Mixtures of g-Priors in Generalized Linear Models
Supplementary Materials: Appendices

A Assumptions, Theoretical Results, and Proofs

A.1 Assumptions and Regularity Conditions

The following assumptions and standard regularity conditions are used throughout the paper unless specified otherwise.

For functions $b(\cdot)$ and $\theta(\cdot)$ in the GLM density (5), their third derivatives exist and are continuous on $\mathbb{R}$. The composite function $b' \circ \theta(\cdot)$, which links $\mathbb{E}(Y)$ and η, is strictly monotonic. The variance function $b'' \circ \theta(\cdot) \geq 0$, and the equality can only occur on the boundary $\pm\infty$.

Finite MLEs $\hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}$ exist and are unique, under all subset models $\mathcal{M}$.

The design matrix $\mathbf{X}$ under the full model is known and has a full column rank p. Here, p is fixed. The column space $C(\mathbf{X})$ does not contain $\mathbf{1}_n$. When studying asymptotics, we assume that for $i = 1, \ldots, n$ the norm of the ith row $\|\mathbf{x}_i\|_2$ is bounded by a constant, and for all n, the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}/n$ is bounded from below by a positive constant. These conditions assure weak consistency (convergence in probability) and asymptotic normality for MLEs (Fahrmeir and Kaufmann 1985).

The true model $\mathcal{M}_T$ is among the 2^p subset models to be selected under consideration. In $\mathcal{M}_T$, true values of the intercept and regression coefficients are denoted by $\alpha^*_{\mathcal{M}_T}$ and $\boldsymbol{\beta}^*_{\mathcal{M}_T}$, respectively.

A.2 Proof of Proposition 1

Proof. We first approximate the likelihood by a second order Taylor expansion at the MLE,

$$
\begin{aligned}
& p(\mathbf{Y} \mid \alpha, \boldsymbol{\beta}_{\mathcal{M}}, \mathcal{M}) \\
\approx\ & p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M}) \\
& \cdot \exp\left\{ -\frac{1}{2} \begin{bmatrix} \alpha - \hat{\alpha}_{\mathcal{M}} \\ \boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}} \end{bmatrix}^T \begin{bmatrix} \mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n & \mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{X}_{\mathcal{M}} \\ \mathbf{X}_{\mathcal{M}}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n & \mathbf{X}_{\mathcal{M}}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{X}_{\mathcal{M}} \end{bmatrix} \begin{bmatrix} \alpha - \hat{\alpha}_{\mathcal{M}} \\ \boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}} \end{bmatrix} \right\} \\
=\ & p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M}) \exp\left\{ -\frac{1}{2} \left(\alpha - \hat{\alpha}_{\mathcal{M}} + \mathbf{m}\right)^T \left(\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n\right) \left(\alpha - \hat{\alpha}_{\mathcal{M}} + \mathbf{m}\right) \right. \\
& \qquad\qquad\qquad\qquad\qquad \left. -\frac{1}{2} \left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right)^T \boldsymbol{\Phi} \left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right) \right\},
\end{aligned}
$$

where the above approximation is precise up to a multiplicative term $[1 + O(n^{-1})]$, $\mathbf{m} = \left(\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n\right)^{-1} \left(\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{X}_{\mathcal{M}}\right) \left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right)$, and

$$
\boldsymbol{\Phi} = \mathbf{X}_{\mathcal{M}}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{X}_{\mathcal{M}} - \left(\mathbf{X}_{\mathcal{M}}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n\right) \left(\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n\right)^{-1} \left(\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{X}_{\mathcal{M}}\right).
$$

In the above approximate likelihood, the matrix $\boldsymbol{\Phi}$ acts like a precision matrix of $\boldsymbol{\beta}_{\mathcal{M}}$. By using the orthogonal projection $\hat{\mathcal{P}}_{\mathbf{1}_n} = \mathbf{1}_n \left(\mathbf{1}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}\right)^{-1} \mathbf{1}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}})$, we can rewrite it as

$$
\begin{aligned}
\boldsymbol{\Phi} &= \mathbf{X}_{\mathcal{M}}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{X}_{\mathcal{M}} - \mathbf{X}_{\mathcal{M}}^T \hat{\mathcal{P}}_{\mathbf{1}_n}^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \hat{\mathcal{P}}_{\mathbf{1}_n} \mathbf{X}_{\mathcal{M}} \\
&= \mathbf{X}_{\mathcal{M}}^T (\mathbf{I}_n - \hat{\mathcal{P}}_{\mathbf{1}_n})^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) (\mathbf{I}_n - \hat{\mathcal{P}}_{\mathbf{1}_n}) \mathbf{X}_{\mathcal{M}} = \mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}).
\end{aligned}
$$

Under the flat prior $p(\alpha) \propto 1$, an integrated Laplace approximation yields the marginal likelihood density conditional on $\boldsymbol{\beta}_{\mathcal{M}}$:

$$
\begin{aligned}
&p(\mathbf{Y} \mid \boldsymbol{\beta}_{\mathcal{M}}, \mathcal{M}) = \int p(\mathbf{Y} \mid \alpha, \boldsymbol{\beta}_{\mathcal{M}}, \mathcal{M}) p(\alpha) d\alpha \\
\propto\ & p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M}) \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right)^T \mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) \left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right)\right\} \\
&\cdot \int \exp\left\{-\frac{1}{2}\left(\alpha - \hat{\alpha}_{\mathcal{M}} + \mathbf{m}\right)^T \left(\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n\right) \left(\alpha - \hat{\alpha}_{\mathcal{M}} + \mathbf{m}\right)\right\} d\alpha \\
\propto\ & p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M}) \left[\mathbf{1}_n^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \mathbf{1}_n\right]^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}\left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right)^T \mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) \left(\boldsymbol{\beta}_{\mathcal{M}} - \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right)\right\}.
\end{aligned}
$$

□

A.3 Asymptotic Behavior of the Observed Information

Lemma A.1. *For any subset model $\mathcal{M}$,*

(1) if $\mathcal{M} \supset \mathcal{M}_T$, then $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) = O_P(n)$ and $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) = O_P(n)$. More specifically, $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})/n - \mathcal{I}_n(\hat{\alpha}_{\mathcal{M}})/n \xrightarrow{P} 0$, and $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n - \mathcal{I}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n \xrightarrow{P} \mathbf{0}$.

(2) if $\mathcal{M} \not\supset \mathcal{M}_T$, then $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) = O_P(n^{\tau_{\mathcal{M}}})$ and $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) = O_P(n^{\tau_{\mathcal{M}}})$, where $0 \le \tau_{\mathcal{M}} \le 1$.

Proof. First, we study the asymptotic of MLEs. The assumptions on the design matrix of the full model $\mathbf{X}$ remain to hold for the design matrix $\mathbf{X}_{\mathcal{M}}$ under all subset models, i.e., $\mathbf{x}_{\mathcal{M},i}$ are bounded for all $i = 1, \ldots, n$, and as n tends to infinity, the smallest eigenvalue of $\mathbf{X}_{\mathcal{M}}^T \mathbf{X}_{\mathcal{M}}/n$ is bounded from below by a positive constant. Since these are stronger than the condition R_c in Fahrmeir and Kaufmann (1985, pp. 355), we have weak consistency and asymptotic normality for MLEs under any $\mathcal{M} \supset \mathcal{M}_T$, i.e., as $n \to \infty$,

$$
\left(\hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right) \xrightarrow{P} \left(\alpha^*_{\mathcal{M}}, \boldsymbol{\beta}^*_{\mathcal{M}}\right), \quad \mathcal{I}_n(\boldsymbol{\beta}^*_{\mathcal{M}})^{\frac{1}{2}} \left(\hat{\boldsymbol{\beta}}_{\mathcal{M}} - \boldsymbol{\beta}^*_{\mathcal{M}_T}\right) \xrightarrow{d} \mathrm{N}(0, \mathbf{I}_{p_{\mathcal{M}}}). \tag{A.1}
$$

Here, $\alpha^*_{\mathcal{M}} = \alpha^*_{\mathcal{M}_T}$, and $\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{\beta}^*_{\mathcal{M}_T}$ in the sense that all entries in $\boldsymbol{\beta}^*_{\mathcal{M}}$ that correspond to predictors not in $\mathcal{M}_T$ are filled with zero. Therefore, if $\mathcal{M} \supset \mathcal{M}_T$, then $\eta^*_{\mathcal{M},i} = \alpha^*_{\mathcal{M}} +$

$\mathbf{x}_{\mathcal{M},i}^T \boldsymbol{\beta}_{\mathcal{M}}^* = \alpha_{\mathcal{M}_T}^* + \mathbf{x}_{\mathcal{M}_T,i}^T \boldsymbol{\beta}_{\mathcal{M}_T}^* = \eta_{\mathcal{M}_T,i}^*$, for all $i = 1, \ldots, n$. On the other hand, if $\mathcal{M} \not\supset \mathcal{M}_T$, Self and Mauritsen (1988) and van der Vaart (2000, pp. 45, Theorem 5.7) suggest that the limits of MLEs still exist, i.e., $\left(\hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}\right) \xrightarrow{P} (\alpha_{\mathcal{M}}^*, \boldsymbol{\beta}_{\mathcal{M}}^*)$, but the linear predictors in the limit $\eta_{\mathcal{M},i}^* \neq \eta_{\mathcal{M}_T,i}^*$.

Under non-canonical links, observed information matrices are functions of $\mathbf{Y}$, therefore we need a weak law of large numbers for independently but non-identically distributed random variables. In Resnick (1999, pp. 205), by Theorem 7.2.1 and the proof of special case (a), we have that for a sequence of independent random variables $Y_1, \ldots, Y_n$, if their variances are bounded, then as $n \to \infty$,

$$\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(Y_i) \xrightarrow{P} 0. \tag{A.2}$$

Next we show asymptotic results for $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})$. In (10), for $i = 1, \ldots, n$, the ith diagonal entry of $\mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}})$ can be rewritten as $d_i = b'' \circ \theta(\hat{\eta}_{\mathcal{M},i}) \left[\theta'(\hat{\eta}_{\mathcal{M},i})\right]^2 + \left[b' \circ \theta(\hat{\eta}_{\mathcal{M},i}) - Y_i\right] \theta''(\hat{\eta}_{\mathcal{M},i})$. Hence, for any model $\mathcal{M}$,

$$\begin{aligned}
&\frac{1}{n}\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) = \frac{1}{n}\mathbf{1}_n \mathcal{J}(\hat{\boldsymbol{\eta}}_{\mathcal{M}})\mathbf{1}_n = \frac{1}{n}\sum_{i=1}^{n} d_i \\
= \;& \frac{1}{n}\left\{\sum_{i=1}^{n} b'' \circ \theta(\hat{\eta}_{\mathcal{M},i}) \left[\theta'(\hat{\eta}_{\mathcal{M},i})\right]^2 + \left[b' \circ \theta(\hat{\eta}_{\mathcal{M},i}) - Y_i\right] \theta''(\hat{\eta}_{\mathcal{M},i})\right\} \\
\xrightarrow{P} \;& \frac{1}{n}\left\{\sum_{i=1}^{n} b'' \circ \theta(\hat{\eta}_{\mathcal{M},i}) \left[\theta'(\hat{\eta}_{\mathcal{M},i})\right]^2 + \left[b' \circ \theta(\hat{\eta}_{\mathcal{M},i}) - b' \circ \theta(\eta_{\mathcal{M}_T,i}^*)\right] \theta''(\hat{\eta}_{\mathcal{M},i})\right\} \\
\xrightarrow{P} \;& \frac{1}{n}\left\{\sum_{i=1}^{n} b'' \circ \theta(\eta_{\mathcal{M},i}^*) \left[\theta'(\eta_{\mathcal{M},i}^*)\right]^2 + \left[b' \circ \theta(\eta_{\mathcal{M},i}^*) - b' \circ \theta(\eta_{\mathcal{M}_T,i}^*)\right] \theta''(\eta_{\mathcal{M},i}^*)\right\},
\end{aligned} \tag{A.3}$$

where the second last line is given by (A.2) and the fact $\mathbb{E}(Y_i) = b' \circ \theta(\eta_{\mathcal{M}_T,i}^*)$, for all $i = 1, \ldots, n$, and the last line is given by the continuous mapping theorem. Since for all $i = 1, \ldots, n$, $\mathbf{x}_i$ is bounded, $\eta_{\mathcal{M},i}^*$ and $\eta_{\mathcal{M}_T,i}^*$ are also bounded. For each term in the summation of (A.3), it is bounded due to the continuity assumptions on the third derivatives of $b(\cdot)$ and $\theta(\cdot)$. Therefore, $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})/n$ is bounded in probability.

If $\mathcal{M} \supset \mathcal{M}_T$, (A.3) becomes

$$\frac{1}{n}\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} b'' \circ \theta(\eta^*_{\mathcal{M}_T,i}) \left[\theta'(\eta^*_{\mathcal{M}_T,i})\right]^2 = \frac{1}{n}\mathcal{I}_n(\alpha^*_{\mathcal{M}}), \tag{A.4}$$

which is also the limit of $\mathcal{I}_n(\hat{\alpha}_{\mathcal{M}})/n$. Because we assume that $b' \circ \theta(\cdot)$ is strictly monotonic, $\theta(\cdot)$ is also strictly monotonic. For each term in the summation of (A.4), it is positive because $\theta'(\cdot) \neq 0$ and $b'' \circ \theta(\eta)$ is positive for finite η. Therefore by (A.4), if $\mathcal{M} \supset \mathcal{M}_T$, then $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})/n$ is positive and bounded in probability, i.e., $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) = O_P(n)$. On the other hand, if $\mathcal{M} \not\supset \mathcal{M}_T$, then only (A.3) holds but not (A.4). Each term in the summation of (A.3) can be either positive, zero, or negative. In this case, by (A.3), $\mathcal{J}_n(\alpha_{\mathcal{M}})/n$ is bounded in probability, and it may equal to zero. Therefore, $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})$ is on the order of $O(n^{\tau_{\mathcal{M}}})$, where $\tau_n \leq 1$, so that it tends to ∞ at a rate no faster than $O_P(n)$.

Last, we show asymptotic results regarding the matrix

$$\begin{aligned}
\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) &= \mathbf{X}^{cT}_{\mathcal{M}}\mathcal{J}(\hat{\boldsymbol{\eta}}_{\mathcal{M}})\mathbf{X}^{c}_{\mathcal{M}} = \mathbf{X}^{T}_{\mathcal{M}}(\mathbf{I}_n - \hat{\mathcal{P}}_{\mathbf{1}_n})^T\mathcal{J}(\hat{\boldsymbol{\eta}}_{\mathcal{M}})(\mathbf{I}_n - \hat{\mathcal{P}}_{\mathbf{1}_n})\mathbf{X}_{\mathcal{M}} \\
&= \mathbf{X}^{T}_{\mathcal{M}}\left[\mathcal{J}(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) - \mathcal{J}(\hat{\boldsymbol{\eta}}_{\mathcal{M}})\mathbf{1}_n\left(\mathbf{1}_n^T\mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}})\mathbf{1}_n\right)^{-1}\mathbf{1}_n^T\mathcal{J}(\hat{\boldsymbol{\eta}}_{\mathcal{M}})\right]\mathbf{X}_{\mathcal{M}}.
\end{aligned}$$

For the (j,k)th entry, $1 \leq j < k \leq p_{\mathcal{M}}$,

$$\begin{aligned}
\frac{1}{n}\left[\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})\right]_{j,k} &= \frac{1}{n}\sum_{i=1}^{n} d_i x_{i,j}x_{i,k} - \frac{1}{n}\left(\sum_{i=1}^{n} d_i x_{i,j}\right)\left(\sum_{i=1}^{n} d_i\right)^{-1}\left(\sum_{i=1}^{n} d_i x_{i,k}\right) \\
&= \frac{1}{n}\sum_{i=1}^{n} d_i x_{i,j}x_{i,k} - \left(\frac{1}{n}\sum_{i=1}^{n} d_i x_{i,j}\right)\left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} d_i x_{i,k}\right)
\end{aligned}$$

is bounded since all $\mathbf{x}_i$ are bounded. Therefore, $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n$ is bounded in probability.

To show that for any $\mathcal{M} \supset \mathcal{M}_T$, $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n$ does not reduce to zero, we will show that it is a positive definite matrix. For any given non-zero vector $\mathbf{a} \in \mathbb{R}^{p_{\mathcal{M}}}$, we denote $\mathbf{X}_{\mathcal{M}}\mathbf{a} = (t_1, \ldots, t_n)^T$, whose entries are all bounded. When $\mathcal{M} \supset \mathcal{M}_T$, by (A.4), all d_i's have a positive lower bound, hence simple calculation gives

$$\frac{1}{n}\mathbf{a}^T\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})\mathbf{a} = \frac{1}{n}\sum_{i=1}^{n} d_i t_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} d_i\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} d_i t_i\right)^2 \geq 0.$$

Here the quality only holds if all t_i's are equal for $i = 1, \ldots, n$, which is impossible here because of the assumption $\mathbf{1}_n \notin C(\mathbf{X}_{\mathcal{M}})$. For large n, the assumption that the smallest eigenvalue of $\mathbf{X}^T\mathbf{X}/n$ being bounded from below by a positive constant suggests that $\mathbf{X}_{\mathcal{M}}^T\mathbf{X}_{\mathcal{M}}/n$ is positive definite, so $\mathbf{a}^T\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})\mathbf{a}/n \not\longrightarrow 0$.

Furthermore, arguing similarly to (A.4), we also have

$$\frac{1}{n}\sum_{i=1}^{n} d_i t_i^k \xrightarrow{P} \frac{1}{n}\sum_{i=1}^{n} b'' \circ \theta(\eta^*_{\mathcal{M}_T,i})\left[\theta'(\eta^*_{\mathcal{M}_T,i})\right]^2 t_i^k,$$

for $k = 0, 1, 2$. Therefore, for any vector $\mathbf{a}$, if $\mathcal{M} \supset \mathcal{M}_T$, then

$$\frac{1}{n}\mathbf{a}^T\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})\mathbf{a} - \frac{1}{n}\mathbf{a}^T\mathcal{I}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})\mathbf{a} \xrightarrow{P} 0.$$

i.e., $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n$ and $\mathcal{I}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n$ are asymptotically the same. □

A.4 Proof of Proposition 2

Proof. We first use proof by contradiction to show that for $\mathcal{M}$, the MLE of the intercept is unique. If both $(\hat{\alpha}_1, \hat{\boldsymbol{\beta}}_1)$ and $(\hat{\alpha}_2, \hat{\boldsymbol{\beta}}_2)$ maximize the likelihood for model $\mathcal{M}$, where $\hat{\alpha}_1 \neq \hat{\alpha}_2$, then

$$\hat{\alpha}_1\mathbf{1}_n + \mathbf{X}_{\mathcal{M}}\hat{\boldsymbol{\beta}}_1 = \hat{\alpha}_2\mathbf{1}_n + \mathbf{X}_{\mathcal{M}}\hat{\boldsymbol{\beta}}_2 \Longrightarrow (\hat{\alpha}_1 - \hat{\alpha}_2)\mathbf{1}_n = \mathbf{X}_{\mathcal{M}}(\hat{\boldsymbol{\beta}}_2 - \hat{\boldsymbol{\beta}}_1),$$

which is contradicted with $\mathbf{1}_n \notin C(\mathbf{X}_{\mathcal{M}})$. Similarly, we can show this MLE is the same as the one for model $\mathcal{M}'$, i.e., $\hat{\alpha}_{\mathcal{M}} = \hat{\alpha}_{\mathcal{M}'}$.

By (25), between the two models $\mathcal{M}$ and $\mathcal{M}'$,

$$\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) = \mathcal{J}_n(\hat{\alpha}_{\mathcal{M}'}), \quad z_{\mathcal{M}} = z_{\mathcal{M}'}.$$

So we just need to show $Q_{\mathcal{M}} = Q_{\mathcal{M}'}$. Since $\hat{\alpha}_{\mathcal{M}} = \hat{\alpha}_{\mathcal{M}'}$, (25) suggests that

$$\mathbf{x}_{\mathcal{M},i}^T \hat{\boldsymbol{\beta}}_{\mathcal{M}} = \mathbf{x}_{\mathcal{M}',i}^T \hat{\boldsymbol{\beta}}_{\mathcal{M}'}, \quad i = 1, \ldots, n.$$

Hence,

$$\begin{aligned}
\mathbf{X}_{\mathcal{M}}^c \hat{\boldsymbol{\beta}}_{\mathcal{M}} &= \mathbf{X}_{\mathcal{M}} \hat{\boldsymbol{\beta}}_{\mathcal{M}} - \left(\sum_{i=1}^n w_i \mathbf{x}_{\mathcal{M},i}^T \hat{\boldsymbol{\beta}}_{\mathcal{M}} \right) \mathbf{1}_n \\
&= \mathbf{X}_{\mathcal{M}'} \hat{\boldsymbol{\beta}}_{\mathcal{M}'} - \left(\sum_{i=1}^n w_i \mathbf{x}_{\mathcal{M}',i}^T \hat{\boldsymbol{\beta}}_{\mathcal{M}'} \right) \mathbf{1}_n = \mathbf{X}_{\mathcal{M}'}^c \hat{\boldsymbol{\beta}}_{\mathcal{M}'},
\end{aligned}$$

where $w_i = d_i / (\sum_{r=1}^n d_r)$. Therefore, we have

$$Q_{\mathcal{M}} = \left[\mathbf{X}_{\mathcal{M}}^c \hat{\boldsymbol{\beta}}_{\mathcal{M}} \right]^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}}) \left[\mathbf{X}_{\mathcal{M}}^c \hat{\boldsymbol{\beta}}_{\mathcal{M}} \right] = \left[\mathbf{X}_{\mathcal{M}'}^c \hat{\boldsymbol{\beta}}_{\mathcal{M}'} \right]^T \mathcal{J}_n(\hat{\boldsymbol{\eta}}_{\mathcal{M}'}) \left[\mathbf{X}_{\mathcal{M}'}^c \hat{\boldsymbol{\beta}}_{\mathcal{M}'} \right] = Q_{\mathcal{M}'}.$$

□

A.5 Proof of Proposition 3

Proof. The marginal likelihood of the mixture of g-priors is obtained by integrating out g from the marginal likelihood of the g-prior, i.e.,

$$p(\mathbf{Y} \mid \mathcal{M}) = \int_0^\infty p(\mathbf{Y} \mid \mathcal{M}, g) p(g) dg$$

Here $p(\mathbf{Y} \mid \mathcal{M}, g)$ is obtained under the integrated Laplace approximation as in (19). Because of the one-to-one mapping between g and u, we rewrite this integral in terms of u.

$$
\begin{aligned}
p(\mathbf{Y} \mid \mathcal{M}) &= \int_0^1 p(\mathbf{Y} \mid \mathcal{M}, u) p(u) du \\
&\propto \int_0^1 p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M}) \mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})^{-\frac{1}{2}} u^{\frac{p_{\mathcal{M}}}{2}} e^{-\frac{Q_{\mathcal{M}}}{2} u} \\
&\quad \cdot \frac{v^{\frac{a}{2}} \exp\left(\frac{s}{2v}\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right) \ \Phi_1\left(\frac{b}{2}, r, \frac{a+b}{2}, \frac{s}{2v}, 1-\kappa\right)} \ \frac{u^{\frac{a}{2}-1}(1-vu)^{\frac{b}{2}-1} e^{-\frac{s}{2}u}}{[\kappa + (1-\kappa)vu]^r} \ \mathbf{1}_{\{0<u<\frac{1}{v}\}} \, du \\
&= p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M}) \mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})^{-\frac{1}{2}} \frac{v^{\frac{a}{2}} \exp\left(\frac{s}{2v}\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right) \ \Phi_1\left(\frac{b}{2}, r, \frac{a+b}{2}, \frac{s}{2v}, 1-\kappa\right)} \\
&\quad \cdot \int_0^1 \frac{u^{\frac{a+p_{\mathcal{M}}}{2}-1}(1-vu)^{\frac{b}{2}-1} e^{-\frac{s+Q_{\mathcal{M}}}{2}u}}{[\kappa + (1-\kappa)vu]^r} \ \mathbf{1}_{\{0<u<\frac{1}{v}\}} \, du.
\end{aligned}
$$

Since the above integrand is proportional to a tCCH density (27) with updated parameters, the above integral equals $B\left(\frac{a+p_{\mathcal{M}}}{2}, \frac{b}{2}\right) \ \Phi_1\left(\frac{b}{2}, r, \frac{a+b+p_{\mathcal{M}}}{2}, \frac{s+Q_{\mathcal{M}}}{2v}, 1-\kappa\right) v^{-\frac{a+p_{\mathcal{M}}}{2}} \exp\left(-\frac{s+Q_{\mathcal{M}}}{2v}\right)$.

□

A.6 Proof of Proposition 4

Proof. The marginal prior on $\boldsymbol{\beta}_{\mathcal{M}}$ after integrating g out is

$$
p(\boldsymbol{\beta}_{\mathcal{M}} \mid \mathcal{M}) \propto \int_0^\infty g^{-\frac{p_{\mathcal{M}}}{2}} \exp\left[-\frac{\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2g}\right] g^{\frac{b}{2}-1} \left(\frac{1}{1+g}\right)^{\frac{a+b}{2}} \exp\left[\frac{sg}{2(1+g)}\right] dg \qquad \text{(A.5)}
$$

We will show that as $\|\boldsymbol{\beta}_{\mathcal{M}}\|_{\mathcal{J}_n} \to \infty$, both a lower bound and an upper bound of (A.5) are proportional to $\left(\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}\right)^{-\frac{a+p_{\mathcal{M}}}{2}}$. Since $s \geq 0$, a lower bound of of the right side of (A.5) is

$$
\int_0^\infty g^{-\frac{p_{\mathcal{M}}}{2}} e^{-\frac{\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2g}} g^{\frac{b}{2}-1} \left(\frac{1}{1+g}\right)^{\frac{a+b}{2}} dg = \int_0^\infty \left(\frac{g}{1+g}\right)^{\frac{a+b}{2}} \left(\frac{1}{g}\right)^{\frac{a+p_{\mathcal{M}}-2}{2}} e^{-\frac{\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2g}} d\left(\frac{1}{g}\right).
$$

Then according to the Watson's Lemma (Olver 1997, pp. 71), as $\|\boldsymbol{\beta}_{\mathcal{M}}\|_{\mathcal{J}_n} \to \infty$, the limit of this lower bound is proportional to $\left(\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}\right)^{-\frac{a+p_{\mathcal{M}}}{2}}$. Next we find an upper bound of the right side of (A.5) as

$$\int_0^\infty g^{-\frac{p_{\mathcal{M}}}{2}} \exp\left[-\frac{\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2(1+g)}\right] g^{\frac{b}{2}-1} \left(\frac{1}{1+g}\right)^{\frac{a+b}{2}} \exp\left[\frac{sg}{2(1+g)}\right] dg$$
$$= e^{-\frac{\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2}} B\left(\frac{b-p_{\mathcal{M}}}{2}, \frac{a+p_{\mathcal{M}}}{2}\right)\ {}_1F_1\left(\frac{b-p_{\mathcal{M}}}{2}, \frac{a+b}{2}, \frac{s+\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2}\right).$$

According to Abramowitz and Stegun (1970) formula (13.1.4),

$$ {}_1F_1(a,b,s) = \frac{\Gamma(b)}{\Gamma(a)} \exp(s) s^{a-b}[1+O(|s|^{-1})], \text{ when } \mathrm{Real}(s) > 0, \tag{A.6}$$

hence as $\|\boldsymbol{\beta}_{\mathcal{M}}\|_{\mathcal{J}_n} \to \infty$, the limit of the above upper bound converges to

$$\exp\left[-\frac{\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2}\right] \Gamma\left(\frac{a+p_{\mathcal{M}}}{2}\right) \exp\left[\frac{s+\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2}\right] \cdot \left(\frac{s+\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}}{2}\right)^{-\frac{a+p_{\mathcal{M}}}{2}} \propto \left(\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}\right)^{-\frac{a+p_{\mathcal{M}}}{2}}.$$

Therefore, as $\|\boldsymbol{\beta}_{\mathcal{M}}\|_{\mathcal{J}_n}$ increases, or equivalently, as $\|\boldsymbol{\beta}_{\mathcal{M}}\|$ increases, both the lower bound and upper bound of $p(\boldsymbol{\beta}_{\mathcal{M}} \mid \mathcal{M})$ are proportional to $\left(\|\boldsymbol{\beta}_{\mathcal{M}}\|^2_{\mathcal{J}_n}\right)^{-\frac{a+p_{\mathcal{M}}}{2}}$. $\square$

A.7 Special Functions: Definition and Useful Properties

We first review a list of special functions, including their definitions and relevant properties, that will be needed in the proof of Proposition 5.

- Confluent hypergeometric function (Abramowitz and Stegun 1970, eq 13.2.1): for $\gamma > \alpha > 0$,

$$ {}_1F_1(\alpha,\gamma,x) = \frac{1}{\mathrm{B}(\gamma-\alpha,\alpha)} \int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1} e^{xu}\, du.$$

  - By (Abramowitz and Stegun 1970, eq 13.2.27): ${}_1F_1(\alpha,\gamma,x) = e^x \cdot {}_1F_1(\gamma-\alpha,\gamma,-x)$.
  - By (Abramowitz and Stegun 1970, eq 6.5.12), the incomplete Gamma function:

$$\gamma(a,s) = \int_0^s t^{a-1}e^{-t}dt = {}_1F_1(a,a+1,-s)\frac{s^a}{a}.$$

- ${}_1F_1(\alpha, \gamma, 0) = 1$.

- Confluent hypergeometric function of two variables (Gordy 1998)[1]: for $\gamma > \alpha > 0$ and $y < 1$,

$$\Phi_1(\alpha, \beta, \gamma, x, y) = \frac{1}{\mathrm{B}(\gamma - \alpha, \alpha)} \int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-yu)^{-\beta} e^{xu}\, du,$$

Special cases:

  - If $x = 0$, then $\Phi_1(\alpha, \beta, \gamma, 0, y) = {}_2F_1(\beta, \alpha; \gamma; y)$.
  - If $\beta = 0$ or $y = 0$, then $\Phi_1(\alpha, 0, \gamma, x, y) = \Phi_1(\alpha, \beta, \gamma, x, 0) = \Phi_1(\alpha, 0, \gamma, x, 0) = {}_1F_1(\alpha, \gamma, x)$.
  - If $x = 0$ and $y = 0$, then $\Phi_1(\alpha, \beta, \gamma, 0, 0) = 1$.

- Hypergeometric function (Abramowitz and Stegun 1970, eq 15.3.1): for $\gamma > \alpha > 0$

$${}_2F_1(\beta, \alpha; \gamma; x) = \frac{1}{\mathrm{B}(\gamma - \alpha, \alpha)} \int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-xu)^{-\beta}\, du.$$

  - By (Abramowitz and Stegun 1970, eq 15.3.3): in the definition of ${}_2F_1$ function above, let $w = \frac{1-u}{1-xu}$, then

$${}_2F_1(\beta, \alpha; \gamma; x) = (1-x)^{\gamma-\beta-\alpha}\ {}_2F_1(\gamma - \beta, \gamma - \alpha; \gamma; x)$$

  - ${}_2F_1(0, \alpha; \gamma, x) = {}_2F_1(\beta, \alpha; \gamma, 0) = 1$
  - ${}_2F_1(\beta, 1; \beta, x) = (1-x)^{-1} {}_2F_1(0, \beta - 1; \beta, x) = (1-x)^{-1}$
  - By (Abramowitz and Stegun 1970, eq 15.3.4): ${}_2F_1(\beta, \alpha; \gamma; x) = (1-x)^{-\beta} {}_2F_1\left(\beta, \gamma - \alpha; \gamma, \frac{x}{x-1}\right)$
  - By (Abramowitz and Stegun 1970, eq 15.3.5): ${}_2F_1(\beta, \alpha; \gamma; x) = (1-x)^{-\alpha} {}_2F_1\left(\alpha, \gamma - \beta; \gamma, \frac{x}{x-1}\right)$

[1] Note: the definition in Gordy (1998) is slightly different from that in Gradshteyn and Ryzhik (2007).

- Hypergeometric function of two variables (Appell function) (Weisstein 2009): for $\gamma > \alpha > 0$,

$$F_1(\alpha; \beta, \beta'; \gamma; x, y) = \frac{1}{\mathrm{B}(\gamma - \alpha, \alpha)} \int_0^1 u^{\alpha-1}(1-u)^{\gamma-\alpha-1}(1-xu)^{-\beta}(1-yu)^{-\beta'}\, du.$$

A.8 Proof of Proposition 5

Proof. To begin we establish that the marginal likelihood conditional on g is well defined under the g-prior when the design matrix is not full rank for a general linear model. We will assume the inner product space defined by the vector space $\mathbb{R}^n$ equipped with inner product $\mathbf{u}^T\mathbf{W}\mathbf{v}$ for two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ where $\mathbf{W}$ is a real, $n \times n$ symmetric positive definite matrix. Similarly, $\|\mathbf{u}\|^2_{\mathbf{W}} \equiv \mathbf{u}^T\mathbf{W}\mathbf{u}$.

For the model

$$\mathbf{Y} = \mathbf{1}_n\beta_0 + \mathbf{X}_{\mathcal{M}}\boldsymbol{\beta}_{\mathcal{M}} + \boldsymbol{\epsilon}, \quad \text{with } \boldsymbol{\epsilon} \mid \phi \sim \mathsf{N}(\mathbf{0}_n, \phi^{-1}\mathbf{W}^{-1}),$$

let $\mathcal{P}_{\mathbf{1}} = \mathbf{1}_n(\mathbf{1}_n^T\mathbf{W}\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{W}$ denote the orthogonal projection onto the column space of $\mathbf{1}_n$ and without loss of generality reparameterize the model

$$\mathbf{Y} = \mathbf{1}_n\alpha + \mathbf{X}^c_{\mathcal{M}}\boldsymbol{\beta}_{\mathcal{M}} + \boldsymbol{\epsilon}$$

where $\mathbf{X}^c_{\mathcal{M}} = (\mathbf{I}_n - \mathcal{P}_{\mathbf{1}_n})\mathbf{X}_{\mathcal{M}}$ and $\alpha \equiv \beta_0 - (\mathbf{1}_n^T\mathbf{W}\mathbf{1}_n)^{-1}\mathbf{1}_n^T\mathbf{W}\mathbf{X}_{\mathcal{M}}\boldsymbol{\beta}_{\mathcal{M}}$. Adopting the g-prior of the form

$$\boldsymbol{\beta}_{\mathcal{M}} \mid \alpha, \phi, g \sim \mathsf{N}\left(0, \frac{g}{\phi}(\mathbf{X}^{cT}_{\mathcal{M}}\mathbf{W}\mathbf{X}^c_{\mathcal{M}})^{-}\right),$$

where $(\mathbf{X}^{cT}_{\mathcal{M}}\mathbf{W}\mathbf{X}^c_{\mathcal{M}})^{-}$ is any generalized inverse, standard normal theory for the linear combination $\mathbf{X}^c_{\mathcal{M}}\boldsymbol{\beta}_{\mathcal{M}} + \boldsymbol{\epsilon}$ can be used to show that $\mathbf{Y}$ is equal in distribution

$$\mathbf{Y} \mid \alpha, \phi, g, \mathcal{M} \sim \mathsf{N}\left(\mathbf{1}_n\alpha, \phi^{-1}(\mathbf{I}_n + g\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}})\mathbf{W}^{-1}\right) \tag{A.7}$$

where $\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}} = \mathbf{X}^c_{\mathcal{M}}(\mathbf{X}^{cT}_{\mathcal{M}}\mathbf{W}\mathbf{X}^c_{\mathcal{M}})^{-}\mathbf{X}^{cT}_{\mathcal{M}}\mathbf{W}$ is the $\rho_{\mathcal{M}} \leq p_{\mathcal{M}}$ orthogonal projection onto the column space $\mathbf{X}^c_{\mathcal{M}}$ in the inner product space. As the projection $\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}}$ does not depend on the choice of generalized inverse, this establishes that the marginal likelihood for the model will not depend on the choice of generalized inverse employed in defining the g-prior.

Continuing with integration with respect to α, ϕ under the independent Jeffreys prior $p(\alpha, \phi) \propto \phi^{-1}$,

$$p(\mathbf{Y} \mid g, \mathcal{M}) = \iint (2\pi)^{-\frac{n}{2}} |\mathbf{I}_n + g\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}}|^{-\frac{1}{2}} |\mathbf{W}|^{\frac{1}{2}} \phi^{\frac{n}{2}-1} e^{-\frac{\phi}{2}\left\{(\mathbf{Y}-\mathbf{1}_n\alpha)^T\mathbf{W}(\mathbf{I}_n - \frac{g}{1+g}\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}})(\mathbf{Y}-\mathbf{1}_n\alpha)\right\}} d\alpha\, d\phi \tag{A.8}$$

rearrangement of terms can be used to show that

$$p(\mathbf{Y} \mid g, \mathcal{M}) = p(\mathbf{Y} \mid \mathcal{M}_{\varnothing})(1+g)^{\frac{n-\rho_{\mathcal{M}}-1}{2}} \left\{1 + g(1 - R^2_{\mathcal{M}})\right\}^{-\frac{n-1}{2}}$$

where $R^2_{\mathcal{M}}$ is defined in (40) and

$$p(\mathbf{Y} \mid \mathcal{M}_{\varnothing}) = (2\pi)^{-\frac{n-1}{2}} \Gamma\left(\frac{n-1}{2}\right) |\mathbf{W}|^{\frac{1}{2}} (\mathbf{1}_n^T\mathbf{W}\mathbf{1}_n)^{-\frac{1}{2}} \left[\frac{\|(\mathbf{I}_n - \mathcal{P}_{\mathbf{1}_n})\mathbf{Y}\|^2_{\mathbf{W}}}{2}\right]^{-\frac{n-1}{2}}$$

is the marginal under the null model. Note that in (A.8), the determinant $|\mathbf{I}_n + g\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}}| = (1+g)^{\rho_{\mathcal{M}}}$, because the eigenvalues of the orthogonal projection $\mathcal{P}_{\mathbf{X}^c_{\mathcal{M}}}$ are one with a multiplicity of $\rho_{\mathcal{M}}$ and zero with a multiplicity of $p_{\mathcal{M}} - \rho_{\mathcal{M}}$. The Bayes Factor for comparing $\mathcal{M}$ to $\mathcal{M}_{\varnothing}$ is thus

$$\mathsf{BF}[\mathcal{M}, \mathcal{M}_{\varnothing}] = (1+g)^{\frac{n-\rho_{\mathcal{M}}-1}{2}} \left\{1 + g(1 - R^2_{\mathcal{M}})\right\}^{-\frac{n-1}{2}},$$

which will be one for any model $\mathcal{M}$ where $R^2_{\mathcal{M}} = 1$ and $\rho_{\mathcal{M}} = n - 1$.

For simplicity in the rest of proof, we omit the subscript $\mathcal{M}$ when there is no ambiguity. We now show part (1). In the tCCH distribution, if $r = 0$ or $\kappa = 1$, then

$$\Phi_1\left(\frac{b}{2}, r, \frac{a+b}{2}, \frac{s}{2v}, 1-\kappa\right) = \Phi_1\left(\frac{b}{2}, 0, \frac{a+b}{2}, \frac{s}{2v}, 0\right) = {}_1F_1\left(\frac{b}{2}, \frac{a+b}{2}, \frac{s}{2v}\right).$$

Then the marginal likelihood becomes

$$
\begin{aligned}
p(\mathbf{Y} \mid \mathcal{M}) &= \frac{p(\mathbf{Y} \mid \mathcal{M}_{\emptyset})\ v^{\frac{a}{2}} \exp\left(\frac{s}{2v}\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{b}{2}, \frac{a+b}{2}, \frac{s}{2v}\right)} \int_0^{1/v} \frac{u^{\frac{a+\rho}{2}-1}(1-vu)^{\frac{b}{2}-1} e^{-\frac{su}{2}}}{[(1-R^2)+R^2u]^{\frac{n-1}{2}}}\ du \\
&= \frac{p(\mathbf{Y} \mid \mathcal{M}_{\emptyset})\ v^{\frac{a}{2}} \exp\left(\frac{s}{2v}\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{b}{2}, \frac{a+b}{2}, \frac{s}{2v}\right)} \int_0^{1/v} \frac{u^{\frac{a+\rho}{2}-1}(1-vu)^{\frac{b}{2}-1} e^{-\frac{su}{2}}}{\left\{\left[1-\left(1-\frac{1}{v}\right)R^2\right]\left[\frac{1-R^2}{1-\left(1-\frac{1}{v}\right)R^2} + \frac{R^2/v}{1-\left(1-\frac{1}{v}\right)R^2}\cdot(vu)\right]\right\}^{\frac{n-1}{2}}}\ du \\
&= \frac{p(\mathbf{Y} \mid \mathcal{M}_{\emptyset})\ v^{\frac{a}{2}} \exp\left(\frac{s}{2v}\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{b}{2}, \frac{a+b}{2}, \frac{s}{2v}\right)} \cdot \frac{\mathrm{B}\left(\frac{a+\rho}{2}, \frac{b}{2}\right)\ \Phi_1\left(\frac{b}{2}, \frac{n-1}{2}, \frac{a+b+\rho}{2}, \frac{s}{2v}, \frac{R^2/v}{1-\left(1-\frac{1}{v}\right)R^2}\right)}{v^{\frac{a+\rho}{2}} \exp\left(\frac{s}{2v}\right)\left[1-\left(1-\frac{1}{v}\right)R^2\right]^{\frac{n-1}{2}}} \\
&= p(\mathbf{Y} \mid \mathcal{M}_{\emptyset}) \cdot \frac{B\left(\frac{a+\rho}{2}, \frac{b}{2}\right)\ \Phi_1\left(\frac{b}{2}, \frac{n-1}{2}, \frac{a+b+\rho}{2}, \frac{s}{2v}, \frac{R^2/v}{1-\left(1-\frac{1}{v}\right)R^2}\right)}{v^{\frac{\rho}{2}}\left[1-\left(1-\frac{1}{v}\right)R^2\right]^{\frac{n-1}{2}} B\left(\frac{a}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{b}{2}, \frac{a+b}{2}, \frac{s}{2v}\right)}.
\end{aligned}
$$

Here the second last equality is given by the propriety of the tCCH density function (27).

Then we show part (2). In the tCCH distribution, when $s = 0$, then

$$
\Phi_1\left(\frac{b}{2}, r, \frac{a+b}{2}, 0, 1-\kappa\right) = {}_2F_1\left(r, \frac{b}{2}; \frac{a+b}{2}; 1-\kappa\right).
$$

Hence, the marginal likelihood becomes

$$
p(\mathbf{Y} \mid \mathcal{M}) = \frac{p(\mathbf{Y} \mid \mathcal{M}_{\emptyset})\ v^{\frac{a}{2}}}{B\left(\frac{a}{2}, \frac{b}{2}\right)\ {}_2F_1\left(r, \frac{b}{2}; \frac{a+b}{2}; 1-\kappa\right)} \int_0^{1/v} \frac{u^{\frac{a+\rho}{2}-1}(1-vu)^{\frac{b}{2}-1}}{[(1-R^2)+R^2u]^{\frac{n-1}{2}}\left[\kappa+(1-\kappa)vu\right]^r}\ du \tag{A.9}
$$

For simplification, we denote $x = 1 - 1/\kappa$ and $w = 1 - (1 - vu)/(1 - xvu)$. By change of variable,

$$
u = \frac{w}{v(1-x+xw)}, \quad \frac{du}{dw} = \frac{1-x}{v(1-x+xw)^2},
$$

and the integral in (A.9) is

$$
\begin{aligned}
&\int_0^{1/v} \frac{u^{\frac{a+\rho}{2}-1}(1-vu)^{\frac{b}{2}-1}}{[(1-R^2)+R^2u]^{\frac{n-1}{2}}\,[\kappa+(1-\kappa)vu]^r}\,du \\
= &\int_0^1 \frac{\left[\frac{w}{v(1-x+xw)}\right]^{\frac{a+\rho}{2}-1}\left[\frac{(1-x)(1-w)}{1-x+xw}\right]^{\frac{b}{2}-1}\frac{1-x}{v(1-x+xw)^2}}{\left\{\frac{(1-R^2)v(1-x)+[(1-R^2)vx+R^2]w}{v(1-x+xw)}\right\}^{\frac{n-1}{2}}\left(\frac{1}{1-x+xw}\right)^r}\,dw \\
= &\frac{(1-x)^{\frac{b}{2}}\;v^{\frac{n-1-a-\rho}{2}}}{[(1-R^2)v(1-x)]^{\frac{n-1}{2}}\;(1-x)^{\frac{a+b+\rho+1-n-2r}{2}}}\int_0^1 \frac{w^{\frac{a+\rho}{2}-1}(1-w)^{\frac{b}{2}-1}}{\left[1-\frac{(1-R^2)vx+R^2}{(1-R^2)v(x-1)}w\right]^{\frac{n-1}{2}}\left(1-\frac{x}{x-1}w\right)^{\frac{a+b+\rho+1-n-2r}{2}}}\,dw \\
= &\frac{\kappa^{\frac{a+\rho-2r}{2}}\;v^{-\frac{a+\rho}{2}}}{(1-R^2)^{\frac{n-1}{2}}}B\left(\frac{a+\rho}{2},\frac{b}{2}\right)\cdot \\
&F_1\left(\frac{a+\rho}{2};\frac{a+b+\rho+1-n-2r}{2},\frac{n-1}{2};\frac{a+b+\rho}{2};1-\kappa,\frac{(1-R^2)v(1-\kappa)-R^2\kappa}{(1-R^2)v}\right).
\end{aligned}
$$

□

A.9 Derivation of (44)

Proof. Similar to (19), we apply integrated Laplace approximation to obtain $p(\mathbf{Y} \mid \phi, \mathcal{M}, g)$, then marginalize ϕ out as follows.

$$
\begin{aligned}
p(\mathbf{Y} \mid \mathcal{M}, g) &= \int_0^\infty p(\mathbf{Y} \mid \phi, \mathcal{M}, g)p(\phi)d\phi \\
\propto &\int_0^\infty p(\mathbf{Y} \mid \hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \phi, \mathcal{M})\left[\phi\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})\right]^{-\frac{1}{2}}(1+g)^{-\frac{p_{\mathcal{M}}}{2}}e^{-\frac{\phi Q_{\mathcal{M}}}{2(1+g)}}\phi^{-1}d\phi \\
\propto &\left[\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})\right]^{-\frac{1}{2}}(1+g)^{-\frac{p_{\mathcal{M}}}{2}}\int_0^\infty \phi^{\frac{n-1}{2}-1}e^{\phi\left\{-\frac{Q_{\mathcal{M}}}{2(1+g)}+\sum_{i=1}^n\left[Y_i(\hat{\theta}_i-t_i)-b(\hat{\theta}_i)+b(t_i)\right]\right\}}d\phi \\
\propto &\left[\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})\right]^{-\frac{1}{2}}(1+g)^{-\frac{p_{\mathcal{M}}}{2}}\left\{\frac{Q_{\mathcal{M}}}{2(1+g)}-\sum_{i=1}^n\left[Y_i(\hat{\theta}_i-t_i)-b(\hat{\theta}_i)+b(t_i)\right]\right\}^{-\frac{n-1}{2}} \\
\propto &\frac{\left[\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})\right]^{-\frac{1}{2}}u^{\frac{p_{\mathcal{M}}}{2}}}{\left\{uQ_{\mathcal{M}}+2\sum_{i=1}^n\left[Y_i(t_i-\hat{\theta}_i)-b(t_i)+b(\hat{\theta}_i)\right]\right\}^{\frac{n-1}{2}}}.
\end{aligned}
$$

Here, the last step replaces g with $u = 1/(1+g)$. □

A.10 Proof of Model Selection Consistency

We first show a lemma about a non-central χ^2 distribution, which is useful to prove some of the following lemmas and theorems. Here the symbol $\chi^2_k(m)$ denotes a non-central χ^2 distribution with degrees of freedom k and non-centrality parameter m.

Lemma A.2. *If a sequence of random variables $\{X_n : n = 1, 2, \ldots\}$ have independent non-central χ^2 distributions: $X_n \sim \chi^2_k(nA_n)$, where random variables $A_n \xrightarrow{D} a_0 \in \mathbb{R}^+ \cup \{0\}$, then as $n \longrightarrow \infty$, $X_n/n \xrightarrow{P} a_0$.*

Proof. For any $n \in \mathbb{N}$, the characteristic function of X_n/n evaluated at $t \in \mathbb{R}$ is

$$\begin{aligned}\phi_{X_n/n}(t) &= \mathbb{E}\left(e^{itX_n/n}\right) = \mathbb{E}_{A_n}\left[\mathbb{E}\left(e^{itX_n/n} \mid A_n\right)\right] \\ &= \mathbb{E}_{A_n}\left[\exp\left(\frac{itA_n}{1-2it/n}\right)(1-2it/n)^{-\frac{k}{2}}\right] = (1-2it/n)^{-\frac{k}{2}} \cdot \mathbb{E}_{A_n}\left[\exp\left(\frac{itA_n}{1-2it/n}\right)\right].\end{aligned}$$

Denote a complex valued random variable $B_n = A_n/(1-2it/n)$. Since the limit of A_n is a constant, for the series $\{A_n : n \in \mathbb{N}\}$, convergence in distribution is equivalent to convergence in probability. Because of the continuous mapping theorem, $B_n \xrightarrow{\mathrm{P}} a_0$, or equivalently, convergence in distribution. Denote the bounded and continuous function $h(B_n) = \exp(itB_n)$, then according to Portmanteau lemma, $\mathbb{E}[h(B_n)] \longrightarrow \mathbb{E}[h(a_0)] = h(a_0)$. So for any $t \in \mathbb{R}$,

$$\lim_{n\to\infty} \phi_{X_n/n}(t) = \lim_{n\to\infty}(1-2it/n)^{-k/2} \cdot \lim_{n\to\infty} \mathbb{E}[h(B_n)] = h(a_0) = \exp(ita_0),$$

where the limit is the characteristic function of a degenerated distribution at a_0. Therefore, X_n/n converge in distribution to a constant a_0, which implies convergence in probability.

$\square$

In order to show the asymptotic performance of the Bayes factor $\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}}$, we first study asymptotic behaviors of the terms in the Bayes factors in the following lemmas. When testing nested models, the log likelihood ratio between $\mathcal{M}_T$ and $\mathcal{M}$ converges in distribution to a central (non-central) χ^2 distribution, when the smaller (larger) model is true. The following lemma studies asymptotic behaviors of the likelihood ratio, which does not require models $\mathcal{M}$ and $\mathcal{M}_T$ to be nested.

Lemma A.3. *Denote the the likelihood ratio by*

$$\Lambda_{\mathcal{M}_T:\mathcal{M}} \triangleq \frac{p(\mathbf{Y}|\hat{\alpha}_{\mathcal{M}_T}, \hat{\boldsymbol{\beta}}_{\mathcal{M}_T}, \mathcal{M}_T)}{p(\mathbf{Y}|\hat{\alpha}_{\mathcal{M}}, \hat{\boldsymbol{\beta}}_{\mathcal{M}}, \mathcal{M})} = \exp\left(\frac{z_{\mathcal{M}_T} - z_{\mathcal{M}}}{2}\right) \tag{A.10}$$

As the sample size n increases,

1) *if $\mathcal{M}_T \subset \mathcal{M}$, then $\Lambda_{\mathcal{M}_T:\mathcal{M}} = O_P(1)$.*

2) *if $\mathcal{M}_T \not\subset \mathcal{M}$, then $\Lambda_{\mathcal{M}_T:\mathcal{M}} = O_P\left(e^{c_{\mathcal{M}} n}\right)$, where $c_{\mathcal{M}}$ is a positive constant.*

Proof. In the first case where $\mathcal{M} \supset \mathcal{M}_T$, from the well-known results of likelihood ratio test, $z_{\mathcal{M}} - z_{\mathcal{M}_T}$ has a central chi-square distribution $\chi^2_{p_{\mathcal{M}} - p_{\mathcal{M}_T}}$. Therefore, the limiting distribution of the log-likelihood ratio does not depend on n, i.e., $\Lambda_{\mathcal{M}_T:\mathcal{M}} = O_P(1)$.

In the second case where $\mathcal{M} \not\supset \mathcal{M}_T$, we first examine the sub-case where $\mathcal{M} \subset \mathcal{M}_T$. According to the power calculation results for GLM in Self et al. (1992) and Shieh (2000), when testing nested models, if the larger model is true, then we have that $z_{\mathcal{M}_T} - z_{\mathcal{M}}$ converges in distribution to a non-central χ^2 of degrees of freedom $p_{\mathcal{M}_T} - p_{\mathcal{M}}$. The non-centrality parameter Ψ is approximately

$$\Psi \approx \sum_{i=1}^{n} b'(\theta^*_{\mathcal{M}_T,i}) \left(\theta^*_{i,\mathcal{M}_T} - \theta^*_{i,\mathcal{M}}\right) - \left[b(\theta^*_{i,\mathcal{M}_T}) - b(\theta^*_{i,\mathcal{M}})\right],$$

where $\theta^*_{i,\mathcal{M}} = \theta(\eta^*_{i,\mathcal{M}})$, for $i = 1, \ldots, n$. By a Taylor expansion, there exist a $\tilde{\theta}_i$ between $\theta^*_{\mathcal{M}_T,i}$ and $\theta^*_{\mathcal{M},i}$, such that $b(\theta^*_{i,\mathcal{M}}) = b(\theta^*_{i,\mathcal{M}_T}) + b'(\theta^*_{\mathcal{M}_T,i}) \left(\theta^*_{i,\mathcal{M}_T} - \theta^*_{i,\mathcal{M}}\right) + b''(\tilde{\theta}_i) \left(\theta^*_{i,\mathcal{M}_T} - \theta^*_{i,\mathcal{M}}\right)^2 /2$. This combined with the assumption $b''(\cdot) > 0$ gives that $\lim_{n\to\infty} \Psi/n$ converges to a positive constant $c_{\mathcal{M}}$. Then by Lemma A.2, $(z_{\mathcal{M}_T} - z_{\mathcal{M}})/n \xrightarrow{P} c_{\mathcal{M}}$, and hence $\Lambda_{\mathcal{M}_T:\mathcal{M}} = O_P(e^{c_{\mathcal{M}} n})$.

In the case where $\mathcal{M}$ and $\mathcal{M}_T$ are not nested, we introduce a third model $\mathcal{M}'$ which includes all the predictors in both $\mathcal{M}$ and $\mathcal{M}_T$. Using a similar method as in Self et al. (1992), we can treat $\mathcal{M}'$ also as the true model (although with some redundant predictors) when comparing with $\mathcal{M}$ and easily show that $\Lambda_{\mathcal{M}':\mathcal{M}}$ also has a non-central χ^2 distribution. Hence we decompose $\Lambda_{\mathcal{M}_T:\mathcal{M}} = \Lambda_{\mathcal{M}_T:\mathcal{M}'} \cdot \Lambda_{\mathcal{M}':\mathcal{M}}$. Since both pairs $(\mathcal{M}_T, \mathcal{M}')$ and $(\mathcal{M}' : \mathcal{M})$ are nested models, we can apply the previous results twice: $\Lambda_{\mathcal{M}_T:\mathcal{M}'} = O_P(1)$ and $\Lambda_{\mathcal{M}':\mathcal{M}} = O_P(e^{c_{\mathcal{M}}n})$. Therefore, we can conclude that $\Lambda_{\mathcal{M}_T:\mathcal{M}} = O_P(1) \cdot O_P(e^{c_{\mathcal{M}}n}) = O_P(e^{c_{\mathcal{M}}n})$. $\square$

The Bayes factors contain the Wald statistics $Q_{\mathcal{M}_T}$ and $Q_{\mathcal{M}}$. We next study their asymptotic behaviors.

Lemma A.4. *The Wald statistic $Q_{\mathcal{M}} = O_P(n^{\xi_{\mathcal{M}}})$, where $0 \leq \xi_{\mathcal{M}} \leq 1$. In particular,*

1) If $\mathcal{M}_T \neq \mathcal{M}_{\o}$, then for any $\mathcal{M} \supset \mathcal{M}_T$, $\xi_{\mathcal{M}} = 1$.

2) if $\mathcal{M}_T = \mathcal{M}_{\o}$, then for any model $\mathcal{M}$, $\xi_{\mathcal{M}} = 0$.

Proof. For any $\mathcal{M} \supset \mathcal{M}_T$, we have shown in the proof of Lemma A.1 that the MLE $\hat{\boldsymbol{\beta}}_{\mathcal{M}}$ converges in probability to the true value $\boldsymbol{\beta}^*_{\mathcal{M}}$, and $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})/n$ is a finite positive definite matrix and converges to $\mathcal{I}_n(\boldsymbol{\beta}^*_{\mathcal{M}})/n$ in probability. By Lemma A.1 and Slutsky's theorem, we can rewrite the asymptotic normality (A.1) as

$$\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})^{\frac{1}{2}} \left(\hat{\boldsymbol{\beta}}_{\mathcal{M}} - \boldsymbol{\beta}^*_{\mathcal{M}}\right) \xrightarrow{d} \mathrm{N}(0, \mathbf{I}_{p_{\mathcal{M}}}).$$

Therefore, $Q_{\mathcal{M}} = \hat{\boldsymbol{\beta}}^T_{\mathcal{M}} \mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) \hat{\boldsymbol{\beta}}_{\mathcal{M}}$ converges in distribution to a non-central χ^2 random variable with degrees of freedom $p_{\mathcal{M}}$ and non-centrality parameter $\boldsymbol{\beta}^{*T}_{\mathcal{M}} \mathcal{I}_n(\boldsymbol{\beta}^*_{\mathcal{M}}) \boldsymbol{\beta}^*_{\mathcal{M}}$, which is $O(n)$ if $\boldsymbol{\beta}^*_{\mathcal{M}} \neq \mathbf{0}$, and zero otherwise. Since $\boldsymbol{\beta}^*_{\mathcal{M}} = \boldsymbol{\beta}^*_{\mathcal{M}_T}$ in the sense that all entries in $\boldsymbol{\beta}^*_{\mathcal{M}}$ that correspond to predictors not in $\mathcal{M}_T$ are filled with zero, $\boldsymbol{\beta}^*_{\mathcal{M}} = \mathbf{0}$ is equivalent to $\mathcal{M}_T = \mathcal{M}_{\o}$. Therefore, by Lemma A.2, if $\mathcal{M}_T \neq \mathcal{M}_{\o}$, then $Q_{\mathcal{M}} = O_P(n)$; if $\mathcal{M}_T = \mathcal{M}_{\o}$, then $Q_{\mathcal{M}} = O_P(1)$.

For any $\mathcal{M} \not\supset \mathcal{M}_T$, since convergence in probability is preserved under addition and multiplication (Resnick 1999, pp. 175), we have $Q_{\mathcal{M}} - \boldsymbol{\beta}^{*T}_{\mathcal{M}} \mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}}) \boldsymbol{\beta}^*_{\mathcal{M}} \xrightarrow{P} 0$, i.e., $Q_{\mathcal{M}}$ is at most on the same order of $\mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})$. By Lemma A.1, we have $\xi_{\mathcal{M}} = \tau_{\mathcal{M}}$ if $\boldsymbol{\beta}^*_{\mathcal{M}} \neq \mathbf{0}$, and $\xi_{\mathcal{M}} = 0$ if $\boldsymbol{\beta}^*_{\mathcal{M}} = \mathbf{0}$. $\square$

Based on the results of Lemma A.4, the next lemma discusses the asymptotic properties of $\Omega^{\text{CH}}_{\mathcal{M}_T:\mathcal{M}}$, a term that appears in the Bayes factor under the CH prior.

Lemma A.5. *Under the CH prior, denote the term in $BF_{\mathcal{M}_T:\mathcal{M}}$:*

$$\Omega^{CH}_{\mathcal{M}_T:\mathcal{M}} \triangleq \frac{B\left(\frac{a+p_{\mathcal{M}_T}}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{a+p_{\mathcal{M}_T}}{2}, \frac{a+b+p_{\mathcal{M}_T}}{2}, -\frac{s+Q_{\mathcal{M}_T}}{2}\right)}{B\left(\frac{a+p_{\mathcal{M}}}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{a+p_{\mathcal{M}}}{2}, \frac{a+b+p_{\mathcal{M}}}{2}, -\frac{s+Q_{\mathcal{M}}}{2}\right)}. \tag{A.11}$$

1) If $\mathcal{M}_T \neq \mathcal{M}_\emptyset$, then as n increases,

$$\Omega^{CH}_{\mathcal{M}_T:\mathcal{M}} = \begin{cases} O_P\left(n^{\frac{\xi_{\mathcal{M}} p_{\mathcal{M}} - p_{\mathcal{M}_T} - a(1-\xi_{\mathcal{M}})}{2}}\right) & \textit{if } b \textit{ is fixed, and } s \textit{ is fixed} \\ O_P\left(n^{\frac{p_{\mathcal{M}} - p_{\mathcal{M}_T}}{2}}\right) & \textit{if } b = O(n), \textit{ or } s = O(n) \end{cases}$$

In particular, if $\mathcal{M} \supset \mathcal{M}_T$, then $\Omega^{CH}_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{\frac{p_{\mathcal{M}} - p_{\mathcal{M}_T}}{2}}\right)$ for all b and s.

2) If $\mathcal{M}_T = \mathcal{M}_\emptyset$, then as n increases,

$$\Omega^{CH}_{\mathcal{M}_T:\mathcal{M}} = \begin{cases} O_P\left(1\right) & \textit{if } b \textit{ is fixed, and } s \textit{ is fixed} \\ O_P\left(n^{\frac{p_{\mathcal{M}} - p_{\mathcal{M}_T}}{2}}\right) & \textit{if } b = O(n), \textit{ or } s = O(n) \end{cases}$$

Proof. We first show Case 1) where $\mathcal{M}_T \neq \mathcal{M}_\emptyset$, by Lemma A.4, $\xi_{\mathcal{M}_T} = 1$. We consider the following three scenarios about parameters b and s being fixed or $O(n)$.

Scenario 1: Both b, s are fixed. By Abramowitz and Stegun (1970) formula (13.1.5),

$${}_1F_1(a, b, s) = \frac{\Gamma(b)}{\Gamma(b-a)}(-s)^{-a}[1 + O(|s|^{-1})], \text{ when } \text{Real}(s) < 0. \tag{A.12}$$

Continuous mapping theorem suggests that for any model $\mathcal{M}$ whose $Q_{\mathcal{M}} = O_P(n^{\xi_{\mathcal{M}}})$,

$$\Omega^{\text{CH}}_{\mathcal{M}_T:\mathcal{M}} \approx \frac{\Gamma\left(\frac{a+p_{\mathcal{M}_T}}{2}\right)\left(\frac{s+Q_{\mathcal{M}_T}}{2}\right)^{-\frac{a+p_{\mathcal{M}_T}}{2}}}{\Gamma\left(\frac{a+p_{\mathcal{M}}}{2}\right)\left(\frac{s+Q_{\mathcal{M}}}{2}\right)^{-\frac{a+p_{\mathcal{M}}}{2}}} \propto \frac{(s+Q_{\mathcal{M}_T})^{-\frac{a+p_{\mathcal{M}_T}}{2}}}{(s+Q_{\mathcal{M}})^{-\frac{a+p_{\mathcal{M}}}{2}}} = O_P\left(n^{\frac{\xi_{\mathcal{M}}p_{\mathcal{M}}-p_{\mathcal{M}_T}-a(1-\xi_{\mathcal{M}})}{2}}\right). \tag{A.13}$$

Scenario 2: b is fixed, and $s = O(n)$. Since $s + Q_{\mathcal{M}_T} = O(n)$ and $s + Q_{\mathcal{M}} = O(n)$, then by (A.13), $\Omega^{\text{CH}}_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right)$.

Scenario 3: $b = O(n)$. Lemma A.4 indicates that $Q_{\mathcal{M}}$ is between $O_P(1)$ and $O_P(n)$. By Slater (1960) formula (4.3.3): if b is large, and a, s are bounded, then

$$ {}_1F_1(a,b,s) = 1 + O(|b|^{-1}) \text{ is bounded;} \tag{A.14}$$

and by Slater (1960) formulas (4.3.7): if b is large, $s = by$, and a, y are bounded, then

$$ {}_1F_1(a,b,s) = (1-y)^{-a}\left[1 - \frac{a(a+1)}{2b}\left(\frac{y}{1-y}\right)^2 + O(|b|^{-2})\right] \text{ is also bounded.} \tag{A.15}$$

Therefore, under the CH prior when parameter $b = O(n)$,

$$\Omega^{\text{CH}}_{\mathcal{M}_T:\mathcal{M}} = \frac{B\left(\frac{a+p_{\mathcal{M}_T}}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{a+p_{\mathcal{M}_T}}{2}, \frac{a+b+p_{\mathcal{M}_T}}{2}, -\frac{s+Q_{\mathcal{M}_T}}{2}\right)}{B\left(\frac{a+p_{\mathcal{M}}}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{a+p_{\mathcal{M}}}{2}, \frac{a+b+p_{\mathcal{M}}}{2}, -\frac{s+Q_{\mathcal{M}}}{2}\right)} \xrightarrow{\text{P}} C \cdot \frac{B\left(\frac{a+p_{\mathcal{M}_T}}{2}, \frac{b}{2}\right)}{B\left(\frac{a+p_{\mathcal{M}}}{2}, \frac{b}{2}\right)}.$$

According to the Stirling's Formula $\Gamma(n) = e^{-n}n^{n-\frac{1}{2}}(2\pi)^{\frac{1}{2}}(1+O(n^{-1}))$, the above ratio becomes $O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right)$.

Next we examine Case 2) where $\mathcal{M}_T = \mathcal{M}_{\emptyset}$. In this case, Lemma A.4 suggests that both $Q_{\mathcal{M}_T}$ and $Q_{\mathcal{M}}$ are on the same order $O_P(1)$. Hence in Scenario 1, where both b and s are fixed, $\Omega^{\text{CH}}_{\mathcal{M}_T:\mathcal{M}} = O_P(1)$; In Scenario 2, since both $s + Q_{\mathcal{M}_T}$ and $s + Q_{\mathcal{M}}$ are on the order of $O_P(n)$, the same deviation and result as in Case 1) Scenario 2 apply. In Scenario 3, both $s + Q_{\mathcal{M}_T}$ and $s + Q_{\mathcal{M}}$ are $O_P(1)$ if s is fixed, and $O_P(n)$ if $s = O(n)$, so the same derivation and result as in Case 1) Scenario 3 apply. $\square$

Lemma A.6. *Under the robust prior, denote the term in $\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}}$:*

$$\Omega^R_{\mathcal{M}_T:\mathcal{M}} \triangleq \left(\frac{p_{\mathcal{M}}+1}{p_{\mathcal{M}_T}+1}\right)^{\frac{1}{2}} \cdot \frac{Q_{\mathcal{M}_T}^{-\frac{p_{\mathcal{M}_T}+1}{2}}}{Q_{\mathcal{M}}^{-\frac{p_{\mathcal{M}}+1}{2}}} \cdot \frac{\gamma\left(\frac{p_{\mathcal{M}_T}+1}{2}, \frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+1)}\right)}{\gamma\left(\frac{p_{\mathcal{M}}+1}{2}, \frac{Q_{\mathcal{M}}(p_{\mathcal{M}}+1)}{2(n+1)}\right)}. \tag{A.16}$$

As the sample size n increases, $\Omega^R_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right)$.

Proof. By Abramowitz and Stegun (1970) formula (6.5.12), the incomplete Gamma function $\gamma(a,s) = \int_0^s t^{a-1}e^{-t}dt$ can be expressed using the ${}_1F_1$ function

$$\gamma(a,s) = {}_1F_1(a, a+1, -s)\frac{s^a}{a}. \tag{A.17}$$

Therefore, (A.16) becomes

$$\left(\frac{p_{\mathcal{M}}+1}{p_{\mathcal{M}_T}+1}\right)^{\frac{1}{2}} \cdot \frac{Q_{\mathcal{M}_T}^{-\frac{p_{\mathcal{M}_T}+1}{2}}}{Q_{\mathcal{M}}^{-\frac{p_{\mathcal{M}}+1}{2}}} \cdot \frac{\left(\frac{p_{\mathcal{M}_T}+1}{2}\right)^{-1}\left(\frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+1)}\right)^{\frac{p_{\mathcal{M}_T}+1}{2}} {}_1F_1\left(\frac{p_{\mathcal{M}_T}+1}{2}, \frac{p_{\mathcal{M}_T}+3}{2}, -\frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+1)}\right)}{\left(\frac{p_{\mathcal{M}}+1}{2}\right)^{-1}\left(\frac{Q_{\mathcal{M}}(p_{\mathcal{M}}+1)}{2(n+1)}\right)^{\frac{p_{\mathcal{M}}+1}{2}} {}_1F_1\left(\frac{p_{\mathcal{M}}+1}{2}, \frac{p_{\mathcal{M}}+3}{2}, -\frac{Q_{\mathcal{M}}(p_{\mathcal{M}}+1)}{2(n+1)}\right)}$$

Since ${}_1F_1(a,b,0) = 1$, and both $Q_{\mathcal{M}_T}/n, Q_{\mathcal{M}}/n$ are bounded, the ratio between the ${}_1F_1$ functions is bounded as n increases. Therefore we further simplify $\Omega^{\mathrm{R}}_{\mathcal{M}_T:\mathcal{M}} \propto (n+1)^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}} = O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right)$. This result holds no matter whether $\mathcal{M}_T = \mathcal{M}_\varnothing$ or not. □

Lemma A.7. *Under the intrinsic prior, denote the term in $\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}}$:*

$$\Omega^I_{\mathcal{M}_T:\mathcal{M}} \triangleq \frac{\left(\frac{n+p_{\mathcal{M}}+1}{p_{\mathcal{M}}+1}\right)^{\frac{p_{\mathcal{M}}}{2}} e^{\frac{Q_{\mathcal{M}}(p_{\mathcal{M}}+1)}{2(n+p_{\mathcal{M}}+1)}} B\left(\frac{p_{\mathcal{M}_T}+1}{2}, \frac{1}{2}\right) \Phi_1\left(\frac{1}{2}, 1, \frac{p_{\mathcal{M}_T}+2}{2}, \frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+p_{\mathcal{M}_T}+1)}, -\frac{p_{\mathcal{M}_T}+1}{n}\right)}{\left(\frac{n+p_{\mathcal{M}_T}+1}{p_{\mathcal{M}_T}+1}\right)^{\frac{p_{\mathcal{M}_T}}{2}} e^{\frac{Q_{\mathcal{M}_T}\left(p_{\mathcal{M}_T}+1\right)}{2\left(n+p_{\mathcal{M}_T}+1\right)}} B\left(\frac{p_{\mathcal{M}}+1}{2}, \frac{1}{2}\right) \Phi_1\left(\frac{1}{2}, 1, \frac{p_{\mathcal{M}}+2}{2}, \frac{Q_{\mathcal{M}}(p_{\mathcal{M}}+1)}{2(n+p_{\mathcal{M}}+1)}, -\frac{p_{\mathcal{M}}+1}{n}\right)}$$

As the sample size n increases, $\Omega^I_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right)$.

Proof. Since $p_{\mathcal{M}_T}, p_{\mathcal{M}}$ are bounded, and $Q_{\mathcal{M}_T}/n, Q_{\mathcal{M}}/n$ are bounded in probability, as $n \to \infty$,

$$\Omega^{\mathrm{I}}_{\mathcal{M}_T:\mathcal{M}} \xrightarrow{\mathrm{P}} C \cdot \frac{\left(\frac{n+p_{\mathcal{M}}+1}{p_{\mathcal{M}}+1}\right)^{\frac{p_{\mathcal{M}}}{2}}}{\left(\frac{n+p_{\mathcal{M}_T}+1}{p_{\mathcal{M}_T}+1}\right)^{\frac{p_{\mathcal{M}_T}}{2}}} = O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right).$$

□

Lemma A.8. *Under the local EB, denote the term in $\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}}$:*

$$\Omega^{LEB}_{\mathcal{M}_T:\mathcal{M}} \triangleq \frac{\max\left\{\exp\left(-\frac{Q_{\mathcal{M}_T}}{2}\right), \left(\frac{Q_{\mathcal{M}_T}}{p_{\mathcal{M}_T}}\right)^{-\frac{p_{\mathcal{M}_T}}{2}} \exp\left(-\frac{p_{\mathcal{M}_T}}{2}\right)\right\}}{\max\left\{\exp\left(-\frac{Q_{\mathcal{M}}}{2}\right), \left(\frac{Q_{\mathcal{M}}}{p_{\mathcal{M}}}\right)^{-\frac{p_{\mathcal{M}}}{2}} \exp\left(-\frac{p_{\mathcal{M}}}{2}\right)\right\}}. \tag{A.18}$$

1) *If $\mathcal{M}_T \neq \mathcal{M}_\o$, then as n increases, $\Omega^{LEB}_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{\frac{\xi_{\mathcal{M}} p_{\mathcal{M}} - p_{\mathcal{M}_T}}{2}}\right)$. In particular, if $\mathcal{M} \supset \mathcal{M}_T$, then $\Omega^{LEB}_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{\frac{p_{\mathcal{M}}-p_{\mathcal{M}_T}}{2}}\right)$.*

2) *If $\mathcal{M}_T = \mathcal{M}_\o$, then as n increases, $\Omega^{LEB}_{\mathcal{M}_T:\mathcal{M}} = O_P(1)$.*

Proof. Case 2) is straightforward, because when $\mathcal{M}_T = \mathcal{M}_\o$, $Q_{\mathcal{M}_T} = O_P(1)$ and $Q_{\mathcal{M}} = O_P(1)$. Now let us focus on Case 1). In (A.18), the numerator equals $\exp(-Q_{\mathcal{M}_T}/2)$ if and only if $Q_{\mathcal{M}_T} \leq p_{\mathcal{M}_T}$, and the denominator follows the same rule when we replacing $\mathcal{M}_T$ with $\mathcal{M}$. Since $\mathcal{M}_T \neq \mathcal{M}_\o$, $Q_{\mathcal{M}} = O_P(n)$ is greater than $p_{\mathcal{M}}$ for large n. Hence the numerator of (A.18) is proportional to $(Q_{\mathcal{M}_T}/p_{\mathcal{M}_T})^{-\frac{p_{\mathcal{M}_T}}{2}} \exp(-p_{\mathcal{M}_T}/2) = O_P(n^{-\frac{p_{\mathcal{M}_T}}{2}})$. For model $\mathcal{M}$ whose $Q_{\mathcal{M}} = O_P(n^{\xi_{\mathcal{M}}})$, if $\xi_{\mathcal{M}} > 0$, then when n is large enough, $Q_{\mathcal{M}} > p_{\mathcal{M}}$, so the denominator is $O_P(n^{-\frac{\xi_{\mathcal{M}} p_{\mathcal{M}}}{2}})$. If $\xi_{\mathcal{M}} = 0$, then the denominator is $O_P(1)$, which can also be written as $O_P(n^{-\frac{\xi_{\mathcal{M}} p_{\mathcal{M}}}{2}})$. □

We now examine the model selection consistency.

Proof of Theorem 1

Proof. By Lemma A.1, $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}}) = O_P(n^{\tau_{\mathcal{M}}})$, where $0 \leq \tau_{\mathcal{M}} \leq 1$, and $\tau_{\mathcal{M}} = 1$ if $\mathcal{M} \supset \mathcal{M}_T$. Hence,

$$\left[\frac{\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}_T})}{\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})}\right]^{-\frac{1}{2}} = O_P\left(n^{-\frac{1-\tau_{\mathcal{M}}}{2}}\right).$$

For the CH prior,

$$\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}} = \left[\frac{\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}_T})}{\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})}\right]^{-\frac{1}{2}} \cdot \Lambda_{\mathcal{M}_T:\mathcal{M}} \cdot \Omega^{\mathrm{CH}}_{\mathcal{M}_T:\mathcal{M}} \cdot [1 + O_P(1/n)]. \tag{A.19}$$

We first consider the case where both b and s are fixed, by using the results in Lemma A.3 and A.5. In the case where $\mathcal{M}_T \neq \mathcal{M}_{\o}$, for any non-true model $\mathcal{M} \supset \mathcal{M}_T$, then $p_{\mathcal{M}} > p_{\mathcal{M}_T}$, $\tau_{\mathcal{M}} = 1$, and $\xi_{\mathcal{M}} = 1$, hence

$$\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}} = O_P(1) \cdot O_P(1) \cdot O_P\left(n^{\frac{p_{\mathcal{M}} - p_{\mathcal{M}_T}}{2}}\right) \cdot [1 + O_P(1/n)] \xrightarrow{\mathrm{P}} \infty.$$

On the other hand, if $\mathcal{M} \not\supset \mathcal{M}_T$, then

$$\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}} = O_P\left(n^{-\frac{1-\tau_{\mathcal{M}}}{2}}\right) \cdot O_P\left(e^{c_{\mathcal{M}} n}\right) \cdot O_P\left(n^{\frac{\xi_{\mathcal{M}} p_{\mathcal{M}} - p_{\mathcal{M}_T} - a(1-\xi_{\mathcal{M}})}{2}}\right) \cdot [1 + O_P(1/n)] \xrightarrow{\mathrm{P}} \infty.$$

In contrast, if $\mathcal{M}_T = \mathcal{M}_{\o}$, then for any model $\mathcal{M}$, since $\mathcal{M} \supset \mathcal{M}_T$, $\tau_{\mathcal{M}} = 1$. So the Bayes factor

$$\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}} = O_P(1) \cdot O_P(1) \cdot O_P(1) \cdot [1 + O_P(1/n)]$$

is bounded, which suggests the selection consistency does not hold when $\mathcal{M}_T = \mathcal{M}_{\o}$.

Next consider the case where $b = O(n)$ or $s = O(n)$. For any model $\mathcal{M} \not\supset \mathcal{M}_T$, the proof is similar as above. If $\mathcal{M} \supset \mathcal{M}_T$, then $\tau_{\mathcal{M}} = 1$ and $p_{\mathcal{M}} > p_{\mathcal{M}_T}$, so

$$\mathsf{BF}_{\mathcal{M}_T:\mathcal{M}} = O_P(1) \cdot O_P(1) \cdot O_P\left(n^{\frac{p_{\mathcal{M}} - p_{\mathcal{M}_T}}{2}}\right) \cdot [1 + O_P(1/n)] \xrightarrow{\mathrm{P}} \infty,$$

which holds even when $\mathcal{M}_T = \mathcal{M}_{\o}$.

For the robust prior, the intrinsic prior, and local EB, their Bayes factor are given by (A.19), with $\Omega^{\mathrm{CH}}_{\mathcal{M}_T:\mathcal{M}}$ replaced by $\Omega^{\mathrm{R}}_{\mathcal{M}_T:\mathcal{M}}$, $\Omega^{\mathrm{I}}_{\mathcal{M}_T:\mathcal{M}}$, and $\Omega^{\mathrm{LEB}}_{\mathcal{M}_T:\mathcal{M}}$, respectively. By Lemma A.6, A.7, and A.8, the proofs are similar to the CH prior, hence omitted. □

A.11 Proof to Proposition 6

Proof. If $b = O(n)$ then by (A.14) or (A.15),

$$\begin{aligned}\mathbb{E}(1/g) &= \frac{B\left(\frac{a}{2}+1, \frac{b}{2}-1\right)\ {}_1F_1\left(\frac{a}{2}+1, \frac{a+b}{2}, -\frac{s}{2}\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right)\ {}_1F_1\left(\frac{a}{2}, \frac{a+b}{2}, -\frac{s}{2}\right)} \\ &\propto \frac{B\left(\frac{a}{2}+1, \frac{b}{2}-1\right)}{B\left(\frac{a}{2}, \frac{b}{2}\right)} \longrightarrow \frac{a}{b-2} = O(1/n).\end{aligned} \tag{A.20}$$

If b is fixed and $s = O(n)$, then by (A.12) and (A.20),

$$\mathbb{E}(1/g) \approx \frac{B\left(\frac{a}{2}+1, \frac{b}{2}-1\right)\Gamma\left(\frac{b}{2}\right)\left(\frac{s}{2}\right)^{\frac{a}{2}}}{B\left(\frac{a}{2}, \frac{b}{2}\right)\Gamma\left(\frac{b}{2}-1\right)\left(\frac{s}{2}\right)^{\frac{a}{2}+1}} \propto \frac{1}{s} = O(1/n).$$

□

A.12 Proof of Proposition 7

Proof. For the CH prior, according to (32), the conditional posterior of $z = 1-u$ is

$$z \mid \mathbf{Y}, \mathcal{M} \xrightarrow{D} \mathrm{CH}\left(\frac{b}{2}, \frac{a+p_{\mathcal{M}}}{2}, -\frac{s+Q_{\mathcal{M}}}{2}\right), \tag{A.21}$$

and its characteristic function is

$$\phi_z(t) = \mathbb{E}\left(e^{itz}\right) = \int \frac{z^{\frac{b}{2}-1}(1-z)^{\frac{a+p_{\mathcal{M}_T}}{2}-1}e^{\left(\frac{s+Q_{\mathcal{M}_T}}{2}+it\right)z}}{B(\frac{b}{2}, \frac{a+p_{\mathcal{M}_T}}{2})\ {}_1F_1(\frac{b}{2}, \frac{a+b+p_{\mathcal{M}_T}}{2}, \frac{s+Q_{\mathcal{M}_T}}{2})}dz = \frac{{}_1F_1(\frac{b}{2}, \frac{a+b+p_{\mathcal{M}_T}}{2}, \frac{s+Q_{\mathcal{M}_T}}{2}+it)}{{}_1F_1(\frac{b}{2}, \frac{a+b+p_{\mathcal{M}_T}}{2}, \frac{s+Q_{\mathcal{M}_T}}{2})}$$

Lemma A.4 shows that if $\mathcal{M}_T \neq \mathcal{M}_{\o}$, then $s + Q_{\mathcal{M}_T} = O_P(n)$. If $b = O(1)$, then by (A.6) and the continuous mapping theorem, for any $t \in \mathbb{R}$, as n goes in to infinity,

$$\phi_z(t) \longrightarrow \frac{\exp(\frac{s+Q_{\mathcal{M}_T}}{2} + it) \cdot (\frac{s+Q_{\mathcal{M}_T}}{2} + it)^{-\frac{a+p_{\mathcal{M}_T}}{2}}}{\exp(\frac{s+Q_{\mathcal{M}_T}}{2}) \cdot (\frac{s+Q_{\mathcal{M}_T}}{2})^{-\frac{a+p_{\mathcal{M}_T}}{2}}} \xrightarrow{P} \exp(it).$$

If $b = O(n)$, then using formula (A.15), we can obtain the same limit.

For the robust prior, we examine the characteristic function of $u = 1 - z$. Based on (37),

$$\begin{aligned}\phi_u(t) = \mathbb{E}\left(e^{itu}\right) &= \frac{\int_0^{\frac{p_{\mathcal{M}_T}+1}{n+1}} u^{\frac{p_{\mathcal{M}_T}+1}{2}-1} e^{\left(it - \frac{Q_{\mathcal{M}_T}}{2}\right)u} du}{\int_0^{\frac{p_{\mathcal{M}_T}+1}{n+1}} u^{\frac{p_{\mathcal{M}_T}+1}{2}-1} e^{-\frac{Q_{\mathcal{M}_T} u}{2}} du} \\ &= \frac{\gamma\left(\frac{p_{\mathcal{M}_T}+1}{2}, \frac{(Q_{\mathcal{M}_T}-2it)(p_{\mathcal{M}_T}+1)}{2(n+1)}\right)}{\gamma\left(\frac{p_{\mathcal{M}_T}+1}{2}, \frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+1)}\right)} \cdot \left(\frac{Q_{\mathcal{M}_T} - 2it}{Q_{\mathcal{M}_T}}\right)^{-\frac{p_{\mathcal{M}_T}+1}{2}}.\end{aligned}$$

Since $Q_{\mathcal{M}_T} = O_P(n)$, for any fixed $t \in \mathbb{R}$, the ratio of the incomplete Gamma functions goes to 1, and so does the second fraction. Therefore, $\phi_u(t) \xrightarrow{\text{P}} 1$, which is the characteristic function of the degenerate distribution at 0.

For the intrinsic prior, by (30) and Table 1, the conditional posterior of u is

$$u \mid \mathbf{Y}, \mathcal{M}_T \sim \text{tCCH}\left(\frac{p_{\mathcal{M}_T}+1}{2}, \frac{1}{2}, 1, \frac{Q_{\mathcal{M}_T}}{2}, \frac{n+p_{\mathcal{M}_T}+1}{p_{\mathcal{M}_T}+1}, \frac{n+p_{\mathcal{M}_T}+1}{n}\right), \tag{A.22}$$

and hence its characteristic function for any $t \in \mathbb{R}$ is

$$\phi_u(t) = \exp\left\{\frac{it(p_{\mathcal{M}_T}+1)}{n+p_{\mathcal{M}_T}+1}\right\} \frac{\Phi_1\left(\frac{1}{2}, 1, \frac{p_{\mathcal{M}_T}+2}{2}, \frac{(Q_{\mathcal{M}_T}-2it)(p_{\mathcal{M}_T}+1)}{2(n+p_{\mathcal{M}_T}+1)}, -\frac{p_{\mathcal{M}_T}+1}{n}\right)}{\Phi_1\left(\frac{1}{2}, 1, \frac{p_{\mathcal{M}_T}+2}{2}, \frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+p_{\mathcal{M}_T}+1)}, -\frac{p_{\mathcal{M}_T}+1}{n}\right)}. \tag{A.23}$$

Since $Q_{\mathcal{M}_T} = O_P(n)$ and

$$\frac{(Q_{\mathcal{M}_T} - 2it)(p_{\mathcal{M}_T}+1)}{2(n+p_{\mathcal{M}_T}+1)} - \frac{Q_{\mathcal{M}_T}(p_{\mathcal{M}_T}+1)}{2(n+p_{\mathcal{M}_T}+1)} \xrightarrow{\text{P}} 0,$$

by continuous mapping theorem, the ratio of the two Φ_1 functions in (A.23) converges to one in probability. Therefore, under the intrinsic prior, $\phi_u(t) \xrightarrow{\text{P}} 1$. $\square$

A.13 Proof of Theorem 2

Proof. For the CH prior, we will prove the BMA estimation consistency in two steps: 1) $\mathcal{M}_T \neq \mathcal{M}_{\o}$ and 2) $\mathcal{M}_T = \mathcal{M}_{\o}$. When $\mathcal{M}_T \neq \mathcal{M}_{\o}$, the model selection consistency always holds, so we just need to show the estimation consistency under the true model $\mathcal{M}_T$. For notation simplicity, we denote $\boldsymbol{\Sigma}_{n,\mathcal{M}} = \mathcal{J}_n(\hat{\boldsymbol{\beta}}_{\mathcal{M}})^{-1}$. According to (17) and (A.21), the characteristic function of the posterior distribution $p(\boldsymbol{\beta}_{\mathcal{M}_T} \mid \mathcal{M}_T, \mathbf{Y})$ is

$$
\begin{aligned}
\phi_{\boldsymbol{\beta}_{\mathcal{M}_T}}(\mathbf{t}) &= \int e^{i\mathbf{t}^T \boldsymbol{\beta}_{\mathcal{M}_T}}\ p(\boldsymbol{\beta}_{\mathcal{M}_T} | \mathcal{M}_T, \mathbf{Y})\ d\boldsymbol{\beta}_{\mathcal{M}_T} \\
&= \int e^{i\mathbf{t}^T \boldsymbol{\beta}_{\mathcal{M}_T}} \left\{ \int p(\boldsymbol{\beta}_{\mathcal{M}_T} | z, \mathcal{M}_T, \mathbf{Y})\ p(z|\mathcal{M}_T, \mathbf{Y}) dz \right\} d\boldsymbol{\beta}_{\mathcal{M}_T} \\
&= \int \left\{ \int e^{i\mathbf{t}^T \boldsymbol{\beta}_{\mathcal{M}_T}}\ p(\boldsymbol{\beta}_{\mathcal{M}_T} | z, \mathcal{M}_T, \mathbf{Y})\ d\boldsymbol{\beta}_{\mathcal{M}_T} \right\} p(z|\mathcal{M}_T, \mathbf{Y}) dz \\
&= \int e^{z(i\mathbf{t}^T \hat{\boldsymbol{\beta}}_{\mathcal{M}_T} - \frac{1}{2}\mathbf{t}^T \boldsymbol{\Sigma}_{n,\mathcal{M}_T} \mathbf{t})}\ p(z|\mathcal{M}_T, \mathbf{Y}) dz
\end{aligned}
$$

In the above calculation, the integrand $e^{i\mathbf{t}^T \boldsymbol{\beta}_{\mathcal{M}_T}}$ has a bounded modulus, so according to Fubini's Theorem, the two integrals (with respect to z and $\boldsymbol{\beta}_{\mathcal{M}_T}$) can be interchanged. Since $Q_{\mathcal{M}_T} = O_P(n)$ and $\boldsymbol{\Sigma}_{n,\mathcal{M}_T} = O_P(n^{-1})$, using methods similar to the proof of Proposition 7 and asymptotic normality of MLE, we can show that for any vector $\mathbf{t}$,

$$
\phi_{\boldsymbol{\beta}_{\mathcal{M}_T}}(\mathbf{t}) \longrightarrow e^{i\mathbf{t}^T \hat{\boldsymbol{\beta}}_{\mathcal{M}_T} - \frac{1}{2}\mathbf{t}^T \boldsymbol{\Sigma}_{n,\mathcal{M}_T} \mathbf{t}} \xrightarrow{P} e^{i\mathbf{t}^T \boldsymbol{\beta}^*_{\mathcal{M}_T}}.
$$

On the other hand, when $\mathcal{M}_T = \mathcal{M}_{\o}$, under the CH prior model selection consistency does not hold if both b and s are fixed. Hence we need to examine the limit of posterior distribution of $\boldsymbol{\beta}_{\mathcal{M}}$ under all models. Under any model $\mathcal{M}$, the true model is nested in it, so the MLE of the coefficient $\hat{\boldsymbol{\beta}}_{\mathcal{M}}$ converges to the true parameters $\mathbf{0}$ in probability as n goes to infinity.

Since the modulus of $e^{i\mathbf{t}^T\boldsymbol{\beta}_{\mathcal{M}}}$ is bounded by a constant 1, which is integrable if regarded as a function of z, so according to the dominated convergence theorem, the characteristic function of the posterior distribution $p(\boldsymbol{\beta}_{\mathcal{M}} \mid \mathbf{Y}, \mathcal{M})$ evaluated at any vector $\mathbf{t} \in \mathbb{R}^p$ is

$$\begin{aligned}\phi_{\boldsymbol{\beta}_{\mathcal{M}}}(\mathbf{t}) &= \int e^{z(i\mathbf{t}^T\hat{\boldsymbol{\beta}}_{\mathcal{M}} - \frac{1}{2}\mathbf{t}^T\boldsymbol{\Sigma}_{n,\mathcal{M}}\mathbf{t})}\, p(z \mid \mathcal{M}, \mathbf{Y})dz \\ &\xrightarrow{P} \int \left[e^{z(i\mathbf{t}^T\mathbf{0} - \frac{1}{2}\mathbf{t}^T\mathbf{0}\mathbf{t})}\right] p(z \mid \mathcal{M}, \mathbf{Y})dz = 1.\end{aligned}$$

For the robust and intrinsic priors, model selection consistency always holds. So we just need to consider under $\mathcal{M}_T$. Based on (37) and (A.22), proofs similar to the above proof of the CH prior can show that either $\mathcal{M}_T \neq \mathcal{M}_{\varnothing}$ or $\mathcal{M}_T = \mathcal{M}_{\varnothing}$, the characteristic function of $p(\boldsymbol{\beta}_{\mathcal{M}_T} \mid \mathcal{M}_T, \mathbf{Y})$ converges to $e^{i\mathbf{t}^T\boldsymbol{\beta}^*_{\mathcal{M}_T}}$ or 1 in probability, respectively. □

B Test-Based Bayes Factors

B.1 Test-Based Bayes Factor under the g-Prior

In Bayesian hypothesis testing, while the traditional Bayes factor computes the ratio between marginal likelihoods of data (referred to as data-based BF, or DBF in short), another type of Bayes factor, defined as the ratio between marginal likelihoods of a test statistic, has also been introduced (Johnson 2005, 2008). In particular, based on the likelihood ratio statistic, the test-based Bayes factor (TBF) has been applied in model selection under the g-prior (Hu and Johnson 2009; Held et al. 2015, 2016), where models with high TBFs are preferable.

To compute the TBF based on the likelihood ratio deviance $z_{\mathcal{M}}$ (22), first, asymptotic theory (Davidson and Lever 1970) suggests that the limit distribution of $z_{\mathcal{M}}$ under the null model $\mathcal{M}_{\varnothing}$ and under a local alternative model $\mathcal{M}$ are central and non-central Chi-squares, respectively,

$$z_{\mathcal{M}} \mid \mathcal{M}_{\varnothing} \sim \chi^2_{p_{\mathcal{M}}}, \quad z_{\mathcal{M}} \mid \mathcal{M} \sim \chi^2_{p_{\mathcal{M}}}(\lambda_{\mathcal{M}}), \text{ where } \lambda_{\mathcal{M}} = \boldsymbol{\beta}^T_{\mathcal{M}}\mathcal{I}_n(\boldsymbol{\beta}_{\mathcal{M}} = \mathbf{0})\boldsymbol{\beta}_{\mathcal{M}}.$$

Then, as $p(z_{\mathcal{M}} \mid \mathcal{M}, \boldsymbol{\beta}_{\mathcal{M}})$ depends on $\boldsymbol{\beta}_{\mathcal{M}}$ through the non-centrality parameter $\lambda_{\mathcal{M}}$, integrating $\boldsymbol{\beta}_{\mathcal{M}}$ out under its prior density yields the marginal likelihood $p(z_{\mathcal{M}} \mid \mathcal{M})$. Last, the TBF is defined as the ratio

$$\mathrm{TBF}_{\mathcal{M}:\mathcal{M}_\varnothing} = \frac{p(z_{\mathcal{M}} \mid \mathcal{M})}{p(z_{\mathcal{M}} \mid \mathcal{M}_\varnothing)} = \frac{\int p(z_{\mathcal{M}} \mid \boldsymbol{\beta}_{\mathcal{M}}, \mathcal{M}) p(\boldsymbol{\beta}_{\mathcal{M}} \mid \mathcal{M}) d\boldsymbol{\beta}_{\mathcal{M}}}{p(z_{\mathcal{M}} \mid \mathcal{M}_\varnothing)}. \tag{B.1}$$

To conduct model selection in GLMs, Held et al. (2015) derive the TBF under the g-prior (8), in whose density, $\boldsymbol{\beta}_{\mathcal{M}}$ appears in the format of $\lambda_{\mathcal{M}}$. Thus the conjugacy permits a tractable marginal likelihood $p(z_{\mathcal{M}} \mid \mathcal{M})$ as a Gamma distribution. Therefore, the resulting TBF has a closed form expression as in (23).

B.2 Comparing Data-Based and Test-Based Bayes Factors

The TBF (23) has a similar expression to the DBF (21). In fact, the two Bayes factors would be the same if $z_{\mathcal{M}} = Q_{\mathcal{M}}$ and $\mathcal{J}_n(\hat{\alpha}_{\mathcal{M}_\varnothing}) = \mathcal{J}_n(\hat{\alpha}_{\mathcal{M}})$. Naturally, it is interesting to examine how different the two Bayes factors are.

We compare DBF (21) and TBF (23) empirically through a logistic regression toy example, with $g = n$ and a single covariate generated from independent standard normal distributions. With the intercept set to $\alpha = 0.5$, three scenarios are studied with different coefficients $\beta = 0, 20/\sqrt{n}, 2$, which correspond to the null, local alternative, and alternative, respectively. To study asymptotics, various sample sizes $n = 100, 500, 1000, 5000$ are taken. For each combination of β and n, 100 independent datasets are generated. To obtain an accurate approximation to the DBF, in addition to the integrated Laplace approximation (ILA) formula (21), we also implement importance sampling (IS), which can be viewed as a gold standard if the number of samples drawn is large. Here we draw $m = 10000$ samples $\alpha^{(t)}, \beta^{(t)}$, independently from Student-t distributions with degrees of freedom 4, with location and scale parameters matching those in the corresponding conditional posteriors (17), (18).

Figure 1 shows that when the null or the local alternative is true, TBF (23) is asymp-

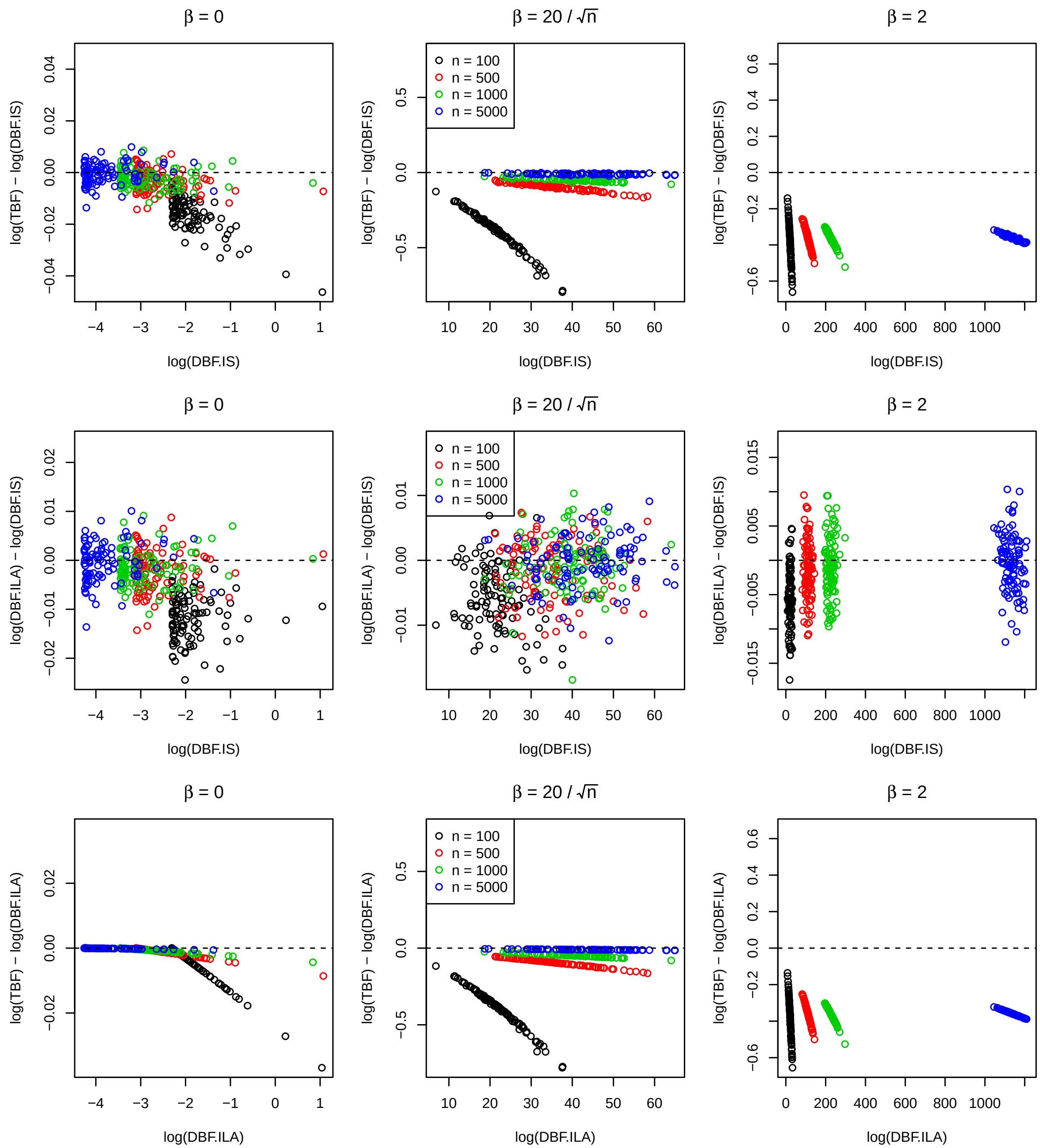

Figure 1: From top to bottom: TBF versus DBF approximated by IS, DBF approximated by ILA vs DBF approximated by IS, and TBF versus DBF approximated by ILA. From left to right: the null, local alternative, and alternative hypotheses.

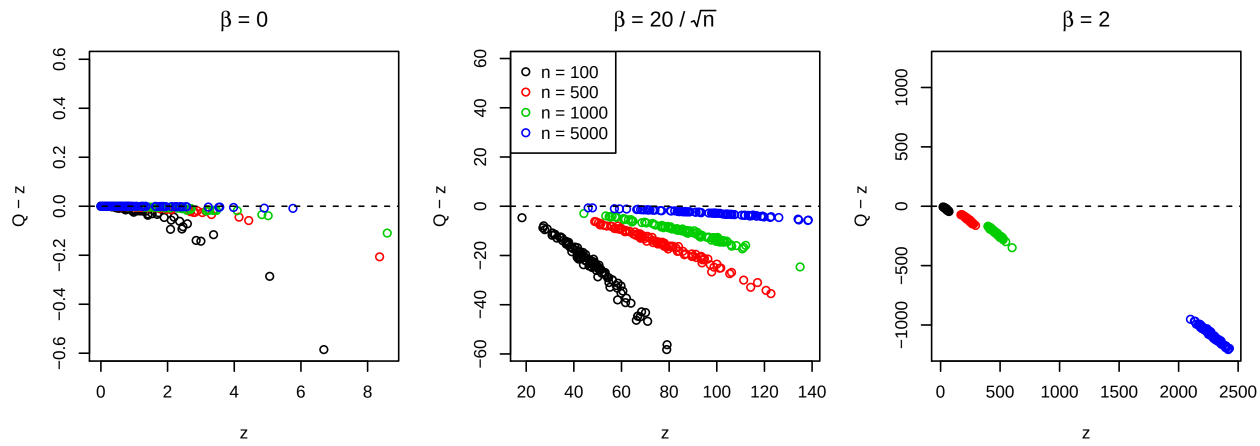

Figure 2: Wald statistic $Q_{\mathcal{M}}$ versus the deviance $z_{\mathcal{M}}$.

totically the same as the DBF computed under either IS or ILA (21). In contrast, when the alternative is true, TBF differs from DBF by a relatively small but systematic amount. Comparison between the Wald statistic $Q_{\mathcal{M}}$ (20) and the deviance $z_{\mathcal{M}}$ (22) suggests a similar phenomenon (Figure 2). They are asymptotically the same under the null or local alternative, but different under the alternative.

In addition to the similarity between the two Bayes factors under g-priors, we notice that as a function of g, the test-based marginal likelihood would have the same kernel $p(z_{\mathcal{M}} \mid \mathcal{M}) \propto (1+g)^{-p_{\mathcal{M}}/2} \exp\left(-z_{\mathcal{M}}/[2(1+g)]\right)$ as its data-based counterpart (19) if $z_{\mathcal{M}} = Q_{\mathcal{M}}$. Therefore, all empirical Bayes and fully Bayes approaches on g, discussed in Section 2.6 and Section 3, can be readily applied to test-based methods with minimal changes. Held et al. (2015) apply local empirical Bayes, $p(z_{\mathcal{M}} \mid \mathcal{M}) = \max_{g \geq 0} p(z_{\mathcal{M}} \mid g, \mathcal{M})$, and fully Bayes, $p(z_{\mathcal{M}} \mid \mathcal{M}) = \int p(z_{\mathcal{M}} \mid g, \mathcal{M}) p(g) dg$ to compute marginal likelihoods for TBFs. However, we find that these optimized and integrated versions of TBF may no longer be coherent, in the sense that results change with the choice of the baseline model. Elaborating, when testing nested models $\mathcal{M}_1 \subset \mathcal{M}_2$,

$$\mathrm{TBF}_{\mathcal{M}_2:\mathcal{M}_1} \neq \frac{\mathrm{TBF}_{\mathcal{M}_2:\mathcal{M}_\emptyset}}{\mathrm{TBF}_{\mathcal{M}_1:\mathcal{M}_\emptyset}},$$

if one computes the left hand side TBF under baseline $\mathcal{M}_1$, but computes the right hand side TBFs under baseline $\mathcal{M}_\emptyset$. The main reason for this incoherence is that for model $\mathcal{M}$, unlike the data-based marginal likelihood which only depends on $\mathcal{M}$ itself, the test statistic $z_\mathcal{M}$ also depends on the baseline model. On the other hand, coherence exists for the TBF (23) under fixed g, since $z_{\mathcal{M}_2:\mathcal{M}_1} = z_{\mathcal{M}_2:\mathcal{M}_\emptyset} - z_{\mathcal{M}_1:\mathcal{M}_\emptyset}$ (Johnson 2008). Hence, change of baseline models does not affect the results of the TBF under fixed g, which is also the case with the DBF.

C Additional Simulation Examples

We first include some additional results from the logistic regression simulation example that are examined in Section 5.1 (see Table C.1) and then introduce a different simulation study on Poisson regressions.

The simulation setup of the Poisson regression example is similar to that of the logistic regression in Section (5.1). True values of coefficients (including the intercept) are set to one-fifth of those in the logistic regression, to avoid occasional extremely large values in $\mathbf{Y}$. Tables C.2-C.4 display model selection and parameter estimation performance. Comparison among priors on $\boldsymbol{\beta}_\mathcal{M}$ leads to similar conclusions to the logistic regression example. For the Poisson regression, overall model selection accuracy is not as high as the logistic regression when $\mathcal{M}_T \neq \mathcal{M}_\emptyset$, which is likely due to the smaller magnitude of coefficients.

Table C.1: Logistic regression simulation example: average size of selected models, out of 100 realizations.

p	20								100			
$p(\mathcal{M})$	Uniform								Uniform		BB(1, 1)	
$p_{\mathcal{M}_T}$	0		5		10		20		5		5	
r	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75
CH($a = 1/2, b = n$)	0	0	5	4	10	8	17	13	17	15	5	3
CH($a = 1, b = n$)	0	0	5	5	10	8	17	13	18	15	5	3
CH($a = 1/2, b = n/2$)	0	0	6	5	10	9	17	14	25	20	5	3
CH($a = 1, b = n/2$)	0	0	6	5	10	9	17	14	26	22	5	3
Beta-prime	0	0	5	4	10	8	17	13	19	15	5	3
ZS adapted	0	0	5	5	10	8	17	13	18	15	5	3
Benchmark	0	0	6	6	11	10	18	15	27	24	5	3
Robust	0	0	6	5	11	9	18	14	34	30	21	10
Intrinsic	0	0	6	5	11	9	18	14	32	30	14	5
Hyper-g/n	0	1	6	5	11	10	18	15	69	56	99	80
DBF, $g = n$	0	0	5	4	9	7	15	11	7	5	5	3
TBF, $g = n$	0	0	5	4	9	7	15	11	7	5	5	3
Jeffreys	3	3	6	6	11	10	18	15	70	60	99	91
Hyper-g	4	4	6	6	11	10	18	15	70	61	100	93
Uniform	4	4	7	6	12	10	18	15	70	61	100	97
Local EB	19	19	6	6	11	10	18	15	71	60	100	96
AIC	3	3	8	7	12	11	18	15	34	34	6	4
BIC	0	0	5	4	9	7	15	11	7	5	5	3

Table C.2: Poisson regression simulation example: number of times the true model are selected out of 100 realizations. Column-wise maximum is in bold type.

p	20								100			
$p(\mathcal{M})$	Uniform								Uniform		BB(1, 1)	
$p_{\mathcal{M}_T}$	0		5		10		20		5		5	
r	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75
CH($a = 1/2, b = n$)	94	92	10	**2**	10	0	0	0	2	0	1	0
CH($a = 1, b = n$)	87	89	10	**2**	10	0	0	0	11	1	1	0
CH($a = 1/2, b = n/2$)	91	89	**11**	**2**	10	0	0	0	3	0	1	0
CH($a = 1, b = n/2$)	82	85	**11**	**2**	9	0	0	0	5	**2**	2	0
Beta-prime	94	92	10	**2**	10	0	0	0	7	0	1	0
ZS adapted	87	89	10	**2**	11	0	0	0	6	0	1	0
Benchmark	**97**	**93**	7	0	12	**1**	0	0	4	0	1	0
Robust	91	89	9	**2**	11	0	0	0	1	0	3	0
Intrinsic	85	88	8	**2**	12	**1**	0	0	1	0	3	0
Hyper-g/n	84	87	9	0	12	**1**	0	0	1	0	3	0
DBF, $g = n$	84	88	7	0	8	0	0	0	11	0	1	0
TBF, $g = n$	84	88	7	0	8	0	0	0	**14**	0	1	0
Jeffreys	0	0	7	1	12	1	0	0	0	0	3	0
Hyper-g	6	7	7	0	**13**	**1**	0	0	0	0	3	0
Uniform	4	2	7	0	**13**	**1**	0	0	1	1	3	0
Local EB	0	0	7	0	**13**	**1**	0	0	0	0	3	0
AIC	4	4	3	0	6	**1**	**1**	0	0	0	**8**	0
BIC	84	88	7	0	8	0	0	0	13	1	1	0

Table C.3: Poisson regression simulation example: average size of selected models, out of 100 realizations.

p	20								100			
$p(\mathcal{M})$	Uniform								Uniform		BB$(1,1)$	
$p_{\mathcal{M}_T}$	0		5		10		20		5		5	
r	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75
CH$(a=1/2, b=n)$	0	0	4	3	9	5	13	7	12	7	3	2
CH$(a=1, b=n)$	0	0	4	3	9	5	13	7	13	8	3	2
CH$(a=1/2, b=n/2)$	0	0	5	3	9	6	13	8	16	10	3	2
CH$(a=1, b=n/2)$	0	0	5	3	9	6	13	8	17	10	3	2
Beta-prime	0	0	4	3	9	5	13	7	13	7	3	2
ZS adapted	0	0	4	3	9	5	13	7	13	7	3	2
Benchmark	0	0	5	4	10	7	14	9	17	7	3	1
Robust	0	0	5	3	9	6	14	8	20	14	3	2
Intrinsic	0	0	5	3	10	6	14	8	22	13	3	2
Hyper-g/n	0	0	5	4	9	7	14	9	24	31	3	4
DBF, $g=n$	0	0	4	2	8	5	12	6	5	3	3	2
TBF, $g=n$	0	0	4	2	8	5	12	6	6	4	3	2
Jeffreys	2	3	5	4	10	7	14	10	29	36	3	18
Hyper-g	3	4	5	5	10	7	15	10	30	37	3	24
Uniform	4	4	6	5	10	7	15	10	30	38	3	34
Local EB	19	19	5	5	10	7	15	10	32	74	3	76
AIC	3	3	7	6	11	8	16	11	30	28	4	2
BIC	0	0	4	2	8	5	12	6	5	3	3	2

Table C.4: Poisson regression simulation example: 1000 times the average SSE $= \sum_{j=1}^{p}(\tilde{\beta}_j - \beta^*_{j,\mathcal{M}_T})^2$ of 100 realizations. Column-wise minimum is in bold type.

p	20								100			
$p(\mathcal{M})$	Uniform								Uniform		BB(1, 1)	
$p_{\mathcal{M}_T}$	0		5		10		20		5		5	
r	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75	0	0.75
CH($a = 1/2, b = n$)	5	8	**24**	**61**	34	120	58	198	66	132	37	103
CH($a = 1, b = n$)	6	9	**24**	**61**	34	120	58	197	66	134	37	98
CH($a = 1/2, b = n/2$)	7	11	**24**	**61**	**33**	116	56	188	75	148	36	97
CH($a = 1, b = n/2$)	7	13	**24**	**61**	**33**	115	55	187	77	135	36	94
Beta-prime	5	8	**24**	**61**	34	120	58	197	66	132	37	103
ZS adapted	6	9	**24**	**61**	34	119	55	197	66	125	37	99
Benchmark	8	18	26	65	**33**	**108**	51	170	74	150	36	133
Robust	7	13	25	63	**33**	115	51	183	88	182	36	97
Intrinsic	8	14	25	63	**33**	115	51	182	90	183	35	94
Hyper-g/n	5	12	25	65	**33**	109	52	172	84	162	36	97
DBF, $g = n$	5	6	25	63	37	132	68	231	**40**	**83**	39	101
TBF, $g = n$	5	6	25	63	37	132	68	231	**40**	84	39	101
Jeffreys	4	9	26	67	**33**	**108**	51	169	87	165	35	97
Hyper-g	4	7	26	68	**33**	**108**	51	**168**	87	164	**34**	112
Uniform	3	7	26	70	**33**	**108**	51	**168**	87	164	**34**	121
Local EB	**2**	**4**	26	71	**33**	**108**	51	**168**	99	256	**34**	222
AIC	17	40	28	74	34	115	**46**	171	120	284	37	**79**
BIC	5	6	25	63	37	132	68	231	**40**	84	39	100

References

Abramowitz, M. and Stegun, I. (1970), *Handbook of Mathematical Functions - with Formulas, Graphs, and Mathematical Tables*, New York: Dover publications.

Davidson, R. R. and Lever, W. E. (1970), "The Limiting Distribution of the Likelihood Ratio Statistic under a Class of Local Alternatives," *Sankhyā: The Indian Journal of Statistics. Series A*, 32, 209–224.

Fahrmeir, L. and Kaufmann, H. (1985), "Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models," *The Annals of Statistics*, 13, 342–368.

Gordy, M. B. (1998), *A Generalization of Generalized Beta Distribution*, Division of Research and Statistics, Division of Monetary Affairs, Federal Reserve Board.

Gradshteyn, I. S. and Ryzhik, I. M. (2007), *Table of Integrals, Series, and Products*, Academic Press, 7th ed.

Held, L., Gravestock, I., and Sabanés Bové, D. (2016), "Objective Bayesian Model Selection for Cox Regression," *Statistics in Medicine.*

Held, L., Sabanés Bové, D., and Gravestock, I. (2015), "Approximate Bayesian Model Selection with the Deviance Statistic," *Statistical Science*, 30, 242–257.

Hu, J. and Johnson, V. E. (2009), "Bayesian Model Selection Using Test Statistics," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71, 143–158.

Johnson, V. E. (2005), "Bayes Factors Based on Test Statistics," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67, 689–701.

— (2008), "Properties of Bayes Factors Based on Test Statistics," *Scandinavian Journal of Statistics*, 35, 354–368.

Olver, F. (1997), *Asymptotics and Special Functions*, A K Peters/CRC Press.

Resnick, S. I. (1999), *A Probability Path*, Birkhäuser.

Self, S. and Mauritsen, R. (1988), "Power/Sample Size Calculations for Generalized Linear Models," *Biometrics*, 44, 79–86.

Self, S., Mauritsen, R., and Ohara, J. (1992), "Power Calculation for Likelihood Ratio Tests in Generalized Linear Models," *Biometrics*, 48, 31–39.

Shieh, G. (2000), "On Power and Sample Size Calculations for Likelihood Ratio Tests in Generalized Linear Models," *Biometrics*, 56, 1192–1196.

Slater, L. (1960), *Confluent Hypergeometric Functions*, Cambridge University Press.

van der Vaart, A. W. (2000), *Asymptotic Statistics*, Cambridge University Press.

Weisstein, E. W. (2009), "Appell Hypergeometric Function," *From MathWorld–A Wolfram Web Resource. Avaiable at http://mathworld.wolfram.com/AppellHypergeometricFunction.html.*