Online Supplement for Inference from Intrinsic Bayes' Procedures Under Model Selection and Uncertainty

Andrew J. Womack   Luis León-Novelo   George Casella

December 31, 2013

Andrew J. Womack is Postdoctoral Researcher, Department of Statistics, University of Florida, Gainesville, FL 32611 (email: ajwomack@ufl.edu). Luis León-Novelo is Assistant Professor, Department of Statistics, University of Louisiana at Lafayette, Lafayette, LA 70504 (email: luis@louisiana.edu). George Casella (1951-2012) was Distinguished Professor, Department of Statistics, University of Florida, Gainesville, FL 32611. Dr. Casella was a dear mentor and friend who contributed substantially to the initial draft of this paper. This research was supported by National Science Foundation Grant DMS-1105127. The authors would like to thank Linda J. Young for her support in editing the manuscript as well as the associate editor and reviewers for many insightful comments and recommendations.

Online Supplement

A1. Derivation of Intrinsic Prior and Proof of Lemma 1

Here, a somewhat different notation is used than in the body of the article. In particular, the base and alternative models are fixed and $(\boldsymbol{\beta}_i, \tau_i)$ is the parameter of model M_i. Let $\tilde{\boldsymbol{y}}$ with predictor matrix $\tilde{\boldsymbol{X}}_i$ be the training data with the distributional assumptions

$$
\begin{aligned}
f(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{X}}_i, \boldsymbol{\beta}_i, \tau_i) &= \mathcal{N}(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{X}}_i\boldsymbol{\beta}_i, \tau_i^{-1}\boldsymbol{I}) \\
\pi_i^R(\boldsymbol{\beta}_i, \tau_i) &= \frac{c_i}{\tau_i}
\end{aligned}
$$

The base model has p_0 covariates in the matrix $\tilde{\boldsymbol{X}}_0$ and the encompassing model has p_1 covariates in the matrix $\tilde{\boldsymbol{X}}_1$. The training sample $\tilde{\boldsymbol{y}}$ has dimension $q > p_1$ and it is assumed that $\tilde{\boldsymbol{X}}_1 = (\tilde{\boldsymbol{X}}_0 \ \tilde{\boldsymbol{X}}_r)$. The reference marginal measures are improper and given by

$$
m_i^R(\tilde{\boldsymbol{y}}) = \frac{c_i \Gamma\left(\frac{q-p_i}{2}\right)}{|\tilde{\boldsymbol{X}}_i'\tilde{\boldsymbol{X}}_i|^{\frac{1}{2}}} \left(\pi \tilde{\boldsymbol{y}}'(\boldsymbol{I} - \tilde{\boldsymbol{H}}_i)\tilde{\boldsymbol{y}}\right)^{-\frac{q-p_i}{2}}
$$

where $\tilde{\boldsymbol{H}}_i$ is the usual hat matrix associated with $\tilde{\boldsymbol{X}}_i$. The intrinsic prior for $(\boldsymbol{\beta}_1, \tau_1)$ is given by

$$
\pi(\boldsymbol{\beta}_1, \tau_1|M_1) = \pi^R(\boldsymbol{\beta}_1, \tau_1|M_1) E^{M_1}_{\boldsymbol{\beta}_1, \tau_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right],
$$

where the notation $E^M_\theta[h(\tilde{\boldsymbol{y}})] = E[h(\tilde{\boldsymbol{y}})|M, \theta]$ is the conditional expectation of $h(\tilde{\boldsymbol{y}})$ given θ and the model M. The only remaining task is to compute the expectation. Expanding

$m_0^R(\tilde{\boldsymbol{y}})$, one can write

$$
\begin{aligned}
E_{\beta_1,\tau_1}^{M_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right] &= \int\int\int \frac{f(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{X}}_0,\boldsymbol{\beta}_0,\tau_0)}{m_1^R(\tilde{\boldsymbol{y}})} f(\tilde{\boldsymbol{y}}|\tilde{\boldsymbol{X}}_1,\boldsymbol{\beta}_1,\tau_1)\mathrm{d}\tilde{\boldsymbol{y}}\pi_0^R(\boldsymbol{\beta}_0,\tau_0)\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\tau_0 \\
&= \int\int\int \frac{|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1|^{\frac{1}{2}}}{c_1\Gamma\left(\frac{q-p_1}{2}\right)}(\pi\tilde{\boldsymbol{y}}(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1)\tilde{\boldsymbol{y}})^{\frac{q-p_1}{2}} \\
&\quad \times \left(\frac{\tau_0}{2\pi}\right)^{\frac{q}{2}}\exp\left(-\frac{\tau_0}{2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{X}}_0\boldsymbol{\beta}_0)'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{X}}_0\boldsymbol{\beta}_0)\right) \\
&\quad \times \left(\frac{\tau_1}{2\pi}\right)^{\frac{q}{2}}\exp\left(-\frac{\tau_1}{2}(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1)'(\tilde{\boldsymbol{y}}-\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1)\right)\mathrm{d}\tilde{\boldsymbol{y}}c_0\tau_0^{-1}\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\tau_0. \\
&= \int\int\int \frac{c_0|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1|^{\frac{1}{2}}\pi^{\frac{q-p_1}{2}}}{c_1\Gamma\left(\frac{q-p_1}{2}\right)(2\pi)^q}\tau_0^{\frac{q}{2}-1}\tau_1^{\frac{q}{2}}(\tilde{\boldsymbol{y}}(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1)\tilde{\boldsymbol{y}})^{\frac{q-p_1}{2}} \\
&\quad \times \exp\left(-\frac{\tau_0+\tau_1}{2}(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}})'(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}})\right) \\
&\quad \times \exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}(\boldsymbol{\beta}_0-\hat{\boldsymbol{\mu}}_0)'\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0(\boldsymbol{\beta}_0-\hat{\boldsymbol{\mu}}_0)\right) \\
&\quad \times \exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}\boldsymbol{\beta}_1'\tilde{\boldsymbol{X}}_1'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1\right)\mathrm{d}\tilde{\boldsymbol{y}}\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\tau_0
\end{aligned}
$$

Gathering terms in the exponent and completing the square to isolate $\tilde{\boldsymbol{y}}$ yields

$$
E_{\beta_1,\tau_1}^{M_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right] \tag{1}
$$

$$
= \frac{c_0|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1|^{\frac{1}{2}}\pi^{\frac{q-p_1}{2}}}{c_1\Gamma\left(\frac{q-p_1}{2}\right)(2\pi)^q}\tau_1^{\frac{q}{2}}\int\tau_0^{\frac{q}{2}-1}\int\left[\int(\tilde{\boldsymbol{y}}(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1)\tilde{\boldsymbol{y}})^{\frac{q-p_1}{2}}\exp\left(-\frac{\tau_0+\tau_1}{2}(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}})'(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}})\right)\mathrm{d}\tilde{\boldsymbol{y}}\right]
$$

$$
\times\exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}(\boldsymbol{\beta}_0-\hat{\boldsymbol{\mu}}_0)'\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0(\boldsymbol{\beta}_0-\hat{\boldsymbol{\mu}}_0)\right)\exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}\boldsymbol{\beta}_1'\tilde{\boldsymbol{X}}_1'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1\right)\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\tau_0, \tag{2}
$$

where $\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}} = \frac{\tau_0\tilde{\boldsymbol{X}}_0\boldsymbol{\beta}_0+\tau_1\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1}{\tau_0+\tau_1}$ and $\hat{\boldsymbol{\mu}}_0 = (\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0)^{-1}\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1$. Integration is first performed over $\tilde{\boldsymbol{y}}$. Transforming $\sqrt{\tau_0+\tau_1}(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}}) \mapsto \boldsymbol{t}$ and noting that $\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}}'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1) = 0$ provide that the integral in square brackets in (1) can be written as

$$
\begin{aligned}
&\int(\tilde{\boldsymbol{y}}'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1)\tilde{\boldsymbol{y}})^{\frac{q-p_1}{2}}\exp\left(-\frac{\tau_0+\tau_1}{2}(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}})'(\tilde{\boldsymbol{y}}-\hat{\boldsymbol{\mu}}_{\tilde{\boldsymbol{y}}})\right)\mathrm{d}\tilde{\boldsymbol{y}} \\
&\quad = \frac{1}{(\tau_0+\tau_1)^{q-p_1/2}}\int[\boldsymbol{t}(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1)\boldsymbol{t}]^{\frac{q-p_1}{2}}\exp(-\boldsymbol{t}'\boldsymbol{t}/2)d\boldsymbol{t} = \frac{(2\pi)^{q/2}}{(\tau_0+\tau_1)^{q-p_1/2}}\frac{\Gamma(q-p_1)2^{(q-p_1)/2}}{\Gamma(\frac{q-p_1}{2})},
\end{aligned}
$$

since $\boldsymbol{t}(\boldsymbol{I}-\tilde{\boldsymbol{H}}_1)\boldsymbol{t} \sim \chi^2_{q-p_1}$, a chi squared random variable with $q-p_1$ degrees of freedom. Thus, the expectation is given by

$$
\begin{aligned}
E^{M_1}_{\beta_1,\tau_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right] &= \int\int \frac{c_0\Gamma\left(q-p_1\right)|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1|^{\frac{1}{2}}}{c_1\Gamma\left(\frac{q-p_1}{2}\right)^2(2\pi)^{\frac{p_1}{2}}}\frac{\tau_0^{\frac{q}{2}-1}\tau_1^{\frac{q}{2}}}{(\tau_0+\tau_1)^{q-\frac{p_1}{2}}}\exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}(\boldsymbol{\beta}_0-\hat{\boldsymbol{\mu}}_0)'\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0(\boldsymbol{\beta}_0-\hat{\boldsymbol{\mu}}_0)\right) \\
&\quad\times\exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}\boldsymbol{\beta}_1'\tilde{\boldsymbol{X}}_1'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1\right)\mathrm{d}\boldsymbol{\beta}_0\mathrm{d}\tau_0 \\
&= \int \frac{c_0\Gamma\left(q-p_1\right)|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1|^{\frac{1}{2}}}{c_1\Gamma\left(\frac{q-p_1}{2}\right)^2|\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0|^{\frac{1}{2}}(2\pi)^{\frac{p_1-p_0}{2}}}\frac{\tau_0^{\frac{q-p_0}{2}-1}\tau_1^{\frac{q-p_0}{2}}}{(\tau_0+\tau_1)^{q-\frac{p_1+p_0}{2}}} \\
&\quad\times\exp\left(-\frac{\tau_0\tau_1}{2(\tau_0+\tau_1)}\boldsymbol{\beta}_1'\tilde{\boldsymbol{X}}_1'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1\right)\mathrm{d}\tau_0
\end{aligned}
$$

Applying the change of variables $\frac{\tau_0}{\tau_0+\tau_1} \mapsto w$ provides the differential identity $w^{-1}(1-w)^{-1}\mathrm{d}w = \tau_0^{-1}\mathrm{d}\tau_0$. This provides

$$
\begin{aligned}
E^{M_1}_{\beta_1,\tau_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right] =& \frac{c_0\Gamma\left(q-p_1\right)|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1|^{\frac{1}{2}}\tau_1^{\frac{p_1-p_0}{2}}}{c_1\Gamma\left(\frac{q-p_1}{2}\right)^2|\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0|^{\frac{1}{2}}(2\pi)^{\frac{p_1-p_0}{2}}} \\
&\times\int w^{\frac{q-p_0}{2}-1}(1-w)^{\frac{q-p_1}{2}-1}\exp\left(-w\frac{\tau_1}{2}\boldsymbol{\beta}_1'\tilde{\boldsymbol{X}}_1'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_1\boldsymbol{\beta}_1\right)\mathrm{d}w
\end{aligned}
$$

Note that

$$
\tilde{\boldsymbol{X}}_1'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_1 = \begin{pmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{X}}_r'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_r \end{pmatrix},
$$

and thus Lemma 1 is established.

Further, $|\tilde{\boldsymbol{X}}_1'\tilde{\boldsymbol{X}}_1| = |\tilde{\boldsymbol{X}}_0'\tilde{\boldsymbol{X}}_0||\tilde{\boldsymbol{X}}_r'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_r|$, $w^{\frac{q-p_0}{2}-1} = w^{\frac{p_1-p_0}{2}}w^{\frac{q-p_1}{2}-1}$, and $\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_0', \boldsymbol{\beta}_r')'$, providing

$$
\begin{aligned}
& E^{M_1}_{\beta_1,\tau_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right] \\
= & \frac{c_0}{c_1}\int\left[\frac{|\tilde{\boldsymbol{X}}_r'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_r|^{1/2}(\tau_1 w)^{\frac{p_1-p_0}{2}}}{(2\pi)^{\frac{p_1-p_0}{2}}}\exp\left(-w\frac{\tau_1}{2}\boldsymbol{\beta}_r'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\boldsymbol{X}_r'\tilde{\boldsymbol{X}}_r\boldsymbol{\beta}_r\right)\right] \\
& \quad\times\left[\frac{\Gamma(q-p_1)}{\Gamma\left(\frac{q-p_1}{2}\right)^2}w^{\frac{q-p_1}{2}-1}(1-w)^{\frac{q-p_1}{2}-1}\right]\mathrm{d}w \\
= & \frac{c_0}{c_1}\int\pi_1^R(\boldsymbol{\beta}_r|\tau_1,w)\pi_1(w)\mathrm{d}w,
\end{aligned}
$$

where $\pi_1^R(\boldsymbol{\beta}_r|\tau_1,w)$ is a multivariate normal density with mean $\mathbf{0}$ and covariance matrix $\tau_1^{-1}w^{-1}(\tilde{\boldsymbol{X}}_r'(\boldsymbol{I}-\tilde{\boldsymbol{H}}_0)\tilde{\boldsymbol{X}}_r)^{-1}$, and $\pi_1(w)$ is a $Beta\left(\frac{q-p_1}{2},\frac{q-p_1}{2}\right)$ density. Therefore the intrinsic prior can be written as

$$
\begin{aligned}
\pi_1^I(\boldsymbol{\beta}_0,\boldsymbol{\beta}_r,\tau_1) &= \pi^R(\boldsymbol{\beta}_0,\boldsymbol{\beta}_r,\tau_1)E^{M_1}_{\beta_1,\tau_1}\left[\frac{m_0^R(\tilde{\boldsymbol{y}})}{m_1^R(\tilde{\boldsymbol{y}})}\right] \\
&= \pi_1^I(\boldsymbol{\beta}_0,\tau_1)\int\pi_1^R(\boldsymbol{\beta}_r|\tau_1,w)\pi_1(w)\mathrm{d}w,
\end{aligned}
$$

because $\pi_1^R(\boldsymbol{\beta}_0,\boldsymbol{\beta}_r,\tau_1)=c_1\tau_1^{-1}$ and $\pi_1^I(\boldsymbol{\beta}_0,\tau_1)=c_0\tau_1^{-1}$. This establishes mixture of g-priors representation of the intrinsic prior.

Three things are immediate. First, the intrinsic prior is a scaled mixture of g-priors with a beta mixing distribution. Second, the prior $\pi_1^I(\boldsymbol{\beta}_r,w|\tau_1)=\pi_1^I(\boldsymbol{\beta}_r|\tau_1,w)\pi_1(w)$ is proper. Third, the prior $\pi_1^I(\boldsymbol{\beta}_0,\tau_1)$ is the reference prior for the base model.

A2. Convergence of the Normal-Gamma Densities

To prove that the Normal-Gamma densities converge to the appropriate point mass, it is proven that each marginal distribution converges to the appropriate point mass when M_T is the true model. The proof is similar for both cases and so it is proven for the lower bound.

For convenience, assume that $\boldsymbol{X}_{r,T} \perp \boldsymbol{X}_0$ and $\tilde{\boldsymbol{X}}_T'\tilde{\boldsymbol{X}} = \frac{q_T}{n}\boldsymbol{X}_T'\boldsymbol{X}$, although the proof follows by similar arguments without the orthogonality assumption.

Proof. A lower bound for the intrinsic posterior is given by

$$
\begin{aligned}
p^I(\boldsymbol{\beta}_T, \tau|\boldsymbol{y}, M_T) &\geq \frac{C\frac{|\tilde{\boldsymbol{X}}_{r,T}'\tilde{\boldsymbol{X}}_{r,T}|^{\frac{1}{2}}\tau^{\frac{p_T-p_0}{2}}}{(2\pi)^{\frac{p_T-p_0}{2}}}\exp\left(-\frac{\tau}{2}\boldsymbol{\beta}_{r,T}'\tilde{\boldsymbol{X}}_{r,T}'\tilde{\boldsymbol{X}}_{r,T}\boldsymbol{\beta}_{r,T}\right)p^R(\boldsymbol{\beta}, \tau|\boldsymbol{y}, M_T)}{\int C\frac{|\tilde{\boldsymbol{X}}_{r,T}'\tilde{\boldsymbol{X}}_{r,T}|^{\frac{1}{2}}\tau^{\frac{p_T-p_0}{2}}}{(2\pi)^{\frac{p_T-p_0}{2}}}p^R(\boldsymbol{\beta}, \tau|\boldsymbol{y}, M_T)\mathrm{d}\boldsymbol{\beta}\mathrm{d}\tau} \qquad (3) \\
&= \left(\frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H}_0 - \boldsymbol{H}_{r,T})\boldsymbol{y}}{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H}_0 - \frac{n}{n+q_T}\boldsymbol{H}_{r,T})\boldsymbol{y}}\right)^{\frac{n-p_0}{2}}\left(\frac{n}{n+q_T}\right)^{-\frac{p_T-p_0}{2}} \\
&\quad \times \mathcal{NG}\left(\boldsymbol{\beta}_T, \tau \,\middle|\, \hat{\boldsymbol{\beta}}_{T,n}, \tau^{-1}\boldsymbol{\Omega}_{T,n}^{-1}, \frac{n-p_0}{2}, \frac{n-p_0}{2\hat{\sigma}_n^2}\right)\prod_{a\notin T}\delta_0(\beta_{r,a})
\end{aligned}
$$

where

$$
\hat{\boldsymbol{\beta}}_{T,n} = \begin{pmatrix} (\boldsymbol{X}_0'\boldsymbol{X}_0)^{-1}\boldsymbol{X}_0'\boldsymbol{y} \\ \frac{n}{n+q_T}(\boldsymbol{X}_{r,T}'\boldsymbol{X}_{r,T})^{-1}\boldsymbol{X}_{r,T}'\boldsymbol{y} \end{pmatrix} \qquad (4)
$$

$$
\boldsymbol{\Omega}_{T,n} = \begin{pmatrix} \boldsymbol{X}_0'\boldsymbol{X}_0 & \boldsymbol{0} \\ \boldsymbol{0} & \frac{n+q_T}{n}\boldsymbol{X}_{r,T}'\boldsymbol{X}_{r,T} \end{pmatrix} \qquad (5)
$$

$$
\hat{\sigma}_{T,n}^2 = \frac{\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{H}_0 - \frac{n}{n+q_T}\boldsymbol{H}_{r,T})\boldsymbol{y}}{n-p_0} \qquad (6)
$$

The marginal for τ is $\mathcal{G}\left(\frac{n-p_0}{2}, \frac{n-p_0}{2\hat{\sigma}_{T,n}^2}\right)$, which converges weakly to δ_{τ^*} since $\hat{\sigma}_{T,n}^2 \to \frac{1}{\tau^*}$. The marginal for β_j is a Student's-t distribution centered at $\left(\hat{\boldsymbol{\beta}}_{T,n}\right)_j$ with scaling factor $\left(\hat{\sigma}_{T,n}^2\left(\boldsymbol{\Omega}_{T,n}^{-1}\right)_{jj}\right)^{\frac{1}{2}}$. This converges weakly to $\delta_{\beta_j^*}$ because the scaling factor converges to 0 — due to the assumption that $\frac{1}{n}\boldsymbol{X}'\boldsymbol{X}$ converges to a positive definite matrix. The coefficient in front of the Normal-Gamma distribution converges to $\exp\left(-\frac{\tau^*}{2}\ell_T\right)$ where $\boldsymbol{y}'\boldsymbol{H}_{r,T}\boldsymbol{y}\left(\frac{q_T}{n+q_T}\right) \to \ell_T$ which is finite because T is the true model. The remainder of the proof follows by similar arguments. ∎

A3. Change of Variables and Direct Sampling

Direct sampling for the intrinsic, Zellner-Siow, and hyper g-over-n priors can be achieved if samples from the posterior of w can be drawn for each prior. This is accomplished through an appropriate change of variables $\theta = \zeta(w)$ for each of the priors where θ has a bounded density on $(0, 1)$. θ can then be sampled using an accept-reject algorithm with an appropriate enveloping density. Such a density can be achieved using a dyadic partition of the interval and piecewise linear interpolation.

These transformations of variables also make the calculation of the Bayes' factor of a model M_A to the base model tractable through the integration of a bounded function over $(0, 1)$. If it is assumed that $\tilde{\boldsymbol{X}}'_A \tilde{\boldsymbol{X}}_A = \frac{p_A+1}{n} \boldsymbol{X}'_A \boldsymbol{X}_A$, then $BF^{\pi}_{A,0}$ is defined by $BF^{\pi}_{A,0}(\boldsymbol{y}) = \int BF_{A,0}(\boldsymbol{y}|w)\pi(w)\mathrm{d}w$ where

$$BF_{A,0}(\boldsymbol{y}|w) = \left(\frac{w(p_A+1)}{n + w(p_A+1)}\right)^{\frac{p_A - p_0}{2}} \left(1 - R^2_A \frac{n}{n + w(p_A+1)}\right)^{-\frac{n-p_0}{2}}$$

The posterior distribution of w is given by $p(w|\boldsymbol{y}, M_A, \pi) = \left(BF^{\pi}_{A,0}(\boldsymbol{y})\right)^{-1} BF_{A,0}(\boldsymbol{y}|w)\pi(w)$. The function $BF_{A,0}(\boldsymbol{y}|w)$ is bounded as a function of w for all $w \in \mathbb{R}^+$. Thus, if the transformation $\theta = \zeta(w)$ makes the prior bounded, then the posterior for θ will be bounded and the integral can be stably computed.

Intrinsic Posterior of w For the intrinsic prior, the change of variables $w = \sin^2\left(\frac{\pi}{2}\theta\right)$ provides a bounded prior for θ on $(0, 1)$. In fact,

$$\pi^I(w)\mathrm{d}w = \frac{I(0 < w < 1)dw}{\pi(w(1-w))^{\frac{1}{2}}} = I(0 < \theta < 1)\mathrm{d}\theta = \pi^I(\theta)\mathrm{d}\theta$$

and so the intrinsic prior for θ is uniform.

Zellner-Siow Posterior of w For the Zellner-Siow prior, with $\pi^{ZS}(w) = \frac{\exp\left(-\frac{w}{2}\right)}{\sqrt{2\pi w}}$, the change of variables $w = \frac{1}{\sin^2\left(\frac{\pi}{2}\theta\right)} - 1$ provides

$$\begin{aligned}
\pi^{ZS}(w)\mathrm{d}w &= I(w>0)\frac{w^{-\frac{1}{2}}}{\sqrt{2\pi}}\exp\left(-\frac{w}{2}\right)\mathrm{d}w \\
&= \frac{\sqrt{\pi}I(0<\theta<1)}{\sqrt{2}\sin^2\left(\frac{\pi}{2}\theta\right)}\exp\left(-\frac{1}{2}\left(\frac{1}{\sin^2\left(\frac{\pi}{2}\theta\right)}-1\right)\right)\mathrm{d}\theta = \pi^{ZS}(\theta)\mathrm{d}\theta
\end{aligned}$$

For any value of $p_A - p_0$, the posterior density of θ converges to 0 as θ converges to 0 or 1.

Hyper g over n Posterior of w The situation is slightly more complicated for the hyper g prior, which is given by

$$\pi^{HG}(w|a,b) = \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{a}{2}\right)}\left(\frac{w}{b}\right)^{\frac{1}{2}}\left(1+\frac{w}{b}\right)^{-\frac{a+1}{2}}\frac{1}{w}$$

The change of variables is based on the value of a:

$$w = \frac{b}{\sin^{2k}\left(\frac{\pi}{2}\theta\right)} - b \quad \text{for } k \geq \frac{1}{a},\ k \in \mathbb{N}$$

This provides the identity:

$$\begin{aligned}
\pi^{HG}(w|a,b)\mathrm{d}w &= \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{a}{2}\right)}\left(\frac{w}{b}\right)^{\frac{1}{2}}\left(1+\frac{w}{b}\right)^{-\frac{a+1}{2}}\frac{I(w>0)\mathrm{d}w}{w} \\
&= \frac{\Gamma\left(\frac{a+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{a}{2}\right)}\left(\sum_{\ell=0}^{k-1}\sin^{2\ell}\left(\frac{\pi}{2}\theta\right)\right)^{-\frac{1}{2}}\left(\sin\left(\frac{\pi}{2}\theta\right)\right)^{ka-1}k\pi I(0<\theta<1)\mathrm{d}\theta
\end{aligned}$$

Any value of $k \geq \frac{1}{a}$ with $k \in \mathbb{N}$ provides a bounded prior density $\pi^{HG}(\theta|a,b,k)$. In particular, for $k = a = 1$, the prior is uniform θ over $(0,1)$.

A4. Skewness of the Posterior Conditioned on a Selected Model

The marginal posterior distribution of a given β is skewed. The skewness is defined as

$$Skew(\beta) = E\left[\left(\frac{\beta - E[\beta|\boldsymbol{y}, M_A, \pi]}{Var(\beta|\boldsymbol{y}, M_A, \pi)^{\frac{1}{2}}}\right)^3 \middle| \boldsymbol{y}, M_A, \pi\right]$$

for $n - p_0 > 3$.

If $M_A \vDash \beta$, then the third central moment of β can be expressed as

$$\begin{aligned}
\kappa_3(\beta) &= E\left[(\beta - E[\beta|\boldsymbol{y}, M_A, \pi])^3 \,|\boldsymbol{y}, M_A, \pi\right] \\
&= E\left[(E[\beta|w, \boldsymbol{y}, M_A, \pi] - E[\beta|\boldsymbol{y}, M_A, \pi])^3 \,|\boldsymbol{y}, M_A, \pi\right] \\
&+ 3E\left[Var[\beta|w, \boldsymbol{y}, M_A, \pi]\;(E[\beta|w, \boldsymbol{y}, M_A, \pi] - E[\beta|\boldsymbol{y}, M_A, \pi])\,|\boldsymbol{y}, M_A, \pi\right]
\end{aligned} \tag{7}$$

The second central moment of β can be expressed as

$$\begin{aligned}
\kappa_2(\beta) &= E\left[(\beta - E[\beta|\boldsymbol{y}, M_A, \pi])^2 \,|\boldsymbol{y}, M_A, \pi\right] \\
&= E\left[(E[\beta|w, \boldsymbol{y}, M_A, \pi] - E[\beta|\boldsymbol{y}, M_A, \pi])^2 \,|\boldsymbol{y}, M_A, \pi\right] \\
&+ E\left[Var[\beta|w, \boldsymbol{y}, M_A, \pi]\;|\boldsymbol{y}, M_A, \pi\right]
\end{aligned} \tag{8}$$

The same expressions hold for $M_0 \vDash \beta$ and any M_A.

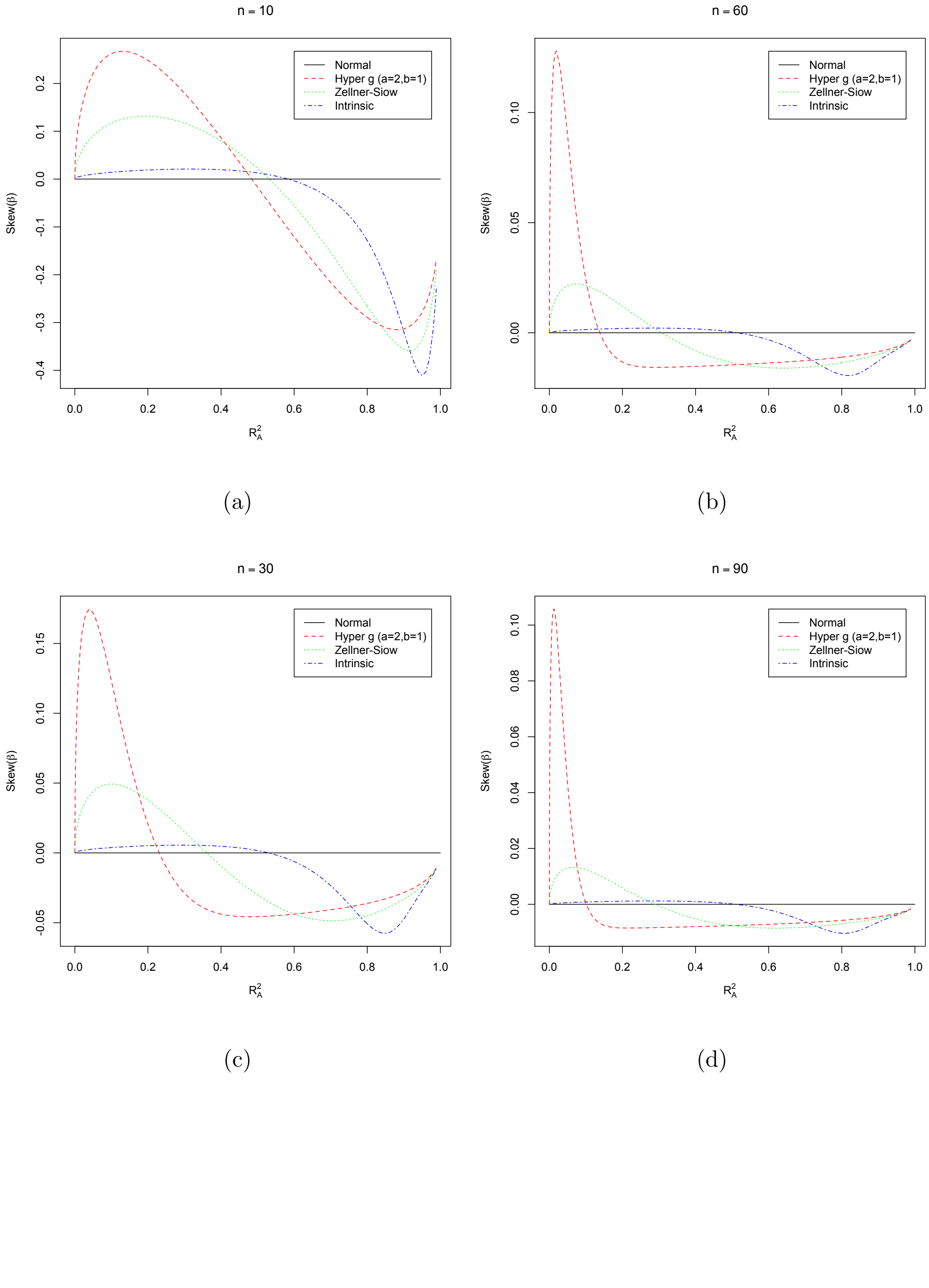
In order to illustrate the skewness, we focus on the case where $y_i = \beta_0 + \beta_1 x_{1,i} + \epsilon_i$ for $i = 1, \ldots, n$, where $\epsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ and ϵ_i is independent of x_i. In such a situation, $\left(\hat{\beta}_1^{MLE}\right)^2 = \frac{R_A^2 \hat{\sigma}_0^2}{Var(\boldsymbol{x}_1)}$ where $\hat{\sigma}_0^2$ is the residual variance under the intercept only model. We can characterize the skewness of β_1 conditioned on the full model in terms of R_A^2, n, and the sign of $\hat{\beta}_1^{MLE}$. In particular, defining $r = \frac{n}{n + w(p_A + 1)}$ and $\hat{r} = E[r|\boldsymbol{y}, M_A, \pi]$ provides

$$\frac{Skew(\beta_1)}{sign\left(\hat{\beta}_1^{MLE}\right)} = \frac{(R_A^2)^{\frac{1}{2}}\left\{E\left[R_A^2\,(r - \hat{r})^3 + \frac{3}{n-3}\,(1 - rR_A^2)\,r\,(r - \hat{r})\middle|\,\boldsymbol{y}, M_A, \pi\right]\right\}}{\left\{E\left[R_A^2\,(r - \hat{r})^2 + \frac{1}{n-3}\,(1 - rR_A^2)\,r\middle|\,\boldsymbol{y}, M_A, \pi\right]\right\}^{\frac{3}{2}}}$$

A similar characterization holds when there is more than one covariate in the model, though the formulas are more complicated. They involve not only R_A^2, but also the percentages of R_A^2 attributed to each regressor.

The skewness of β is non-zero ($\boldsymbol{y}$ a.e.) and converges to 0 as the sample size increases. Figure 1 shows $Skew(\beta_1)$ for $sign\left(\hat{\beta}_1^{MLE}\right) = 1$ as a function of R_A^2 for $n = 10, 30, 60, 90$. In general, the skewness is positive for small values of R_A^2 and negative for large values of R_A^2. Additionally, the intrinsic prior produces generally less shrinkage (in absolute value) than the Zellner-Siow and Hyper g-priors, though this can be violated. Figure 2 shows the skewness as a function of n for $R_A^2 = 0.2, 0.4, 0.6, 0.8$ and demonstrates that the skewness converges to 0 and $n \to \infty$. and that the skewness of the intrinsic prior converges to 0 faster than the skewness for the Zellner-Siow and Hyper g-priors unless R_A^2 is large

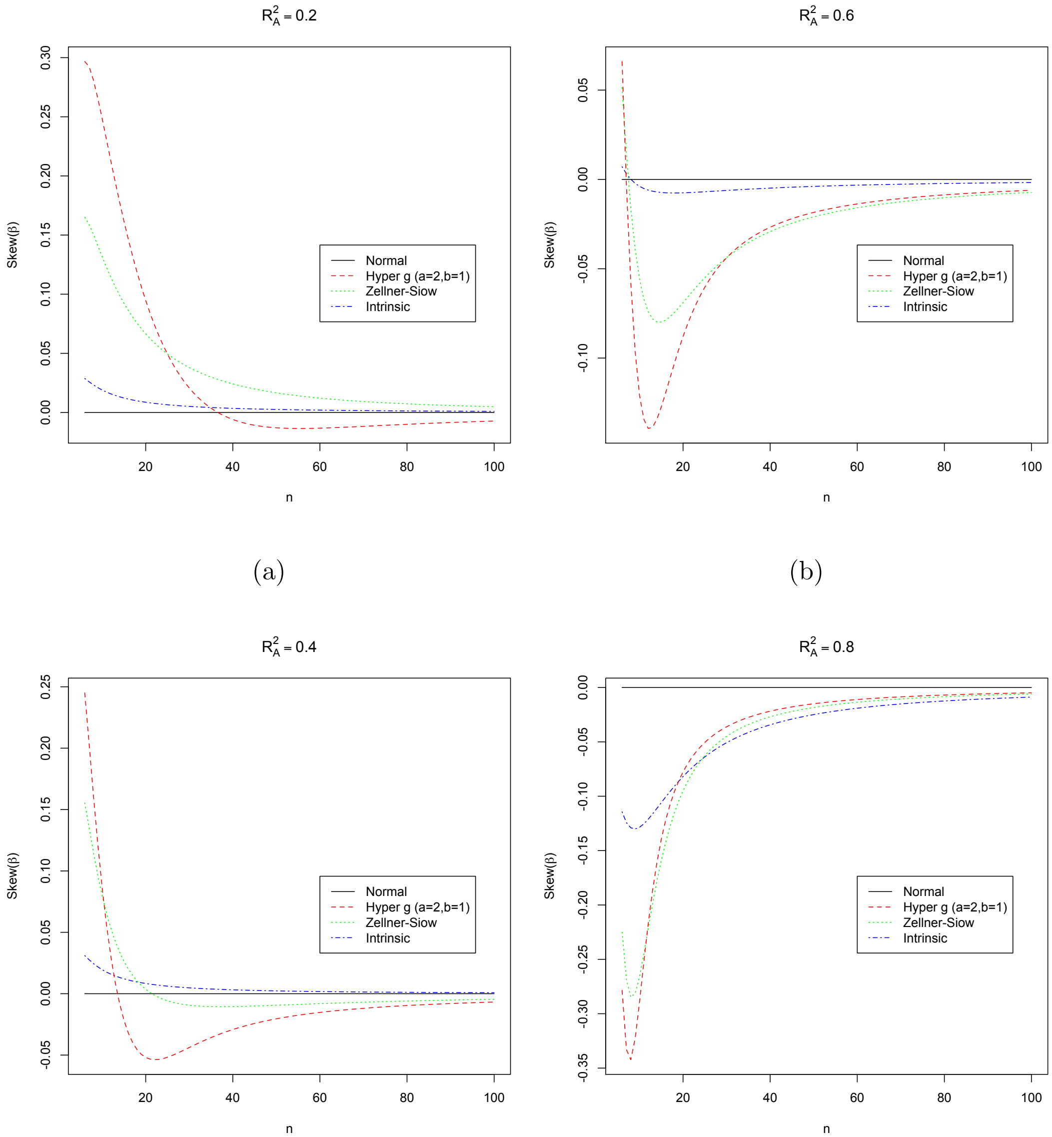
Figure 1: Plots of $Skew(\beta_1)$ as a function of R_A^2 for the linear model with one covariate.



(a) (b)

(c) (d)

Figure 2: Plots of $Skew(\beta_1)$ as a function of n for the linear model with one covariate.



(a) (b)

(c) (d)

A5. Simulation Study

We performed a simulation study to investigate the properties of the intrinsic prior and different priors on the model space. For this simulation, the base model is the intercept only model, there are $K = 40$ test covariates, and $n = 100$ observations. The test covariates $\boldsymbol{X}$ are generated as $\boldsymbol{X} = \boldsymbol{ZC}$ where $\boldsymbol{Z}$ is an $n \times K$ matrix of $N(0, \sigma^2)$ deviates and $\boldsymbol{C}$ is a $K \times K$ matrix with ones on the diagonal and ρ off of the diagonal. We let $\sigma \in \{1, 1.5, 2, 3\}$ and $\rho \in \{0, .25\}$. The true value of the regression coefficients are $\beta_0 = 1$, $\beta_{r,a} = 0.6 * a$ for $a = 1, \ldots, 5$, and $\beta_{r,a} = 0$ for $a > 5$, which allows us to compare the behavior of the posterior distribution of coefficients of covariates with varying signal strength. The simulation involved 1000 replicates for each combination of σ and ρ. Posterior probabilities were computed via renormalization, taking $p(M|\boldsymbol{y}) \propto p(\boldsymbol{y}|M)p(M)$ and normalizing over the models considered in the random walk. The selected model is the model maximizing its estimated posterior probability. The model space was explored using a random walk with 10000 iterations, initialized at the base model.

In addition to investigating the behavior of the posterior across these scenarios, we also compare the prediction and model selection properties of three priors on the model space. The first prior is the standard uniform prior. The second prior is the multiplicity correction prior where the probability of a model M with k covariates is $\pi(M) = \frac{1}{K+1}\binom{K}{k}^{-1}$. This prior is obtained by treating the inclusion of variables as independent Bernoulli random variables with probability p. This probability is then given a uniform prior on $(0, 1)$. Though this prior does provide a correction for the combinatorial complexity of the model space, it does not explicitly penalize models for the number of covariates they use. Increasing the number of covariates in a model decreases the model's probability until $K/2$ covariates are added, at which point the probabilities increase. An alternative is a prior that provides greater penalization than the multiplication prior. An example of this is to let the prior for p to be $Beta(1, K)$. This prior has the desirable property that $M \subset M'$ provides that the prior probability of M is at least that of M'.

We consider coverage and length of intervals computed under the uniform prior, the predictive sum of squared errors under each of the model space priors, as well as the model selection properties of the three priors.

A5.1 Coverage and Length of Intervals

Tables 1 and 2 show the average coverage (and length) of three different 95% intervals for $\rho = 0$ and 0.25, respectively. We computed the intervals for each $\beta_{r,a}$ for $a = 0, \ldots, 6$. The three intervals we consider are the quantile based credible set under model averaging, the quantile based credible interval conditioned on the selected model, and the standard frequentist confidence interval for the MLE conditioned on the selected model. The uniform prior on the model space was used to compute the credible set under model uncertainty as well as performing model selection.

Three facts are apparent from the tables. First, with few exceptions, the credible sets under model uncertainty have higher coverage than those conditioned on the selected model. The better coverage of the credible sets under model uncertainty is explained by their longer Lebesgue measure, which can be considerably larger than the length of intervals conditioned on the selected model. Second, as the signal strength decreases (either by increasing σ of decreasing the true value of the regression coefficient), the coverage suffers, failing to reach the nominal level. This is most pronounced for the intervals conditioned on the selected model. Third, the coverage also decreases when the covariates are correlated. As we will see later in this example, this poor behavior is most likely caused by the poor behavior of the model posterior probabilities under the uniform prior.

σ	Type	β_0	β_1	β_2	β_3	β_4	β_5	β_6
1	MA CS	0.945	0.947	0.947	0.965	0.958	0.936	0.999
		(0.385)	(0.414)	(0.411)	(0.413)	(0.415)	(0.413)	(0.05)
	SM CI	0.947	0.938	0.937	0.955	0.95	0.931	0.986
		(0.388)	(0.401)	(0.401)	(0.402)	(0.405)	(0.402)	(0.005)
	SM FI	0.946	0.937	0.94	0.952	0.946	0.931	0.986
		(0.386)	(0.399)	(0.399)	(0.4)	(0.403)	(0.4)	(0.005)
1.5	MA CS	0.945	0.918	0.951	0.946	0.945	0.938	0.997
		(0.569)	(0.599)	(0.61)	(0.611)	(0.61)	(0.607)	(0.094)
	SM CI	0.946	0.898	0.94	0.937	0.93	0.929	0.978
		(0.572)	(0.546)	(0.592)	(0.592)	(0.591)	(0.588)	(0.013)
	SM FI	0.944	0.896	0.937	0.936	0.923	0.927	0.978
		(0.569)	(0.544)	(0.59)	(0.589)	(0.588)	(0.584)	(0.013)
2	MA CS	0.944	0.869	0.942	0.956	0.937	0.941	0.995
		(0.762)	(0.734)	(0.825)	(0.825)	(0.824)	(0.823)	(0.15)
	SM CI	0.945	0.719	0.931	0.939	0.933	0.927	0.968
		(0.765)	(0.587)	(0.793)	(0.794)	(0.794)	(0.793)	(0.024)
	SM FI	0.944	0.715	0.93	0.933	0.928	0.929	0.968
		(0.759)	(0.587)	(0.792)	(0.792)	(0.79)	(0.787)	(0.024)
3	MA CS	0.945	0.729	0.926	0.933	0.918	0.902	0.993
		(1.144)	(0.85)	(1.21)	(1.236)	(1.232)	(1.248)	(0.339)
	SM CI	0.942	0.413	0.907	0.924	0.902	0.902	0.941
		(1.145)	(0.528)	(1.116)	(1.183)	(1.18)	(1.192)	(0.078)
	SM FI	0.942	0.402	0.896	0.92	0.91	0.905	0.938
		(1.134)	(0.532)	(1.122)	(1.186)	(1.179)	(1.185)	(0.078)

Table 1: Average coverage (and length) of three different 95% intervals. MA CS is the quantile based credible set under model uncertainty. SM CI is the quantile based credible interval under the selected model. SM FI is the standard frequentist confidence interval under the selected model. The prior on the model space is uniform and $\rho = 0$.

σ	Type	β_0	β_1	β_2	β_3	β_4	β_5	β_6
1	MA CS	0.952	0.951	0.951	0.93	0.945	0.947	1
		(0.387)	(0.461)	(0.457)	(0.456)	(0.456)	(0.457)	(0.041)
	SM CI	0.955	0.942	0.942	0.919	0.937	0.935	0.99
		(0.39)	(0.436)	(0.437)	(0.437)	(0.436)	(0.437)	(0.004)
	SM FI	0.953	0.938	0.94	0.917	0.936	0.93	0.99
		(0.388)	(0.434)	(0.435)	(0.434)	(0.434)	(0.435)	(0.004)
1.5	MA CS	0.948	0.893	0.945	0.954	0.959	0.955	0.999
		(0.581)	(0.656)	(0.694)	(0.694)	(0.698)	(0.693)	(0.082)
	SM CI	0.95	0.821	0.926	0.943	0.941	0.937	0.986
		(0.585)	(0.556)	(0.655)	(0.656)	(0.658)	(0.654)	(0.009)
	SM FI	0.95	0.819	0.926	0.935	0.939	0.934	0.986
		(0.581)	(0.554)	(0.652)	(0.654)	(0.655)	(0.651)	(0.009)
2	MA CS	0.947	0.811	0.942	0.936	0.945	0.954	0.999
		(0.764)	(0.737)	(0.91)	(0.913)	(0.911)	(0.911)	(0.137)
	SM CI	0.951	0.591	0.923	0.919	0.914	0.937	0.98
		(0.767)	(0.533)	(0.849)	(0.857)	(0.856)	(0.856)	(0.017)
	SM FI	0.945	0.588	0.922	0.917	0.909	0.929	0.98
		(0.762)	(0.532)	(0.847)	(0.855)	(0.853)	(0.852)	(0.017)
3	MA CS	0.951	0.663	0.89	0.955	0.936	0.942	0.992
		(1.148)	(0.806)	(1.316)	(1.384)	(1.38)	(1.394)	(0.295)
	SM CI	0.951	0.323	0.849	0.93	0.918	0.916	0.95
		(1.152)	(0.465)	(1.141)	(1.286)	(1.289)	(1.301)	(0.065)
	SM FI	0.949	0.313	0.843	0.929	0.907	0.923	0.95
		(1.143)	(0.465)	(1.141)	(1.285)	(1.287)	(1.297)	(0.065)

Table 2: Average coverage (and length) of three different 95% intervals. MA CS is the quantile based credible set under model uncertainty. SM CI is the quantile based credible interval under the selected model. SM FI is the standard frequentist confidence interval under the selected model. The prior on the model space is uniform and $\rho = 0.25$.

A5.2 Predicted Sum of Squared Errors

Table 3 provides the predictive sum of squared errors — defined in Section 4.4 — for different model space priors. We compare the model averaged means, the model averaged MLEs, the selected model means, and the selected model MLE for point estimates of $\boldsymbol{\beta}$.

Most striking from this example is the fact that the uniform prior produces much larger predictive error than the multiplicity correction and penalization priors — the latter two providing essentially the same predictive error. This gain is most pronounced when σ is large. It is also apparent from this example that model averaging provides smaller predictive error than conditioning on the highest probability model. There is a slight increase in predictive error when $\rho = 0.25$ versus when $\rho = 0$ when using the multiplicity or penalization priors, which is caused by the increased marginal variance in each x_i introduced by ρ.

Prior	Estimator	(0,1)	(0,1.5)	(0,2)	(0,3)	(0.25,1)	(0.25,1.5)	(0.25,2)	(0.25,3)
Uniform	MA Mean	8.92	23.10	49.81	136.29	8.06	22.23	44.84	125.07
	SM Mean	9.37	24.35	56.14	153.41	8.06	23.62	50.46	141.48
	MA MLE	8.97	23.53	50.78	142.60	8.08	22.44	45.40	128.02
	SM MLE	9.42	24.74	57.24	160.54	8.08	23.81	51.03	144.92
Multiplicity	MA Mean	6.23	18.13	36.69	98.29	6.38	19.03	36.50	95.17
	SM Mean	6.09	19.60	41.32	106.28	6.34	21.67	41.69	110.66
	MA MLE	6.25	18.37	36.93	100.07	6.38	19.14	36.73	95.69
	SM MLE	6.11	19.85	41.66	108.44	6.35	21.80	41.98	111.65
Penalizing	MA Mean (P)	5.95	18.48	36.33	96.25	6.46	19.98	37.21	97.80
	SM Mean (P)	5.89	20.94	41.64	110.07	6.45	22.98	42.30	116.67
	MA MLE (P)	5.97	18.68	36.45	97.03	6.47	20.08	37.39	97.95
	SM MLE (P)	5.91	21.17	41.90	111.88	6.46	23.11	42.56	117.50

Table 3: Comparison of average predictive sum of squared errors under different model space priors. The columns describe (ρ, σ).

A5.3 Model Selection Properties

Inclusion Probabilities for non-zero covariates In Table 4 we compare the average inclusion of the true non-zero covariates in the selected model under each of the three model space priors. Differences are most pronounced for β_1 and β_2, which correspond to the covariates with the weakest signals. The penalization prior produces the largest average inclusion of each variable in the selected model under all circumstances. Increasing σ or introducing ρ decrease the probability of including these variables in the selected model, as expected as the signal weakens under both scenarios.

Covariate	Prior	(0,1)	(0,1.5)	(0,2)	(0,3)	(0.25,1)	(0.25,1.5)	(0.25,2)	(0.25,3)
x_1	Uniform	0.977	0.711	0.41	0.295	0.954	0.586	0.317	0.136
	Multiplicity	0.989	0.794	0.489	0.31	0.972	0.668	0.391	0.163
	Penalizing	0.997	0.927	0.753	0.461	0.997	0.858	0.63	0.367
x_2	Uniform	1.000	1.000	0.989	0.832	1.000	0.999	0.979	0.669
	Multiplicity	1.000	1.000	0.992	0.872	1.000	0.999	0.987	0.727
	Penalizing	1.000	1.000	1.000	0.95	1.000	1.000	0.994	0.888
x_3	Uniform	1.000	1.000	1.000	0.968	1.000	1.000	1.000	0.987
	Multiplicity	1.000	1.000	1.000	0.986	1.000	1.000	1.000	0.991
	Penalizing	1.000	1.000	1.000	0.999	1.000	1.000	1.000	0.998
x_4	Uniform	1.000	1.000	1.000	0.998	1.000	1.000	1.000	0.997
	Multiplicity	1.000	1.000	1.000	0.999	1.000	1.000	1.000	1.000
	Penalizing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
x_5	Uniform	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Multiplicity	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Penalizing	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

Table 4: Comparison of average inclusion probabilities for "true" covariates in the selected model under different model space priors. The columns describe (ρ, σ).

Properties of Selected and True Model The final comparison we make regards the properties of the selected and true models under each of the three priors, which is presented in Table 5. We track the probability of the selected and true models, the rank of the true model, and the number of true and false positives. This is where the differences in the priors on the model space are the most striking. Under the uniform prior, the posterior probabilities of all of the models are quite small. In contrast, the multiplicity correction and penalization priors concentrate mass on parsimonious models that fit the data reasonably well. Most striking is the rank of the true model under the different scenarios. As the signal to noise ratio decreases (by increasing σ or introducing ρ), the rank of the true model decreases dramatically under the uniform prior. In contrast, the other priors provide a high rank for the true model. The trade off is between true and false positive rates. The uniform prior has a larger true positive rate, at the expense of increasing the number of false positives. The penalizing prior is more conservative, sacrificing weak true positives in order to decrease false positives.

Scenario	Prior	P(selected model)	TP	FP	Rank(true model)	P(true model)
(0,1)	Uniform	0.044	4.997	0.599	44.089	0.036
	Multiplicity	0.461	4.989	0.081	1.255	0.447
	Penalizing	0.675	4.977	0.03	1.078	0.66
(0,1.5)	Uniform	0.019	4.927	0.8	197.392	0.014
	Multiplicity	0.336	4.794	0.124	2.492	0.274
	Penalizing	0.55	4.711	0.057	1.872	0.428
(0,2)	Uniform	0.008	4.753	1.368	969.306	0.004
	Multiplicity	0.256	4.481	0.171	5.285	0.136
	Penalizing	0.484	4.399	0.084	3.511	0.228
(0,3)	Uniform	0.005	4.41	2.171	3440.147	0.002
	Multiplicity	0.263	4.167	0.433	24.156	0.089
	Penalizing	0.491	4.093	0.321	9.513	0.162
(0.25,1)	Uniform	0.084	4.997	0.371	5.592	0.075
	Multiplicity	0.55	4.972	0.057	1.111	0.534
	Penalizing	0.721	4.954	0.024	1.084	0.7
(0.25,1.5)	Uniform	0.033	4.858	0.595	51.446	0.025
	Multiplicity	0.399	4.667	0.089	2.000	0.289
	Penalizing	0.585	4.585	0.036	1.888	0.387
(0.25,2)	Uniform	0.018	4.624	0.855	378.498	0.01
	Multiplicity	0.34	4.378	0.133	6.162	0.143
	Penalizing	0.542	4.296	0.065	4.653	0.192
(0.25,3)	Uniform	0.006	4.253	1.576	1962.423	0.002
	Multiplicity	0.261	3.881	0.252	29.323	0.039
	Penalizing	0.467	3.789	0.173	20.72	0.054

Table 5: Comparison of model selection properties under different model space priors. Scenario describes (ρ, σ). The columns from left to right are the probability of the selected model, the number of true positives, the number of false positives, the rank of the true model and the probability of the true model.