Supplementary materials for "Data transforming augmentation for heteroscedastic models"

A. Iterative algorithms for univariate linear mixed models in Section 3.1

A.1. DTA-based iterative algorithms

To sample the full posterior $p(A, \boldsymbol{\beta}|y^{\text{obs}}) \propto L(A, \boldsymbol{\beta}; y^{\text{obs}})I_{A>0}$, Kelly (2014) proposes the following DTA-based Gibbs-type algorithm that iteratively samples $[y^{\text{aug}}|y^{\text{obs}}, A, \boldsymbol{\beta}]$ and $[A, \boldsymbol{\beta}|y^{\text{aug}}]$. These two conditional distributions can be directly sampled from standard family distributions:

$$\begin{aligned}
[y_i^{\text{aug}} \mid y_i^{\text{obs}}, A, \boldsymbol{\beta}] &\sim \text{N}_1\big(\ (1 - w_i B_i) y_i^{\text{obs}} + w_i B_i x_i^\top \boldsymbol{\beta},\ w_i V_{\min} + w_i^2 V_i (1 - B_i)\ \big)\,, \\
[A \mid y^{\text{aug}}] &\sim \text{IG}\Big((k - m - 2)/2,\ (y^{\text{aug}} - X\hat{\beta})^\top (y^{\text{aug}} - X\hat{\beta})/2\Big) \ \text{ for } A > V_{\min}, \\
[\boldsymbol{\beta} \mid A, y^{\text{aug}}] &\sim \text{N}_m\Big(\hat{\boldsymbol{\beta}},\ (A + V_{\min})(X^\top X)^{-1}\Big)\,,
\end{aligned} \tag{20}$$

where $w_i = 1 - V_{\min}/V_i$, $B_i = V_i/(V_i + A)$, $y^{\text{aug}} = (y_1^{\text{aug}}, \ldots, y_k^{\text{aug}})^\top$, $\text{IG}(a, b)$ denotes the inverse-Gamma distribution with shape parameter a and scale parameter b, X is a k by m matrix whose row vector is $\boldsymbol{x}_i^\top$, and $\hat{\boldsymbol{\beta}} = (X^\top X)^{-1} X^\top y^{\text{aug}}$. The second step in (20) can be achieved by repeatedly sampling A from the inverse-Gamma distribution until $A > V_{\text{mis}}$ or by an inverse CDF sampling method if its cumulative distribution function and quantile function are available; also see Appendix of Everson and Morris (2000).

Kelly (2014) also derives the corresponding DTA-based EM algorithm for the posterior modes (or maximum likelihood estimates) of A and $\boldsymbol{\beta}$ by constructing the following

Q function in the E-step: Since $[y_i^{\rm aug}|A,\boldsymbol{\beta}] \sim {\rm N}_1(\boldsymbol{x}_i^\top\boldsymbol{\beta}, A+V_{\min})$ under DTA,

$$
\begin{aligned}
Q(A,\beta \mid A^*,\boldsymbol{\beta}^*) &= \sum_{i=1}^k E(\log(f(y_i^{\rm aug} \mid A,\boldsymbol{\beta})) \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*),\\
&= -\frac{k}{2}\log(A+V_{\min}) - \frac{\sum_{i=1}^k E\left((y_i^{\rm aug}-\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*\right)}{2(A+V_{\min})},
\end{aligned}
$$

where A^* and $\boldsymbol{\beta}^*$ are the values that have maximized the Q function in the previous iteration. The conditional expectation in the second equality can be computed by

$$
E\left((y_i^{\rm aug}-\boldsymbol{x}_i^\top\boldsymbol{\beta})^2 \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*\right) = \left(E(y_i^{\rm aug} \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*) - \boldsymbol{x}_i^\top\boldsymbol{\beta}\right)^2 + {\rm Var}(y_i^{\rm aug} \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*),
$$

where the conditional mean and variance on the right-hand side, i.e., $E(y_i^{\rm aug} \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*)$ and ${\rm Var}(y_i^{\rm aug} \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*)$, are specified in (20). Maximizing this Q function with respect to $\boldsymbol{\beta}$ and A results in the following M-step with closed-form updates for $\boldsymbol{\beta}$ and A:

$$
\begin{aligned}
&\text{Step 1: } \boldsymbol{\beta'} \leftarrow (X^\top X)^{-1}X^\top E(y^{\rm aug} \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*),\\
&\text{Step 2: } A' \leftarrow \max\left\{\frac{1}{k}\sum_{i=1}^k E\left((y_i^{\rm aug}-\boldsymbol{x}_i^\top\boldsymbol{\beta'})^2 \mid y^{\rm obs}, A^*, \boldsymbol{\beta}^*\right) - V_{\min},\ 0\right\},\\
&\text{Step 3: } (\boldsymbol{\beta}^*, A^*) \leftarrow (\boldsymbol{\beta'}, A').
\end{aligned}
$$

A.2. DA-based iterative algorithms

We treat the random effects $\theta = (\theta_1,\ldots,\theta_k)^\top$ as missing data, which is typical in fitting hierarchical or mixed-effects models via DA-based iterative algorithms (van Dyk, 2000). The resulting DA-based Gibbs algorithm iteratively samples

$$
\begin{aligned}
[\theta_i \mid y_i^{\rm obs}, A, \boldsymbol{\beta}] &\sim {\rm N}_1\big(\ (1-B_i)y_i^{\rm obs} + B_i\boldsymbol{x}_i^\top\boldsymbol{\beta},\ V_i(1-B_i)\ \big),\\
[A \mid \theta, y^{\rm obs}] &\sim {\rm IG}\Big((k-m-2)/2,\ (\theta - X\hat{\boldsymbol{\beta}}_{\rm DA})^\top(\theta - X\hat{\boldsymbol{\beta}}_{\rm DA})/2\Big),\\
[\boldsymbol{\beta} \mid A, \theta, y^{\rm obs}] &\sim {\rm N}_m\Big(\hat{\boldsymbol{\beta}}_{\rm DA},\ A(X^\top X)^{-1}\Big),
\end{aligned}
\tag{21}
$$

where $\hat{\boldsymbol{\beta}}_{\text{DA}} = (X^\top X)^{-1} X^\top \theta$. Kelly (2014) also shows a DA-based EM algorithm that corresponds to the Gibbs sampler in (21). Its E-step computes the Q function as follows:

$$\begin{aligned} Q(A, \boldsymbol{\beta} \mid A^*, \boldsymbol{\beta}^*) &= \sum_{i=1}^{k} E(\log(f(\theta_i \mid A, \boldsymbol{\beta})) \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*) \\ &= -\frac{k}{2}\log(A) - \frac{\sum_{i=1}^{k} E\left((\theta_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*\right)}{2A}, \end{aligned}$$

where

$$E\left((\theta_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*\right) = \left(E(\theta_i \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*) - \boldsymbol{x}_i^\top \boldsymbol{\beta}\right)^2 + \text{Var}(\theta_i \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*). \quad (22)$$

The conditional mean and variance on the right-hand side of (22) are specified in (21). The resulting M-step sets $\boldsymbol{\beta}^*$ and A^* to the values that maximize this Q function and these values are also closed-form updates as follows:

$$\begin{aligned} &\text{Step 1: } \boldsymbol{\beta}' \leftarrow (X^\top X)^{-1} X^\top E(\theta \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*), \\ &\text{Step 2: } A' \leftarrow \frac{1}{k}\sum_{i=1}^{k} E\left((\theta_i - \boldsymbol{x}_i^\top \boldsymbol{\beta})^2 \mid y^{\text{obs}}, A^*, \boldsymbol{\beta}^*\right), \\ &\text{Step 3: } (\boldsymbol{\beta}^*, A^*) \leftarrow (\boldsymbol{\beta}', A'). \end{aligned}$$

B. Iterative algorithms for multivariate linear mixed models in Section 3.2

B.1. DTA-based iterative algorithms

The joint posterior distribution $p(\boldsymbol{A}, \boldsymbol{\beta} | \boldsymbol{y}^{\text{aug}})$ factors into the following two conditional distributions, $p(\boldsymbol{A}, \boldsymbol{\beta} | \boldsymbol{y}^{\text{aug}}) = p_1(\boldsymbol{A}, | \boldsymbol{y}^{\text{aug}}) p_2(\boldsymbol{\beta} | \boldsymbol{A}, \boldsymbol{y}^{\text{aug}})$, and these can be directly sampled via inverse-Wishart and multivariate Gaussian distributions in a homoscedastic case. A DTA-based Gibbs-type algorithm specified in (3) iteratively samples the following three conditional distributions.

$$
\begin{aligned}
[\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}_i^{\text{obs}}, \boldsymbol{A}, \boldsymbol{\beta}] &\sim \mathrm{N}_p\big((1 - W_i B_i)\boldsymbol{y}_i^{\text{obs}} + W_i B_i X_i \boldsymbol{\beta},\ \boldsymbol{V}_{\min} W_i^\top + W_i(1 - B_i)\boldsymbol{V}_i W_i^\top\big)\,, \\
[\boldsymbol{A} + \boldsymbol{V}_{\min} \mid \boldsymbol{y}^{\text{aug}}] &\sim \mathrm{IW}\left(k - m - p - 1,\ \sum_{i=1}^{k} (\boldsymbol{y}_i^{\text{aug}} - X_i \hat{\boldsymbol{\beta}}_{\text{DTA}})(\boldsymbol{y}_i^{\text{aug}} - X_i \hat{\boldsymbol{\beta}}_{\text{DTA}})^\top\right), \\
[\boldsymbol{\beta} \mid \boldsymbol{A}, \boldsymbol{y}^{\text{aug}}] &\sim \mathrm{N}_{mp}\left(\hat{\boldsymbol{\beta}}_{\text{DTA}},\ \left(\sum_{i=1}^{k} X_i^\top (\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} X_i\right)^{-1}\right),
\end{aligned}
\tag{23}
$$

where $B_i = \boldsymbol{V}_i(\boldsymbol{V}_i + \boldsymbol{A})^{-1}$, $\mathrm{IW}(a, b)$ indicates the inverse-Wishart distribution with a degrees of freedom and scale matrix b, and

$$
\hat{\boldsymbol{\beta}}_{\text{DTA}} = \left(\sum_{i=1}^{k} X_i^\top (\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} X_i\right)^{-1} \sum_{i=1}^{k} X_i^\top (\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} \boldsymbol{y}_i^{\text{aug}}.
$$

To sample $\boldsymbol{A}$ instead of $\boldsymbol{A} + \boldsymbol{V}_{\min}$ in the middle of (23), we repeatedly draw a random sample K from the inverse-Wishart distribution in (23) until $|K - \boldsymbol{V}_{\min}| > 0$, and then set $\boldsymbol{A}$ to $K - \boldsymbol{V}_{\min}$.

We specify the corresponding DTA-based EM algorithm by constructing the Q function for the E-step, using the marginal distribution $[\boldsymbol{y}_i^{\text{aug}} | \boldsymbol{A}, \boldsymbol{\beta}] \sim \mathrm{N}_p(\boldsymbol{X}_i \boldsymbol{\beta},\ \boldsymbol{A} + \boldsymbol{V}_{\min})$:

$$
\begin{aligned}
&Q(\boldsymbol{A}, \boldsymbol{\beta} \mid \boldsymbol{A}^*, \boldsymbol{\beta}^*) = \sum_{i=1}^{k} E\left(\log(f(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{A}, \boldsymbol{\beta})) \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*\right) \\
&= -\frac{k}{2}\log(|\boldsymbol{A} + \boldsymbol{V}_{\min}|) - \frac{1}{2}\sum_{i=1}^{k} E\left((\boldsymbol{y}_i^{\text{aug}} - X_i\boldsymbol{\beta})^\top (\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} (\boldsymbol{y}_i^{\text{aug}} - X_i\boldsymbol{\beta}) \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*\right).
\end{aligned}
\tag{24}
$$

The conditional expectation of a quadratic form in the second equality is equivalent to

$$
\begin{aligned}
&E\left((\boldsymbol{y}_i^{\text{aug}} - X_i\boldsymbol{\beta})^\top (\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} (\boldsymbol{y}_i^{\text{aug}} - X_i\boldsymbol{\beta}) \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*\right) \\
&\quad = \left(E(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}\right)^\top (\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} \left(E(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}\right) \\
&\qquad + \mathrm{trace}\big[(\boldsymbol{A} + \boldsymbol{V}_{\min})^{-1} \mathrm{Cov}(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*)\big]\,.
\end{aligned}
\tag{25}
$$

The conditional expectation and covariance of $\boldsymbol{y}_i^{\text{aug}}$ on the right-hand side in (25), i.e., $E(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*)$ and $\mathrm{Cov}(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*)$, are specified in (23).

The M-step updates $\boldsymbol{A}^*$ and $\boldsymbol{\beta}^*$ by the values that maximize the Q function in (24), which results in the following four steps for closed-form updates:

$$\begin{aligned}
&\text{Step 1: } \boldsymbol{\beta}' \leftarrow \left(\sum_{i=1}^{k} X_i^\top X_i\right)^{-1} \sum_{i=1}^{k} X_i^\top E(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*). \\
&\text{Step 2: } \boldsymbol{A}_{\text{temp}} \leftarrow \frac{1}{k}\sum_{i=1}^{k} \Big\{ \big(E(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}'\big)\big(E(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}'\big)^\top \\
&\qquad\qquad\qquad\qquad + \text{Cov}\left(\boldsymbol{y}_i^{\text{aug}} \mid \boldsymbol{y}^{\text{obs}}, \boldsymbol{A}^*, \boldsymbol{\beta}^*\right) \Big\} - \boldsymbol{V}_{\min}. \\
&\text{Step 3: } \boldsymbol{A}' \leftarrow \boldsymbol{A}_{\text{temp}} \ \text{ if } \ |\boldsymbol{A}_{\text{temp}}| > 0 \ \text{ and } \ \boldsymbol{A}^* \leftarrow 0_p \ \text{ otherwise.} \\
&\text{Step 4: } (\boldsymbol{\beta}^*, \boldsymbol{A}^*) \leftarrow (\boldsymbol{\beta}', \boldsymbol{A}').
\end{aligned}$$

The notation 0_p in Step 3 indicates a p by p matrix filled with zeros.

B.2. DA-based iterative algorithms

In a typical DA scheme, we treat the random effects, $\boldsymbol{\theta} \equiv \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_k\}$, as missing data, and thus the augmented data in this case are $\boldsymbol{y}^{\text{aug}} = (\boldsymbol{y}^{\text{obs}}, \boldsymbol{\theta})$. The full posterior density function of $[\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{A} | \boldsymbol{y}^{\text{obs}}]$ can be derived up to a constant multiplication, i.e.,

$$\pi(\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{A} \mid \boldsymbol{y}^{\text{obs}}) \propto \prod_{i=1}^{k} f(\boldsymbol{y}_i^{\text{obs}} \mid \boldsymbol{\theta}_i) p(\boldsymbol{\theta}_i \mid \boldsymbol{\beta}, \boldsymbol{A}) I_{|\boldsymbol{A}|>0},$$

where $[\boldsymbol{y}_i^{\text{obs}} | \boldsymbol{\theta}_i]$ and $[\boldsymbol{\theta}_i | \boldsymbol{A}, \boldsymbol{\beta}]$ are defined in (11). Similarly to the DTA-based Gibbs sampler in (23), the DA-based one iteratively samples the following conditional distributions:

$$\begin{aligned}
[\boldsymbol{\theta}_i \mid \boldsymbol{y}_i^{\text{obs}}, \boldsymbol{A}, \boldsymbol{\beta}] &\sim \text{N}_p\big(\ (1 - B_i)\boldsymbol{y}_i^{\text{obs}} + B_i X_i \boldsymbol{\beta},\ (1 - B_i)\boldsymbol{V}_i\ \big), \\
[\boldsymbol{A} \mid \boldsymbol{\theta}, \boldsymbol{y}^{\text{obs}}] &\sim \text{IW}\left(\ k - m - p - 1,\ \sum_{i=1}^{k} (\boldsymbol{\theta}_i - X_i\hat{\boldsymbol{\beta}}_{\text{DA}})(\boldsymbol{\theta}_i - X_i\hat{\boldsymbol{\beta}}_{\text{DA}})^\top \right), \\
[\boldsymbol{\beta} \mid \boldsymbol{A}, \boldsymbol{\theta}, \boldsymbol{y}^{\text{obs}}] &\sim \text{N}_{mp}\left(\hat{\boldsymbol{\beta}}_{\text{DA}},\ \left(\sum_{i=1}^{k} X_i^\top \boldsymbol{A}^{-1} X_i\right)^{-1}\right),
\end{aligned} \tag{26}$$

where $B_i = \boldsymbol{V}_i(\boldsymbol{V}_i + \boldsymbol{A})^{-1}$ and

$$\hat{\boldsymbol{\beta}}_{\text{DA}} = \left(\sum_{i=1}^{k} X_i^\top \boldsymbol{A}^{-1} X_i\right)^{-1} \sum_{i=1}^{k} X_i^\top \boldsymbol{A}^{-1} \boldsymbol{\theta}_i.$$

The corresponding DA-based EM algorithm adopts the following Q function for the E-step, using the distribution of missing data, $[\boldsymbol{\theta}_i|\boldsymbol{A},\boldsymbol{\beta}] \sim \mathrm{N}_p(\boldsymbol{X}_i\boldsymbol{\beta},\ \boldsymbol{A})$:

$$
\begin{aligned}
Q(\boldsymbol{A},\boldsymbol{\beta}\,|\boldsymbol{A}^*,\boldsymbol{\beta}^*) &= \sum_{i=1}^{k} E(\log(f(\boldsymbol{\theta}_i \mid \boldsymbol{A},\boldsymbol{\beta})) \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*),\\
&= -\frac{k}{2}\log(|\boldsymbol{A}|) - \frac{1}{2}\sum_{i=1}^{k} E\left((\boldsymbol{\theta}_i - X_i\boldsymbol{\beta})^\top \boldsymbol{A}^{-1}(\boldsymbol{\theta}_i - X_i\boldsymbol{\beta}) \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*\right).
\end{aligned}
\tag{27}
$$

The conditional expectation of a quadratic form on the right-hand side in (27) can be computed by

$$
\begin{aligned}
&E\left((\boldsymbol{\theta}_i - X_i\boldsymbol{\beta})^\top A^{-1}(\boldsymbol{\theta}_i - X_i\boldsymbol{\beta}) \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*\right)\\
&\quad = \left(E(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}\right)^\top \boldsymbol{A}^{-1}\left(E(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}\right)\\
&\qquad + \mathrm{trace}\left[\boldsymbol{A}^{-1}\mathrm{Cov}(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*)\right].
\end{aligned}
\tag{28}
$$

The conditional expectation and covariance of $\boldsymbol{\theta}_i$ given $\boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A},\boldsymbol{\beta}$ in (28) are specified in (26).

Like the M-step under DTA in (24), the M-step under DA updates $\boldsymbol{A}^*$ and $\boldsymbol{\beta}^*$ by the values that maximize the Q function in (27) via the following three steps:

Step 1: $\boldsymbol{\beta}' \leftarrow \left(\sum_{i=1}^{k} X_i^\top X_i\right)^{-1} \sum_{i=1}^{k} X_i^\top E(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*)$.

Step 2: $\boldsymbol{A}' \leftarrow \frac{1}{k}\sum_{i=1}^{k}\Big\{ \left(E(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\beta^*) - X_i\boldsymbol{\beta}'\right)\left(E(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*) - X_i\boldsymbol{\beta}'\right)^\top + \mathrm{Cov}\big(\boldsymbol{\theta}_i \mid \boldsymbol{y}^{\mathrm{obs}},\boldsymbol{A}^*,\boldsymbol{\beta}^*\big)\Big\}$.

Step 3: $(\boldsymbol{\beta}^*,\boldsymbol{A}^*) \leftarrow (\boldsymbol{\beta}',\boldsymbol{A}')$.

C. The DTA scheme for the Beta-Binomial model in Section 4

Given the homoscedastic augmented data y^{aug}, we reproduce the approximate marginal posterior density $p^*(\beta \mid y^{\mathrm{aug}})$ from (17): With $g(l) = nk + c + l$,

$$
p^*(\beta \mid y^{\mathrm{aug}}) = \sum_{i=s_1}^{s_t}\sum_{j=f_1}^{f_t}\sum_{l=0}^{km_1+m_2} a_i b_j c_l^* B(g(l) - i - 1,\ i+1)\frac{\beta^j}{(\beta + n_{\max})^{g(l)-i-1}},
$$

where s_1 denotes the number of groups with at least one success, s_t is the total number of successes ($s_t = \sum_{i=1}^k y_i^{\text{aug}}$), f_1 indicates the number of groups with at least one failure, and f_t is the total number of failures ($f_t = \sum_{i=1}^k (n - y_i^{\text{aug}})$). If we transform β into $B = \beta/(\beta + n_{\max})$, the corresponding density function with respect to B is as follows: With the Jacobean $J = n_{\max}/(1-B)^2$,

$$p^*(B \mid y^{\text{aug}}) = \sum_{i=s_1}^{s_t} \sum_{j=f_1}^{f_t} \sum_{l=0}^{km_1+m_2} a_i b_j c_l^* B(g(l)-i-1,\ i+1) \frac{B^j (1-B)^{g(l)-i-j-3}}{n_{\max}^{g(l)-i-j-1}}.$$

Thus, each mixture component is composed of the Beta$(j+1,\ g(l)-i-j-2)$ density function and the corresponding coefficient (that is proportional to its weight) equal to

$$a_i b_j c_l^* B(g(l)-i-1,\ i+1) B(j+1,\ g(l)-i-j-2) n_{\max}^{-(g(l)-i-j-1)}. \tag{29}$$

We can easily generate $B = \beta/(\beta + n_{\max})$ via a two-step procedure; (i) we randomly choose a combination of (i, j, l) according to its weight defined in (29), and (ii) given the selected values of (i, j, l), we can generate B from the Beta$(j+1,\ g(l)-i-j-2)$ distribution. Finally, we set $\beta = n_{\max} B/(1-B)$ that is a random number generated from $p^*(\beta | y^{\text{aug}})$.

Given the random number from $p^*(\beta | y^{\text{aug}})$, we need to sample $p^*(\alpha | y^{\text{aug}}, \beta)$ that is proportional to $p^*(\alpha, \beta | y^{\text{aug}})$ in (16), i.e.,

$$p^*(\alpha \mid y^{\text{aug}}, \beta) \propto \sum_{i=s_1}^{s_t} \sum_{j=f_1}^{f_t} \sum_{l=0}^{km_1+m_2} a_i b_j c_l^* \frac{\alpha^i \beta^j}{(\alpha + \beta + n_{\max})^{g(l)}}.$$

Once we transform α to $A = \alpha/(\alpha + \beta + n_{\max})$, we obtain the following density function with the Jacobean $J = (\beta + n_{\max})/(1-A)^2$:

$$\begin{aligned} p^*(A \mid y^{\text{aug}}, \beta) &\propto \sum_{i=s_1}^{s_t} \sum_{j=f_1}^{f_t} \sum_{l=0}^{km_1+m_2} a_i b_j c_l^* \frac{\beta^j}{(\beta + n_{\max})^{g(l)-i-1}} A^i (1-A)^{g(l)-i-2} \\ &= \left(\sum_{j=f_1}^{f_t} b_j \beta^j \right) \left(\sum_{i=s_1}^{s_t} \sum_{l=0}^{km_1+m_2} a_i c_l^* \frac{1}{(\beta + n_{\max})^{g(l)-i-1}} A^i (1-A)^{g(l)-i-2} \right) \\ &\propto \sum_{i=s_1}^{s_t} \sum_{l=0}^{km_1+m_2} a_i c_l^* \frac{1}{(\beta + n_{\max})^{l-i-1}} A^i (1-A)^{g(l)-i-2}. \end{aligned} \tag{30}$$

Then, the density with respect to A is a mixture of the Beta$(i+1,\ g(l)-i-1)$ densities with its coefficient (weight) equal to

$$a_i c_l^* \frac{B(i+1,\ g(l)-i-1)}{(\beta+n_{\max})^{l-i-1}}. \tag{31}$$

Sampling α from $p^*(\alpha \mid y^{\text{aug}}, \beta)$ is a three-step procedure; (i) a combination of (i, l) is randomly selected according to its weight in (31), (ii) a value of A is randomly generated from the Beta$(i+1,\ g(l)-i-1)$ distribution given the chosen values of (i, l), and finally (iii) α is set to $(n_{\max}+\beta)A/(1-A)$.

Therefore, the proposed augmentation scheme for heteroscedastic Binomial data results in a Gibbs-type algorithm that iterates for following five steps:

Step 1: Sample $\theta_i \sim \text{Beta}(y_i^{\text{obs}}+\alpha,\ n_i - y_i^{\text{obs}}+\beta)$ for $i = 1, \ldots, k$.
Step 2: Sample $y_i^{\text{mis}} \sim \text{Bin}(n_{\max}-n_i,\ \theta_i)$ for $i = 1, \ldots, k$.
Step 3: Set $y_i^{\text{aug}} = y_i^{\text{obs}} + y_i^{\text{mis}}$ for $i = 1, \ldots, k$.
Step 4: Sample β from $p^*(\beta \mid y^{\text{aug}})$.
Step 5: Sample α from $p^*(\alpha \mid y^{\text{aug}}, \beta)$.

We note that Steps 1–3 are corresponding to the first two steps of (2), and Steps 4–5 are related to the last step of (2).

References

Everson, P. J. and Morris, C. N. (2000). Simulation from Wishart Distributions with Eigenvalue Constraints. *Journal of Computational and Graphical Statistics*, 9(2):380–389.

Kelly, J. (2014). *Advances in the Normal-Normal Hierarchical Model*. PhD thesis, Harvard University.

van Dyk, D. A. (2000). Fitting Mixed-Effects Models Using Efficient EM-Type Algorithms. *Journal of Computational and Graphical Statistics*, 9(1):78–98.