WEB APPENDIX FOR

A Nonparametric Prior for Simultaneous Covariance Estimation

J. T. Gaskins and M. J. Daniels

Appendix 1: Derivation of Theoretical Properties

This appendix contains proof for the properties presented in Section 5.

A1.1 Sparsity Grouping Prior

The proofs for properties 1.–3. can be found in the Appendix of Dunson et al. (2008).

4.

$$
\begin{aligned}
Pr(\phi_{mj} = \phi_{m'j}) &= Pr(\phi_{mj} = \phi_{m'j} \neq 0) + Pr(\phi_{mj} = \phi_{m'j} = 0) \\
&= \mathrm{E}\left\{ \sum_h \pi_{mjh}\pi_{m'jh}\delta_{\xi_{jh}}(\mathcal{R}\backslash 0) \right\} + \mathrm{E}\left\{ \sum_h \pi_{mjh}\delta_{\xi_{jh}}(0) \times \sum_i \pi_{m'ji}\delta_{\xi_{ji}}(0) \right\} \\
&= \mathrm{E}\left\{ \sum_h \pi_{mjh}\pi_{m'jh}\delta_{\xi_{jh}}(\mathcal{R}\backslash 0) \right\} + \mathrm{E}\left\{ \sum_h \pi_{mjh}\pi_{m'jh}\delta_{\xi_{jh}}(0) \right\} \\
&\quad + 2\mathrm{E}\left\{ \sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'ji}\delta_{\xi_{jh}}(0)\delta_{\xi_{ji}}(0) \right\} \\
&= \mathrm{E}\left\{ \sum_h \pi_{mjh}\pi_{m'jh}\delta_{\xi_{jh}}(\mathcal{R}) \right\} + 2\mathrm{E}\left\{ \sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'ji}\delta_{\xi_{jh}}(0)\delta_{\xi_{ji}}(0) \right\} \\
&= \mathrm{E}\left\{ \sum_h \pi_{mjh}\pi_{m'jh} \right\} + 2\epsilon^2\mathrm{E}\left\{ \sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'ji} \right\} = \text{(I)} + 2\epsilon^2\text{(II)},
\end{aligned}
$$

where expressions (I) and (II) are calculated below.

$$
\begin{aligned}
\text{(I)} &= \mathrm{E}\left\{\sum_h U_{mh}U_{m'h}X_{jh}^2 \prod_{l<h}\left(1 - U_{ml}X_{jl} - U_{m'l}X_{jl} + U_{ml}U_{m'l}X_{jl}^2\right)\right\} \\
&= \sum_h \mathrm{E}U_{mh}\mathrm{E}U_{m'h}\mathrm{E}X_{jh}^2 \prod_{l<h}\left(1 - 2\mathrm{E}U_{ml}\mathrm{E}X_{jl} + \mathrm{E}U_{ml}\mathrm{E}U_{m'l}\mathrm{E}X_{jl}^2\right) \\
&= \sum_h \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\right]^{h-1} \\
&= \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\left[\frac{2}{(1+\alpha)(1+\beta)} - \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\right]^{-1} \\
&= \frac{1}{(1+\alpha)(2+\beta)-1}.
\end{aligned}
$$

$$
\begin{aligned}
\text{(II)} &= \mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} U_{mh}X_{jh}\left(1 - U_{m'h}X_{jh}\right)\prod_{l<h}\left(1 - U_{ml}X_{jl}\right)\left(1 - U_{m'l}X_{jl}\right)U_{m'i}X_{ji}\right. \\
&\qquad \left.\times \prod_{l=h+1}^{i-1}\left(1 - U_{m'l}X_{jl}\right)\right\} \\
&= \sum_h \sum_{i=h+1}^{\infty}\left[\mathrm{E}U_{mh}\mathrm{E}X_{jh} - \mathrm{E}U_{mh}\mathrm{E}U_{m'h}\mathrm{E}X_{jh}^2\right]\mathrm{E}U_{m'i}\mathrm{E}X_{ji} \\
&\qquad \times \prod_{l<h}\left[1 - 2\mathrm{E}U_{ml}\mathrm{E}X_{jl} + \mathrm{E}U_{ml}\mathrm{E}U_{m'l}\mathrm{E}X_{jl}^2\right]\prod_{l=h+1}^{i-1}\left(1 - \mathrm{E}U_{m'l}\mathrm{E}X_{jl}\right) \\
&= \sum_h \sum_{i=h+1}^{\infty}\left[\frac{1}{(1+\alpha)(1+\beta)} - \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\right]\frac{1}{(1+\alpha)(1+\beta)} \\
&\qquad \times \left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\right]^{h-1}\left[1 - \frac{1}{(1+\alpha)(1+\beta)}\right]^{i-h-1} \\
&= \sum_h \frac{1}{(1+\alpha)(1+\beta)}\left[1 - \frac{2}{(1+\alpha)(2+\beta)}\right] \\
&\qquad \times \left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\right]^{h-1} \\
&= \frac{1}{(1+\alpha)(1+\beta)}\left[1 - \frac{2}{(1+\alpha)(2+\beta)}\right]\left[\frac{2}{(1+\alpha)(1+\beta)} - \frac{2}{(1+\alpha)^2(1+\beta)(2+\beta)}\right]^{-1}
\end{aligned}
$$

$$= \frac{1}{2}\left[1 - \frac{1}{(1+\alpha)(2+\beta)-1}\right].$$

Using (I) and (II), we have

$$Pr(\phi_{mj} = \phi_{m'j}) \;=\; \text{(I)} + 2\epsilon^2\text{(II)} \;=\; \epsilon^2 + \frac{1-\epsilon^2}{(1+\alpha)(2+\beta)-1}.$$

5. To compute the correlation, we first obtain the expected value of the product of the distributions.

$$\begin{aligned}
\mathrm{E}\left(F_{mj}(A)F_{mj'}(A)\right) &= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\delta_{\xi_{jh}}(A)\delta_{\xi_{j'h}}(A)\right\} \\
&\quad +2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{mj'i}\delta_{\xi_{jh}}(A)\delta_{\xi_{j'i}}(A)\right\} \\
&= \Psi(A)^2\,\mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\right\} + 2\Psi(A)^2\,\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{mj'i}\right\} \\
&= \Psi(A)^2\,\text{(III)} + 2\Psi(A)^2\,\text{(IV)},
\end{aligned}$$

where (III) and (IV) follow.

$$\begin{aligned}
\text{(III)} &= \mathrm{E}\left\{\sum_h U_{mh}^2 X_{jh}X_{j'h}\prod_{l<h}\left(1 - U_{ml}X_{jl} - U_{ml}X_{j'l} + U_{ml}^2 X_{jl}X_{j'l}\right)\right\} \\
&= \sum_h \mathrm{E}U_{mh}^2\mathrm{E}X_{jh}\mathrm{E}X_{j'h}\prod_{l<h}\left(1 - 2\mathrm{E}U_{ml}\mathrm{E}X_{jl} + \mathrm{E}U_{ml}^2\mathrm{E}X_{jl}\mathrm{E}X_{j'l}\right) \\
&= \sum_h \frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\right]^{h-1} \\
&= \frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\left[\frac{2}{(1+\alpha)(1+\beta)} - \frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\right]^{-1} \\
&= \frac{1}{(2+\alpha)(1+\beta)-1}.
\end{aligned}$$

$$
\begin{aligned}
\text{(IV)} &= \mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} U_{mh}X_{jh}\left(1-U_{mh}X_{j'h}\right)\prod_{l<h}\left(1-U_{ml}X_{jl}\right)\left(1-U_{ml}X_{j'l}\right)U_{mi}X_{j'i}\right. \\
&\qquad \left.\times \prod_{l=h+1}^{i-1}\left(1-U_{ml}X_{j'l}\right)\right\} \\
&= \sum_h \sum_{i=h+1}^{\infty}\left[\mathrm{E}U_{mh}\mathrm{E}X_{jh} - \mathrm{E}U_{mh}^2\mathrm{E}X_{jh}\mathrm{E}X_{j'h}\right]\mathrm{E}U_{mi}\mathrm{E}X_{j'i} \\
&\qquad \times \prod_{l<h}\left[1-2\mathrm{E}U_{ml}\mathrm{E}X_{jl}+\mathrm{E}U_{ml}^2\mathrm{E}X_{jl}\mathrm{E}X_{j'l}\right]\prod_{l=h+1}^{i-1}\left(1-\mathrm{E}U_{m'l}\mathrm{E}X_{jl}\right) \\
&= \sum_h \sum_{i=h+1}^{\infty}\left[\frac{1}{(1+\alpha)(1+\beta)} - \frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\right]\frac{1}{(1+\alpha)(1+\beta)} \\
&\qquad \times \left[1-\frac{2}{(1+\alpha)(1+\beta)}+\frac{2}{((1+\alpha)(2+\alpha)(1+\beta)^2}\right]^{h-1}\left[1-\frac{1}{(1+\alpha)(1+\beta)}\right]^{i-h-1} \\
&= \sum_h \frac{1}{(1+\alpha)(1+\beta)}\left[1-\frac{2}{(1+\alpha)(2+\beta)}\right] \\
&\qquad \times \left[1-\frac{2}{(1+\alpha)(1+\beta)}+\frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\right]^{h-1} \\
&= \frac{1}{(1+\alpha)(1+\beta)}\left[1-\frac{2}{(1+\alpha)(2+\beta)}\right]\left[\frac{2}{(1+\alpha)(1+\beta)}-\frac{2}{(1+\alpha)(2+\alpha)(1+\beta)^2}\right]^{-1} \\
&= \frac{1}{2}\left[1-\frac{1}{(2+\alpha)(1+\beta)-1}\right].
\end{aligned}
$$

Thus,

$$
\mathrm{E}\left(F_{mj}(A)F_{mj'}(A)\right) = \Psi(A)^2\,\text{(III)} + 2\Psi(A)^2\,\text{(IV)} = \Psi(A)^2 = \mathrm{E}F_{mj}(A)\,\mathrm{E}F_{mj'}(A),
$$

and $F_{mj}(A)$ and $F_{mj'}(A)$ are uncorrelated.

The proof of $\mathrm{Cov}(F_{mj}(A)F_{mj'}(A))$ proceeds similarly; see expressions (V) and (VI) from Appendix A1.2.

6. $Pr(\phi_{mj} = \phi_{mj'}) = Pr(\phi_{mj} = \phi_{mj'} \neq 0) + Pr(\phi_{mj} = \phi_{mj'} = 0) = 0 + \epsilon_q\epsilon_{q'}$.

A1.2 Lag-block Sparsity Grouping Prior

Properties 1.–4. follow as in Appendix A1.1.

5. Let $q = q(j) = q(j')$. Making use of the previously derived formulas (III) and (IV),

$$\begin{aligned}
\mathrm{E}\left(F_{mj}(A)F_{mj'}(A)\right) &= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\delta_{\xi_{qh}}(A)\right\} + 2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{mj'i}\delta_{\xi_{qh}}(A)\delta_{\xi_{q'i}}(A)\right\} \\
&= \Psi(A)\,\mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\right\} + 2\Psi(A)^2\,\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{mj'i}\right\} \\
&= \Psi(A)\,\textbf{(III)} + 2\Psi(A)^2\,\textbf{(IV)} \\
&= \frac{1}{(2+\alpha)(1+\beta)-1}\Psi(A)\left[1-\Psi(A)\right] + \Psi(A)^2,
\end{aligned}$$

which gives the correlation stated.

If $q = q(j) \neq q' = q(j')$, then

$$\begin{aligned}
\mathrm{E}\left(F_{mj}(A)F_{mj'}(A)\right) &= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\delta_{\xi_{qh}}(A)\delta_{\xi_{q'h}}(A)\right\} \\
&\quad +2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{mj'i}\delta_{\xi_{qh}}(A)\delta_{\xi_{q'i}}(A)\right\} \\
&= \Psi(A)^2\,\textbf{(III)} + 2\Psi(A)^2\,\textbf{(IV)} = \Psi(A)^2.
\end{aligned}$$

6. Let $q = q(j) = q(j')$.

$$\begin{aligned}
Pr(\phi_{mj} = \phi_{mj'}) &= Pr(\phi_{mj} = \phi_{mj'} \neq 0) + Pr(\phi_{mj} = \phi_{mj'} = 0) \\
&= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\delta_{\xi_{qh}}(\mathcal{R}\backslash 0)\right\} + \mathrm{E}\left\{\sum_h \pi_{mjh}\delta_{\xi_{qh}}(0) \times \sum_i \pi_{mj'i}\delta_{\xi_{qi}}(0)\right\} \\
&= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{mj'h}\delta_{\xi_{qh}}(\mathcal{R})\right\} + 2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{mj'i}\delta_{\xi_{qh}}(0)\delta_{\xi_{qi}}(0)\right\} \\
&= \textbf{(III)} + 2\epsilon^2\textbf{(IV)} = \epsilon_q^2 + \frac{1-\epsilon_q^2}{(2+\alpha)(1+\beta)-1}.
\end{aligned}$$

If $q = q(j) \neq q' = q(j')$,

$$Pr(\phi_{mj} = \phi_{mj'}) = Pr(\phi_{mj} = \phi_{mj'} \neq 0) + Pr(\phi_{mj} = \phi_{mj'} = 0) = 0 + \epsilon_q\epsilon_{q'}.$$

7. Let $q = q(j) = q(j')$. Then,

$$
\begin{aligned}
\mathrm{E}\left(F_{mj}(A)F_{m'j'}(A)\right) &= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{m'j'h}\delta_{\xi_{qh}}(A)\right\} + 2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'j'i}\delta_{\xi_{qh}}(A)\delta_{\xi_{qi}}(A)\right\} \\
&= \Psi(A)\,\mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{m'j'h}\right\} + 2\Psi(A)^2\,\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'j'i}\right\} \\
&= \Psi(A)\,(\mathrm{V}) + 2\Psi(A)^2\,(\mathrm{VI}),
\end{aligned}
$$

where

$$
\begin{aligned}
(\mathrm{V}) &= \mathrm{E}\left\{\sum_h U_{mh}U_{m'h}X_{jh}X_{j'h}\prod_{l<h}\left(1 - U_{ml}X_{jl} - U_{ml}X_{j'l} + U_{ml}U_{m'l}X_{jl}X_{j'l}\right)\right\} \\
&= \sum_h \frac{1}{(1+\alpha)^2(1+\beta)^2}\left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{1}{(1+\alpha)^2(1+\beta)^2}\right]^{h-1} \\
&= \frac{1}{(1+\alpha)^2(1+\beta)^2}\left[\frac{2}{(1+\alpha)(1+\beta)} - \frac{1}{(1+\alpha)^2(1+\beta)^2}\right]^{-1} \\
&= \frac{1}{2(1+\alpha)(1+\beta) - 1}
\end{aligned}
$$

and

$$
\begin{aligned}
(\mathrm{VI}) &= \mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} U_{mh}X_{jh}\left(1 - U_{m'h}X_{j'h}\right)\prod_{l<h}\left(1 - U_{ml}X_{jl}\right)\left(1 - U_{m'l}X_{j'l}\right)U_{m'i}X_{j'i}\right. \\
&\qquad \left.\times \prod_{l=h+1}^{i-1}\left(1 - U_{m'l}X_{j'l}\right)\right\} \\
&= \sum_h \sum_{i=h+1}^{\infty} \frac{1}{(1+\alpha)^2(1+\beta)^2}\left[1 - \frac{1}{(1+\alpha)(1+\beta)}\right]\left[1 - \frac{1}{(1+\alpha)(1+\beta)}\right]^{i-h-1} \\
&\qquad \times \left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{1}{(1+\alpha)^2(1+\beta)^2}\right]^{h-1} \\
&= \sum_h \frac{1}{(1+\alpha)(1+\beta)}\left[1 - \frac{1}{(1+\alpha)(1+\beta)}\right]\left[1 - \frac{2}{(1+\alpha)(1+\beta)} + \frac{1}{(1+\alpha)^2(1+\beta)^2}\right]^{h-1} \\
&= \frac{1}{(1+\alpha)(1+\beta)}\left[1 - \frac{1}{(1+\alpha)(1+\beta)}\right]\left[\frac{2}{(1+\alpha)(1+\beta)} - \frac{1}{(1+\alpha)^2(1+\beta)^2}\right]^{-1} \\
&= \frac{1}{2}\left[1 - \frac{1}{2(1+\alpha)(1+\beta) - 1}\right].
\end{aligned}
$$

Using expressions (V) and (VI), we obtain the stated correlation in Property 7.

For $q = q(j) \neq q' = q(j')$.

$$\begin{aligned}
\mathrm{E}\left(F_{mj}(A)F_{m'j'}(A)\right) &= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{m'j'h}\delta_{\xi_{qh}}(A)\delta_{\xi_{q'h}}(A)\right\} \\
&\quad +2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'j'i}\delta_{\xi_{qh}}(A)\delta_{\xi_{q'i}}(A)\right\} \\
&= \Psi(A)^2\,\mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{m'j'h}\right\} + 2\Psi(A)^2\,\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'j'i}\right\} \\
&= \Psi(A)^2\,(\mathrm{V}) + 2\Psi(A)^2\,(\mathrm{VI}) = \Psi(A)^2.
\end{aligned}$$

8. Let $q = q(j) = q(j')$.

$$\begin{aligned}
Pr(\phi_{mj} = \phi_{m'j'}) &= Pr(\phi_{mj} = \phi_{m'j'} \neq 0) + Pr(\phi_{mj} = \phi_{m'j'} = 0) \\
&= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{m'j'h}\delta_{\xi_{qh}}(\mathcal{R}\backslash 0)\right\} + \mathrm{E}\left\{\sum_h \pi_{mjh}\delta_{\xi_{qh}}(0) \times \sum_i \pi_{m'j'i}\delta_{\xi_{qi}}(0)\right\} \\
&= \mathrm{E}\left\{\sum_h \pi_{mjh}\pi_{m'j'h}\delta_{\xi_{qh}}(\mathcal{R})\right\} + 2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \pi_{mjh}\pi_{m'j'i}\delta_{\xi_{qh}}(0)\delta_{\xi_{qi}}(0)\right\} \\
&= (\mathrm{V}) + 2\epsilon^2(\mathrm{VI}) = \epsilon_q^2 + \frac{1-\epsilon_q^2}{2(1+\alpha)(1+\beta)-1}.
\end{aligned}$$

If $q = q(j) \neq q' = q(j')$,

$$Pr(\phi_{mj} \;=\; \phi_{m'j'}) = Pr(\phi_{mj} = \phi_{m'j'} \neq 0) + Pr(\phi_{mj} = \phi_{m'j'} = 0) \;=\; 0 + \epsilon_q\epsilon_{q'}.$$

A1.3 Innovation Variance Properties

Properties 1.–4. follow as in Dunson et al. (2008).

5. For a common value of α and β, the distributions of U_{mh} and W_{mh}, as well as X_{jh} and Z_{jh}, are the same. Hence, the set $\{\tau_{mjh}\}$ will be distributed the same as the set $\{\pi_{mjh}\}$, and we may continue to use the expressions (I)–(VI) to obtain expectations of the IV stick-breaking

weights.

$$
\begin{aligned}
\mathrm{E}\left(G_{mj}(A)G_{mj'}(A)\right) &= \mathrm{E}\left\{\sum_h \tau_{mjh}\tau_{mj'h}\delta_{\eta_{jh}}(A)\delta_{\eta_{j'h}}(A)\right\} \\
&\quad +2\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \tau_{mjh}\tau_{mj'i}\delta_{\eta_{jh}}(A)\delta_{\eta_{j'i}}(A)\right\} \\
&= \mathrm{E}\left\{\delta_{\eta_{jh}}(A)\delta_{\eta_{j'h}}(A)\right\}\,\mathrm{E}\left\{\sum_h \tau_{mjh}\tau_{mj'h}\right\} \\
&\quad +2\left(\mathrm{E}\delta_{\eta_{jh}}(A)\right)\left(\mathrm{E}\delta_{\eta_{j'h}}(A)\right)\mathrm{E}\left\{\sum_h \sum_{i=h+1}^{\infty} \tau_{mjh}\tau_{mj'i}\right\} \\
&= \mathrm{E}\left\{\delta_{\eta_{jh}}(A)\delta_{\eta_{j'h}}(A)\right\}\,\text{(III)} + 2\left(\mathrm{E}\delta_{\eta_{jh}}(A)\right)\left(\mathrm{E}\delta_{\eta_{j'h}}(A)\right)\text{(IV)} \\
&= \frac{1}{(2+\alpha)(1+\beta)-1}\Big[\mathrm{E}\left\{\delta_{\omega_{j1}}(\log A)\delta_{\omega_{j'1}}(\log A)\right\} \\
&\quad -\mathrm{E}\delta_{\omega_{j1}}(\log A)\,\mathrm{E}\delta_{\omega_{j'1}}(\log A)\Big] + \mathrm{E}\delta_{\omega_{j1}}(\log A)\,\mathrm{E}\delta_{\omega_{j'1}}(\log A) \\
&= \frac{1}{(2+\alpha)(1+\beta)-1}\mathrm{Cov}\left(\delta_{\omega_{j1}}(\log A),\ \delta_{\omega_{j'1}}(\log A)\right) + \Phi(\log A)^2 \\
&= \frac{1}{(2+\alpha)(1+\beta)-1}\mathrm{Cov}\left(I\{\omega_{j1}\in \log A\},\ I\{\omega_{j'1}\in \log A\}\right) + \Phi(\log A)^2.
\end{aligned}
$$

Applying $\mathrm{Var}\{\delta_{\omega_{j1}}(\log A)\} = \Phi(\log A)(1-\Phi(\log A))$ and properties 1. and 2. gives the final result.

6. The proof of property 6. follows the same as above, except one uses expressions (V) and (VI) in place of (III) and (IV).

7. Follows from the observation that $\omega_{jh} \neq \omega_{j'h'}$ almost surely as a consequence of the multivariate normal distribution with a non-degenerate correlation.

Appendix 2: MCMC Details

As mentioned in Section 6.2, we introduce several latent variables to facilitate the MCMC simulation from the distributions $F_{mj}(\cdot)$ and $G_{mj}(\cdot)$ in equations (2) and (4), following the algorithm of

Dunson et al. (2008). We will draw the random variables R_{mj} and A_{mj} from multinomial distributions with respective probabilities of $\{\pi_{mjh}\}_h$ and $\{\tau_{mjh}\}_h$. To this end, first consider the following four sets of binary dummy variables, for all m, j, h:

$$
\begin{aligned}
u_{mjh} &\sim \text{Bern}(U_{mh}),\ \ m = 1, \ldots, M,\ j = 1, \ldots, J,\ h = 1, \ldots, H_\phi; \\
x_{mjh} &\sim \text{Bern}(X_{jh}),\ \ m = 1, \ldots, M,\ j = 1, \ldots, J,\ h = 1, \ldots, H_\phi; \\
w_{mjh} &\sim \text{Bern}(W_{mh}),\ \ m = 1, \ldots, M,\ j = 1, \ldots, p,\ h = 1, \ldots, H_\gamma; \\
z_{mjh} &\sim \text{Bern}(Z_{jh}),\ \ m = 1, \ldots, M,\ j = 1, \ldots, p,\ h = 1, \ldots, H_\gamma.
\end{aligned}
$$

Now define $R_{mj} = \min\{h : 1 = u_{mjh} = x_{mjh}\}$ and $A_{mj} = \min\{h : 1 = w_{mjh} = z_{mjh}\}$. These R_{mj}'s and A_{mj}'s are distributed according to the appropriate multinomial distributions. We let R_{mj} designate which ξ_{jh} to choose as ϕ_{mj}, and likewise, A_{mj} gives the η_{jh} to select as γ_{mj}. Hence, Φ is determined by $\{R_{mj}\}$ and $\{\xi_{jh}\}$ and Γ by $\{A_{mj}\}$ and $\{\eta_{jh}\}$. Thus, after sampling the values of $\{R_{mj}\}$, $\{\xi_{jh}\}$, $\{A_{mj}\}$, and $\{\eta_{jh}\}$, the values of Φ and Γ are determined.

Now we calculate the conditional distributions that we will need for our Gibbs sampler for each of the grouping priors. Notationally, we denote the conditional distribution for a random variable, say C, conditional on the remaining random variables by $C|-$.

A2.1 Posterior Computations for Sparsity/InvGamma Grouping Prior

1. Conditional for ξ_{jh} for $j = 1, \ldots, J$ and $h = 1, \ldots, H_\phi$:

   It is important to recall the definition of the GARP parameters. For instance, the first parameter ϕ_{m1} is the regression coefficient for y_{mi1} onto y_{mi2} with innovation variance γ_{m1}. Likewise, ϕ_{m2} and ϕ_{m3} are the coefficients of y_{mi1} and y_{mi2} for modeling y_{mi3} with variance γ_{m2}. For fixed j, we let x^*_{mi} denote the component of y_{mi} that corresponds to the j^{th} GARP parameter regressor, e.g. $x^*_{mi} = y_{mi1}$ for $j = 1, 2$ and $x^*_{mi} = y_{mi2}$ for $j = 3$. Similarly, we let γ^*_m denote the relevant innovation variance. For $j = 1$, $\gamma^*_m = \gamma_{m1}$, and for j =2 and 3,

$\gamma_m^* = \gamma_{m2}$. Finally, we define e_{mi}^* to be the residual for the regression equation, excluding the contribution of x_{mi}^*. That is, for $j = 1$, $e_{mi}^* = y_{mi2}$, for $j = 2$, $e_{mi}^* = y_{mi3} - \phi_{m3}y_{mi2}$, and for $j = 3$, $e_{mi}^* = y_{mi3} - \phi_{m2}y_{mi1}$. In general the *-variables are defined in the natural way for each j so that $e_{mi}^* \sim \mathrm{N}(\phi_{mj}x_{mi}^*, \gamma_m^*)$. Having established the necessary notation, we see that the contribution to the distribution of the Y_{mi}'s from ϕ_{mj} is proportional to

$$\exp\left\{-\frac{1}{2\gamma_m^*}\sum_{i=1}^{n_m}\left(e_{mi}^* - \phi_{mj}x_{mi}^*\right)^2\right\}.$$

However, we do not draw the ϕ_{mj}'s but ξ_{jh}. The contribution from Y about ξ_{jh} is

$$\exp\left\{\sum_{m:R_{mj}=h}\frac{-1}{2\gamma_m^*}\sum_{i=1}^{n_m}(e_{mi}^* - \xi_{jh}x_{mi}^*)^2\right\}. \tag{6}$$

This summation over m such that $R_{mj} = h$ means that we are only including the samples whose jth GARP parameter is drawn from cluster h. From this observation, we have that the conditional distribution of ξ_{jh} is

$$\begin{aligned}\pi(\xi_{jh}|-) &\propto \exp\left\{\sum_{m:R_{mj}=h}\frac{-1}{2\gamma_m^*}\sum_{i=1}^{n_m}(e_{mi}^* - \xi_{jh}x_{mi}^*)^2\right\} \\ &\qquad \times\left(\epsilon_{q(j)}\delta_0(\xi_{jh}) + (1-\epsilon_{q(j)})(2\pi\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{\xi_{jh}^2}{2\sigma^2}\right\}\right) \\ &\propto \epsilon_{q(j)}\delta_0(\xi_{jh}) + (1-\epsilon_{q(j)})\frac{\sigma^*}{\sigma}\exp\left\{\frac{(\mu^*)^2}{2\sigma^{2*}}\right\}\mathrm{N}(\mu^*, \sigma^{2*}),\end{aligned} \tag{7}$$

where

$$\mu^* = \sigma^{2*}\sum_{m:R_{mj}=h}\sum_{i=1}^{n_m}\frac{e_{mi}^*x_{mi}^*}{\gamma_m^*} \text{ and } \sigma^{2*} = \left(\frac{1}{\sigma^2} + \sum_{m:R_{mj}=h}\sum_{i=1}^{n_m}\frac{(x_{mi}^*)^2}{\gamma_m^*}\right)^{-1}. \tag{8}$$

Thus, to sample from this conditional, we set ξ_{jh} to zero with probability

$$\frac{\epsilon_{q(j)}}{\epsilon_{q(j)} + (1-\epsilon_{q(j)})\frac{\sigma^*}{\sigma}\exp\left\{\frac{(\mu^*)^2}{2\sigma^{2*}}\right\}},$$

and draw from the specified $N(\mu^*, \sigma^{2*})$ distribution otherwise. Note that if there are no groups with $R_{mj} = h$ then $\mu^* = 0$ and $\sigma^{2*} = \sigma^2$, and so (7) simplifies to the original prior for ξ_{jh} given by (3).

2. Conditional for $\{R_{mj}\}$, $\{u_{mjh}\}$, and $\{x_{mjh}\}$:

   First, we draw R_{mj} from the marginal over $\{u_{mjh}, x_{mjh}\}_h$ of the conditional distribution of the three. Define $\gamma_m^*, e_{mi}^*, x_{mi}^*$ as in step 1. Then we have

$$P\left(R_{mj} = h \middle| - \backslash \{u_{mjh}, x_{mjh}\}_h\right) \quad \propto \quad \pi_{mjh} \times \exp\left\{-\frac{1}{2\gamma_m^*}\sum_{i=1}^{n_m}(e_{mi}^* - \xi_{jh}x_{mi}^*)^2\right\}. \quad (9)$$

   Hence, we draw R_{mj} from the multinomial distribution with probabilities from (9), normalized to sum to one. Given the value of R_{mj}, we can draw the set $\{u_{mjh}, x_{mjh}\}_h$ to require that R_{mj} is the first occasion where both u_{mjh} and x_{mjh} are one. For $h > R_{mj}$ draw $u_{mjh} \sim$ Bern(U_{mh}) and $x_{mjh} \sim$ Bern(X_{jh}), and when $h = R_{mj}$, $1 = u_{mjh} = x_{mjh}$. For $h < R_{mj}$, then we jointly draw u_{mjh} and x_{mjh} in accordance to the following probabilities

$$\begin{aligned} P\left(u_{mjh} = 0, x_{mjh} = 0\right) &= (1 - U_{mh})(1 - X_{jh})/(1 - U_{mh}X_{jh}), \\ P\left(u_{mjh} = 1, x_{mjh} = 0\right) &= U_{mh}(1 - X_{jh})/(1 - U_{mh}X_{jh}), \\ P\left(u_{mjh} = 0, x_{mjh} = 1\right) &= (1 - U_{mh})X_{jh}/(1 - U_{mh}X_{jh}). \end{aligned}$$

3. Conditional for U_{mh} and X_{jh}:

   Given the values of the u_{mjh}'s and the other variables, the conditional for U_{mh} for $h < H_\phi$ is

$$U_{mh} \sim \text{Beta}\left(1 + \sum_{j=1}^{J} u_{mjh},\ \alpha_\phi + \sum_{j=1}^{J}(1 - u_{mjh})\right).$$

   Likewise, for $h < H_\phi$,

$$X_{jh} \sim \text{Beta}\left(1 + \sum_{m=1}^{M} x_{mjh},\ \beta_\phi + \sum_{m=1}^{M}(1 - x_{mjh})\right).$$

   U_{mH_ϕ} and X_{jH_ϕ} are drawn from distribution degenerate at 1.

   One should recognize that this is slightly different from the specification of Dunson et al. (2008). This is because the authors only define u_{mjh} and x_{mjh} for $R_{mj} \geq h$, and so the above conditional has shape parameters determined by summing over j (or m) where $R_{mj} \geq h$. We choose to include latent variable for each combination of m, j, h for clarity, but one may follow Dunson et al.'s choice as well.

4. Conditional for ϵ_q, $q = 1, \ldots, p-1$:

   By placing a Beta(α_q, β_q) prior on ϵ_q, the conditional for ϵ_q is

$$\epsilon_q|- \sim \text{Beta}\left(\alpha_q + \sum_{j:q(j)=q} \sum_{h=1}^{H_\phi} \delta_0(\xi_{jh}),\ \beta_q + \sum_{j:q(j)=q} \sum_{h=1}^{H_\phi} (1 - \delta_0(\xi_{jh}))\right),$$

   where the sum over $j : q(j) = q$ is simply the sum over the j corresponding to the lag-q GARPs. It is necessary to specify the values of α_q and β_q. We recommend using $\alpha_q = \beta_q = 1$ for all q, which gives a Unif(0,1) prior for each ϵ_q. Alternatively, one could choose the values of α_q and β_q to more aggressively shrink ϵ_q for lower lags toward zero and ϵ_q for higher lags toward one.

5. Conditional for η_{jh} for $j = 1, \ldots, J$ and $h = 1, \ldots, H_\gamma$:

   Let $\tilde{e}_{mi}$ be the residual obtained from the difference of y_{mij} and the previous components of y_{mi} multiplied by the appropriate GARP. For instance, when $j = 1$ $\tilde{e}_{mi} = y_{mi1}$, and for $j = 2$ $\tilde{e}_{mi} = y_{mi2} - \phi_{m1} y_{mi1}$, and so on. Note that this is a different definition of these $\tilde{e}$-residuals from the e^*-residuals used in the ξ_{jh} step. For each value of j, this yields $\tilde{e}_{mi} \sim \mathrm{N}(0, \gamma_{mj})$. The contribution to the likelihood from $Y_{mi} \sim \mathrm{N}\left(0, \Sigma(\Phi_m, \Gamma_m)\right)$ is proportional to

$$\eta_{jh}^{-\frac{1}{2}} \exp\left\{-\frac{\tilde{e}_{mi}^2}{2\eta_{jh}}\right\} \delta_h(A_{mj}).$$

   Hence, the conditional for each η_{jh} is

$$\eta_{jh}|- \sim \text{InvGamma}\left(\lambda_1 + \frac{1}{2} \sum_{m:A_{mj}=h} n_m,\ \lambda_2 + \frac{1}{2} \sum_{m:A_{mj}=h} \sum_{i=1}^{n_m} \tilde{e}_{mi}^2\right).$$

6. Conditional for $\{A_{mj}\}$, $\{w_{mjh}\}$, and $\{z_{mjh}\}$:

   To draw A_{mj} we will proceed similarly to step 2 by looking at the conditional marginally over $\{w_{mjh}, z_{mjh}\}_h$.

$$P\left(A_{mj} = h| - \setminus \{w_{mjh}, z_{mjh}\}_h\right) \propto \tau_{mjh} \times \eta_{jh}^{-\frac{1}{2} n_m} \exp\left\{-\frac{1}{2\eta_{jh}} \sum_{i=1}^{n_m} \tilde{e}_{mi}^2\right\}. \quad (10)$$

Hence, we draw A_{mj} from the multinomial distribution with probabilities from (10), normalized to sum to one. As before, we simulate the sets w_{mjh} and z_{mjh} conditional on A_{mj} being the first occasion where both w_{mjh} and z_{mjh} are one. For $h > A_{mj}$ draw $w_{mjh} \sim \text{Bern}(W_{mh})$ and $z_{mjh} \sim \text{Bern}(Z_{jh})$, and when $h = A_{mj}$, $1 = w_{mjh} = z_{mjh}$. For $h < A_{mj}$, we jointly draw w_{mjh} and z_{mjh} in accordance to the following probabilities

$$\begin{aligned} P\left(w_{mjh}=0, z_{mjh}=0\right) &= (1-W_{mh})(1-Z_{jh})/(1-W_{mh}Z_{jh}), \\ P\left(w_{mjh}=1, z_{mjh}=0\right) &= W_{mh}(1-Z_{jh})/(1-W_{mh}Z_{jh}), \\ P\left(w_{mjh}=0, z_{mjh}=1\right) &= (1-W_{mh})Z_{jh}/(1-W_{mh}Z_{jh}). \end{aligned}$$

7. Conditional for W_{mh} and Z_{jh}:

   Proceeding identically to step 3, we get the following conditionals for $h < H_\gamma$,

$$\begin{aligned} W_{mh}|- &\sim \text{Beta}\left(1+\sum_{j=1}^{J} w_{mjh},\ \alpha_\gamma + \sum_{j=1}^{J}(1-w_{mjh})\right), \\ Z_{jh}|- &\sim \text{Beta}\left(1+\sum_{m=1}^{M} z_{mjh},\ \beta_\gamma + \sum_{m=1}^{M}(1-z_{mjh})\right), \end{aligned}$$

   and $W_{mH_\gamma}, Z_{jH_\gamma} \sim \delta_1$.

We now look at some of the issues involved in dealing with the hyperparameters. In practice, it will generally be infeasible to specify values for these quantities, so we wish to choose reasonable, disperse prior distributions for them.

8. The first hyperparameter of interest is the variance σ^2 from the normal component of the ξ_{jh}'s in equation (3). We choose the InvGamma(a, b) family of distributions for the prior, so that we will have conjugacy. This yields the following conditional distribution for σ^2,

$$\sigma^2|- \sim \text{InvGamma}\left(a + \frac{1}{2}\sum_{j,h}(1-\delta_0(\xi_{jh})),\ b + \frac{1}{2}\sum_{j,h}\xi_{jh}^2\right).$$

   One must now specify the values of a, b. We recommend InvGamma(0.1, 0.1), so that our prior approximates the commonly-used improper prior $\pi(\sigma^2) \propto \sigma^{-2}$.

9. The α_ϕ and β_ϕ control the amount of clustering for the GARP parameters. It is not intuitively obvious where these parameters would congregate, so we require priors that will not too strongly inform the posterior. Following the example for Dunson et al. (2008), we choose a Gamma(1,1) prior for both α_ϕ and β_ϕ. Then the conditional for α_ϕ is

$$\alpha_\phi|- \sim \text{Gamma}\left(M(H_\phi - 1) + 1,\ 1 - \sum_{m=1}^{M} \sum_{h=1}^{H_\phi - 1} \log(1 - U_{mh})\right).$$

Likewise,

$$\beta_\phi|- \sim \text{Gamma}\left(J(H_\phi - 1) + 1,\ 1 - \sum_{j=1}^{J} \sum_{h=1}^{H_\phi - 1} \log(1 - X_{jh})\right).$$

Clearly, we can choose a different Gamma(a, b) prior instead of Gamma(1,1), and we will maintain the Gamma-Gamma conjugacy.

10. The λ_1 and λ_2 parameters control the distribution of the η_{jh}. We place independent Gamma(1,1) priors on each. The conditional for λ_2 is

$$\lambda_2|- \sim \text{Gamma}\left(\lambda_1 p H_\gamma + 1,\ 1 + \sum_{j,h} \eta_{jh}^{-1}\right).$$

The conditional for λ_1 is

$$\pi(\lambda_1|-) \propto \Gamma(\lambda_1)^{-pH_\gamma} \lambda_2^{-\lambda_1 p H_\gamma} \exp\left\{-\lambda_1 \left(1 + \sum_{j,h} \log(\eta_{jh})\right)\right\},$$

but this is not a standard distribution to use in the Gibbs sampler. So it becomes necessary to implement an alternative sampling method, and we choose to introduce a Metropolis in Gibbs step to approximately simulate from the conditional of λ_1. Draw the candidate value λ_1^* to replace the current value λ_1 from the N(λ_1, ζ) distribution, and accept the move to λ_1^* with probability

$$\min\left\{1,\ \left[\exp\left\{\log\left(\frac{\Gamma(\lambda_1)}{\Gamma(\lambda_1^*)}\right) + \frac{1}{pH_\gamma}(\lambda_1 - \lambda_1^*)\left(1 + \sum_{j,h} \log(\eta_{jh}) + pH_\gamma \log(\lambda_2)\right)\right\}\right]^{pH_\gamma} I(\lambda_1^* > 0)\right\},$$

It is necessary to prespecify a candidate variance ζ such that the acceptance rate is 20 to 40% (Gelman et al., 1996).

11. The α_γ and β_γ parameters control the amount of clustering for the innovation variance parameters. As in step 2, we put a Gamma(1,1) prior on both, and we have the following conditionals:

$$\begin{aligned}\alpha_\gamma|- &\sim \text{Gamma}\left(M(H_\gamma-1)+1,\ 1-\sum_{m=1}^{M}\sum_{h=1}^{H_\gamma}\log(1-W_{mh})\right),\\ \beta_\gamma|- &\sim \text{Gamma}\left(M(H_\gamma-1)+1,\ 1-\sum_{j=1}^{J}\sum_{h=1}^{H_\gamma}\log(1-Z_{jh})\right).\end{aligned}$$

Having specified all of the necessary conditionals for the model, the MCMC algorithm is implemented by sampling the parameters from each set in order.

A2.2 Posterior Computations for the Non-sparse Grouping Prior

Most of the parameters of the non-sparse prior yield identical conditional distribution to those from the sparsity grouping prior. Hence, we only discuss those parameters with diverging distributions.

1. Because the prior distribution of the ξ_{jh} does not incorporate a zero point mass for the GARP parameters, the conditional will no longer be a mixture of a zero point mass and normal. We have $\xi_{jh} \sim \text{N}(\mu^*, \sigma^{2*})$, where the normal parameters come from Equation (8).

4. There are no longer any ϵ's in the non-sparse prior, so this is an empty step.

8. The distribution of the variance for the GARP candidates is

$$\sigma^2|- \sim \text{InvGamma}\left(a+\frac{1}{2}JH_\phi,\ b+\frac{1}{2}\sum_{j,h}\xi_{jh}^2\right),$$

where the prior for σ^2 is InvGamma(a, b).

A2.3 Posterior Computations for the Lag-block Prior

1. The conditional for ξ_{qh} will again be a mixture of a point mass at zero and a normal distribution. Let $\mathcal{P}_{qh}$ denotes the set of (m,j) such that $q(j)=q$ and $R_{mj}=h$, which is the set of group-GARP pairs that contribute to the estimation of ξ_{qh}. For each $(m,j)\in\mathcal{P}_{qh}$, we let $e^*_{mij}, x^*_{mij}, \gamma^*_{mj}$ be the residual, GARP-regressor, and IV such that $e^*_{mij}\sim \mathrm{N}(\phi_{mj}x^*_{mij}, \gamma^*_{mj})$, as described in the step 1 for the sparsity grouping prior. Defining

$$\mu^* = \sigma^{2*}\sum_{(m,j)\in\mathcal{P}_{qh}}\sum_{i=1}^{n_m}\frac{e^*_{mij}x^*_{mij}}{\gamma^*_{mj}} \text{ and } \sigma^{2*} = \left(\frac{1}{\sigma^2}+\sum_{(m,j)\in\mathcal{P}_{qh}}\sum_{i=1}^{n_m}\frac{(x^*_{mij})^2}{\gamma^*_{mj}}\right)^{-1},$$

we have that $\xi_{qh}|-$ is a mixture of zero and the $\mathrm{N}(\mu^*,\sigma^{2*})$ distribution, where we draw the point mass at 0 with probability

$$\frac{\epsilon_q}{\epsilon_q+(1-\epsilon_q)\frac{\sigma^*}{\sigma}\exp\left\{\frac{(\mu^*)^2}{2\sigma^{2*}}\right\}}.$$

Note if $\mathcal{P}_{qh}$ is empty, then the conditional is $\epsilon_q\delta_0+(1-\epsilon_q)\mathrm{N}(0,\sigma^2)$.

2. The lag-block conditional for R_{mj} marginalized over $\{u_{mjh}, x_{mjh}\}_h$ is multinomial with probabilities proportional to

$$P\left(R_{mj}=h|-\backslash\{u_{mjh},x_{mjh}\}_h\right) \propto \pi_{mjh}\times\exp\left\{-\frac{1}{2\gamma^*_m}\sum_{i=1}^{n_m}(e^*_{mi}-\xi_{q(j)h}x^*_{mi})^2\right\}.$$

The conditionals for $\{u_{mjh}, x_{mjh}\}_h$ are the same as the sparsity grouping case.

4. With a Beta(α_q,β_q) prior on ϵ_q, the conditional is

$$\epsilon_q|-\sim \text{Beta}\left(\alpha_q+\sum_{h=1}^{H_\phi}\delta_0(\xi_{qh}),\ \beta_q+\sum_{h=1}^{H_\phi}(1-\delta_0(\xi_{qh}))\right).$$

8. With the prior for σ^2 of InvGamma(a,b), we have the conditional distribution

$$\sigma^2|-\sim \text{InvGamma}\left(a+\frac{1}{2}\sum_{q,h}(1-\delta_0(\xi_{qh}))\,,\ b+\frac{1}{2}\sum_{q,h}\xi^2_{qh}\right).$$

A2.4 Posterior Computations for the Correlated-logNormal Prior

5. Instead of considering the conditional for η_{jh}, we instead choose to look in terms of $\omega_{jh} = \log \eta_{jh}$. For each sampling set, we partition ω_h into $(\omega_{hA}, \omega_{hB})$ so that ω_{hA} contains the collection of ω_{jh} such that $A_{mj} = h$ for at least one m. This divides ω_h into the ω_{hB}, which can be drawn easily through a conjugate distribution, and the ω_{hA}, which require a more advanced sampling method.

   To sample ω_{hB} given the remaining variables, we let a denote the length of ω_{hA} and $b = p-a$ denote the length of ω_{hB}. Define R_{AA} to be the submatrix of $R(\rho)$ corresponding to the elements of ω_{hA}, R_{BB} corresponding to the elements of ω_{hB}, and R_{BA} contain the elements of the rows of ω_{hB} and columns of ω_{hA}. Then, using standard multivariate normal results,

   $$\omega_{hB}|\omega_{hA}, - \sim \mathrm{N}_b\left(\psi\mathbf{1}_b + R_{BA}R_{AA}^{-1}(\omega_{hA} - \psi\mathbf{1}_a),\ \Omega(R_{BB} - R_{BA}R_{AA}^{-1}R'_{BA})\right).$$

   Jointly drawing the vector ω_{hB} leads to better mixing than drawing each component separately.

   To sample ω_{hA}, we cycle through the components $\omega_{h\alpha}$ of ω_{hA} for $\alpha = 1, \ldots, a$. We recognize that the contribution to the conditional of $\omega_{h\alpha}$ from the prior is

   $$\exp\left\{-\frac{1}{2\Omega^*}(\omega_{h\alpha} - \psi^*)^2\right\},$$

   where

   $$\psi^* = \psi + R_{\alpha,(-\alpha)}R^{-1}_{(-\alpha),(-\alpha)}(\omega_{h(-\alpha)} - \psi\mathbf{1}_{p-1}),\ \ \Omega^* = \Omega\left(1 - R_{\alpha,(-\alpha)}R^{-1}_{(-\alpha),(-\alpha)}R'_{\alpha,(-\alpha)}\right),$$

   $\omega_{h(-\alpha)}$ is the ω_h vector after removing $\omega_{h\alpha}$, $R_{(-\alpha),(-\alpha)}$ is the $R(\rho)$ matrix formed by removing the row and column corresponding to α, and $R_{\alpha,(-\alpha)}$ is the vector defined by taking the α row of $R(\rho)$ and removing the α component. We view this equivalently as $\eta_{h\alpha} = \exp(\omega_{h\alpha}) \sim$

logNormal(ψ^*, Ω^*), and calculate the conditional distribution in terms of $\eta_{h\alpha}$. This gives

$$\pi(\eta_{h\alpha}|\eta_{h(-\alpha)}, -) \propto \eta_{h\alpha}^{-\frac{1}{2}\sum_m n_m \delta_h(A_{mj})-1} \exp\left\{-\frac{1}{2\eta_{h\alpha}}\sum_m \sum_{i=1}^{n_m} (\tilde{e}_{mij})^2 \delta_h(A_{mj}) - \frac{1}{2\Omega^*}(\log \eta_{h\alpha} - \psi^*)^2\right\}.$$

Sampling from this distribution requires an approximate sampling step. We recommend slice sampling (Neal 2003), although an alternative sampling strategy could be used.

10. With the correlated-logNormal prior, we no longer have the hyperparameters λ_1, λ_2, but we now have ψ, Ω, ρ. Choosing $\Omega \sim$ InvGamma(a, b) and $\psi|\Omega \sim$ N$(0,c^2\Omega)$ as priors for the two hyperparameters yields the following conditionals

$$\begin{aligned}
\psi|\Omega, \rho, - &\sim \text{N}\left(\frac{\mathbf{1}_p' R(\rho)^{-1}\sum_h \omega_h}{c^{-2} + H_\gamma \mathbf{1}_p' R(\rho)^{-1}\mathbf{1}_p}, \frac{\Omega}{c^{-2} + H_\gamma \mathbf{1}_p' R(\rho)^{-1}\mathbf{1}_p}\right), \\
\Omega|\psi, \rho, - &\sim \text{InvGamma}\Bigg(a + \frac{1}{2}(pH_\gamma + 1), \\
&\qquad b + \frac{\psi^2}{2c^2} + \frac{1}{2}\sum_h (\omega_h - \psi \mathbf{1}_p)' R(\rho)^{-1}(\omega_h - \psi \mathbf{1}_p)\Bigg).
\end{aligned}$$

In the simulation and data example, we use $a = b = .1, c^2 = 1000$. As mentioned in Section 6.2, it has been our experience that sampling ρ leads to instability, and we generally recommend fixing it.

A2.5 Final Comments about MCMC Computations

We finally note that one can view our grouping priors in a hierarchical fashion with multiple levels. As is often the case in hierarchical models, there may be little information about the parameters in the lowest levels. We have often found this to be the case for the grouping priors resulting in poor mixing for some of the model parameters. While the values of the GARPs and IVs tend to mix well, as evidenced by trace and autocorrelation plots, the stick-breaking parameters α_ϕ, β_ϕ, α_γ, and β_γ do not mix as well. While the GARPs/IVs show minimal autocorrelation within ten iterations, the stick-breaking parameters require more than fifty. As we are usually not interested in directly performing inference on α, β and due to the previously mentioned concerns about the computational time, we recommend selecting a thinning value that accommodates good mixing of the GARPs and IVs. We also encourage the user to consider the trace plot formed by the log density of the data given the values of the mean (if non-zero) and covariance parameters. An alternative solution is to run a short initial chain and fix the values of the stick-breaking parameters at their posterior means/modes for use in the full MCMC analysis.

When using the correlated-logNormal grouping prior, we similarly observe problems with the sampling for the ω correlation ρ. In many cases, ρ will alternate between values close to 1 and -1, which does not correspond with our intuition about the IVs. Hence, we opt to treat ρ as a tuning parameter. We recommend specifying a default value such as $\rho = 0.75$, possibly trying a few other choice and selecting the value with the superior DIC. As shown in the depression data study (see Table 5), the three choices of ρ=0.5, 0.75, and 0.9 lead to similar model fits as measured by the deviance. Based on our simulation studies, we believe that the correlated-logNormal prior is fairly robust to the choice of ρ.

Appendix 3: Additional Risk Simulation Details

Here we include details about some additional risk simulations beyond those discussed in Section 7 of the article.

A3.1 Risk Simulation A1

We perform another risk simulation similar to the first with five groups and $p = 4$. The true covariance matrices are given by

$$
\begin{aligned}
\Phi_1 = \Phi_2 &= (1,\ 0.5,\ 1,\ 0.5,\ 0.5,\ 1), & \Gamma_1 = \Gamma_3 &= (2,\ 2,\ 2,\ 2),\\
\Phi_3 = \Phi_4 &= (1,\ -0.5,\ 1,\ -0.5,\ -0.5,\ 1), & \Gamma_2 = \Gamma_4 &= (4,\ 4,\ 4,\ 4),\\
\Phi_5 &= (2,\ -1.0,\ 2,\ -0.5,\ -1.0,\ 1), & \Gamma_5 &= (2,\ 2,\ 1,\ 1).
\end{aligned}
$$

As in the article, we create fifty datasets and use the same sample sizes $n_1 = \ldots = n_4 = 30, n_5 = 15$. There should be a large amount of clustering in this case, since there is a great deal of commonality among GARPs and IVs for different samples. These covariance matrices also do not have any conditional independence relationships to exploit since each of the GARPs are nonzero. We now specify $H_\phi = H_\gamma = 30$ for the grouping priors and use the same hyperpriors as before.

Risk estimates are shown in Table 1. As in the previous risk simulation, the lag-block/correlated-logNormal prior produces the best risk (15% and 20% lower than the top naive prior NB2/NB). For this specification of Σ, we see that the priors that do not promote zeros in the $T(\Phi_m)$ matrices (NB2 and non-sparse grouping) perform better than their sparsity-inducing counterparts (NB1 and sparsity grouping). This is not unexpected because this choice of GARPs does not have any conditional independence relationships. The lag-block again is the top prior for the GARPs because it allows for sharing information across all GARP parameters of a common lag $q(j)$, instead of only the GARPs at a common j. As before modeling the innovation variances is improved from the naive Bayes prior to the InvGamma prior to the correlated-logNormal prior. For this particular choice of Σ, we again see that the grouping priors significantly improved the estimation of the covariance matrices with risk improvements ranging from 20–36% for L_1 and 15–30% for L_2 over the group-specific flat prior.

Risk simulations with these covariance specifications and a doubled sample size for each group produced the similar results to these. The lag-block and grouping priors continue to dominate over the flat prior and the naive Bayes estimates.

A3.2 Risk Simulation A2

We explore how the estimates obtained from the proposed priors perform with an increase to the dimension of the covariance matrices and the number of groups as in Risk Simulation 2 of the article. Here we allow for $M = 8$ groups and consider 6×6 covariance matrices, defined by the

GARP and IV parameters in Table 2. This choice for Φ incorporates commonality both within lag and across groups, as well as possessing many conditional independence relationships among the higher lag terms. We choose a sample size of thirty for the first five groups and fifteen for the final three groups, and thirty clusters for the grouping priors. The estimated risk associated with estimating the covariance matrices for each of the two loss functions is shown in Table 3.

With the increased values of p and M, all of the grouping priors beat the naive priors. The ability to borrow strength across groups improves the estimation such that even the non-sparse grouping prior, which does not allow the correct independence relationships, beats the NB1 prior, which correctly incorporates the potential independence. The lag-block/correlated-logNormal prior continues to beat the remainder of the grouping priors, with a risk improvement of 30 and 23% over the NB1/NB prior and 64 and 51% over the group-specific flat prior. From these and other simulation studies, we believe that as the number of groups M and the dimension of the covariance matrix p increases, the grouping estimators for Σ will outperform the naive Bayes estimators and the margin by which they do so increases. This is particularly important since the number of possible models increases as p and M increase.

Web Appendix References

Dunson, D. B., Xue, Y., and Carin, L. (2008). The matrix stick-breaking process: Flexible Bayes meta-analysis. *Journal of the American Statistical Association*, 103(481):317–327.

Gelman, A., Roberts, G. O., and Gilks, W. R. (1996). Efficient Metropolis jumping rules. In Bernardo, J., Berger, J., Dawid, A., and Smith, A., editors, *Bayesian Statistics 5 – Proceedings of the Fifth Valencia International Meeting*, pages 599–607.

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, 31(3):705–767.

Priors		Estimated Risk	
GARP	IV	Loss Fcn 1	Loss Fcn 2
Lag-block	Corr-logNormal	0.297	0.492
Lag-block	InvGamma	0.318	0.520
Non-sparse	InvGamma	0.342	0.552
Sparsity	Corr-logNormal	0.348	0.563
NB2	NB	0.351	0.563
Sparsity	InvGamma	0.371	0.596
NB1	NB	0.385	0.612
Group-specific flat		0.464	0.701
Common-Σ flat		4.815	48.081

Table 1: Risk Estimates for Simulation A1

Φ_1=(	0.7,	0.2,	0.7,	0,	0.2,	0.7,	0,	0,	0.2,	0.7,	0,	0,	0,	0.2,	0.7)
Φ_2=(	0.7,	0.2,	0.7,	0.1,	0.2,	0.7,	0,	0.1,	0.2,	0.7,	0,	0,	0.1,	0.2,	0.7)
Φ_3=(	0.3,	0,	0.3,	0,	0,	0.3,	0,	0,	0,	0.3,	0,	0,	0,	0,	0.3)
Φ_4=(	0.3,	0,	0.3,	-0.1,	0,	0.3,	0,	-0.1,	0,	0.3,	0,	0,	-0.1,	0,	0.3)
Φ_5=(	1,	-0.5,	1,	0,	-0.5,	1,	0,	0 ,	-0.5,	1,	0,	0,	0,	-0.5,	1)
Φ_6=(	1,	-0.5,	1,	0.3,	-0.5,	1,	0,	0.3,	-0.5,	1,	0,	0,	0.3,	-0.5,	1)
Φ_7=(	1,	-0.2,	1,	-0.2,	-0.2,	1,	-0.2,	-0.2,	-0.2,	1,	-0.2,	-0.2,	-0.2,	-0.2,	1)
Φ_8=(	1,	-0.2,	1,	-0.2,	-0.2,	1,	-0.2,	-0.2,	-0.2,	1,	-0.2,	-0.2,	-0.2,	-0.2,	1)

Γ_1=Γ_2=(	1,	1,	1,	1,	1,	1)
Γ_3=Γ_8=(	3.4,	3.1,	2.8,	2.5,	2.2,	1.8)
Γ_4=(	3,	3,	2,	2,	2,	1)
Γ_5=(	5,	3,	3,	4,	4,	4)
Γ_6=(	5,	5,	3,	3,	2,	2)
Γ_7=(	2,	1.8,	1.6,	1.4,	1.2,	1)

Table 2: Parameter Values for Simulation A2

Priors		Estimated Risk	
GARP	IV	Loss Fcn 1	Loss Fcn 2
Lag-block	Corr-logNormal	0.468	0.781
Lag-block	InvGamma	0.492	0.816
Sparsity	Corr-logNormal	0.556	0.904
Sparsity	InvGamma	0.583	0.939
Non-sparse	InvGamma	0.602	0.963
NB1	NB	0.664	1.013
NB2	NB	0.761	1.121
Group-specific flat		1.300	1.584
Common-Σ flat		3.036	14.149

Table 3: Risk Estimates for Simulation A2