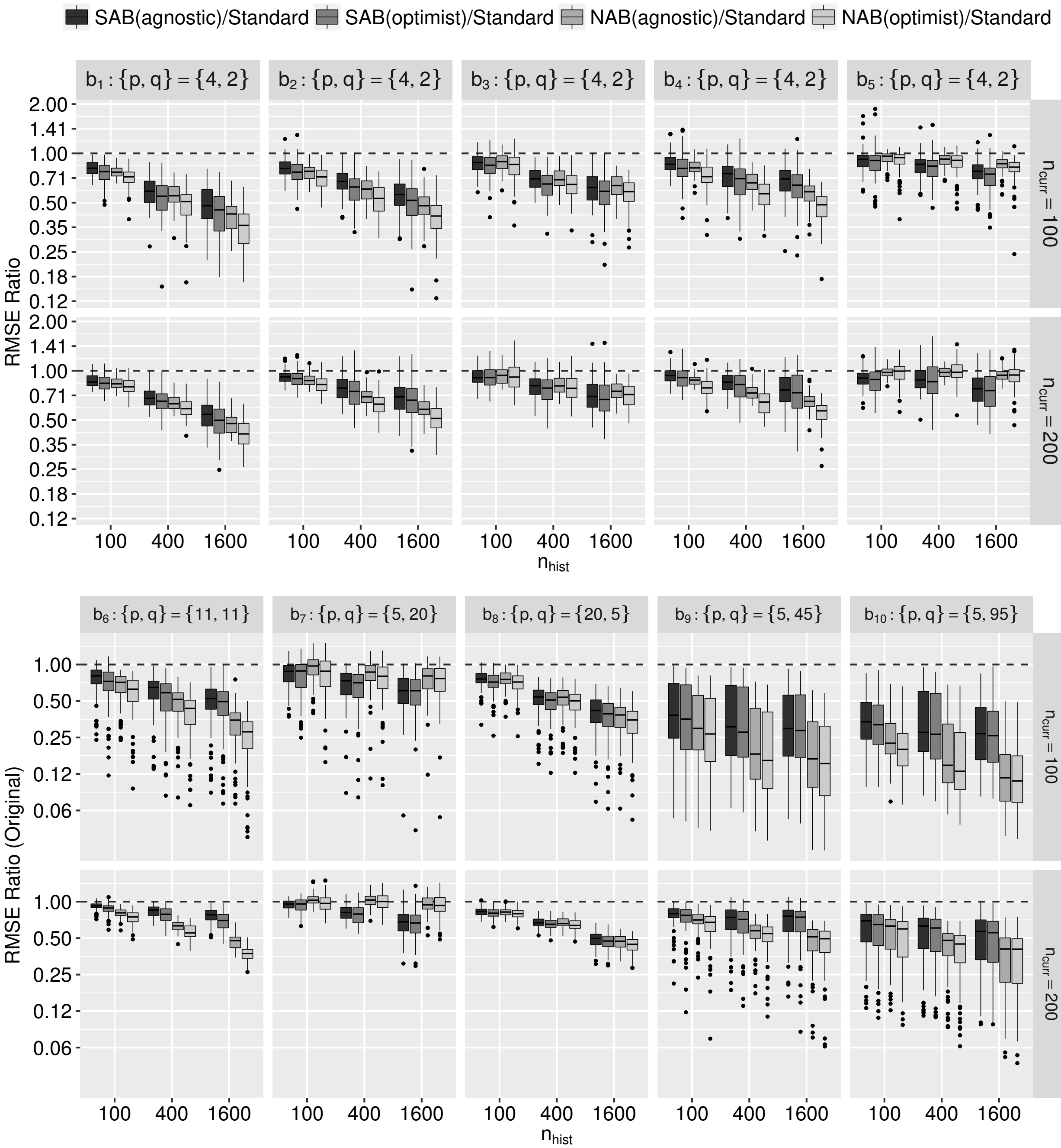

Figure S1: Boxplots of RMSE ratios **for β^o only** (y-axis, on the $\log_2$-scale) comparing four adaptive priors that make use of the historical information against a standard hierarchical shrinkage prior against varying sample sizes of the historical data (n_{hist}; x-axis) for ten true values of the regression coefficients (b_k, $k = 1, \ldots, 10$; columns) taken from Table 3 in the manuscript and varying sample sizes of the current data (n_{curr}; rows). Each boxplot compares the posteriors across 80 independent datasets. Smaller ratios indicate that better performance of the corresponding adaptive prior.

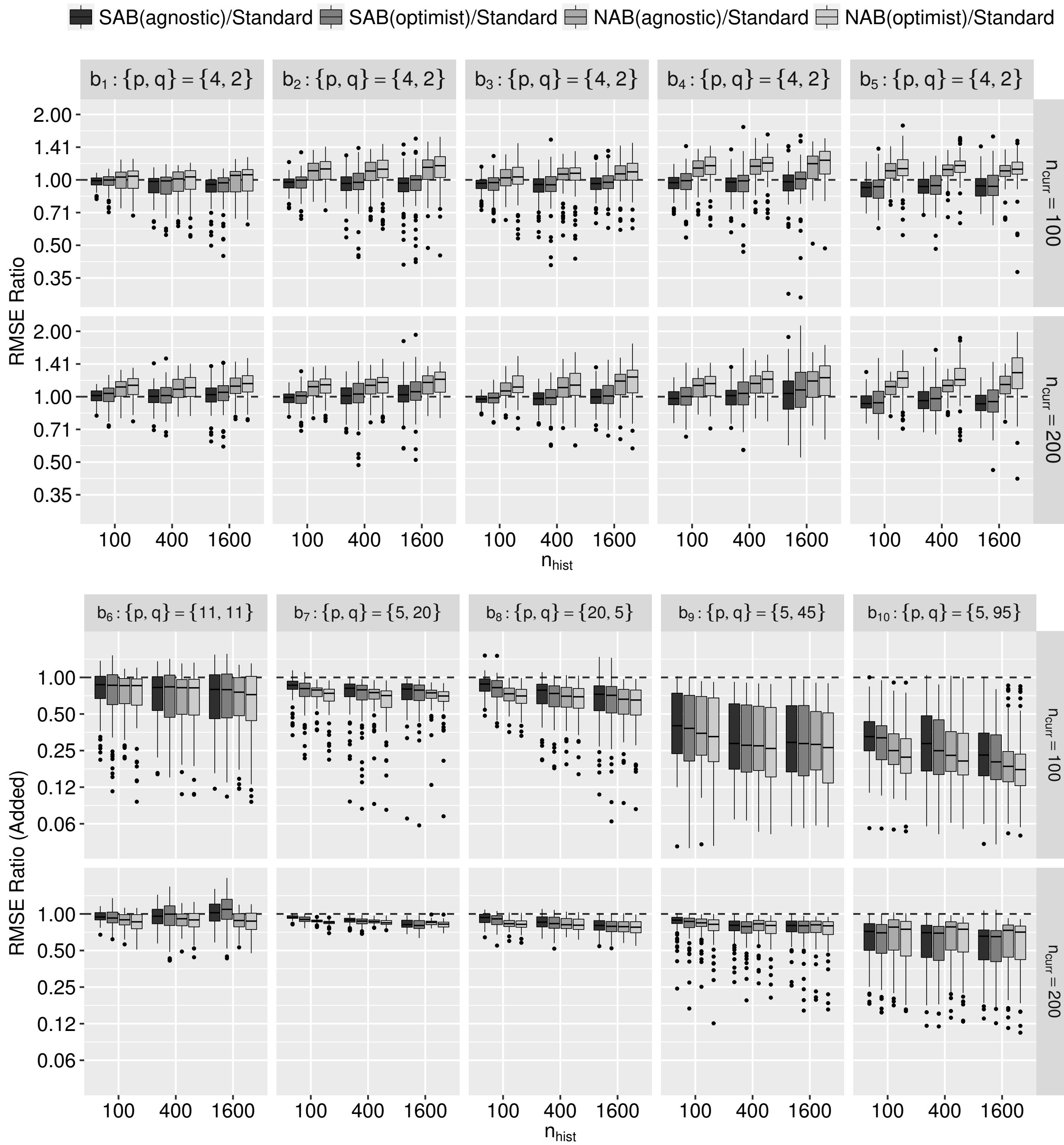

Figure S2: Boxplots of RMSE ratios **for β^a only** (y-axis, on the $\log_2$-scale) comparing four adaptive priors that make use of the historical information against a standard hierarchical shrinkage prior against varying sample sizes of the historical data (n_{hist}; x-axis) for ten true values of the regression coefficients (b_k, $k = 1, \ldots, 10$; columns) taken from Table 3 in the manuscript and varying sample sizes of the current data (n_{curr}; rows). Each boxplot compares the posteriors across 80 independent datasets. Smaller ratios indicate that better performance of the corresponding adaptive prior.