Supplemental Appendix: Bayesian Hierarchical Models with Conjugate Full-Conditional Distributions for Dependent Data from the Natural Exponential Family

Jonathan R. Bradley[1], Scott H. Holan[23], Christopher K. Wikle[2]

Introduction

In Supplemental Appendix A, we provide additional discussion surrounding the DY and CM distributions introduced in the main text. In Supplemental Appendix B, we provide the proofs of the results, propositions, and theorems stated in the main text. Details surrounding the collapsed Gibbs sampler are provided in Supplemental Appendix C. Additional simulation results are presented in Supplemental Appendix D.

[1](to whom correspondence should be addressed) Department of Statistics, Florida State University, 117 N. Woodward Ave, Tallahassee, Fl 32306, bradley@stat.fsu.edu

[2]Department of Statistics, University of Missouri, 146 Middlebush Hall, Columbia, MO 65211-6100

[3]U.S. Census Bureau, 4600 Silver Hill Road, Washington, D.C., 20233-9100

Appendix A: Additional Discussion on the DY and CM Distribution

Appendix A.i: Example Univariate Distributions

In Table 1, we give examples of ψ, EF$(Y;\psi)$, and $K(\alpha,\kappa)$.

Data Model	Natural Parameter	Log Partition Function (i.e., ψ and b)	Normalizing Constant	How to Simulate From the DY Distribution
Gamma(a,k) $f(Z\vert\alpha,\kappa)=\frac{1}{\Gamma(a)k^{a}}\exp(-Z/k)$ $a>0,k>0,Z>0$	Negative Reciprocal: $Y=-\frac{1}{k}$.	$\psi_1(Y)=\log\left(-\frac{1}{Y}\right)$ $b=a$	$K(\alpha,\kappa)=\frac{\alpha^{\kappa+1}}{\Gamma(\kappa+1)}$	Let $W\sim$ Gamma$(\kappa+1,1/\alpha)$, where $\alpha>0$, and $\kappa>0$. Then, $-W\sim$ DY$(\alpha,\kappa;\,\psi_1)$.
Bin(t,p) $f(Z\vert t,p)=\binom{t}{Z}p^{Z}(1-p)^{t-Z}$ $0<p<1,t=1,2,\ldots,Z=0,\ldots,t$	Logit: $Y=\log\left(\frac{p}{1-p}\right)$	$\psi_2(Y)=\log(1+\exp(Y))$ $b=t$	$K(\alpha,\kappa)=\frac{\Gamma(\kappa)}{\Gamma(\alpha)\Gamma(\kappa-\alpha)}$	Let $W\sim$ Beta$(\alpha,\kappa-\alpha)$, where $\kappa>\alpha>0$ and "Beta$(\alpha,\kappa-\alpha)$" is a shorthand for the beta distribution with shape parameter α and scale parameter $\kappa-\alpha$. Then, $\log\left(\frac{W}{1-W}\right)\sim$ DY$(\alpha,\kappa;\,\psi_2)$.
NegBin(t,p) $f(Z\vert t,p)=\binom{Z+t-1}{Z}p^{Z}(1-p)^{t}$ $0\leq p\leq 1,t=1,2,\ldots,Z=0,1,\ldots,$	Logit: $Y=\log\left(\frac{p}{1-p}\right)$	$\psi_2(Y)=\log(1+\exp(Y))$ $b=t+Z$	$K(\alpha,\kappa)=\frac{\Gamma(\kappa)}{\Gamma(\alpha)\Gamma(\kappa-\alpha)}$	Let $W\sim$ Beta$(\alpha,\kappa-\alpha)$, where $\kappa>\alpha>0$. Then, $\log\left(\frac{W}{1-W}\right)\sim$ DY$(\alpha,\kappa;\,\psi_2)$.
Pois(μ) $f(Z\vert\mu)=\frac{\mu^{Z}\exp(-\mu)}{Z!}$ $\mu\in\mathbb{R}^{+},Z=0,1,2,\ldots$	Log $Y=\log(\mu)$	$\psi_3(Y)=\exp(Y)$ $b=1$	$K(\alpha,\kappa)=\frac{\kappa^{\alpha}}{\Gamma(\alpha)}$	Let $W\sim$ Gamma$(\alpha,1/\kappa)$, where $\alpha>0$ and $\kappa>0$. Then, $\log(W)\sim$ DY$(\alpha,\kappa;\,\psi_3)$.
Norm(μ,s) $f(Z\vert\mu,s)=\left(\frac{1}{2\pi s^{2}}\right)^{1/2}\exp\left(-\frac{-(Z-\mu)^{2}}{2s^{2}}\right)$ $\mu\in\mathbb{R},s\in\mathbb{R}^{+},Z\in\mathbb{R}$	Linear: $Y=\frac{\mu}{s^{2}}$	$\psi_4(Y)=Y^{2}$ $b=\frac{s^{2}}{2}$	$K(\alpha,\kappa)=\left(\frac{\kappa}{\pi}\right)^{1/2}\exp(-\frac{\alpha^{2}}{4\kappa})$	Let W be a normal random variable with mean $\frac{\alpha}{2\kappa}$ and variance $\frac{1}{2\kappa}$. Then, $W\sim$ DY$(\alpha,\kappa;\,\psi_4)$.

Table 1: Univariate Distributions: The first column has the data model, the second column has the natural parameter, the third column contains quantities that define the log partition function, the fourth column has the normalizing constant, and the fifth column has instructions on how to simulate from the DY random variable with the corresponding ψ. Let $\mathbb{R}^{+}=\{x:x>0\}$.

Appendix A.ii: A Metropolis-Hastings Approach to the Conditional CM distribution

To use the affine transformation (i.e., $\mathbf{q} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{w}$) as a means to generate from a pdf proportional to CM_c, one does not necessarily have to marginalize across $\boldsymbol{\mu}$. This is because the unnormalized CM distribution is proportional to the marginal distribution from an improper extension of $\mathbf{q}$. Specifically, let ρ be an unnormalized CM distribution with mean $\mathbf{V}\boldsymbol{\mu}$ and covariance parameter $\mathbf{V}^{-1} = [\mathbf{H}, \frac{1}{\sigma_2}\mathbf{Q}_2]$, where $\mathbf{Q}_2$ is the $n \times (n-r)$ orthonormal basis for the null space of $\mathbf{H}$. Then we introduce a latent $(n-r)$-dimensional random vector $\mathbf{q}_2$ and augment the distribution of $\mathbf{q}$ with,

$$\begin{aligned}\rho(\mathbf{q}, \mathbf{q}_2 | \mathbf{c} = \mathbf{V}\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) &= \exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q} - \boldsymbol{\kappa}'\psi(\mathbf{H}\mathbf{q} - \boldsymbol{\mu})\right\}\\ &= g(\mathbf{q}|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) g(\mathbf{q}_2|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}),\end{aligned}$$

where

$$g(\mathbf{q}_1|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = \exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q} - \boldsymbol{\kappa}'\psi(\mathbf{H}\mathbf{q} - \boldsymbol{\mu})\right\} \propto f(\mathbf{q}_1|\mathbf{q}_2 = \mathbf{0}_{n-r,1}, \boldsymbol{\mu}, \mathbf{H}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \tag{A.1}$$

$$g(\mathbf{q}_2|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) = 1. \tag{A.2}$$

Thus, the Metropolis-Hastings ratio with update $\mathbf{q} = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{w}$ is one in the limit. That is, the following Metropolis-Hastings ratio approaches one as σ_2 increases,

$$\frac{\exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q}^* - \boldsymbol{\kappa}'\psi(\mathbf{H}\mathbf{q}^* - \boldsymbol{\mu})\right\}}{\exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q}^{[m]} - \boldsymbol{\kappa}'\psi(\mathbf{H}\mathbf{q}^{[m]} - \boldsymbol{\mu})\right\}} \frac{\exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q}^{[m]} + \frac{1}{\sigma_2}\boldsymbol{\alpha}'\mathbf{Q}_2\mathbf{q}_2^{[m]} - \boldsymbol{\kappa}'\psi(\mathbf{H}\mathbf{q}^{[m]} + \frac{1}{\sigma_2}\mathbf{Q}_2\mathbf{q}_2^{[m]} - \boldsymbol{\mu})\right\}}{\exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q}^* + \frac{1}{\sigma_2}\boldsymbol{\alpha}'\mathbf{Q}_2\mathbf{q}_2^* - \boldsymbol{\kappa}'\psi(\mathbf{H}\mathbf{q}^* + \frac{1}{\sigma_2}\mathbf{Q}_2\mathbf{q}_2^* - \boldsymbol{\mu})\right\}},$$

where $\mathbf{q}^*$ and $\mathbf{q}_2^*$ are a proposed values of $\mathbf{q}$ and $\mathbf{q}_2$, and $\mathbf{q}^{[m]}$ and $\mathbf{q}_2^{[m]}$ are the previous values in the Markov chain. The argument in (A.1) and (A.2) is very similar to a result in Bradley et al. (2018, cf. Theorem 2), which was clarified in the rejoinder of Bradley et al. (2018). Although the CM_c is proper, it is crucial that we recognize that $\mathbf{q}$ follows an unnormalized CM_c and is extended by an improper $\mathbf{q}_2$. This *improper extension* results in a lack of Kolmogorov consistency (Daniell, 1919; Kolmogorov, 1933; Bradley et al., 2018). However, proper extensions of the CM distribution are Kolmogorov consistent (see Theorem 4).

Appendix B: Proofs

In this appendix we provide proofs for the technical results stated in the paper.

Proof of Theorem 1(i): From (2) of the main text we see that the distribution of the random vector $\mathbf{w}$ in (7) is given by,

$$\left(\prod_{i=1}^{n} K(\alpha_i, \kappa_i)\right) \exp\left\{\boldsymbol{\alpha}'\mathbf{w} - \boldsymbol{\kappa}'\psi(\mathbf{w})\right\}; \quad \mathbf{w} \in \mathbb{R}^n.$$

The inverse of the transform of (7) is given by $\mathbf{w} = \mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu})$, and the Jacobian is given by $|\det(\mathbf{V}^{-1})|$. Then, by a change-of-variables (e.g., see Casella and Berger, 2002), we have that the pdf of $\mathbf{Y}$ is given by,

$$\det(\mathbf{V}^{-1}) \left(\prod_{i=1}^{n} K(\alpha_i, \kappa_i)\right) \exp\left[\boldsymbol{\alpha}'\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu}) - \boldsymbol{\kappa}'\psi\left\{\mathbf{V}^{-1}(\mathbf{Y} - \boldsymbol{\mu})\right\}\right]; \; \mathbf{Y} \in \mathscr{M}^n.$$

This completes the proof of Theorem 1(i).

Proof of Theorem 2: It follows from Proposition 1(i) that the conditional distribution is given by

$$
\begin{aligned}
f(\mathbf{Y}_1|\mathbf{Y}_2,\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa}) &= \frac{[f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})]_{\mathbf{Y}_2=\mathbf{d}}}{[\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1]_{\mathbf{Y}_2=\mathbf{d}}},\\
&\propto \exp\left[\boldsymbol{\alpha}'(\mathbf{H}\ \ \mathbf{B})\begin{pmatrix}\mathbf{Y}_1\\ \mathbf{d}\end{pmatrix} - \boldsymbol{\kappa}'\psi\left\{(\mathbf{H}\ \ \mathbf{B})\begin{pmatrix}\mathbf{Y}_1\\ \mathbf{d}\end{pmatrix} - \mathbf{V}^{-1}\boldsymbol{\mu}\right\}\right],\\
&\propto \exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{Y}_1 - \boldsymbol{\kappa}'\psi\left(\mathbf{H}\mathbf{Y}_1 + \mathbf{B}\mathbf{d} - \mathbf{V}^{-1}\boldsymbol{\mu}\right)\right\},\\
&= \exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{Y}_1 - \boldsymbol{\kappa}'\psi\left(\mathbf{H}\mathbf{Y}_1 - \boldsymbol{\mu}^*\right)\right\};\ \ \mathbf{Y}_1 \in \mathbb{R}^n,
\end{aligned}
$$

which proves the result. The normalizing constant can be found using a change of variables

$$
M = \frac{\det(\mathbf{V}^{-1})\left\{\prod_{i=1}^n K(\alpha_i,\kappa_i)\right\}\exp\left(\boldsymbol{\alpha}'\mathbf{B}\mathbf{d} - \boldsymbol{\alpha}'\mathbf{V}^{-1}\boldsymbol{\mu}\right)}{[\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1]_{\mathbf{Y}_2=\mathbf{d}}}. \tag{B.1}
$$

Although we do not find the expression of the integral $[\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1]_{\mathbf{Y}_2=\mathbf{d}}$, and consequently M, we know that M is non-zero and finite. To see this, let $\mathscr{N}_1 = \{\mathbf{Y}_2 : [\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1]_{\mathbf{Y}_2} = 0\}$; then, by the definition of the CM distribution for $\mathbf{Y} \in \mathscr{M}^n$ and $\mathbf{Y}_2 \in \mathscr{N}_1$

$$
f\left(\begin{bmatrix}\mathbf{Y}_1\\ \mathbf{Y}_2\end{bmatrix}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa}\right) > 0.
$$

Taking the integral with respect to $\mathbf{Y}_1$ on both sides of the inequality gives $0 > 0$, which is a false statement. Thus, we have that $[\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1]_{\mathbf{Y}_2=\mathbf{d}}$ is non-zero, and hence, M is finite. Similarly, let $\mathscr{N}_2 = \{\mathbf{Y}_2 : [\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1]_{\mathbf{Y}_2} = \infty\}$ be non-empty, and let $\mathscr{N}_2^c$ denote the set complement of $\mathscr{N}_2$. Then, if $\mathbf{w} \sim \mathrm{CM}(\mathbf{0}_{n,1},\mathbf{I}_n,\boldsymbol{\alpha},\boldsymbol{\kappa})$, a change of variables within the integral (see

Proposition 1) gives,

$$
\begin{aligned}
1 &= \int f\left(\mathbf{w}|\boldsymbol{\mu}=\mathbf{0}_{n,1},\mathbf{V}=\mathbf{I}_n,\boldsymbol{\alpha},\boldsymbol{\kappa}\right)d\mathbf{w} = \int f\left(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa}\right)d\mathbf{Y} = \int\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1 d\mathbf{Y}_2 \\
&= \int_{\mathscr{N}_2}\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1 d\mathbf{Y}_2 + \int_{\mathscr{N}_2^c}\int f(\mathbf{Y}|\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_1 d\mathbf{Y}_2 \\
&= \infty,
\end{aligned}
$$

which is a contradiction. Thus, we have that the conditional distribution of $\mathbf{Y}_1|\mathbf{Y}_2,\boldsymbol{\mu},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa}$ is proper.

Proof of Theorem 3: Consider the transformation $Q = \left(\frac{\psi''(0)}{\psi'(0)}\right)^{1/2}\alpha^{1/2}W$, where W follows an unnormalized DY $\left(\alpha, \frac{\alpha}{\psi'(0)};\, \psi\right)$. Then we have that

$$
f(Q|\alpha,\kappa) \propto \exp\left[\left(\frac{\psi'(0)}{\psi''(0)}\right)^{1/2}\alpha^{1/2}Q - \frac{\alpha}{\psi'(0)}\,\psi\left\{\left(\frac{\psi'(0)}{\psi''(0)}\right)^{1/2}\alpha^{-1/2}Q\right\}\right],
$$

and using the Taylor Series expansion of $\psi(x)$ we have

$$
\begin{aligned}
& f(Q|\alpha,\kappa) \\
& \propto \exp\left[\left(\frac{\psi'(0)}{\psi''(0)}\right)^{1/2}\alpha^{1/2}Q\right. \\
& \left.-\frac{\alpha}{\psi'(0)}\left\{\psi'(0)\left(\frac{\psi'(0)}{\psi''(0)}\right)^{1/2}\alpha^{-1/2}Q + \psi''(0)\left(\frac{\psi'(0)}{\psi''(0)}\right)\alpha^{-1}\frac{Q^2}{2} + O\left(\frac{\psi'(0)^{3/2}}{\psi''(0)^{3/2}}\alpha^{-3/2}Q^3\right)\right\}\right],
\end{aligned}
$$

where "$O(\cdot)$" is the "Big-O" notation (e.g., see Lehmann, 1999, among others). Then, letting α go to infinity yields,

$$
\lim_{\alpha\to\infty} f(Q|\alpha,\kappa) \propto \exp\left(-\frac{Q^2}{2}\right) \propto \text{Normal}(0,1).
$$

Thus, Q converges in distribution to a standard normal distribution as α goes to infinity. Now suppose $\mathbf{w} = (w_1, \ldots, w_n)'$ follows an unnormalized CM $\left(\mathbf{0}_n, \alpha^{1/2}\mathbf{I}_n, \alpha\mathbf{J}_{n,1}, \frac{\alpha}{\psi'(0)}\mathbf{J}_{n,1};\, \psi\right)$. Then it follows from the result above that $\left(\frac{\alpha}{\psi'(0)}\right)^{1/2}\mathbf{w}$ converges to a standard multivariate Gaussian distribution. Now, define the transformation $\mathbf{Y} = \boldsymbol{\mu} + \mathbf{V}(\alpha^{1/2}\mathbf{w})$. It follows from Theorem 5.1.8 of Lehmann (1999), and the fact that $\frac{\alpha^{1/2}}{\psi'(0)}\mathbf{w}$ converges to a standard Gaussian distribution, that $\mathbf{Y}$ converges in distribution to a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{V}\mathbf{V}'$.

Proof of Theorem 4: In the main-text we stated that the CM distribution is Kolmogorov consistent. We now prove that result. To prove Kolmogorov consistency we need to show the following:

1. For any finite set $\{1, \ldots, n\}$ and for a generic permutation $\{i_1, \ldots, i_n\}$, we have $f\left\{(Y_{i_1}, \ldots, Y_{i_n})' \,|\mathbf{c}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}\right\} = f\left\{(Y_1, \ldots, Y_n)' \,|\mathbf{c}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}\right\}$.

2. Let $\{j_1, \ldots, j_n\}$ be a generic permutation of $\{1, \ldots, n\}$ and let $m < n$. Then we have that the marginal density $f\left(Y_{j_1}, \ldots, Y_{j_m}|\mathbf{c}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}\right) = \int_{\mathscr{M}} \ldots \int_{\mathscr{M}} f\left(Y_1, \ldots, Y_n|\mathbf{c}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}\right) dY_{j_{m+1}} \ldots dY_{j_n}$ exists.

Note that the conditions of the Kolmogorov extension theorem do not require that probability density functions exist. However, from Proposition 1(i), we have an expression of the pdf of $\mathbf{Y}$, which will be useful in our proof; hence, we can simplify the conditions of the Kolmogorov extension theorem to the setting where the joint probability density function exists.

For Item 1, define a $n \times n$ permutation matrix $\mathbf{\Pi}$ such that $(Y_{i_1}, \ldots, Y_{i_n})' \equiv \mathbf{Y}_\pi = \mathbf{\Pi}\mathbf{Y}$. Recall that permutation matrices have the following properties: $\mathbf{\Pi}\mathbf{\Pi}' = \mathbf{\Pi}'\mathbf{\Pi} = \mathbf{I}_n$ and $\mathbf{\Pi}^{-1} = \mathbf{\Pi}'$. From Equation (7) of the main text we have that,

$$\mathbf{Y}_\pi = \mathbf{\Pi}\boldsymbol{c} + \mathbf{\Pi}\mathbf{V}\mathbf{w}, \tag{B.2}$$

where $\mathbf{w}$ consist of mutually independent DY random variables with respective shape and scale parameters organized into the n-dimensional vectors $\boldsymbol{\alpha}$ and $\boldsymbol{\kappa}$.

From Proposition 1(i),

$$
\begin{aligned}
&f(\mathbf{Y}_\pi|\boldsymbol{c},\boldsymbol{V},\boldsymbol{\alpha},\boldsymbol{\kappa})\\
&=\det(\mathbf{V}^{-1})\left(\prod_{i=1}^{n}K(\alpha_i,\kappa_i)\right)\exp[\boldsymbol{\alpha}'\boldsymbol{V}^{-1}\boldsymbol{\Pi}'(\mathbf{Y}_\pi-\boldsymbol{\Pi}\mathbf{c})-\boldsymbol{\kappa}'\psi\{\boldsymbol{V}^{-1}\boldsymbol{\Pi}'(\mathbf{Y}_\pi-\boldsymbol{\Pi}\mathbf{c})\}]\\
&=f(\mathbf{Y}|\boldsymbol{c},\boldsymbol{V},\boldsymbol{\alpha},\boldsymbol{\kappa}),
\end{aligned}
$$

where the last equality holds since $\boldsymbol{\Pi}'\boldsymbol{\Pi}=\mathbf{I}_n$ and $\boldsymbol{\Pi}'\mathbf{Y}_\pi=\boldsymbol{\Pi}'\boldsymbol{\Pi}\mathbf{Y}=\mathbf{Y}$. Thus, permutation holds.

We now need to show that the marginal distribution stays the same regardless of what the "extended" proper joint distribution is defined as. Without loss of generality (due to Item 1) set $\mathbf{P}_m'=[\mathbf{I}_m,\mathbf{0}_{m,n-m}]$ where $\mathbf{0}_{m,n-m}$ is a $m\times(n-m)$ matrix of zeros. Then define $\mathbf{V}=[\mathbf{M},\mathbf{C}]'$, $\mathbf{M}'$ to be a $m\times n$ is a real-valued matrix, $\mathbf{C}$ to be any $n\times(n-m)$ real-valued matrix such that $\mathbf{V}$ is invertible, $\mathbf{Y}\in\mathbb{R}^n$, $\mathbf{Y}=\mathbf{c}+\mathbf{V}\mathbf{w}=(\mathbf{Y}_1',\mathbf{Y}_2')'$, $\mathbf{Y}_1'$ is m-dimensional, and $\mathbf{Y}_2$ is $(n-m)$-dimensional.

The joint distribution is determined by $\mathbf{V}=[\mathbf{M},\mathbf{C}]$, $\mathbf{c}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\kappa}$. Thus, we need to show that joint probability density functions with different values of $\mathbf{C}$ and $\mathbf{c}$ results in the *same* marginal probability density function upon integrating the joint probability density function. Let $\mathbf{C}_1$ denote a generic real-valued matrix such that $\mathbf{V}_1=[\mathbf{M},\mathbf{C}_1]'$ is invertible and $\mathbf{C}\neq\mathbf{C}_1$. Let $\mathbf{c}_1\in\mathbb{R}^n$. Define $\mathbf{Y}^{(1)}=\mathbf{c}_1+\mathbf{V}_1\mathbf{w}=(\mathbf{Y}_1^{(1)\prime},\mathbf{Y}_2^{(1)\prime})'$, where $\mathbf{Y}_1^{(1)}$ is m-dimensional, and $\mathbf{Y}_2^{(1)}$ is $(n-m)$-dimensional. Then we have that

$$
f(\mathbf{Y}_1^{(1)}|\mathbf{c}_1,\mathbf{V}_1,\boldsymbol{\alpha},\boldsymbol{\kappa})=\int f(\mathbf{Y}^{(1)}|\mathbf{c}_1,\mathbf{V}_1,\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{q}_2^{(1)}, \tag{B.3}
$$

and a change of variables $\mathbf{Y}=\mathbf{V}\mathbf{V}_1^{-1}\mathbf{Y}^{(1)}-\mathbf{V}\mathbf{V}_1^{-1}\mathbf{c}_1+\mathbf{c}$ within (B.3) gives,

$$
\begin{aligned}
&f(\mathbf{Y}_1^{(1)}|\mathbf{c}_1,\mathbf{V}_1,\boldsymbol{\alpha},\boldsymbol{\kappa})=\int f(\mathbf{Y}^{(1)}|\mathbf{c}_1,\mathbf{V}_1,\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_2^{(1)}=\int f(\mathbf{Y}|\mathbf{c},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa})d\mathbf{Y}_2\\
&=f(\mathbf{Y}_1|\mathbf{c},\mathbf{V},\boldsymbol{\alpha},\boldsymbol{\kappa}).
\end{aligned}
$$

This completes the proof.

Proof of Theorem 5: The distribution of $\mathbf{q}$ is equal to $\mathrm{CM}_c(\mathbf{c} = -\mathbf{B}\mathbf{q}_2 + \boldsymbol{\mu}, \mathbf{V} = (\mathbf{H},\mathbf{B})^{-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ $h(\mathbf{q}_2|\boldsymbol{\mu}, \mathbf{V} = (\mathbf{H},\mathbf{B})^{-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$, where recall we have reparameterized $\mathbf{c} = -\mathbf{B}\mathbf{q}_2 + \boldsymbol{\mu}$ and $f(\mathbf{q}_2|\boldsymbol{\mu}, \mathbf{V} = (\mathbf{H},\mathbf{B})^{-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \propto 1$. Thus,

$$
\begin{aligned}
f(\mathbf{q}_1, \mathbf{q}_2|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) &\propto \exp\left\{\boldsymbol{\alpha}'\mathbf{H}\mathbf{q}_1 + \boldsymbol{\alpha}'\mathbf{B}\mathbf{q}_2 - \boldsymbol{\alpha}'\boldsymbol{\mu} - \boldsymbol{\kappa}'\psi\left(\mathbf{H}\mathbf{q}_1 + \mathbf{B}\mathbf{q}_2 - \boldsymbol{\mu}\right)\right\} \\
&= \exp\left\{\boldsymbol{\alpha}'\mathbf{V}^{-1}(\mathbf{q} - \mathbf{V}\boldsymbol{\mu}) - \boldsymbol{\kappa}'\psi\left(\mathbf{V}^{-1}(\mathbf{q} - \mathbf{V}\boldsymbol{\mu}\right)\right\}.
\end{aligned}
$$

Integrating out $\mathbf{q}_2$ we obtain,

$$
f(\mathbf{q}_1|\boldsymbol{\mu}, \mathbf{V}, \boldsymbol{\alpha}, \boldsymbol{\kappa}) \propto \int \exp\left\{\boldsymbol{\alpha}'\mathbf{V}^{-1}(\mathbf{q} - \mathbf{V}\boldsymbol{\mu}) - \boldsymbol{\kappa}'\psi\left(\mathbf{V}^{-1}(\mathbf{q} - \mathbf{V}\boldsymbol{\mu}\right)\right\} d\mathbf{q}_2. \tag{B.4}
$$

Thus, $\mathbf{q}_1$ is the marginal random vector associated with $\mathrm{CM}(\mathbf{V}\boldsymbol{\mu}, \mathbf{V} = (\mathbf{H},\mathbf{B})^{-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$. Thus, we are left to show that $\mathbf{q}_1 = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{w}$ is a sample from this marginal distribution.

Denote the QR decomposition of $\mathbf{H} = \mathbf{Q}\mathbf{R}$, where the $M \times r$ matrix $\mathbf{Q}$ satisfies $\mathbf{Q}'\mathbf{Q} = \mathbf{I}_r$ and $\mathbf{R}$ is a $r \times r$ upper triangular matrix. Now recall the definition of the $M \times (M-r)$ matrix $\mathbf{B}$, which satisfies $\mathbf{B}'\mathbf{B} = \mathbf{I}_{M-r}$ and $\mathbf{B}'\mathbf{Q} = \mathbf{0}_{M-r,r}$. Then $\mathbf{V}^{-1}$ can be written as

$$
\mathbf{V}^{-1} = \begin{bmatrix} \mathbf{Q} & \mathbf{B} \end{bmatrix} \begin{bmatrix} \mathbf{R} & \mathbf{0}_{r,M-r} \\ \mathbf{0}_{M-r,r} & \mathbf{I}_{M-r}, \end{bmatrix}. \tag{B.5}
$$

It follows that

$$
\mathbf{V} = \begin{bmatrix} \mathbf{R}^{-1} & \mathbf{0}_{r,M-r} \\ \mathbf{0}_{M-r,r} & \mathbf{I}_{M-r}, \end{bmatrix} \begin{bmatrix} \mathbf{Q}' \\ \mathbf{B}', \end{bmatrix} = \begin{bmatrix} (\mathbf{H}^{*\prime}\mathbf{H}^{*})^{-1}\mathbf{H}^{*\prime} \\ \mathbf{B}' \end{bmatrix},
$$

where the last equality in the above can be verified by substituting $\mathbf{H} = \mathbf{Q}\mathbf{R}$ into $(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'$.Then, $\mathbf{q}$ is distributed according to $\mathrm{CM}(\mathbf{V}\boldsymbol{\mu}, \mathbf{V} = (\mathbf{H},\mathbf{B})^{-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ and can be written as

$$
\begin{bmatrix} \mathbf{q}_1 \\ \mathbf{q}_2 \end{bmatrix} = \begin{bmatrix} (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{w} \\ \mathbf{B}'\mathbf{w} \end{bmatrix}, \tag{B.6}
$$

where the n-dimensional random vector $\mathbf{w}$ is distributed according to $\text{CM}(\boldsymbol{\mu}, \mathbf{V} = \mathbf{I}_M, \boldsymbol{\alpha}, \boldsymbol{\kappa})$. Multiplying both sides of (B.6) by $[\mathbf{I}_r, \mathbf{0}_{r,M-r}]$ we have

$$\mathbf{q}_1 = (\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{w}, \tag{B.7}$$

and hence the distribution associated with $(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\mathbf{w}$ is the marginal distribution associated with $\text{CM}(\mathbf{V}\boldsymbol{\mu}, \mathbf{V} = (\mathbf{H}, \mathbf{B})^{-1}, \boldsymbol{\alpha}, \boldsymbol{\kappa})$ as desired.

Appendix C: The Collapsed Gibbs Sampler

Adding a small number to the data to avoid zero counts changes the priors in the LCM stated in Section 5, and results in a considerable amount of bookkeeping. In Appendices C.i and C.ii, we give these technical details. While the model structure is complicated, it's implementation is computationally straightforward. In Appendix C.iii, we outline the steps involved for the collapsed Gibbs sampler for the model in Appendix C.i.

Appendix C.i: Adding a Small Number to Zero Counts

The version of the LCM model that allows for zero counts, can be written as the product of the following conditional and marginal distributions:

$$\text{Data Model}: Z_i|\boldsymbol{\beta},\boldsymbol{\eta},\xi_i,b \overset{\text{ind}}{\sim} \text{EF}\left(\mathbf{x}_i'\boldsymbol{\beta}+\boldsymbol{\phi}_i'\boldsymbol{\eta}+\xi_i+\mathbf{b}_{\beta,i}'\mathbf{q}_\beta+\mathbf{b}_{\eta,i}'\mathbf{q}_\eta+\mathbf{b}_{\xi,i}'\mathbf{q}_\xi;\,\psi_j\right)\zeta_\beta(\mathbf{q}_\beta)\zeta_\eta(\mathbf{q}_\eta)\zeta_\xi(\mathbf{q}_\xi);$$

$$\text{Process Model 1}: \boldsymbol{\eta}|\mathbf{V},\alpha_\eta,\kappa_\eta \sim \text{CM}_\text{c}\left(-\mathbf{B}_\eta\mathbf{q}_\eta,\mathbf{M},\boldsymbol{\alpha}_\eta,\boldsymbol{\kappa}_\eta;\,\psi_k\right);$$

$$\text{Process Model 2}: \boldsymbol{\xi}|\boldsymbol{\alpha}_\xi,\boldsymbol{\kappa}_\xi \sim \text{CM}_\text{c}\left(-\mathbf{B}_\xi\mathbf{q}_\xi,\mathbf{M}_\xi,\boldsymbol{\alpha}_\xi,\boldsymbol{\kappa}_\xi;\,\psi_k\right);$$

$$\text{Parameter Model 1}: b|\alpha_b,\kappa_b \sim \text{CM}(0,1,\alpha_b,\kappa_b;\,\psi_k)I(b>0)$$

$$\text{Parameter Model 2}: \boldsymbol{\beta}|\alpha_\beta,\kappa_\beta \sim \text{CM}_\text{c}\left(-\mathbf{B}_\beta\mathbf{q}_\beta,\mathbf{M}_\beta,\boldsymbol{\alpha}_\beta,\boldsymbol{\kappa}_\beta;\,\psi_k\right)$$

$$\text{Parameter Model 3}: c|\alpha_c,\kappa_c \sim \text{CM}(0,1,\alpha_c,\kappa_c;\,\psi_k);$$

$$\text{Parameter Model 4}: c_\xi|\alpha_c,\kappa_c \sim \text{CM}(0,1,\alpha_c,\kappa_c;\,\psi_k);$$

$$\text{Parameter Model 5}: c_\beta|\alpha_c,\kappa_c \sim \text{CM}(0,1,\alpha_c,\kappa_c;\,\psi_k);$$

$$\text{Parameter Model 6}: \mathbf{v}_i \overset{\text{ind}}{\sim} \text{CM}(\mathbf{0},\sigma_v\mathbf{I}_{i-1},\alpha_v\mathbf{J}_{i-1,1},\kappa_v\mathbf{J}_{i-1,1};\,\psi_k);\, i=2,\ldots,r,k=1,2,3,4;$$

$$\text{Parameter Model 7}: f(\alpha_\beta,\kappa_\beta|\gamma_{\beta,1},\gamma_{\beta,2},\rho_\beta) \propto \exp\left[\gamma_{\beta,1}\alpha_\beta+\gamma_{\beta,2}\kappa_\eta-\rho_\beta\log\left\{\frac{1}{K\left(\alpha_\beta,\kappa_\eta\right)}\right\}\right];$$

$$\text{Parameter Model 8}: f(\alpha_\eta,\kappa_\eta|\gamma_{\eta,1},\gamma_{\eta,2},\rho_\eta) \propto \exp\left[\gamma_{\eta,1}\alpha_{\eta,m}+\gamma_{\eta,2}\kappa_{\eta,m}-\rho_\eta\log\left\{\frac{1}{K\left(\alpha_{\eta,m},\kappa_{\eta,m}\right)}\right\}\right];$$

$$\text{Parameter Model 9}: f(\alpha_\xi,\kappa_\xi|\gamma_{\xi,1},\gamma_{\xi,2},\rho_\xi) \propto \exp\left[\gamma_{\xi,1}\alpha_\xi+\gamma_{\xi,2}\kappa_\xi-\rho_\xi\log\left\{\frac{1}{K\left(\alpha_\xi,\kappa_\xi\right)}\right\}\right];$$

$$\text{Parameter Model 10}: f(\alpha_v,\kappa_v|\gamma_{v,1},\gamma_{v,2},\rho_v) \propto \exp\left[\gamma_{\beta,1}\alpha_\beta+\gamma_{\beta,2}\kappa_\eta-\rho_\beta\log\left\{\frac{1}{K\left(\alpha_\beta,\kappa_\beta\right)}\right\}\right];$$

$$\text{Parameter Model 11}: f(\mathbf{q}_\beta)=1;$$

$$\text{Parameter Model 12}: f(\mathbf{q}_\eta)=1;$$

$$\text{Parameter Model 13}: f(\mathbf{q}_\xi)=1;$$

$$\text{Parameter Model 14}: f(\mathbf{q}_{v,i})=1; \quad i=1,\ldots,n,j=1,2,3,4,,$$

where ψ_j and ψ_k (for $j,k = 1,\ldots,4$) are defined in Table 1 and the elements of n-dimensional vector $\mathbf{Z} \equiv (Z_1,\ldots,Z_n)'$ represent data that can be reasonably modeled using a member from the natural exponential family. Additionally for each i, $\mathbf{x}_i$ is a known p-dimensional vector of covariates, $\boldsymbol{\beta} = (\beta_1,\ldots,\beta_p)' \in \mathbb{R}^p$ is an unknown vector interpreted as fixed effects, $\boldsymbol{\phi}_i$ is a known r-dimensional real-valued vector (see Section 3 for examples), and the r-dimensional vector $\boldsymbol{\eta} = (\eta_1,\ldots,\eta_r)'$ and n-dimensional vector $\boldsymbol{\xi} \equiv (\xi_1,\ldots,\xi_n)'$ are interpreted as real-valued random effects. The hyperparameters and variance parameters are as follows: define the $(n+p)$-dimensional vector $\boldsymbol{\alpha}_\beta = (\varepsilon_\alpha,\ldots,\varepsilon_\alpha,\alpha_{\beta,1},\ldots,\alpha_{\beta,p})'$, the $(n+r)$-dimensional vector $\boldsymbol{\alpha}_\eta = (\varepsilon_\alpha,\ldots,\varepsilon_\alpha,\alpha_{\eta,1},\ldots,\alpha_{\eta,r})'$, the $(2n)$-dimensional vector $\boldsymbol{\alpha}_\xi = (\varepsilon_\alpha,\ldots,\varepsilon_\alpha,\alpha_{\xi,1},\ldots,\alpha_{\xi,n})'$, the $(n+p)$-dimensional vector $\boldsymbol{\kappa}_\beta = (\varepsilon_{\kappa,1},\ldots,\varepsilon_{\kappa,n},\kappa_{\beta,1},\ldots,\kappa_{\beta,p})'$, the $(n+r)$-dimensional vector $\boldsymbol{\kappa}_\eta = (\varepsilon_{\kappa,1},\ldots,\varepsilon_{\kappa,n},\kappa_{\eta,1},\ldots,\kappa_{\eta,r})'$, the $2n$-dimensional vector $\boldsymbol{\kappa}_\xi = (\varepsilon_{\kappa,1},\ldots,\varepsilon_{\kappa,n},\kappa_{\xi,1},\ldots,\kappa_{\xi,n})'$, the $(n+p)\times p$ real-valued matrix $\mathbf{M}_\beta = (\mathbf{X}',\mathbf{V}_\beta')'$, the $(n+r)\times r$ real-valued matrix $\mathbf{M} = (\boldsymbol{\Phi}',\mathbf{V}_\eta')'$, the $(2n)\times n$ real-valued matrix $\mathbf{M}_\xi = (\mathbf{I}_n,\mathbf{V}_\xi')'$, $\mathbf{V}_\beta \in \mathbb{R}^p \times \mathbb{R}^p$, $\mathbf{V}_\eta \in \mathbb{R}^r \times \mathbb{R}^r$, and $\mathbf{V}_\xi \in \mathbb{R}^n \times \mathbb{R}^n$, where to ensure propriety (see Section 2.5) $\alpha_{\beta,i}/\kappa_{\beta,i} \in \mathscr{Y}$, $\alpha_{\eta,j}/\kappa_{\eta,j} \in \mathscr{Y}$, $\alpha_{\xi,k}/\kappa_{\xi,k} \in \mathscr{Y}$, $\kappa_{\beta,i} > 0$, $\kappa_{\eta,j} > 0$, and $\kappa_{\xi,k} > 0$; $i = 1,\ldots,p$, $j = 1,\ldots,r$, $k = 1,\ldots,n$.

We have additionally assumed that $\alpha_{\beta,i} \equiv \alpha_\beta$, $\alpha_{\eta,i} \equiv \alpha_\eta$, $\alpha_{\xi,i} \equiv \alpha_\xi$, $\kappa_{\beta,i} \equiv \kappa_\beta$, $\kappa_{\eta,i} \equiv \kappa_\eta$, and $\kappa_{\xi,i} \equiv \kappa_\xi$. Using Theorem 3 from the main text, we argue that large values of α_c, $\alpha_{\beta,c}$, $\alpha_{\xi,c}$, κ_b, κ_c, $\kappa_{\beta,c}$, $\kappa_{\xi,c}$, and κ_b imply a roughly normal prior on c, c_β, and c_ξ. Also, in our implementation we have assumed that $\mathbf{V}_\beta = \mathbf{I}_p$ and $\mathbf{V}_\xi = \mathbf{I}_n$, and that $\mathbf{V}_\eta$ is a lower unit triangular matrix with i-th row $\mathbf{v}_i$.

There are two specifications of the vectors $\mathbf{b}_{\beta,i}$, $\mathbf{b}_{\eta,i}$, and $\mathbf{b}_{\xi,i}$. The first specification involves defining a real-valued $n\times n$ matrix $\mathbf{B}_{\beta,1} = (\mathbf{b}_{\beta,1}',\ldots,\mathbf{b}_{\beta,n}')'$, $n\times n$ matrix $\mathbf{B}_{\eta,1} = (\mathbf{b}_{\eta,1}',\ldots,\mathbf{b}_{\eta,n}')'$, and $n\times n$ matrix $\mathbf{B}_{\xi,1} = (\mathbf{b}_{\xi,1}',\ldots,\mathbf{b}_{\xi,n}')'$. Thus, in this setting $\mathbf{q}_\beta$ is n-dimensional, $\mathbf{q}_\eta$ is n-dimensional, and $\mathbf{q}_\xi$ is n-dimensional. The second specification, increases the row and column dimensions, and involves defining a real-valued $n\times(2n)$ matrix $\mathbf{B}_{\beta,1} = (\mathbf{b}_{\beta,1}',\ldots,\mathbf{b}_{\beta,n}')'$, $n\times(2n)$ matrix $\mathbf{B}_{\eta,1} = (\mathbf{b}_{\eta,1}',\ldots,\mathbf{b}_{\eta,n}')'$, and $n\times(2n)$ matrix $\mathbf{B}_{\xi,1} = (\mathbf{b}_{\xi,1}',\ldots,\mathbf{b}_{\xi,n}')'$. In this setting $\mathbf{q}_\beta$ is $(2n)$-dimensional, $\mathbf{q}_\eta$ is $(2n)$-dimensional, and $\mathbf{q}_\xi$ is $2n$-dimensional. The exact specifications of $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, and $\mathbf{B}_{\xi,1}$, will be given in Appendix C.ii. The random vector $\mathbf{q}_{v,i}$ is i-dimensional.

In a similar manner there are two specifications of $\mathbf{B}_\beta$, $\mathbf{B}_\eta$, and $\mathbf{B}_\xi$. In the first setting, $\mathbf{B}_\beta$ has dimensions $(n+p)\times n$, $\mathbf{B}_\eta$ has dimension $(n+r)\times n$, and $\mathbf{B}_\xi$ has dimension $(2n)\times n$. Additionally, the first n rows of $\mathbf{B}_\beta$, $\mathbf{B}_\eta$, and $\mathbf{B}_\xi$ are defined to be $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, and $\mathbf{B}_{\xi,1}$, respectively. In the second setting, $\mathbf{B}_\beta$ has dimensions $(n+p)\times(2n)$, $\mathbf{B}_\eta$ has dimension $(n+r)\times(2n)$, and $\mathbf{B}_\xi$ has dimension $(2n)\times(2n)$. The exact specifications of $\mathbf{B}_\beta$, $\mathbf{B}_\eta$, and $\mathbf{B}_\xi$, will be given in Appendix C.ii.

The functions $\zeta_\beta : \mathbb{R}^{n+p} \to \mathbb{R}$, $\zeta_\eta : \mathbb{R}^{n+r} \to \mathbb{R}$, and $\zeta_\xi : \mathbb{R}^{2n} \to \mathbb{R}$ are defined in Appendix C.ii, and have the property that $\zeta_\beta(\mathbf{0}_{a,1}) = 1$, $\zeta_\eta(\mathbf{0}_{a,1}) = 1$, and $\zeta_\xi(\mathbf{0}_{a,1}) = 1$, where $a = n$ or $2n$ depending on the specifications of $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, $\mathbf{B}_{\xi,1}$, $\mathbf{B}_\beta$, $\mathbf{B}_\eta$, and $\mathbf{B}_\xi$. These functions are needed so that ε_α and $\varepsilon_{\kappa,i}$ can be introduced and a Collapsed Gibbs sampler, similar to the one outlined in the Pseudo-Code in the main text, can be used. Recall, the values of $\varepsilon_\alpha > 0$ and $\varepsilon_{\kappa,i} > 0$ are needed to account for the case where Z_i is equal to a boundary value on it's support (e.g., a zero Poisson count). Other solutions to this boundary value problem exist in the Poisson setting (Bradley et al., 2018), however we have found more consistent results using the approach in this paper. We perform inference using samples from the distribution of $\boldsymbol{\beta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$ given the data $\mathbf{Z}$ and the events $\mathbf{q}_\beta = \mathbf{0}_{a,1}$, $\mathbf{q}_\eta = \mathbf{0}_{a,1}$, $\mathbf{q}_\xi = \mathbf{0}_{a,1}$, and $\mathbf{q}_{v,i} = \mathbf{0}_{i,1}$. To simulate from this conditional distribution we implement a collapsed Gibbs sampler similar to the one outlined in Section 2.5 of the main text. The derivation of this collapsed Gibbs sampler is given in Appendix C.ii.

Appendix C.ii: Derivation of the Full-Conditional Distributions within a Collapsed Gibbs Sampler

We assume $j = k$ in Appendix C.i and drop the subscript on the log partition function ψ. Let the $n\times p$ matrix $\mathbf{X} \equiv (\mathbf{x}_1,\ldots,\mathbf{x}_n)'$, the $n\times r$ matrix $\boldsymbol{\Phi} \equiv (\boldsymbol{\phi}_1,\ldots,\boldsymbol{\phi}_n)'$, and $\underset{\mathbf{Z}}{\propto}$ denotes the "proportional to as a function of **Z**" symbol. It follows that

$$f(\mathbf{Z}|\cdot,\mathbf{q}_\beta,\mathbf{q}_\eta=\mathbf{0}_{a,1},\mathbf{q}_\xi=\mathbf{0}_{a,1},\mathbf{q}_{v,i}=\mathbf{0}_{i,1})$$
$$\underset{\boldsymbol{\beta}}{\propto}\exp\left(\mathbf{Z}'\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}'\mathbf{B}_{\beta,1}\mathbf{q}_\beta-b\mathbf{J}_{n,1}'\psi\left(\mathbf{X}\boldsymbol{\beta}+\mathbf{B}_{\beta,1}\mathbf{q}_\beta+\boldsymbol{\Phi}\boldsymbol{\eta}+\boldsymbol{\xi}\right)\right)\zeta_\beta(\mathbf{q}_\beta)h \tag{C.1}$$
$$f(\mathbf{Z}|\cdot,\mathbf{q}_\beta=\mathbf{0}_{a,1},\mathbf{q}_\eta,\mathbf{q}_\xi=\mathbf{0}_{a,1},\mathbf{q}_{v,i}=\mathbf{0}_{i,1})$$
$$\underset{\boldsymbol{\eta}}{\propto}\exp\left(\mathbf{Z}'\boldsymbol{\Phi}\boldsymbol{\eta}+\mathbf{Z}'\mathbf{B}_{\eta,1}\mathbf{q}_\eta-b\mathbf{J}_{n,1}'\psi\left(\boldsymbol{\Phi}\boldsymbol{\eta}+\mathbf{B}_{\eta,1}\mathbf{q}_\eta+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\xi}\right)\right)\zeta_\eta(\mathbf{q}_\eta)h \tag{C.2}$$
$$f(\mathbf{Z}|\cdot,\mathbf{q}_\beta=\mathbf{0}_{a,1},\mathbf{q}_\eta=\mathbf{0}_{a,1},\mathbf{q}_\xi,\mathbf{q}_{v,i}=\mathbf{0}_{i,1})$$
$$\underset{\boldsymbol{\xi}}{\propto}\exp\left(\mathbf{Z}'\boldsymbol{\xi}+\mathbf{Z}'\mathbf{B}_{\xi,1}\mathbf{q}_\xi-b\mathbf{J}_{n,1}'\psi\left(\boldsymbol{\xi}+\mathbf{B}_{\xi,1}\mathbf{q}_\xi+\boldsymbol{\Phi}\boldsymbol{\eta}+\mathbf{X}\boldsymbol{\beta}\right)\right)\zeta_\xi(\mathbf{q}_\xi)h, \tag{C.3}$$

where $h=\left\{\prod_{i=1}^n I(\mathbf{x}_i'\boldsymbol{\beta}+\boldsymbol{\psi}_i'\boldsymbol{\eta}+\xi_i+\mathbf{b}_{\beta,i}'\mathbf{q}_\beta+\mathbf{b}_{\eta,i}'\mathbf{q}_\eta+\mathbf{b}_{\xi,i}'\mathbf{q}_\xi\in\mathscr{Y})\right\}$, and $a=n$ or $2n$ depending on the specifications of $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, $\mathbf{B}_{\xi,1}$, $\mathbf{B}_\beta$, $\mathbf{B}_\eta$, and $\mathbf{B}_\xi$. We have that

$$f(\boldsymbol{\beta},\mathbf{q}_\beta|\mathbf{V}_\beta,\boldsymbol{\alpha}_\beta,\boldsymbol{\kappa}_\beta,\mathbf{q}_\eta=\mathbf{0}_{a,1},\mathbf{q}_\xi=\mathbf{0}_{a,1},\mathbf{q}_{v,i}=\mathbf{0}_{i,1}) \tag{C.4}$$
$$\propto\exp\left\{\boldsymbol{\alpha}_\beta'\mathbf{M}_\beta\boldsymbol{\beta}+\boldsymbol{\alpha}_\beta'\mathbf{B}_\beta\mathbf{q}_\beta-\boldsymbol{\kappa}_\beta'\psi\left(\mathbf{M}_\beta\boldsymbol{\beta}+\mathbf{B}_\beta\mathbf{q}_\beta-c_\beta\mathbf{J}_{n+p,1}\right)\right\},$$
$$f(\boldsymbol{\eta},\mathbf{q}_\eta|\mathbf{V}_\eta,\boldsymbol{\alpha}_\eta,\boldsymbol{\kappa}_\eta,\mathbf{q}_\beta=\mathbf{0}_{a,1},\mathbf{q}_\xi=\mathbf{0}_{a,1},\mathbf{q}_{v,i}=\mathbf{0}_{i,1}) \tag{C.5}$$
$$\propto\exp\left\{\boldsymbol{\alpha}_\eta'\mathbf{M}\boldsymbol{\eta}+\boldsymbol{\alpha}_\eta'\mathbf{B}_\eta\mathbf{q}_\eta-\boldsymbol{\kappa}_\eta'\psi\left(\mathbf{M}\boldsymbol{\eta}+\mathbf{B}_\eta\mathbf{q}_\eta-c_\eta\mathbf{J}_{n+r,1}\right)\right\},$$
$$f(\boldsymbol{\xi},\mathbf{q}_\xi|\mathbf{V}_\xi,\boldsymbol{\alpha}_\xi,\boldsymbol{\kappa}_\xi,\mathbf{q}_\eta=\mathbf{0}_{a,1},\mathbf{q}_\beta=\mathbf{0}_{a,1},\mathbf{q}_{v,i}=\mathbf{0}_{i,1}) \tag{C.6}$$
$$\propto\exp\left\{\boldsymbol{\alpha}_\xi'\mathbf{M}_\xi\boldsymbol{\xi}+\boldsymbol{\alpha}_\xi'\mathbf{B}_\xi\mathbf{q}_\xi-\boldsymbol{\kappa}_\xi'\psi\left(\mathbf{M}_\xi\boldsymbol{\xi}+\mathbf{B}_\xi\mathbf{q}_\xi-c_\xi\mathbf{J}_{2n,1}\right)\right\}.$$

Using (C.1) and (C.4) we have that

$$
\begin{aligned}
& f(\boldsymbol{\beta}, \mathbf{q}_\beta | \cdot, \mathbf{q}_\eta = \mathbf{0}_{a,1}, \mathbf{q}_\xi = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1}) \underset{\boldsymbol{\beta}}{\propto} f(\mathbf{Z}|\cdot) f(\boldsymbol{\beta}|\mathbf{V}_\beta, \boldsymbol{\alpha}_\beta, \boldsymbol{\kappa}_\beta, c_\beta) f(\mathbf{q}_\beta) \\
& \underset{\boldsymbol{\beta}}{\propto} \exp\Big\{ \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \mathbf{Z}'\mathbf{B}_{\beta,1}\mathbf{q}_\beta + \boldsymbol{\alpha}_\beta' \mathbf{M}_\beta \boldsymbol{\beta} + \boldsymbol{\alpha}_\beta' \mathbf{B}_\beta \mathbf{q}_\beta \\
& - \boldsymbol{\kappa}_\beta' \psi\left(\mathbf{M}_\beta \boldsymbol{\beta} + \mathbf{B}_\beta \mathbf{q}_\beta - c_\beta \mathbf{J}_{n+p,1}\right) - b\mathbf{J}_{n,1}' \psi\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{B}_{\beta,1}\mathbf{q}_\beta + \boldsymbol{\Phi}\boldsymbol{\eta} + \boldsymbol{\xi}\right) \Big\} \zeta_\beta(\mathbf{q}_\beta) h \\
& = \exp\Big\{ \mathbf{Z}'\mathbf{X}\boldsymbol{\beta} + \varepsilon \mathbf{J}_{n,1}' \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\alpha}_{\beta,-\varepsilon}' \mathbf{V}_\beta^{-1} \boldsymbol{\beta} \\
& - \boldsymbol{\kappa}_\beta' \psi\left(\mathbf{M}_\beta \boldsymbol{\beta} + \mathbf{B}_\beta \mathbf{q}_\beta - c_\beta \mathbf{J}_{n+p,1}\right) - b\mathbf{J}_{n,1}' \psi\left(\mathbf{X}\boldsymbol{\beta} + \mathbf{B}_{\beta,1}\mathbf{q}_\beta + \boldsymbol{\Phi}\boldsymbol{\eta} + \boldsymbol{\xi}\right) \Big\} \zeta_\beta(\mathbf{q}_\beta) \omega_\beta(\mathbf{q}_\beta) h \\
& \propto \mathrm{CM_c} \Big\{ \boldsymbol{\mu}_\beta, \mathbf{V}_\beta^*, \boldsymbol{\alpha}_\beta^*, \boldsymbol{\kappa}_\beta^*; \psi \Big\} h,
\end{aligned}
$$

where

$$
\begin{aligned}
\omega_\beta(\mathbf{q}_\beta) &= \exp(\mathbf{Z}'\mathbf{B}_{\beta,1}\mathbf{q}_\beta + \boldsymbol{\alpha}_\beta' \mathbf{B}_\beta \mathbf{q}_\beta) \\
\zeta_\beta(\mathbf{q}_\beta) &= \frac{1}{\omega_\beta(\mathbf{q})} \exp(\boldsymbol{\alpha}_\beta^{*\prime} \mathbf{Q}_\beta \mathbf{q}_\beta),
\end{aligned}
$$

$\mathbf{V}_\beta^* = (\mathbf{H}_\beta, \mathbf{Q}_\beta)^{-1}$, $\mathbf{Q}_\beta$ is the null basis for $\mathbf{H}_\beta$, $\boldsymbol{\alpha}_{\beta,-\varepsilon} = \left(\alpha_{\beta,1}, \ldots, \alpha_{\beta,p}\right)'$, and $\boldsymbol{\mu}_\beta$, $\mathbf{H}_\beta$, $\boldsymbol{\alpha}_\beta^*$, and $\boldsymbol{\kappa}_\beta^*$ are defined in Table 2.

Recall from Appendix C.i there are two specifications of $\mathbf{B}_\beta$ and $\mathbf{B}_{\beta,1}$. When $\mathbf{H}_\beta$ is $(n+p) \times p$ (as defined in the first and third columns of Table 2), we use the first specification, and let $\mathbf{B}_{\beta,1}$ be the first n rows of $\mathbf{B}_\beta$, and $\mathbf{B}_\beta$ is set equal to the $(n+p) \times n$ matrix $\mathbf{Q}_\beta$. When $\mathbf{H}_\beta$ is $(2n+p) \times p$ (as defined in the second column of Table 2), we use the second specification of $\mathbf{B}_\beta$ and $\mathbf{B}_{\beta,1}$, and let the matrix $(\mathbf{B}_{\beta,1}', \mathbf{B}_\beta')'$ be set equal to the $(2n+p) \times 2n$ matrix $\mathbf{Q}_\beta$.

In a similar manner to Equations (15) through (17) of the main text, a sample from $f(\boldsymbol{\beta}|\cdot, \mathbf{q}_\eta = \mathbf{0}_{a,1}, \mathbf{q}_\xi = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1})$ can be easily obtained with,

$$
\boldsymbol{\beta} = (\mathbf{H}_\beta' \mathbf{H}_\beta)^{-1} \mathbf{H}_\beta' \boldsymbol{\mu}_\beta + (\mathbf{H}_\beta' \mathbf{H}_\beta)^{-1} \mathbf{H}_\beta' \mathbf{w}, \tag{C.7}
$$

where $\mathbf{w} \sim \mathrm{CM}(\mathbf{0}_{g,1}, \mathbf{I}_g, \boldsymbol{\alpha}_\beta^*, \boldsymbol{\kappa}_\beta^*)$, g is the number of rows in $\mathbf{H}_\beta$, and $a = n$ or $2n$ depending on the specifications of $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, $\mathbf{B}_{\xi,1}$, $\mathbf{B}_{\beta}$, $\mathbf{B}_{\eta}$, and $\mathbf{B}_{\xi}$.

We can find the full conditional distributions associated with $\boldsymbol{\eta}$ and $\mathbf{q}_{\eta}$, and $\boldsymbol{\xi}$ and $\mathbf{q}_{\xi}$ in a similar manner. Using (C.2) and (C.5) we have that

$$
\begin{aligned}
&f(\boldsymbol{\eta},\mathbf{q}_{\eta}|\cdot,\mathbf{q}_{\beta}=\mathbf{0}_{a,1},\mathbf{q}_{\xi}=\mathbf{0}_{a,1},\mathbf{q}_{v,i}=\mathbf{0}_{i,1}) \underset{\boldsymbol{\eta}}{\propto} f(\mathbf{Z}|\cdot)f(\boldsymbol{\eta}|\mathbf{V}_{\eta},\boldsymbol{\alpha}_{\eta},\boldsymbol{\kappa}_{\eta},c_{\eta})f(\mathbf{q}_{\eta})\\
&\underset{\boldsymbol{\eta}}{\propto} \exp\left\{\mathbf{Z}'\boldsymbol{\Phi}\boldsymbol{\eta}+\mathbf{Z}'\mathbf{B}_{\eta,1}\mathbf{q}_{\eta}+\boldsymbol{\alpha}_{\eta}'\mathbf{M}\boldsymbol{\eta}+\boldsymbol{\alpha}_{\eta}'\mathbf{B}_{\eta}\mathbf{q}_{\eta}\right.\\
&\left.-\boldsymbol{\kappa}_{\eta}'\psi\left(\mathbf{M}\boldsymbol{\eta}+\mathbf{B}_{\eta}\mathbf{q}_{\eta}-c_{\eta}\mathbf{J}_{n+r,1}\right)-b\mathbf{J}_{n,1}'\psi\left(\boldsymbol{\Phi}\boldsymbol{\eta}+\mathbf{B}_{\eta,1}\mathbf{q}_{\eta}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\xi}\right)\right\}\zeta_{\eta}(\mathbf{q}_{\eta})h\\
&=\exp\left\{\mathbf{Z}'\boldsymbol{\Phi}\boldsymbol{\eta}+\varepsilon\mathbf{J}_{n,1}'\boldsymbol{\Phi}\boldsymbol{\eta}+\boldsymbol{\alpha}_{\eta,-\varepsilon}'\mathbf{V}_{\eta}^{-1}\boldsymbol{\eta}\right.\\
&\left.-\boldsymbol{\kappa}_{\eta}'\psi\left(\mathbf{M}\boldsymbol{\eta}+\mathbf{B}_{\eta}\mathbf{q}_{\eta}-c_{\eta}\mathbf{J}_{n+r,1}\right)-b\mathbf{J}_{n,1}'\psi\left(\boldsymbol{\Phi}\boldsymbol{\eta}+\mathbf{B}_{\eta,1}\mathbf{q}_{\eta}+\mathbf{X}\boldsymbol{\beta}+\boldsymbol{\xi}\right)\right\}\zeta_{\eta}(\mathbf{q}_{\eta})\omega_{\eta}(\mathbf{q}_{\eta})h\\
&\propto \mathrm{CM_c}\left\{\boldsymbol{\mu}_{\beta},\mathbf{V}_{\beta}^{*},\boldsymbol{\alpha}_{\beta}^{*},\boldsymbol{\kappa}_{\beta}^{*};\,\psi\right\}h,
\end{aligned}
$$

where

$$
\begin{aligned}
\omega_{\eta}(\mathbf{q}_{\eta})&=\exp(\mathbf{Z}'\mathbf{B}_{\eta,1}\mathbf{q}_{\eta}+\boldsymbol{\alpha}_{\eta}'\mathbf{B}_{\eta}\mathbf{q}_{\eta})\\
\zeta_{\eta}(\mathbf{q}_{\eta})&=\frac{1}{\omega_{\eta}(\mathbf{q}_{\eta})}\exp(\boldsymbol{\alpha}_{\eta}^{*\prime}\mathbf{Q}_{\eta}\mathbf{q}_{\eta}),
\end{aligned}
$$

$\mathbf{V}_{\eta}^{*}=(\mathbf{H}_{\eta},\mathbf{Q}_{\eta})^{-1}$, $\mathbf{Q}_{\eta}$ is the null basis for $\mathbf{H}_{\eta}$, $\boldsymbol{\alpha}_{\eta,-\varepsilon}=\left(\alpha_{\eta,1},\ldots,\alpha_{\eta,p}\right)'$, and $\boldsymbol{\mu}_{\eta}$, $\mathbf{H}_{\eta}$, $\boldsymbol{\alpha}_{\eta}^{*}$, and $\boldsymbol{\kappa}_{\eta}^{*}$ are defined in Table 2.

Recall from Appendix C.i there are two specifications of $\mathbf{B}_{\eta}$ and $\mathbf{B}_{\eta,1}$. When $\mathbf{H}_{\eta}$ is $(n+r)\times r$ (as defined in the first and third columns of Table 2), we use the first specification, and let $\mathbf{B}_{\eta,1}$ be the first n rows of $\mathbf{B}_{\eta}$, and $\mathbf{B}_{\eta}$ is set equal to the $(n+r)\times n$ matrix $\mathbf{Q}_{\eta}$. When $\mathbf{H}_{\eta}$ is $(2n+r)\times r$ (as defined in the second column of Table 2), we use the second specification of $\mathbf{B}_{\eta}$ and $\mathbf{B}_{\eta,1}$, and let the matrix $(\mathbf{B}_{\eta,1}',\mathbf{B}_{\eta}')'$ be set equal to the $(2n+r)\times 2n$ matrix $\mathbf{Q}_{\eta}$.

In a similar manner to Equations (15) through (17) of the main text, a sample from $f(\boldsymbol{\eta}|\cdot,\mathbf{q}_{\beta}=\mathbf{0}_{a,1},\mathbf{q}_{\xi}=\mathbf{0}_{a,1})$ can be easily obtained with,

$$
\boldsymbol{\eta}=(\mathbf{H}_{\eta}'\mathbf{H}_{\eta})^{-1}\mathbf{H}_{\eta}'\boldsymbol{\mu}_{\eta}+(\mathbf{H}_{\eta}'\mathbf{H}_{\eta})^{-1}\mathbf{H}_{\eta}'\mathbf{w}, \tag{C.8}
$$

where $\mathbf{w} \sim \mathrm{CM}(\mathbf{0}_{g,1}, \mathbf{I}_g, \boldsymbol{\alpha}_\eta^*, \boldsymbol{\kappa}_\eta^*)$, g is the number of rows in $\mathbf{H}_\eta$, and $a = n$ or $2n$ depending on the specifications of $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, $\mathbf{B}_{\xi,1}$, $\mathbf{B}_\beta$, $\mathbf{B}_\eta$, and $\mathbf{B}_\xi$.

Using (C.3) and (C.6) we have that

$$
\begin{aligned}
&f(\boldsymbol{\xi}, \mathbf{q}_\xi | \cdot, \mathbf{q}_\beta = \mathbf{0}_{a,1}, \mathbf{q}_\eta = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1}) \underset{\boldsymbol{\xi}}{\propto} f(\mathbf{Z}|\cdot) f(\boldsymbol{\xi}|\mathbf{V}_\xi, \boldsymbol{\alpha}_\xi, \boldsymbol{\kappa}_\xi, c_\xi) f(\mathbf{q}_\xi) \\
&\underset{\boldsymbol{\xi}}{\propto} \exp\Big\{\mathbf{Z}'\boldsymbol{\xi} + \mathbf{Z}'\mathbf{B}_{\xi,1}\mathbf{q}_\xi + \boldsymbol{\alpha}_\xi'\mathbf{M}_\xi\boldsymbol{\xi} + \boldsymbol{\alpha}_\xi'\mathbf{B}_\xi\mathbf{q}_\xi \\
&-\boldsymbol{\kappa}_\xi'\psi\left(\mathbf{M}_\xi\boldsymbol{\xi} + \mathbf{B}_\xi\mathbf{q}_\xi - c_\xi\mathbf{J}_{2n,1}\right) - b\mathbf{J}_{n,1}'\psi\left(\boldsymbol{\xi} + \boldsymbol{\Phi}\boldsymbol{\eta} + \mathbf{B}_{\xi,1}\mathbf{q}_\xi + \mathbf{X}\boldsymbol{\beta}\right)\Big\} \zeta_\xi(\mathbf{q}_\xi) h \\
&= \exp\Big\{\mathbf{Z}'\boldsymbol{\xi} + \varepsilon\mathbf{J}_{n,1}'\boldsymbol{\xi} + \boldsymbol{\alpha}_{\xi,-\varepsilon}'\mathbf{V}_\xi^{-1}\boldsymbol{\xi} \\
&-\boldsymbol{\kappa}_\xi'\psi\left(\mathbf{M}_\xi\boldsymbol{\xi} + \mathbf{B}_\xi\mathbf{q}_\xi - c_\xi\mathbf{J}_{2n,1}\right) - b\mathbf{J}_{n,1}'\psi\left(\boldsymbol{\xi} + \boldsymbol{\Phi}\boldsymbol{\eta} + \mathbf{B}_{\xi,1}\mathbf{q}_\xi + \mathbf{X}\boldsymbol{\beta}\right)\Big\} \zeta_\xi(\mathbf{q}_\xi)\omega_\xi(\mathbf{q}_\xi) h \\
&\propto \mathrm{CM_c}\Big\{\boldsymbol{\mu}_\xi, \mathbf{V}_\xi^*, \boldsymbol{\alpha}_\xi^*, \boldsymbol{\kappa}_\xi^*;\, \psi\Big\} h,
\end{aligned}
$$

where

$$
\begin{aligned}
\omega_\xi(\mathbf{q}_\xi) &= \exp(\mathbf{Z}'\mathbf{B}_{\xi,1}\mathbf{q}_\xi + \boldsymbol{\alpha}_\xi'\mathbf{B}_\xi\mathbf{q}_\xi) \\
\zeta_\xi(\mathbf{q}_\xi) &= \frac{1}{\omega_\xi(\mathbf{q}_\xi)}\exp(\boldsymbol{\alpha}_\xi^{*\prime}\mathbf{Q}_\xi\mathbf{q}_\xi),
\end{aligned}
$$

$\mathbf{V}_\xi^* = (\mathbf{H}_\xi, \mathbf{Q}_\xi)^{-1}$, $\mathbf{Q}_\xi$ is the null basis for $\mathbf{H}_\xi$, $\boldsymbol{\alpha}_{\xi,-\varepsilon} = \left(\alpha_{\xi,1}, \ldots, \alpha_{\xi,p}\right)'$, and $\boldsymbol{\mu}_\xi$, $\mathbf{H}_\xi$, $\boldsymbol{\alpha}_\xi^*$, and $\boldsymbol{\kappa}_\xi^*$ are defined in Table 2.

Recall from Appendix C.i there are two specifications of $\mathbf{B}_\xi$ and $\mathbf{B}_{\xi,1}$. When $\mathbf{H}_\xi$ is $(2n) \times n$ (as defined in the first and third columns of Table 2), we use the first specification, and let $\mathbf{B}_{\xi,1}$ be the first n rows of $\mathbf{B}_\xi$, and $\mathbf{B}_\xi$ is set equal to the $(2n) \times n$ matrix $\mathbf{Q}_\xi$. When $\mathbf{H}_\xi$ is $(3n) \times n$ (as defined in the second column of Table 2), we use the second specification of $\mathbf{B}_\xi$ and $\mathbf{B}_{\xi,1}$, and let the matrix $(\mathbf{B}_{\xi,1}', \mathbf{B}_\xi')'$ be set equal to the $(3n) \times 2n$ matrix $\mathbf{Q}_\xi$.

In a similar manner to Equations (15) through (17) of the main text, a sample from $f(\boldsymbol{\xi}|\cdot, \mathbf{q}_\beta = \mathbf{0}_{a,1}, \mathbf{q}_\xi = \mathbf{0}_{a,1})$ can be easily obtained with,

$$
\boldsymbol{\xi} = (\mathbf{H}_\xi'\mathbf{H}_\xi)^{-1}\mathbf{H}_\xi'\boldsymbol{\mu}_\xi + (\mathbf{H}_\xi'\mathbf{H}_\xi)^{-1}\mathbf{H}_\xi'\mathbf{w}, \tag{C.9}
$$

where $\mathbf{w} \sim \mathrm{CM}(\mathbf{0}_{g,1}, \mathbf{I}_g, \boldsymbol{\alpha}_{\xi}^{*}, \boldsymbol{\kappa}_{\xi}^{*})$, g is the number of rows in $\mathbf{H}_{\xi}$, and $a = n$ or $2n$ depending on the specifications of $\mathbf{B}_{\beta,1}$, $\mathbf{B}_{\eta,1}$, $\mathbf{B}_{\xi,1}$, $\mathbf{B}_{\beta}$, $\mathbf{B}_{\eta}$, and $\mathbf{B}_{\xi}$.

If b is unknown (e.g., the negative-binomial distribution) a prior for b is introduced in Appendix C.i. The full-conditional distribution is given by,

$$
\begin{aligned}
& f(b|\cdot, \mathbf{q}_{\beta} = \mathbf{0}_{a,1}, \mathbf{q}_{\eta} = \mathbf{0}_{a,1}, \mathbf{q}_{\xi} = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1}) \\
& \propto K_{EF}(b, \mathbf{Z}) \exp\left\{\alpha_b b - \kappa_b \psi(b) - b\mathbf{J}_{n,1}' \psi(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\Phi}\boldsymbol{\eta} + \boldsymbol{\xi})\right\} I\left\{b > \max(\mathbf{Z}) + 1\right\},
\end{aligned}
$$

where $K_{EF}(b, \mathbf{Z})$ is the normalizing constant associated with the distribution of $\mathbf{Z}$, and $\max(\mathbf{Z})$ returns the maximum element of the vector $\mathbf{Z}$. A slice sampler can be used to generate from this full-conditional distribution.

The full-conditional distribution for c is given by,

$$
\begin{aligned}
& f(c|\cdot, \mathbf{q}_{\beta} = \mathbf{0}_{a,1}, \mathbf{q}_{\eta} = \mathbf{0}_{a,1}, \mathbf{q}_{\xi} = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1}) \\
& \underset{c}{\propto} f(c) f(\boldsymbol{\eta} | \mathbf{V}, \boldsymbol{\alpha}_{\beta}, \boldsymbol{\kappa}_{\beta}, c) \\
& \underset{c}{\propto} \exp\left\{\alpha_c c + \boldsymbol{\alpha}_{\eta}' \mathbf{M}\boldsymbol{\eta} - \boldsymbol{\kappa}_{\eta}' \psi(\mathbf{M}_{\eta}\boldsymbol{\eta} - c\mathbf{J}_{n+r,1}) - \kappa_c \psi(c)\right\} I(c \in \mathscr{Y}) h \\
& \propto \mathrm{CM_c}\left\{\boldsymbol{\mu}_c, \mathbf{H}_c^{*}, \boldsymbol{\alpha}_c^{*}, \boldsymbol{\kappa}_c^{*};\ \psi\right\} I(c \in \mathscr{Y}),
\end{aligned}
$$

where $\boldsymbol{\mu}_c$, $\mathbf{H}_c^{*}$, $\boldsymbol{\alpha}_c^{*}$, and $\boldsymbol{\kappa}_c^{*}$ are defined in Table 2. The full-conditional distributions for c_{ξ} and c_{β} are found in a similar way. That is,

$$
\begin{aligned}
f(c_{\beta}|\cdot, \mathbf{q}_{\beta} = \mathbf{0}_{a,1}, \mathbf{q}_{\eta} = \mathbf{0}_{a,1}, \mathbf{q}_{\xi} = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1}) \propto \mathrm{CM_c}\left\{\boldsymbol{\mu}_{c,\beta}, \mathbf{H}_{c,\beta}^{*}, \boldsymbol{\alpha}_{c,\beta}^{*}, \boldsymbol{\kappa}_{c,\beta}^{*};\ \psi\right\} I(c_{\beta} \in \mathscr{Y}) \\
f(c_{\xi}|\cdot, \mathbf{q}_{\beta} = \mathbf{0}_{a,1}, \mathbf{q}_{\eta} = \mathbf{0}_{a,1}, \mathbf{q}_{\xi} = \mathbf{0}_{a,1}, \mathbf{q}_{v,i} = \mathbf{0}_{i,1}) \propto \mathrm{CM_c}\left\{\boldsymbol{\mu}_{c,\xi}, \mathbf{H}_{c,\xi}^{*}, \boldsymbol{\alpha}_{c,\xi}^{*}, \boldsymbol{\kappa}_{c,\xi}^{*};\ \psi\right\} I(c_{\xi} \in \mathscr{Y}),
\end{aligned}
$$

where $\boldsymbol{\mu}_{c,\beta}$, $\mathbf{H}_{c,\beta}^{*}$, $\boldsymbol{\alpha}_{c,\beta}^{*}$, $\boldsymbol{\kappa}_{c,\beta}^{*}$, $\boldsymbol{\mu}_{c,\xi}$, $\mathbf{H}_{c,\xi}^{*}$, $\boldsymbol{\alpha}_{c,\xi}^{*}$, and $\boldsymbol{\kappa}_{c,\xi}^{*}$ are defined in Table 2. One could use the argument in Appendix A.ii to update c, c_{β}, and c_{ξ}. However, it is rather straightforward to update these parameters using the slice sampler.

Quantities Needed for Gibbs Sampling

No Boundary Adjustments	ψ_2	ψ_3
$\boldsymbol{\mu}_\beta = (-\boldsymbol{\eta}'\boldsymbol{\Phi}' - \boldsymbol{\xi}', c_\beta \mathbf{J}_{1,p})'$	$\boldsymbol{\mu}_\beta = (c_\beta \mathbf{J}_{1,n}, -\boldsymbol{\eta}'\boldsymbol{\Phi}' - \boldsymbol{\xi}', c_\beta \mathbf{J}_{1,p})'$	$\boldsymbol{\mu}_\beta = \mathbf{0}_{n+p,1}$
$\boldsymbol{\mu}_\eta = (-\boldsymbol{\beta}'\mathbf{X}' - \boldsymbol{\xi}', c\mathbf{J}_{1,r})'$	$\boldsymbol{\mu}_\eta = (c\mathbf{J}_{1,n}, -\boldsymbol{\beta}'\mathbf{X}' - \boldsymbol{\xi}', c\mathbf{J}_{1,r})'$	$\boldsymbol{\mu}_\eta = \mathbf{0}_{n+r,1}$
$\boldsymbol{\mu}_\xi = (-\boldsymbol{\beta}'\mathbf{X}' - \boldsymbol{\eta}'\boldsymbol{\Phi}', c_\xi \mathbf{J}_{1,n})'$	$\boldsymbol{\mu}_\xi = (c_\xi \mathbf{J}_{1,n}, -\boldsymbol{\beta}'\mathbf{X}' - \boldsymbol{\eta}'\boldsymbol{\Phi}', c_\xi \mathbf{J}_{1,n})'$	$\boldsymbol{\mu}_\xi = \mathbf{0}_{2n,1}$
$\boldsymbol{\mu}_{\gamma,i} = (\eta_i, \mathbf{0}_{1,i-1})'; \quad i = 2,\ldots,r$	$\boldsymbol{\mu}_{\gamma,i} = (\eta_i, \mathbf{0}_{1,i-1})'; \quad i = 2,\ldots,r$	$\boldsymbol{\mu}_{\gamma,i} = (\eta_i, \mathbf{0}_{1,i-1})'; \quad i = 2,\ldots,r$
$\boldsymbol{\mu}_c = (\boldsymbol{\eta}'\boldsymbol{\Phi}', 0)'$	$\boldsymbol{\mu}_c = (\boldsymbol{\eta}'\boldsymbol{\Phi}', 0)'$	$\boldsymbol{\mu}_c = (\boldsymbol{\eta}'\boldsymbol{\Phi}', 0)'$
$\boldsymbol{\mu}_{c,\xi} = (\boldsymbol{\xi}', 0)'$	$\boldsymbol{\mu}_{c,\xi} = (\boldsymbol{\xi}', 0)'$	$\boldsymbol{\mu}_{c,\xi} = (\boldsymbol{\xi}', 0)'$
$\boldsymbol{\mu}_{c,\beta} = (\boldsymbol{\beta}'\mathbf{X}', 0)'$	$\boldsymbol{\mu}_{c,\beta} = (\boldsymbol{\beta}'\mathbf{X}', 0)'$	$\boldsymbol{\mu}_{c,\beta} = (\boldsymbol{\beta}'\mathbf{X}', 0)'$
$\mathbf{H}_\beta = (\mathbf{X}', \mathbf{V}_\beta^{-1\prime})'$	$\mathbf{H}_\beta = (\mathbf{X}', \mathbf{X}', \mathbf{V}_\beta^{-1\prime})'$	$\mathbf{H}_\beta = (\mathbf{X}', \mathbf{V}_\beta^{-1\prime})'$
$\mathbf{H}_\eta = (\boldsymbol{\Phi}', \mathbf{V}_\eta^{-1\prime})'$	$\mathbf{H}_\eta = (\boldsymbol{\Phi}', \boldsymbol{\Phi}', \mathbf{V}_\eta^{-1\prime})'$	$\mathbf{H}_\eta = (\boldsymbol{\Phi}', \mathbf{V}_\eta^{-1\prime})'$
$\mathbf{H}_\xi = (\mathbf{I}_n, \mathbf{V}_\xi^{-1\prime})'$	$\mathbf{H}_\xi = (\mathbf{I}_n, \mathbf{I}_n, \mathbf{V}_\xi^{-1\prime})'$	$\mathbf{H}_\xi = (\mathbf{I}_n, \mathbf{V}_\xi^{-1\prime})'$
$\mathbf{H}_{\gamma,i} = \left\{(\eta_1,\ldots,\eta_{i-1},)', \mathbf{C}_i'\right\}'; \quad i = 2,\ldots,r$	$\mathbf{H}_{\gamma,i} = \left\{(\eta_1,\ldots,\eta_{i-1},)', \mathbf{C}_i'\right\}'; \quad i = 2,\ldots,r$	$\mathbf{H}_{\gamma,i} = \left\{(\eta_1,\ldots,\eta_{i-1},)', \mathbf{C}_i'\right\}'; \quad i = 2,\ldots,r$
$\mathbf{H}_c^* = -\mathbf{J}_{n+r,1}$	$\mathbf{H}_c^* = -\mathbf{J}_{n+r,1}$	$\mathbf{H}_c^* = -\mathbf{J}_{n+r,1}$
$\mathbf{H}_{c,\xi}^* = -\mathbf{J}_{2n,1}$	$\mathbf{H}_{c,\xi}^* = -\mathbf{J}_{2n,1}$	$\mathbf{H}_{c,\xi}^* = -\mathbf{J}_{2n,1}$
$\mathbf{H}_{c,\beta}^* = -\mathbf{J}_{n+p,1}$	$\mathbf{H}_{c,\beta}^* = -\mathbf{J}_{n+p,1}$	$\mathbf{H}_{c,\beta}^* = -\mathbf{J}_{n+p,1}$
$\boldsymbol{\alpha}_\beta^* = (\mathbf{Z}', \alpha_{\beta,1},\ldots,\alpha_{\beta,p})'$	$\boldsymbol{\alpha}_\beta^* = (\frac{1}{2}\mathbf{Z}' + \frac{\varepsilon_\alpha}{2}\mathbf{J}_{1,n}, \frac{1}{2}\mathbf{Z}' + \frac{\varepsilon_\alpha}{2}\mathbf{J}_{1,n}, \alpha_{\beta,1},\ldots,\alpha_{\beta,p})'$	$\boldsymbol{\alpha}_\beta^* = (\mathbf{Z}' + \varepsilon_\alpha \mathbf{J}_{1,n}, \alpha_{\beta,1},\ldots,\alpha_{\beta,p})'$
$\boldsymbol{\alpha}_\eta^* = (\mathbf{Z}', \alpha_{\eta,1},\ldots,\alpha_{\eta,r})'$	$\boldsymbol{\alpha}_\eta^* = (\frac{1}{2}\mathbf{Z}' + \frac{\varepsilon_\alpha}{2}\mathbf{J}_{1,n}, \frac{1}{2}\mathbf{Z}' + \frac{\varepsilon_\alpha}{2}\mathbf{J}_{1,n}, \alpha_{\eta,1},\ldots,\alpha_{\eta,r})'$	$\boldsymbol{\alpha}_\eta^* = (\mathbf{Z}' + \varepsilon_\alpha \mathbf{J}_{1,n}, \alpha_{\eta,1},\ldots,\alpha_{\eta,r})'$
$\boldsymbol{\alpha}_\xi^* = (\mathbf{Z}',, \alpha_{\xi,1},\ldots,\alpha_{\xi,n})'$	$\boldsymbol{\alpha}_\xi^* = (\frac{1}{2}\mathbf{Z}' + \frac{\varepsilon_\alpha}{2}\mathbf{J}_{1,n}, \frac{1}{2}\mathbf{Z}' + \frac{\varepsilon_\alpha}{2}\mathbf{J}_{1,n}, \alpha_{\xi,1},\ldots,\alpha_{\xi,n})'$	$\boldsymbol{\alpha}_\xi^* = (\mathbf{Z}' + \varepsilon_\alpha \mathbf{J}_{1,n}, \alpha_{\xi,1},\ldots,\alpha_{\xi,n})'$
$\boldsymbol{\alpha}_{\gamma,i} = (\alpha_{\eta,i}, \boldsymbol{\alpha}_i')'; \quad i = 2,\ldots,r$	$\boldsymbol{\alpha}_{\gamma,i} = (\alpha_{\eta,i}, \boldsymbol{\alpha}_i')'; \quad i = 2,\ldots,r$	$\boldsymbol{\alpha}_{\gamma,i} = (\alpha_{\eta,i}, \boldsymbol{\alpha}_i')'; \quad i = 2,\ldots,r$
$\boldsymbol{\alpha}_c^* = \alpha_c \mathbf{J}_{n+r,1}$	$\boldsymbol{\alpha}_c^* = \alpha_c \mathbf{J}_{n+r,1}$	$\boldsymbol{\alpha}_c^* = \alpha_c \mathbf{J}_{n+r,1}$
$\boldsymbol{\alpha}_{c,\xi}^* = \alpha_{c,\xi} \mathbf{J}_{2n,1}$	$\boldsymbol{\alpha}_{c,\xi}^* = \alpha_{c,\xi} \mathbf{J}_{2n,1}$	$\boldsymbol{\alpha}_{c,\xi}^* = \alpha_{c,\xi} \mathbf{J}_{2n,1}$
$\boldsymbol{\alpha}_{c,\beta}^* = \alpha_{c,\beta} \mathbf{J}_{n+p,1}$	$\boldsymbol{\alpha}_{c,\beta}^* = \alpha_{c,\beta} \mathbf{J}_{n+p,1}$	$\boldsymbol{\alpha}_{c,\beta}^* = \alpha_{c,\beta} \mathbf{J}_{n+p,1}$
$\boldsymbol{\kappa}_\beta^* = (\mathbf{b}', \kappa_{\beta,1},\ldots,\kappa_{\beta,p})'$	$\boldsymbol{\kappa}_\beta^* = (\mathbf{b}', \mathbf{b}', \kappa_{\beta,1},\ldots,\kappa_{\beta,p})'$	$\boldsymbol{\kappa}_\beta^* = (\exp(\boldsymbol{\eta}'\boldsymbol{\Phi}' + \boldsymbol{\xi}') + \boldsymbol{\varepsilon}_\kappa', \kappa_{\beta,1},\ldots,\kappa_{\beta,p})'$
$\boldsymbol{\kappa}_\eta^* = (\mathbf{b}', \kappa_{\eta,1},\ldots,\kappa_{\eta,r})'$	$\boldsymbol{\kappa}_\eta^* = (\mathbf{b}', \mathbf{b}', \kappa_{\eta,1},\ldots,\kappa_{\eta,r})'$	$\boldsymbol{\kappa}_\eta^* = (\exp(\boldsymbol{\beta}'\mathbf{X}' + \boldsymbol{\xi}') + \boldsymbol{\varepsilon}_\kappa', \kappa_{\eta,1},\ldots,\kappa_{\eta,r})'$
$\boldsymbol{\kappa}_\xi^* = (\mathbf{b}', \kappa_{\xi,1},\ldots,\kappa_{\xi,n})'$	$\boldsymbol{\kappa}_\xi^* = (\mathbf{b}', \mathbf{b}', \kappa_{\xi,1},\ldots,\kappa_{\xi,n})'$	$\boldsymbol{\kappa}_\xi^* = (\exp(\boldsymbol{\beta}'\mathbf{X}' + \boldsymbol{\eta}'\boldsymbol{\Phi}') + \boldsymbol{\varepsilon}_\kappa', \kappa_{\xi,1},\ldots,\kappa_{\xi,n})'$
$\boldsymbol{\kappa}_{\gamma,i} = (\kappa_{\eta,i}, \boldsymbol{\kappa}_i')'; \quad i = 2,\ldots,r$	$\boldsymbol{\kappa}_{\gamma,i} = (\kappa_{\eta,i}, \boldsymbol{\kappa}_i')'; \quad i = 2,\ldots,r$	$\boldsymbol{\kappa}_{\gamma,i} = (\kappa_{\eta,i}, \boldsymbol{\kappa}_i')'; \quad i = 2,\ldots,r$
$\boldsymbol{\kappa}_c^* = (\boldsymbol{\kappa}_\eta', \kappa_c)'$	$\boldsymbol{\kappa}_c^* = (\boldsymbol{\kappa}_\eta', \kappa_c)'$	$\boldsymbol{\kappa}_c^* = (\boldsymbol{\kappa}_\eta', \kappa_c)'$
$\boldsymbol{\kappa}_{c,\xi}^* = (\boldsymbol{\kappa}_\xi', \kappa_{c,\xi})'$	$\boldsymbol{\kappa}_{c,\xi}^* = (\boldsymbol{\kappa}_\xi', \kappa_{c,\xi})'$	$\boldsymbol{\kappa}_{c,\xi}^* = (\boldsymbol{\kappa}_\xi', \kappa_{c,\xi})'$
$\boldsymbol{\kappa}_{c,\beta}^* = (\boldsymbol{\kappa}_\beta', \kappa_{c,\beta})'$	$\boldsymbol{\kappa}_{c,\beta}^* = (\boldsymbol{\kappa}_\beta', \kappa_{c,\beta})'$	$\boldsymbol{\kappa}_{c,\beta}^* = (\boldsymbol{\kappa}_\beta', \kappa_{c,\beta})'$

Table 2: A comprehensive list of matrices, vectors, and constants to define the full-conditional distributions in Theorem 3. If Z_i does not lay on the boundary of it's support then use the left-hand column. The other columns should be used when $j = k = 2$ and $j = k = 3$ and when there exists Z_i on the boundary of it's support (i.e., there exists an i such that $Z_i = 0$ or t_i for $j = k = 2$ and $Z_i = 0$ for $j = k = 3$). The i-th element of $\mathbf{b}$ is the value of b associated with Z_i, where we note that this value is assumed to be the same for all i. In the left-most column $\varepsilon_\alpha = \varepsilon_{\kappa,i} \equiv 0$. In the middle column ε_α is chosen to be "small" and $(\varepsilon_{\kappa,1},\ldots,\varepsilon_{\kappa,n})' = \mathbf{b}$. In the third column the elements of $\boldsymbol{\varepsilon}_\kappa \equiv (\varepsilon_{\kappa,1},\ldots,\varepsilon_{\kappa,n})'$ and ε_α are chosen to be "small." When $\psi = \psi_3$, set $c = c_\eta = c_\xi = 0$.

Unit Log Partition Function	Form of the Prior Distribution on α and κ (i.e., $f(\alpha,\kappa\vert\gamma_1,\gamma_2,\rho)$)	Suggested Hyperparameters	Special Case of the Prior Distribution
$\psi_1(Y)=\log\left(-\frac{1}{Y}\right)$	$\exp\{\gamma_1\alpha+\gamma_2\kappa-\rho\log(\Gamma(\kappa+1))-\rho(\kappa+1)\log(\alpha)\}$ $=\frac{1}{\Gamma(\kappa+1)^\rho}\left(\alpha^{-\rho}\exp(\gamma_2)\right)^{\kappa+1}\exp(\gamma_1\alpha)$	$\gamma_1=-1000$ $\gamma_2=1000$ $\rho=10^{-15}$	If κ is integer-valued then the conditional distribution of $\kappa\vert\alpha$ is Conway-Maxwell-Poisson with parameters $\alpha^\rho\exp(\gamma_2)$ and ρ, and the conditional distribution of $\alpha\vert\kappa$ is Gamma$((\kappa+1)\rho+1,-1/\gamma_1)$ provided that $\gamma_2\in\mathbb{R}$, γ_1 is negative, and $\rho\geq 0$.
$\psi_2(Y)=\log(1+\exp(Y))$	$\exp[\gamma_1\alpha+\gamma_2\kappa+\rho\log\{\Gamma(\kappa)\}-\rho\log\{\Gamma(\alpha)\}-\rho\log\{\Gamma(\kappa-\alpha)\}]$ $=\left(\frac{\Gamma(\kappa)}{\Gamma(\alpha)\Gamma(\kappa-\alpha)}\right)^\rho\exp(\gamma_1)^\alpha\exp(\gamma_2)^\kappa$	$\gamma_1=0$ $\gamma_2=-1000$ $\rho=1$	Let $\rho=1$, $\gamma_1\in\mathbb{R}$, and $\gamma_2<0$. If α and κ are integer-valued, then the conditional distribution of $(\alpha-1)\vert\kappa$ is binomial with κ number of Bernoulli trials, and probability of success $\exp(\gamma_1)/(1+\exp(\gamma_1))$. Also, $(\kappa-\alpha-1)\vert\alpha$ follows a negative binomial distribution with $\alpha+1$ number of successful Bernoulli trials, and probability of success $\exp(\gamma_2)$.
$\psi_3(Y)=\exp(Y)$	$\exp\{\gamma_1\alpha+\gamma_2\kappa-\rho\log(\Gamma(\alpha))-\rho(\alpha)\log(\kappa)\}=\frac{1}{\Gamma(\alpha)^\rho}\left(\kappa^{-\rho}\exp(\gamma_1)\right)^\alpha\exp(\gamma_2\kappa)$	$\gamma_1=1$ $\gamma_2=-10^{-15}$ $\rho=1$	If α is integer-valued then the conditional distribution of $(\alpha-1)\vert\kappa$ is Conway-Maxwell-Poisson with parameters $\kappa^\rho\exp(\gamma_1)$ and ρ, and the conditional distribution of $\kappa\vert\alpha$ is Gamma$(\alpha\rho+1,-1/\gamma_2)$ provided that $\gamma_1\in\mathbb{R}$, γ_2 is negative, and $\rho\geq 0$.
$\psi_4(Y)=Y^2$	$\exp\left(\gamma_1\alpha+\gamma_2\kappa+\frac{\rho}{2}\log\kappa-\frac{\alpha^2}{4\kappa}\right)=\kappa^{\rho/2+1-1}\exp(\gamma_2\kappa)\exp(-\frac{(\alpha-2\kappa\gamma_1)^2}{4\kappa})$	Set $\alpha=0$ $\gamma_1=0$ $\gamma_2=-\frac{1}{2}$ $\rho=2$	We have that κ is distributed as Gamma$(\rho/2+1,-1/\gamma_2)$ and is independent of α, which is distributed as normal with mean $2\kappa\gamma_1$ and variance 2κ. The suggested hyperparameters result in an inverse-gamma prior distribution on the variance of a normal random variable with shape 2 and scale 1, which yields mean 1 and variance infinity.

Table 3: Special Cases: We list the form of the the prior distribution in Equation (2) of the main text by ψ_j for $j=1,\ldots,4$. The first column has the unit log partition function, the second column has the form of the prior distributions (up to a proportionality constant), the third column gives suggested hyperparameters, and the fourth column gives special cases of the conditional distributions $\alpha\vert\kappa$ and $\kappa\vert\alpha$.

Using induction we find that,

$$
\begin{aligned}
&f(\boldsymbol{\eta}|c\mathbf{J}_{a,1},\mathbf{M},\boldsymbol{\alpha}_\eta,\boldsymbol{\kappa}_\eta) \underset{\mathbf{V}}{\propto} \exp\left[\alpha_\eta \mathbf{J}_{r,1}'\mathbf{V}^{-1}\boldsymbol{\eta} - \kappa_\eta \mathbf{J}_{r,1}'\psi\left\{\mathbf{V}^{-1}\boldsymbol{\eta} - c\mathbf{J}_{r,1}\right\}\right]\\
&\propto \exp\left[\sum_{i=2}^{r}\sum_{j=1}^{i-1}\alpha_\eta v_{i,j}\eta_j - \sum_{i=2}^{r}\kappa_\eta\psi\left(\sum_{j=1}^{i-1}\eta_j v_{i,j} + \eta_i - c\right)\right]\\
&= \prod_{i=2}^{r}\exp\left\{\alpha_\eta \boldsymbol{\Sigma}_i\mathbf{v}_i - \kappa_\eta\psi(\boldsymbol{\Sigma}_i\mathbf{v}_i + Y_i - c)\right\}
\end{aligned}
$$

where $\boldsymbol{\Sigma}_i' = \left(\eta_j : j = 1,\ldots,i-1\right)'$. Thus, the full conditional distribution is given by

$$
\begin{aligned}
&f(\mathbf{v}_2,\ldots,\mathbf{v}_r|\cdot) \underset{\mathbf{V}}{\propto} f(\boldsymbol{\eta}|c\mathbf{J}_a,\mathbf{V},\boldsymbol{\alpha}_\eta,\boldsymbol{\kappa}_\eta)\prod_{i=2}^{r}f(\mathbf{v}_i)\\
&\underset{\mathbf{V}}{\propto} \prod_{i=2}^{r}\exp\left[\alpha_\eta \mathbf{J}_{i,1}'\mathbf{H}_{\gamma,i}\mathbf{v}_i - \kappa_\eta \mathbf{J}_{i,1}'\psi\left\{\mathbf{H}_{\gamma,i}\mathbf{v}_i - \boldsymbol{\mu}_{\gamma,i}\right\}\right],\\
&\underset{\mathbf{V}}{\propto} \prod_{i=2}^{n}\mathrm{CM_c}\left(\boldsymbol{\mu}_{\gamma,i},\mathbf{H}_{\gamma,i},\alpha_\eta\mathbf{J}_{i,1},\kappa_\eta\mathbf{J}_{i,1};\ \psi\right),
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{H}_{\gamma,i} &\equiv \begin{bmatrix} \boldsymbol{\Sigma}_i \\ \sigma_v \mathbf{I}_{i-1} \end{bmatrix},\\
\boldsymbol{\mu}_{\gamma,i} &= (\eta_i,\ \mathbf{0}_{1,i-1})'.
\end{aligned}
$$

The Metropolis-Hasting algorithm in Appendix A.ii provides a way to sample $\mathbf{v}_i$ from $f(\mathbf{v}_i|\cdot,\mathbf{q}_\beta = \mathbf{0}_{a,1},\mathbf{q}_\eta = \mathbf{0}_{a,1},\mathbf{q}_\xi = \mathbf{0}_{a,1})f(\mathbf{q}_{v,i})$, which leads to the following update for $\mathbf{v}_i$,

$$
\mathbf{v}_i = (\mathbf{H}_{\gamma,i}'\mathbf{H}_{\gamma,i})^{-1}\mathbf{H}_{\gamma,i}'\boldsymbol{\mu}_{\gamma,i} + (\mathbf{H}_{\gamma,i}'\mathbf{H}_{\gamma,i})^{-1}\mathbf{H}_{\gamma,i}'\mathbf{w}, \tag{C.10}
$$

where $\mathbf{w} \sim \mathrm{CM}(\mathbf{0}_{i,1},\mathbf{I}_i,\alpha_\eta\mathbf{J}_{i,1},\kappa_\eta\mathbf{J}_{i,1};\psi)$.

Appendix C.iii: Step-by-Step Implementation

The Gibbs sampler associated with (20) of the main text requires one to compute certain quantities. These values are listed in Table 2. To aid the reader, we provide step-by-step instructions for implementing the Gibbs sampler associated with (20) of the main text as follows.

1. Initialize $\boldsymbol{\beta}$, $\boldsymbol{\eta}$, $\boldsymbol{\xi}$, c, c_β, c_ξ, $\{\mathbf{v}_i\}$, α_η, α_ξ, κ_η, and κ_ξ. Denote these initializations with $\boldsymbol{\beta}^{[0]}$, $\boldsymbol{\eta}^{[0]}$, $\boldsymbol{\xi}^{[0]}$, $c^{[0]}$, $c_\beta^{[0]}$, $c_\xi^{[0]}$, $\{\mathbf{v}_i^{[0]}\}$, $\alpha_\eta^{[0]}$, $\alpha_\xi^{[0]}$, $\kappa_\eta^{[0]}$, and $\kappa_\xi^{[0]}$. Set $m = 1$.

2. Set $\boldsymbol{\beta}^{[m]}$ equal to the right hand side of (C.7). The matrix $\mathbf{H}_\beta$ and the vector μ_β are defined in Table 2. The r-dimensional vector $\boldsymbol{\eta}$ is set equal to $\boldsymbol{\eta}^{[m-1]}$, the n-dimensional vector $\boldsymbol{\xi}$ is set equal to $\boldsymbol{\xi}^{[m-1]}$, α_β is set equal to $\alpha_\beta^{[m-1]}$, and κ_β is set equal to $\kappa_\beta^{[m-1]}$.

3. Set $\boldsymbol{\eta}^{[m]}$ equal to the right hand side of (C.8). The matrix $\mathbf{H}_\eta$ and the vector μ_η are defined in Table 2. The p-dimensional vector $\boldsymbol{\beta}$ is set equal to $\boldsymbol{\beta}^{[m]}$, the n-dimensional vector $\boldsymbol{\xi}$ is set equal to $\boldsymbol{\xi}^{[m-1]}$, α_η is set equal to $\alpha_\eta^{[m-1]}$, κ_η is set equal to $\kappa_\eta^{[m-1]}$, and for each i the i-dimensional vector $\mathbf{v}_i$ is set equal to $\mathbf{v}_i^{[m-1]}$.

4. Set $\boldsymbol{\xi}^{[m]}$ equal to the right hand side of (C.9). The matrix $\mathbf{H}_\beta$ and the vector μ_β are defined in Table 2. The r-dimensional vector $\boldsymbol{\eta}$ is set equal to $\boldsymbol{\eta}^{[m]}$, the p-dimensional vector $\boldsymbol{\beta}$ is set equal to $\boldsymbol{\beta}^{[m]}$, α_ξ is set equal to $\alpha_\xi^{[m-1]}$, and κ_ξ is set equal to $\kappa_\xi^{[m-1]}$.

5. For $i = 2, \ldots, r$ set $\mathbf{v}_i^{[m]}$ equal to a value generated to the right hand side of (C.10). The matrix $\mathbf{H}_{\gamma,i}$ and the vector $\mu_{\gamma,i}$ are defined in Table 2. The r-dimensional vector $\boldsymbol{\eta}$ is set equal to $\boldsymbol{\eta}^{[m]}$.

6. Set $c^{[m]}$ equal to a draw from $\mathrm{CM_c}(\boldsymbol{\mu}_c, \mathbf{H}_c^*, \boldsymbol{\alpha}_c^*, \kappa_c^*)$ using a slice sampler, where $\boldsymbol{\mu}_c$, $\mathbf{H}_c^*$, $\boldsymbol{\alpha}_c^*$, and $\boldsymbol{\kappa}_c^*$ are computed using Table 2 and the most current values of the remaining parameters. We have found that c is weakly identifiable, and hence, truncating the support of the prior or using an informative prior often leads to better results.

7. Set $c_\beta^{[m]}$ equal to a draw from $\mathrm{CM_c}(\boldsymbol{\mu}_{c,\beta}, \mathbf{H}_{c,\beta}^*, \boldsymbol{\alpha}_{c,\beta}^*, \kappa_{c,\beta}^*)$ using a slice sampler, where $\boldsymbol{\mu}_{c,\beta}$, $\mathbf{H}_{c,\beta}^*$, $\boldsymbol{\alpha}_{c,\beta}^*$, and $\boldsymbol{\kappa}_{c,\beta}^*$ are computed using Table 2 and the most current values of the remaining

parameters. We have found that c_β is weakly identifiable, and hence, truncating the support of the prior or using an informative prior often leads to better results.

8. Set $c_\xi^{[m]}$ equal to a draw from $\mathrm{CM_c}(\boldsymbol{\mu}_{c,\xi}, \mathbf{H}^*_{c,\xi}, \boldsymbol{\alpha}^*_{c,\xi}, \kappa^*_{c,\xi})$ using a slice sampler, where $\boldsymbol{\mu}_{c,\xi}$, $\mathbf{H}^*_{c,\xi}$, $\boldsymbol{\alpha}^*_{c,\xi}$, and $\boldsymbol{\kappa}^*_{c,\xi}$ are computed using Table 2 and the most current values of the remaining parameters. We have found that c_ξ is weakly identifiable, and hence, truncating the support of the prior or using an informative prior often leads to better results.

9. Use a slice sampler (or Metropolis) to set $\alpha_\beta^{[m]}$ and $\kappa_\beta^{[m]}$ to a value generated from the pdf:

$$
\begin{aligned}
& f\left(\alpha_\beta, \kappa_\beta | \cdot\right) \\
& \propto \exp\left[(\gamma_{\beta,1} + \mathbf{J}_{1,p}\mathbf{V}_\beta^{[m]-1}\boldsymbol{\beta}^{[m]})\alpha_\beta + \left\{\gamma_{\beta,2} - \mathbf{J}_{1,g}\boldsymbol{\psi}(\mathbf{M}_\beta^{[m]}\boldsymbol{\beta}^{[m]} - c_\beta^{[m]}\mathbf{J}_{g,1})\right\}\kappa_\beta \right. \\
& \qquad\qquad\qquad \left. -(\rho_\beta + g)\log\left\{\frac{1}{K\left(\alpha_\beta, \kappa_\beta\right)}\right\}\right],
\end{aligned}
$$

where $g = p$ if no boundary value update is needed, $g = n + p$ if $\psi = \psi_3$, and $g = 2n + p$ if $\psi = \psi_2$.

10. Use a slice sampler (or Metropolis) to set $\alpha_\eta^{[m]}$ and $\kappa_\eta^{[m]}$ to a value generated from the pdf:

$$
\begin{aligned}
& f\left(\alpha_\eta, \kappa_\eta | \cdot\right) \\
& \propto \exp\left[(\gamma_{\eta,1} + \mathbf{J}_{1,r}\mathbf{V}^{[m]-1}\boldsymbol{\eta}^{[m]})\alpha_\eta + \left\{\gamma_{\eta,2} - \mathbf{J}_{1,g}\boldsymbol{\psi}(\mathbf{M}^{[m]}\boldsymbol{\eta}^{[m]} - c^{[m]}\mathbf{J}_{g,1})\right\}\kappa_\eta \right. \\
& \qquad\qquad\qquad \left. -(\rho_\eta + g)\log\left\{\frac{1}{K\left(\alpha_\eta, \kappa_\eta\right)}\right\}\right],
\end{aligned}
$$

where $g = r$ if no boundary value update is needed, $g = n + r$ if $\psi = \psi_3$, and $g = 2n + r$ if $\psi = \psi_2$.

11. Using a slice sampler (or Metropolis) to set $\alpha_\xi^{[m]}$ and $\kappa_\xi^{[m]}$ equal to values generated from the

pdf:

$$
\begin{aligned}
&f\left(\alpha_{\xi},\kappa_{\xi}|\cdot\right)\\
&\propto \exp\left[(\gamma_{\xi,1}+\mathbf{J}_{1,n}\boldsymbol{\xi})\alpha_{\xi}+\left\{\gamma_{\xi,2}-\mathbf{J}_{1,2n}\psi(\mathbf{M}_{\xi}\boldsymbol{\xi}-c_{\xi}^{[m]}\mathbf{J}_{g,1})\right\}\kappa_{\xi}\right.\\
&\qquad\qquad\qquad\qquad \left. -(\rho_{\xi}+g)\log\left\{\frac{1}{K\left(\alpha_{\xi},\kappa_{\xi}\right)}\right\}\right],
\end{aligned}
$$

where $g=n$ if no boundary value update is needed, $g=2n$ if $\psi=\psi_3$, and $g=3n$ if $\psi=\psi_2$.

12. Set $m=m+1$.

13. Repeat steps 2 through 12 until convergence of the Gibbs sampler.

It is straightforward to adjust this Gibbs sampler in variety of ways to be more appropriate for a particular problem. For example, one could consider different hyperparameters, different basis functions $\{\boldsymbol{\phi}_j\}$, update the shape and scale of the prior on $\mathbf{V}^{-1}$, and assume heterogeneous DY parameters associated with $\boldsymbol{\beta}$, $\boldsymbol{\eta}$, and $\boldsymbol{\xi}$.

It is important to note that many software packages have built in functions to simulate from beta and gamma distributions, which are needed when $j=k=2$ and $j=k=3$, respectively. However, it is common for the Gibbs sampler to produce small values of shape and scale parameters, which may lead to computational errors when simulating from a beta or a gamma distribution. In this setting, we simulate beta and gamma random variables using strategies outlined in Devroye (1986, pgs. 181, 182, and 419). Additionally, if the shape and scale parameters are so small (i.e., close to zero) that it is not possible to simulate the beta and gamma random variables using the techniques in Devroye (1986), we reject the proposed sample. However, after a sufficient burn-in period of the Gibbs sampler, the acceptance rate is approximately equal to one.

Finally, the updates for shape and rate parameters can be simplified in many settings. These simplifications often require additional assumptions, such as, the shape parameter is assumed to be integer-valued. We refer the reader to Table 3 to see a list of special cases by log-partition function.

Appendix D: The ANOVA Table for the Simulation Study in Section 3.2 of the Main Text

The ANOVA Table associated with the simulation study in Section 3.2 is given in Table 4. The assumptions for this ANOVA may not hold, and hence, we interpret large F statistics subjectively.

Source	DF	SS	MS	F
Factor 1	1	8.174	8.174	**941.13**
Factor 2	1	≈ 0	≈ 0	0.05
Factor 3	1	0.01	0.007	0.86
Factor 4	2	1487.85	743.926	**85649.6**
Factor 5	1	≈ 0	≈ 0	≈ 0
Factor 6	1	0.07	0.066	7.63
Factor 1 $\times$ Factor 2	1	0.12	0.117	13.51
Factor 1 $\times$ Factor 3	1	≈ 0	≈ 0	0.01
Factor 1 $\times$ Factor 4	2	8.08	4.041	**465.24**
Factor 1 $\times$ Factor 5	1	0.04	0.039	4.53
Factor 1 $\times$ Factor 6	1	≈ 0	0.001	0.16
Factor 2 $\times$ Factor 3	1	0.01	0.011	1.25
Factor 2 $\times$ Factor 4	2	0.01	0.003	0.3
Factor 2 $\times$ Factor 5	1	≈ 0	≈ 0	0.01
Factor 2 $\times$ Factor 6	1	≈ 0	≈ 0	0.01
Factor 3 $\times$ Factor 4	2	≈ 0	0.001	0.17
Factor 3 $\times$ Factor 5	1	≈ 0	≈ 0	0.04
Factor 3 $\times$ Factor 6	1	0.03	0.035	3.99
Factor 4 $\times$ Factor 5	2	≈ 0	≈ 0	0.04
Factor 4 $\times$ Factor 6	2	≈ 0	0.002	0.28
Factor 5 $\times$ Factor 6	1	0.01	0.011	1.31
Residual	925	8.03	0.009	

Table 4: Analysis of variance (ANOVA). The response in this experiment is the log total prediction error in (21) of the main text. The six factors are listed in Section 3.2 of the main text, for up to two-way interactions. In the table, the column "Source" contains the source of variability; "DF" stands for degrees of freedom; "SS" denotes the sum of squared error; "MS" stands for mean squared error; and "F" denotes the F-statistic. There are 96 factor-level combinations each containing 10 replicates. We denote "approximately equal to zero" with "≈ 0." Large F-statistics are bold.

Acknowledgments

We would like to express our sincere gratitude to the editors, the associate editor, and the referees for their very helpful comments that improved this manuscript. We would also like to thank Drs. Matthew Simpson of SAS Inc. and Erin Schliep at the University of Missouri for helpful discussions. This research was partially supported by the U.S. National Science Foundation (NSF) and the U.S. Census Bureau under NSF grant SES-1132031, funded through the NSF-Census Research Network (NCRN) program. This article is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed are those of the authors and not those of the NSF or the U.S. Census Bureau.

References

Bradley, J., Holan, S., and Wikle, C. (2018). "Computationally Efficient Distribution Theory for Bayesian Inference of High-Dimensional Dependent Count-Valued Data (with Discussion)." *Bayesian Analysis*, 13, 253 – 310.

Casella, G. and Berger, R. (2002). *Statistical Inference*. Pacific Grove, CA: Duxbury.

Daniell, P. J. (1919). "Integrals in an Infinite Number of Dimensions." *Annals of Mathematics*, 20, 281–288.

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. New York, NY: Springer-Verlag.

Kolmogorov, A. N. (1933). *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin: Springer.

Lehmann, E. (1999). *Elements of Large-Sample Theory*. New York, NY: Springer.