Bayesian Covariate-Varying Interaction Analysis for Multivariate Count Data: Application to Microbiome Studies

Shuangjie Zhang *
Department of Statistics, Texas A & M University;

Michael Patnode
Department of Microbiology and Environmental Toxicology, University of California Santa Cruz

Juhee Lee
Department of Statistics, University of California Santa Cruz

Abstract

Understanding covariate-dependent interdependencies among features is of great interest in various applications. Motivated by a dataset of multivariate counts from a microbiome study, where microbial abundance and interaction patterns may change with environmental factors, we develop a Bayesian covariate-dependent factor model that flexibly estimates heteroscedasticity in the covariance matrix due to covariates. Our approach employs covariance regression through linear regression on a lower-dimensional factor loading matrix. This formulation, combined with joint sparsity induced by the Dirichlet-Horseshoe prior for the factor loadings, provides robust estimation of covariate-dependent covariance in high-dimensional settings. The model uses a regression approach to the mean abundance and addresses the varying mean and covariance structure with covariates. Furthermore, the model tackles significant statistical challenges such as discreteness, over-dispersion, compositionality, and high dimensionality that are common in microbiome data analyses, using a flexible nonparametric Bayesian approach. We thoroughly explore the properties of the model and perform extensive simulation studies to examine its performance. Real data examples from microbiome studies are used for illustration.

Keywords: Covariate-dependent interdependencies, Microbiome data analysis, Multivariate counts, Bayesian factor model, Covariance regression, Heteroscedasticity.

*Address for Correspondence: 155 Ireland St College Station, TX 77843-3143. E-mail: szhan209@tamu.edu.

1 Introduction

Covariance estimation is a fundamental task in multivariate statistical analysis, critical for understanding the relationships between variables. Covariance matrices play a pivotal role in various statistical methods, including principal component analysis (Pearson, 1901), factor analysis (Rummel, 1988), and canonical-correlation analysis (Hotelling, 1992). Conventional methods for covariance estimation, such as the sample covariance matrix and covariance estimators with structural assumptions Fan et al. (2013), typically assume that the data is identically and independently distributed (i.i.d.). However, this assumption is often violated in real-world scenarios, where data often exhibits heteroscedasticity and covariance changes with covariates. Ignoring these covariate dependencies can result in inaccurate models and misleading conclusions, necessitating the development of methods that account for these dependencies to provide sensible estimates of the interrelationships between variables. The problem is more challenging for high-dimensional data, whose dimension can be much larger than the sample size.

Covariance regression has gained significant attention over a long history due to its ability to incorporate covariate information, thereby enhancing the accuracy and interpretability of covariance estimates. Carroll and Ruppert (1982) first considered a heteroscedastic model in which the variances were given by a parametric function of the mean. More methods such as a linear model for the standard deviation(Rutemiller and Bowers, 1968) and a generalized model with a link function to allow non-negativity of variance (Smyth, 1989) were also developed for uni-variate cases. When it comes to multivariate heteroscedasticity, Leonard and Hsu (1992), Chiu et al. (1996) and Pourahmadi (2011) modeled the logarithm of elements of the covariance matrix as a linear function of known matrices to guarantee the positive definiteness of the covariance matrix. However, it is difficult to interpret parameters of covariate effects in the log scale and the number of parameters to estimate can be quite large in high-dimensional data. More recently, sparse and low-rank methods for covariate-

dependent covariance estimation or its inversion (precision matrix) have been considered to manage high-dimensional data where traditional methods are often inadequate. These approaches leverage structural assumptions, sparsity and low rankness, to enhance estimation accuracy and interpretability. Zhang and Leng (2012) modeled Cholesky decomposition of covariance matrix as linear functions of covariates with a moving average construction. Hoff and Niu (2012) expressed the covariance as a baseline covariance matrix plus a rank-1 positive definite matrix which depends on covariates. They further extended to allow the deviation of each covariate-dependent covariance from the baseline to be any rank. Fox and Dunson (2015) put a Gaussian process prior on the latent factor model and induced a flexible Bayesian nonparametric covariance regression model. The predictor-dependent framework was characterized as a combination of Gaussian process random functions of covariates. Ni et al. (2019) proposed a graphical regression method that estimates directed acyclic graphs for the precision matrix in heterogenous data with additional subject-level covariates. Niu et al. (2023) further modeled continuously varying undirected graphs with additional assistance from any general covariates for underlying heterogeneous multivariate observations.

Besides modeling the covariance matrix with covariates, joint modeling for means and covariances allows for the simultaneous exploration of covariate effects on the mean and the covariance of a variable(s) of interest. Pourahmadi (1999) provided a joint mean-covariance model with applications to longitudinal data. In the context of temporal heteroscedasticity, Fong et al. (2006) studied multivariate autoregressive conditionally heteroscedastic (ARCH) models in the financial data. Niu and Hoff (2019) extended their model in Hoff and Niu (2012) to a joint mean and covariance model, studying the covariate effects on both mean and covariance in the application of multiple health outcome measures. Moran et al. (2021) used a parametric covariance regression model to analyze verbal autopsy data. It was designed specifically for cause of death denoted covariance. However, the above

approaches are built for continuous data, and they can be inappropriate for analyzing multivariate count data. With the advent of high-throughput sequencing (HTS) sequencing technologies, multivariate count tables arise in various biological applications for statistical analyses. Especially in microbiome studies, 16S ribosomal RNA (16S rRNA) sequencing uses similarity-based clustering algorithms to group 16S rRNA sequences into Operational Taxonomic Units (OTUs), producing multivariate count tables for downstream analysis. An OTU represents a group of organisms classified together based on genetic sequence similarity. Analyzing OTU tables and detecting the structure of microbial interactions is essential for more accurately characterizing microbial communities. Popular methods in microbiome studies such as SparCC (Friedman and Alm, 2012), CCLasso (Fang et al., 2015) and SPIEC-EASI (Kurtz et al., 2015) adopt log-transformed counts or log-transformed ratio for analysis of interactions. Specifically, SparCC (Friedman and Alm, 2012) adds pseudo counts and then divides the raw counts by the sample's total counts for normalization. It models log-transformed ratios of these normalized counts to infer correlations between OTUs through sparse networks. Similarly, CCLasso in Fang et al. (2015) uses ℓ_1 penalty to estimate the correlation network of log-transformed counts. SPIEC-EASI (Kurtz et al., 2015) uses graphical lasso (Friedman et al., 2008), a popular penalized method outputting the association of undirected graphs, to obtain a robust precision matrix estimate. The raw OTU counts are also first centered by log-ratio (clr) transformation. See REBECCA (Ban et al., 2015), COAT (Cao et al., 2019) and MOFA (Argelaguet et al., 2018) for more. However, most methodologies above simply subtract the sample mean and assume the mean centered at 0. In addition, those covariance estimates remain the same across any covariates.

To circumvent the challenges described above and address the effect of covariates on microbial interactions, we propose a Bayesian covariate-dependent sparse factor model with a rounded kernel. The model assesses interrelationships between OTUs varying as a func-

tion of covariates. Furthermore, it simultaneously performs model-based normalization and flexibly accommodates large variability in count data. Specifically, we use nonparametric mixtures of rounded multivariate log-normal kernels to introduce latent continuous random variables. The covariance matrix of the kernels is then allowed to vary with covariates, characterizing the covariate-varying interrelationships among OTUs. We adopt the low-rank structure factor model and place a Dir-HS prior (Zhang et al., 2025) on the factor loading matrix to effectively learn a high-dimensional covariance structure despite a limited sample size. Moreover, we use a Dirichlet process (DP) prior on relative abundances to obtain a flexible joint distribution of count vectors. This nonparametric approach that models the mean counts flexibly handles excess zeros and overdispersion, common in microbiome data. We also relate covariates to the mean to detect different OTU abundances under covariates.

In the rest of the paper, we describe the model and its applications. § 2 and § 3 describe the covariate-dependent rounded multivariate log-normal kernel model, its prior specification and posterior computation. § 4 shows the results of simulation studies to evaluate the performance of our method. § 5 has results from the model applied to the real dataset, and § 6 concludes with some discussion of the results and areas of future research.

2 Model and Prior Specification

In this section, we first construct a Bayesian sparse factor model that allows factor loadings to vary with covariates, enabling the estimation of covariate-dependent covariance matrices. We then develop a rounded kernel model with a mean regression function to accommodate the discreteness of count data. This model employs a Bayesian nonparametric approach, providing a flexible joint distribution for multivariate count responses.

2.1 Sparse Covariate-dependent Factor Model

Let $\tilde{\boldsymbol{Y}}^\star = (\tilde{Y}_1^\star, \ldots, \tilde{Y}_J^\star) \in \mathbf{R}^J$ be a J-dimensional normal vector,

$$\tilde{\boldsymbol{Y}}^\star \mid \boldsymbol{\mu}(\boldsymbol{x}), \Sigma(\boldsymbol{x}) \overset{indep}{\sim} \mathrm{N}_J(\boldsymbol{\mu}(\boldsymbol{x}), \Sigma(\boldsymbol{x})), \tag{1}$$

where $\boldsymbol{x} = (x_1, \ldots, x_P)$ is a P-dimensional vector of covariates for the sample, with $x_1 = 1$ representing the intercept. We first build a prior probability model for the covariate-dependent covariance $\Sigma(\boldsymbol{x})$, the main parameter of interest. The covariates in $\boldsymbol{\mu}$ may differ from those in Σ depending on the context of the problem. We will discuss a model for $\boldsymbol{\mu}(\boldsymbol{x})$ later in § 2.2. To overcome the difficulty posed by high dimensionality, particularly when the sample size is much smaller than the number of features, i.e., $N \ll J$, we extend the spiked covariance structure assumption (Johnstone, 2001). Specifically, we decompose $\Sigma(\boldsymbol{x})$ into a low-rank matrix and a diagonal matrix, and construct a factor model with factor loadings that vary with $\boldsymbol{x}$;

$$\Sigma(\boldsymbol{x}) = \Lambda(\boldsymbol{x})\Lambda'(\boldsymbol{x}) + \sigma^2 \mathrm{I}_J, \tag{2}$$

where $\Lambda(\boldsymbol{x}_i) = [\lambda_{jk}(\boldsymbol{x})]$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$, is a $J \times K$ covariate-dependent factor loading matrix. Here, K is the dimension of the subspace that is assumed to capture statistical variability, and typically we have $K \ll J$. Similar to Bhattacharya and Dunson (2011) and Xie et al. (2018), we do not impose any constraints on $\Lambda(\boldsymbol{x})$, such as column orthogonality, nor do we seek to interpret latent factors, as our primary focus is on the inference of $\Sigma(\boldsymbol{x})$. We further express $\lambda_{jk}(\boldsymbol{x})$ as

$$\lambda_{jk}(\boldsymbol{x}) = q_{jk} \boldsymbol{f}_k' \boldsymbol{x}. \tag{3}$$

Here, $\boldsymbol{f}_k$ is a P-dimensional coefficient vector that accommodates covariate effects for factor k. While the effect $\boldsymbol{f}_k'\boldsymbol{x}$ is shared by all features, q_{jk} adjusts it for feature j. When the local effect q_{jk} is close to 0, the corresponding λ_{jk} becomes small. If $f_{kp} \approx 0$, $\lambda_{jk}(\boldsymbol{x})$ do not vary with x_p for all j's. If f_{kp} is small for all k and p, $\Sigma(\boldsymbol{x})$ does not change with $\boldsymbol{x}$ much. With (2) and (3), we can rewrite

$$\Sigma(\boldsymbol{x}) = \sum_{k=1}^{K} (\boldsymbol{q}_k \boldsymbol{f}_k' \boldsymbol{x})(\boldsymbol{x}' \boldsymbol{f}_k \boldsymbol{q}_k') + \sigma^2 \mathrm{I}_J, \tag{4}$$

where $\boldsymbol{q}_k = (q_{1k}, \ldots, q_{Jk})'$. The covariance between features j and j' is a sum of K quadratic functions of $\boldsymbol{x}$;

$$\Sigma_{jj'}(\boldsymbol{x}) = \sum_{k=1}^{K} q_{jk} q_{j'k} (\boldsymbol{f}_k' \boldsymbol{x})^2 + \sigma^2 1(j = j'), \tag{5}$$

and it can flexibly capture various shapes in a reasonable range of $\boldsymbol{x}$. The structure in (4) has $JK + KP$ unknown parameters and significantly reduces the number of parameters to estimate compared to a case of estimating a $J \times J$ matrix Σ for each $\boldsymbol{x}$. This reduction is crucial in a high-dimensional setting.

In addition to the spiked covariance structural assumption, we further assume joint sparsity on $\Sigma(\boldsymbol{x})$ by employing a Dirichlet-Horseshoe (Dir-HS) prior in Zhang et al. (2025) for $\boldsymbol{q}_k$; for $k = 1, \cdots, K$,

$$\begin{aligned}
\tau_k \mid a_\tau, b_\tau &\overset{iid}{\sim} \mathrm{Ga}(a_\tau, b_\tau / J), \\
\boldsymbol{\phi}_k = (\phi_{1k}, \ldots, \phi_{Jk}) \mid a_\phi &\overset{iid}{\sim} \mathrm{Dir}(a_\phi, \ldots, a_\phi), \\
\zeta_{jk} &\overset{iid}{\sim} \mathrm{C}^+(0, 1), \qquad j = 1, \ldots, J, \\
q_{jk} \mid \phi_{jk}, \tau_k, \zeta_{jk} &\overset{indep}{\sim} \mathrm{N}(0, \zeta_{jk}^2 \phi_{jk} \tau_k), \qquad j = 1, \ldots, J.
\end{aligned} \tag{6}$$

where $\mathrm{C}^+(0, 1)$ is the half-Cauchy distribution for $\mathbb{R}^+$ with location and scale parameters 0 and 1, and $\mathrm{Ga}(a, b)$ represents the gamma distribution with mean a/b. Under the model in (4) with the prior in (6), ϕ_{jk} locally shrinks q_{jk} towards zero, which in turn shrinks

$\lambda_{jk}(\boldsymbol{x})$. On the other hand, τ_k controls the global shrinkage for each factor and performs an effective truncation of the number of latent factors. The joint sparsity assumption leads to obtaining a reliable estimate of the structure with a small sample size and achieving good theoretical properties (Cai et al., 2015; Xie et al., 2018). Zhang et al. (2025) discussed theoretical properties of the Dir-HS distribution that achieves heavy tails with a sharp peak at zero and showed that it efficiently induces joint sparsity and performs the robust estimation of a covariance matrix in a high-dimensional setting. Compared to the covariance-regression models discussed earlier (Fox and Dunson, 2015; Moran et al., 2021; Hoff and Niu, 2012), our model is more parsimonious due to (3), while the form in (4) remains flexible enough to capture non-traditional patterns. Furthermore, the structural assumption with joint sparsity enables robust estimation in a high-dimensional setting more efficiently. We demonstrate this in § 4.2 with simulation studies, where the true underlying covariance matrix is an arbitrary function of a covariate.

To complete the prior specification for $\Sigma(\boldsymbol{x})$, we assume a conditionally conjugate prior on σ^2, $\sigma^2 \sim \text{inv-Ga}(a_\sigma, b_\sigma)$ with fixed a_σ and b_σ and a standard normal prior on $f_{kp} \overset{iid}{\sim} \text{N}(0,1)$.

2.2 A Flexible Model for Multivariate Count Responses

Next, we use the latent factor models with covariate-dependent factor loadings in § 2.1 as a building block and build a Bayesian nonparametric mixture model with a rounded kernel to obtain a flexible multivariate count distribution whose mean and covariance vary with $\boldsymbol{x}$.

We have N samples, for which we have observed J-dimensional random count vectors $\boldsymbol{Y}_i = (Y_{i1}, \ldots, Y_{iJ})$, $i = 1, \ldots, N$, of the features with covariates $\boldsymbol{x}_i$. We introduce a latent multivariate log-normal vector $\boldsymbol{Y}_i^\star = \exp(\tilde{\boldsymbol{Y}}_i^\star) \in \mathbf{R}_+^J$ and assume $\tilde{\boldsymbol{Y}}_i^\star$'s are an independent sample from the model in (1). We next build a joint distribution for count vector $\boldsymbol{Y}_i$ using

a rounded kernel in Canale and Dunson (2011);

$$\mathrm{P}(\boldsymbol{Y}_i = \boldsymbol{y} \mid \boldsymbol{\mu}(\boldsymbol{x}_i), \Sigma(\boldsymbol{x}_i)) = \int_{A(\boldsymbol{y})} f_{\boldsymbol{y}^\star}(\boldsymbol{y}^\star \mid \boldsymbol{\mu}(\boldsymbol{x}_i), \Sigma(\boldsymbol{x}_i)) d\boldsymbol{y}^\star, \tag{7}$$

where the region of integration $A(\boldsymbol{y}) = \{\boldsymbol{y}^\star \mid y_1 \leq y_1^\star < y_1 + 1, \ldots, y_J \leq y_J^\star < y_J + 1\}$ and $f_{\boldsymbol{y}^\star}(\cdot)$ is the probability density function of a $J-$dimensional log-normal distribution with parameters $\boldsymbol{\mu}(\boldsymbol{x}_i)$ and $\Sigma(\boldsymbol{x}_i)$. The multivariate log-normal density is zero for a vector with negative values, and the kernel defines a valid multivariate count distribution for $\boldsymbol{Y}_i$. $\exp(\mu_{ij})$ is the median of $Y_{ij}^\star$ and provides an inference on the abundance of feature j in sample i. We consider the model in § 2.1 for $\Sigma(\boldsymbol{x}_i)$. We have the mean $\mathrm{E}(Y_{ij}^\star) = \exp(\mu_{ij} + \frac{1}{2}\Sigma_{jj}(\boldsymbol{x}_i))$ and $\mathrm{Cov}(Y_{ij}^\star, Y_{ij'}^\star) = \mathrm{E}(Y_{ij}^\star)\mathrm{E}(Y_{ij'}^\star)\{\exp(\Sigma_{jj'}(\boldsymbol{x}_i)) - 1\}$. When $\Sigma_{jj'}(\boldsymbol{x}_i) = 0$, it implies there are no microbial interactions between features. In terms of the count distributions, the mean and covariance of $\boldsymbol{Y}_i$ can be easily verified finite and computed through probability mass function defined in (7). We illustrate the distributions of $\boldsymbol{y}$ obtained under the rounded kernel model with specific examples in Supp. S 2.

We relate $\boldsymbol{\mu}_i(\boldsymbol{x}_i)$ to covariates $\boldsymbol{x}_i$ through regression;

$$\mu_{ij}(\boldsymbol{x}_i) = \mu_{ij} = r_i + \alpha_j + \tilde{\boldsymbol{x}}_i' \boldsymbol{\beta}_j, \tag{8}$$

where $\tilde{\boldsymbol{x}}_i$ is $\boldsymbol{x}_i$ without intercept x_{i1}. r_i is the sample (library) size factor, normalizing counts across samples. α_j represents the normalized baseline abundance of feature j. Regression coefficients β_{jp} quantify the change in the abundance of feature j from its baseline abundance α_j by covariate x_{ip}. We consider the conditionally conjugate prior distribution for β_{jp} and let $\beta_{jp} \overset{iid}{\sim} \mathrm{N}(0, u_\beta^2)$ with fixed u_β^2.

We take a Bayesian nonparametric approach and construct a flexible prior model for α_j to account for the large variability among features. This flexible modeling of mean abundances may further enhance the estimation of $\Sigma(\boldsymbol{x}_i)$. While $\boldsymbol{\beta}_j$ and $r_i + \alpha_j$ are identifiable,

the individual parameters r_i and α_j in (8) are not identifiable due to the multiplicative structure, $\mathrm{E}(y^{\star}_{ij} \mid r_i, \alpha_j) \propto \exp(r_i + \alpha_j)$. However, since the primary inferential goal is to estimate β_{jp}, this non-identifiability does not impact the inference of interest. To mitigate potential computational issues, we impose a mean-constrained Dirichlet process prior for α_j as follows;

$$\alpha_j \mid G \overset{iid}{\sim} G = \sum_{l=1}^{\infty} \psi_l^{\alpha} \left\{ \omega_l^{\alpha} \delta_{\xi_l^{\alpha}} + (1 - \omega_l^{\alpha}) \delta_{\left(\frac{\nu^{\alpha} - \omega_l^{\alpha} \xi_l^{\alpha}}{1 - \omega_l^{\alpha}} \right)} \right\}, \quad j = 1, \ldots, J, \tag{9}$$

where δ_{ξ} is a point mass centered at ξ. We let $\xi_l^{\alpha} \mid \nu^{\alpha}, u_{\alpha}^2 \overset{iid}{\sim} \mathrm{N}(\nu^{\alpha}, u_{\alpha}^2)$, $l = 1, 2, \ldots$, with fixed ν^{α} and u_{α}^2. The outer mixture weights ψ_l^{α} in (9) are constructed using a stick-breaking process (Sethuraman, 1994); let $\psi_1^{\alpha} = V_1^{\alpha}$ and $\psi_l^{\alpha} = V_l^{\alpha} \prod_{l'=1}^{l-1} (1 - V_{l'}^{\alpha})$, $l > 1$ with $V_l^{\alpha} \mid c^{\alpha} \overset{iid}{\sim} \mathrm{Be}(1, c^{\alpha})$, where the concentration parameter c^{α} is fixed. Assume inner mixture weights $\omega_l^{\alpha} \mid a_{\omega}^{\alpha}, b_{\omega}^{\alpha} \overset{iid}{\sim} \mathrm{Be}(a_{\omega}^{\alpha}, b_{\omega}^{\alpha})$, where a_{ω}^{α} and b_{ω}^{α} are fixed. Under (9), the prior and posterior means of α_j are fixed at ν^{α}. We will impose a similar constraint on the prior of r_i below to achieve soft identifiability. Shuler et al. (2021) and Zhang et al. (2025) showed that overall means can be well estimated under the mean-constrained prior. Under the model in (1) and (9), we obtain a Dirichlet process mixture model for $\boldsymbol{Y}_i^{\star}$,

$$\boldsymbol{Y}_i^{\star} \mid \boldsymbol{\mu}(\tilde{\boldsymbol{x}}_i), \Sigma(\boldsymbol{x}_i) \overset{indep}{\sim} \int \text{log-N}_J(\boldsymbol{y}^{\star} \mid r_i \mathbf{1}_J + \boldsymbol{\alpha} + \boldsymbol{\beta} \tilde{\boldsymbol{x}}_i', \Sigma(\boldsymbol{x}_i)) \prod_j dG(\alpha_j), \tag{10}$$

where $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_J]'$ and $\boldsymbol{\beta}$ is a $J \times P$ coefficient matrix with $\boldsymbol{\beta}_j$ in the rows. With random mixture weights, ω_l^{α} and ψ_l^{α}, and random locations ξ_l^{α} in $G(\alpha)$, the mixture model in (10) can flexibly capture various shapes of a distribution and accommodate variability in the count distribution. We also consider an extension of the model in (8)-(10) to accommodate inter-subject heterogeneity. We illustrate it in Simulations 2 and 3 in detail.

Similar to (9), we consider a flexible infinite mixture model for r_i;

$$r_i \mid \psi_l^r, \omega_l^r \overset{iid}{\sim} \sum_{l=1}^{\infty} \psi_l^r \left\{ \omega_l^r \mathrm{N}(\xi_l^r, u_r^2) + (1-\omega_l^r)\mathrm{N}\left(\frac{\nu^r - \omega_l^r \xi_l^r}{1-\omega_l^r}, u_r^2 \right) \right\}, \tag{11}$$

where ν^r and u_r^2 are fixed. The prior and posterior expectations of r_i are fixed ν^r in (11). We jointly specify values of ν^α and ν^r using observed counts. For example, we first fix ν^r at the average of the logarithm of the total count, $\nu^r = \frac{1}{N}\sum_{i=1}^N \log\left(\sum_{j=1}^J y_{ij}\right)$, and set $\nu^\alpha = \frac{1}{NJ}\sum_{i=1}^N \sum_{j=1}^J \{\log(y_{ij}+0.01) - \nu^r\}$. We consider similar following priors for ψ_l^r, ω_l^r and ξ_l^r; assume $\xi_l^r \mid \nu^r, u_{\xi^r}^2 \overset{iid}{\sim} \mathrm{N}(\nu^r, u_{\xi^r}^2)$, $\omega_l^r \mid a_\omega^r, b_\omega^r \overset{iid}{\sim} \mathrm{Be}(a_\omega^r, b_\omega^r)$, $\psi_1^r = V_1^r$ and $\psi_l^r = V_l^r \prod_{l'=1}^{l-1}(1-V_{l'}^r)$, $l > 1$, where $V_l^r \mid c^r \overset{iid}{\sim} \mathrm{Be}(1, c^r)$. Here, $u_{\xi^r}^2$, a_ω^r, b_ω^r, and c^r are fixed.

3 Prior Calibration and Posterior Computation

The prior of $\Sigma(\boldsymbol{x})$ in (6) requires the specification of the fixed hyperparameters K, a_ϕ, a_τ and b_τ. Selecting the dimension of the latent space K can be challenging. K determines the number of parameters, and a model with a random K requires complicated algorithms such as a reversible jump Markov Chain Monte Carlo (MCMC) method (Green and Hastie, 2009) for posterior simulation. Instead, we set K at a reasonably large value for computational convenience. For example, we empirically determine a value of K by performing principle component analysis (PCA) on the sample covariance matrix of log-transformed normalized counts and fixing K such that the K largest eigenvalues explain 95% of the total variance. With a sufficiently large K, the model can let some τ_k small for redundant latent factors. If desired, a geometric or truncated Poisson prior distribution can be placed on K to achieve an optimal posterior contraction rate (Pati et al., 2014). In terms of the hyperparameters a_ϕ, a_τ and b_τ, we follow the setup in Zhang et al. (2025) and let $a_\phi = 1/(0.2 \times J)$, $a_\tau = 0.1$ and $b_\tau = 1/J$. From simulation studies, we observed that a too small value of a_ϕ tends to overly shrink q_{jk} toward zero, resulting in a poor estimate of $\Sigma(\boldsymbol{x})$. We also

examined sensitivity to the specifications of those hyperparameters and found that the model's performance remains robust within a reasonable range of those values.

Let $\boldsymbol{\theta} = \{q_{jk}, \phi_{jk}, \tau_k, \zeta_{jk}, f_{kp}, \sigma^2, \alpha_j, \omega_l^\alpha, V_l^\alpha, \xi_{jl}^\alpha, r_i, \omega_l^r, V_l^r, \xi_l^r, \beta_{jp}\}$ a vector of all random parameters. We use Markov Chain Monte Carlo (MCMC) to sample $\boldsymbol{\theta}$ from their posterior distribution. To facilitate the posterior computation, we introduce a latent normal vector $\boldsymbol{\eta}_i \overset{iid}{\sim} \mathrm{N}_K(0, \mathrm{I}_K)$. We then have $Y_{ij}^\star \mid \mu_{ij}(\boldsymbol{x}_i), \boldsymbol{\lambda}_j(\boldsymbol{x}_i), \boldsymbol{\eta}_i, \sigma^2 \overset{indep}{\sim} \text{log-N}(\mu_{ij}(\boldsymbol{x}_i) + \boldsymbol{\lambda}_j'(\boldsymbol{x}_i)\boldsymbol{\eta}_i, \sigma^2)$ as independent log-normal variables, which results in significant computational efficiency. The joint posterior distribution of the augmented model is

$$p(\boldsymbol{\theta}, \boldsymbol{Y}^\star, \boldsymbol{\eta} \mid \boldsymbol{y}, \boldsymbol{x}) \propto \prod_{i=1}^{N} \prod_{j=1}^{J} p(y_{ij} \le Y_{ij}^\star < y_{ij} + 1 \mid \boldsymbol{\eta}_i, \boldsymbol{\theta}) \prod_{i=1}^{N} p(\boldsymbol{\eta}_i \mid \boldsymbol{\theta}) p(\boldsymbol{\theta}). \tag{12}$$

We use the blocked Gibbs sampling algorithm (Ishwaran and James, 2001) by considering a finite-dimensional truncation of the stick-breaking processes in (9) and (11). We set the truncation levels L^r and L^α to sufficiently large values. Given the latent variables, all parameters except $\boldsymbol{\phi}_k$ can be updated through Gibbs steps. Although f_{kp} has a conjugate full conditional distribution, we found the mixing over f_{kp} could be poor, and used an adaptive MH algorithm (Haario et al., 2001) for an efficient update of $\boldsymbol{\phi}_k$ and f_{kp}. Details of the MCMC algorithm are in Supp. § 1. The codes can be found at https://github.com/shuang-jie/BCAIA.

4 Simulation Studies

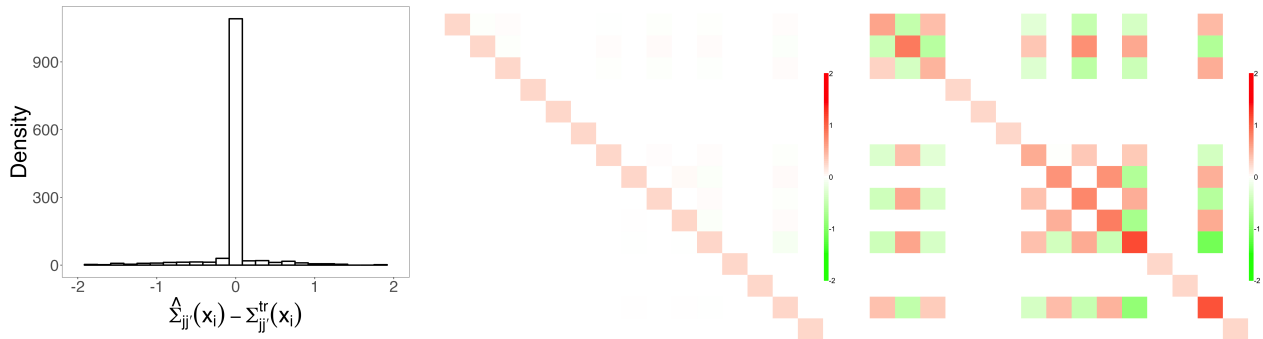
4.1 Simulation 1

For Simulation 1, we considered a setting similar to the mice gut microbiome dataset in § 5. In real data, we have two categorical variables, one with two levels and the other with three levels, resulting in a total of six experimental conditions. Using indicator variables with main effects, we constructed $\boldsymbol{x}$ with $P = 4$. We assumed $J = 15$ OTUs and generated

five samples per condition, resulting in a total of $N = 30$ samples. To specify $\Sigma^{\text{tr}}(\boldsymbol{x})$, we set $K^{\text{tr}} = 2$. With probability 0.5, we set $q_{jk}^{\text{tr}} = 0$. For the remaining probability, we simulated a random variable from $\text{N}(0, 1)$, shifted it away from zero by 1, and set it to q_{jk}^{tr}. We also assume that latent factors contribute differently to each categorical level, and set $f_{kp}^{\text{tr}} \overset{iid}{\sim} \text{Unif}(-1, 1)$ and set $f_{11} = -f_{13}, f_{21} = -f_{22}$, which results in factor 1 having no contribution to level $(1, 0, 1, 0)$ and factor 2 having no contribution to level $(1, 1, 0, 0)$. Then, $\Sigma^{\text{tr}}(\boldsymbol{x}) = \Lambda^{\text{tr}}(\boldsymbol{x})\Lambda^{\text{tr},\prime}(\boldsymbol{x}) + \sigma^{\text{tr},2}\text{I}_J$ with $\lambda_{jk}^{\text{tr}}(\boldsymbol{x}) = q_{jk}^{\text{tr}}\boldsymbol{f}_k^{\text{tr},\prime}\boldsymbol{x}$ and $\sigma^{\text{tr},2} = 0.5^2$. The true covariance matrix $\Sigma^{\text{tr}}(\boldsymbol{x}_i)$ is illustrated in the upper triangles of the heatmaps in Fig 1(b)-(c) for two samples with $\boldsymbol{x} = (1, 1, 0, 0)$ and $\boldsymbol{x} = (1, 1, 0, 1)$. For the mean count abundance, we set $r_i^{\text{tr}} \overset{iid}{\sim} \text{Unif}(0, 2)$, $i = 1, \ldots, N$ and $\alpha_j^{\text{tr}} \overset{iid}{\sim} 0.3\text{N}(2.5, 0.5^2) + 0.7\text{N}(5, 0.5^2)$, $j = 1, \ldots, J$. And we generated β_{jp}^{tr}, similar to q_{jk}^{tr}; we first let β_{jp}^{tr} 0 with probability 0.5. With the remaining probability, we sampled a random variable from $\text{N}(0, 1)$, shifted away from 0 by 1, and set it to be β_{jp}^{tr}. Finally, we generated $\boldsymbol{Y}_i^{\star,\text{tr}}$ from $\text{log-N}_J(\boldsymbol{\mu}^{\text{tr}}(\boldsymbol{x}_i), \Sigma^{\text{tr}}(\boldsymbol{x}_i))$, where $\mu_{ij}^{\text{tr}} = r_i^{\text{tr}} + \alpha_j^{\text{tr}} + \boldsymbol{\beta}_j^{\text{tr}}\tilde{\boldsymbol{x}}_i$, and obtained count vectors $\boldsymbol{Y}_i = \lfloor \boldsymbol{Y}_i^{\star,\text{tr}} \rfloor$.

We specified the values of the fixed hyperparameters as discussed in § 3, $K = 8$, $c^r = c^\alpha = 3$, $L^r = 30, L^\alpha = 35$, $a_\sigma = b_\sigma = 3$, $a_\omega^r = b_\omega^r = a_\omega^\alpha = b_\omega^\alpha = 5$ to fit the model for the simulated dataset. We ran MCMC for 160,000 iterations and discarded the first half for burn-in. The computation took 13 minutes on an Apple M1 chip laptop.

Fig. 1(a) presents a histogram of the differences $\hat{\Sigma}_{jj'}(\boldsymbol{x}_i) - \Sigma_{jj'}^{\text{tr}}(\boldsymbol{x}_i)$ from all samples, $j \leq j'$. The differences are tightly centered around 0, indicating that the model provides good estimates of the covariance that varies with $\boldsymbol{x}$. Figs. 1(b) and (c) compare posterior median estimates of $\Sigma_{jj'}(\boldsymbol{x})$ to their true values $\Sigma_{jj'}^{\text{tr}}(\boldsymbol{x})$ for two randomly selected samples, $i = 16$ and 26, which have $\boldsymbol{x}_{16} = (1, 1, 0, 0)'$ and $\boldsymbol{x}_{26} = (1, 1, 0, 1)$, respectively. For sample 16, all OTU pairs have a small covariance in the truth, representing inactive OTU communities. While in sample 26, some OTU pairs exhibit strong covariance. Our method produces reasonable estimates for both levels.



(a) $\hat{\Sigma}_{jj'}(\boldsymbol{x}) - \Sigma^{\text{tr}}_{jj'}(\boldsymbol{x})$ (b) $\hat{\Sigma}(\boldsymbol{x}_{16})$ vs $\Sigma^{\text{tr}}(\boldsymbol{x}_{16})$ (c) $\hat{\Sigma}(\boldsymbol{x}_{26})$ vs $\Sigma^{\text{tr}}(\boldsymbol{x}_{26})$

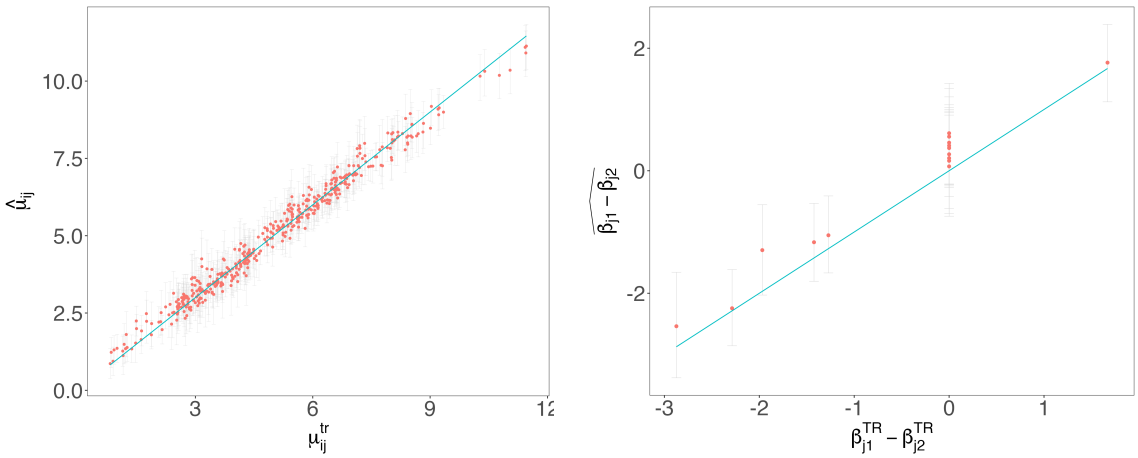
Figure 1: [Simulation 1] Panel (a) has a histogram of differences between $\hat{\Sigma}_{jj'}(\boldsymbol{x}_i)$ and $\Sigma^{\text{tr}}_{jj'}(\boldsymbol{x}_i)$ under six levels, $j \leq j'$. Panels (b) and (c) compare $\hat{\Sigma}^{\text{tr}}(\boldsymbol{x})$ (lower triangular) to its posterior median estimates $\hat{\Sigma}(\boldsymbol{x})$ (upper triangular) for two arbitrarily selected samples, $i = 16$ and 26, having $\boldsymbol{x}_{16} = (1, 1, 0, 0)'$ and $\boldsymbol{x}_{26} = (1, 1, 0, 1)$.

Fig. 2(a) compares the posterior median estimates of μ_{ij} to their true values, with red dots and grey vertical lines representing the posterior median estimates and 95% credible interval estimates, respectively. The means of $y^{\star}_{ij}$ are well estimated overall. Figs. 2(b)-(d) illustrate the posterior median estimates of covariate effects on the mean abundance. The difference $\beta^{\text{tr}}_{j1} - \beta^{\text{tr}}_{j2}$ quantifies the effect of the binary covariate. The parameters, $\beta^{\text{tr}}_{j3} - \beta^{\text{tr}}_{j4}$, $\beta^{\text{tr}}_{j3} - \beta^{\text{tr}}_{j5}$, and $\beta^{\text{tr}}_{j4} - \beta^{\text{tr}}_{j5}$ quantify the differences in μ_{ij} between pairs of the three levels of the ternary covariate. Supp Fig 3 shows the estimates of σ^2, q_{jk} and τ_k. The traceplot shows the convergence and good mixing of the MCMC chain. From these figures, our model obtained reasonable estimates of the covariate effects on the abundances.

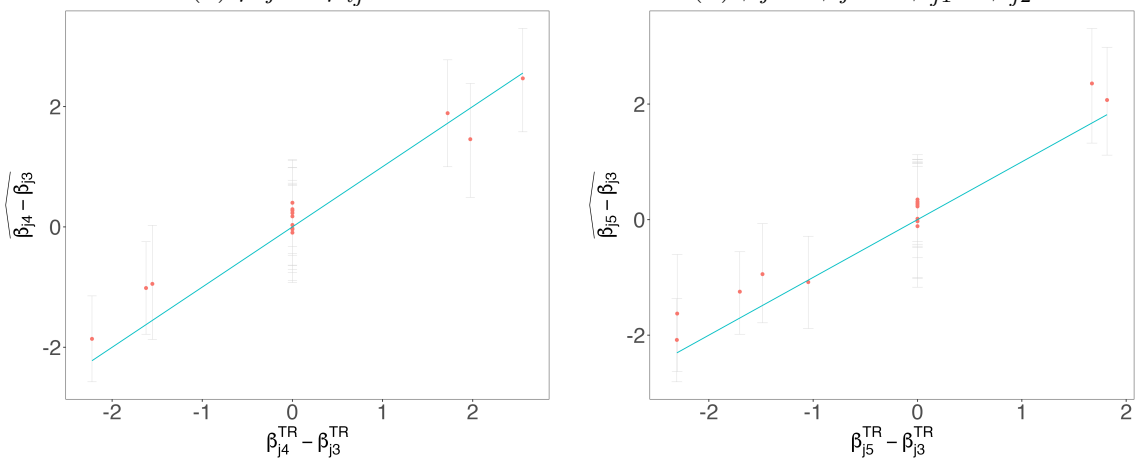
Most recently, Ahn and Datta (2024) introduce a jackknife pseudo-value method to adjust microbiome network connectivity differential analysis for clinical covariates, enabling comparisons of microbial interactions between groups.

4.2 Simulation 2

Simulation 2 considers a dataset with repeated samples that typically has large inter-subject variability. Suppose we have multiple samples from each in a set of subjects, $\{1, \ldots, S\}$. We let $s_i \in \{1, \ldots, S\}$ denote the subject from which sample i is taken. To accommodate inter-



(a) $\hat{\mu}_{ij}$ vs μ_{ij}^{tr} (b) $\widehat{\beta_{j1}-\beta_{j2}}$ vs $\beta_{j1}^{\text{tr}}-\beta_{j2}^{\text{tr}}$



(c) $\widehat{\beta_{j4}-\beta_{j3}}$ vs $\beta_{j4}^{\text{tr}}-\beta_{j3}^{\text{tr}}$ (d) $\widehat{\beta_{j5}-\beta_{j3}}$ vs $\beta_{j5}^{\text{tr}}-\beta_{j3}^{\text{tr}}$

Figure 2: [Simulation 1] The posterior median estimate of mean abundance μ_{ij} is plotted against the truth in panels (a). Panels (b)-(d) plot posterior point estimates of $\beta_{j1}-\beta_{j2}$, $\beta_{j3}-\beta_{j4}$, $\beta_{j3}-\beta_{j5}$, and $\beta_{j4}-\beta_{j5}$ against their true values. The red dots represent posterior median estimates, and the grey vertical lines 95% posterior credible interval estimates.

subject variability, we extend the model in a fashion similar to that in Zhang et al. (2025); we let the normalized mean abundance indexed by subjects, that is, $\alpha_{s_i,j}$ and assumed the following model for $\boldsymbol{\alpha}_{s_i} = (\alpha_{s_i1}, \ldots, \alpha_{s_iJ})'$;

$$\begin{aligned} \boldsymbol{\alpha}_{s_i} \mid G &\overset{iid}{\sim} G(\boldsymbol{\alpha}),\ s_i \in \{1, \ldots, S\}, \\ G(\boldsymbol{\alpha}) = \prod_{j=1}^{J} G_j(\alpha_j) &= \prod_{j=1}^{J} \left[\sum_{l=1}^{\infty} \psi_l^{\alpha} \left\{ \omega_l^{\alpha} \delta_{\xi_{jl}^{\alpha}} + (1-\omega_l^{\alpha}) \delta_{\left(\frac{\nu_j^{\alpha} - \omega_l^{\alpha} \xi_{jl}^{\alpha}}{1-\omega_l^{\alpha}} \right)} \right\} \right]. \end{aligned} \tag{13}$$

That is, a common α_j is shared by the samples from a subject, and the prior distribution of $\boldsymbol{\alpha}_s$ has different mixing locations ξ^{α}_{jl} for each OTU.

To generate a dataset, we let $S = 25$ subjects and $J = 100$ OTUs. We included a continuous covariate $x^c_{s_i} \overset{iid}{\sim} \mathrm{N}(0,1)$ and a binary covariate, $x^d_i \in \{0,1\}$, that represents two experimental conditions. Assuming that a sample is obtained from each condition for a subject, we had $N = 50$ samples. Adapting the common factor model in De Vito et al. (2019, 2021), we let $\Sigma^{\mathrm{tr}}(\boldsymbol{x}) = \Lambda^{\mathrm{tr}}_0 \Lambda^{\mathrm{tr},\prime}_0 + \Lambda^{\mathrm{tr}}(\boldsymbol{x})\Lambda^{\mathrm{tr},\prime}(\boldsymbol{x}) + \sigma^{2,\mathrm{tr}}\mathrm{I}_J$, where Λ^{tr}_0 is $J \times K_0$ matrix of common factor loadings with $K_0 = 2$, and $\Lambda^{\mathrm{tr}}_0 \Lambda^{\mathrm{tr},\prime}_0$ is a baseline covariance. We simulated random variables from $\mathrm{N}(0,1)$ and shifted them away from zero by $1/2$ for $\lambda^{\mathrm{tr}}_{0,jk}$ of OTUs 1-25 and 51-100 to ensure that those OTUs have baseline interactions. We let $\lambda^{\mathrm{tr}}_{0,jk} = 0$ for the remaining OTUs. We next let $\Lambda^{\mathrm{tr}}(\boldsymbol{x})$ a $J \times K_1$ factor loading matrix with $K_1 = 3$ and let $\lambda^{\mathrm{tr}}_{1,jk}(\boldsymbol{x}) = q^{\mathrm{tr}}_{jk} \boldsymbol{f}^{\mathrm{tr},\prime}_k \boldsymbol{x}$. We then simulated random variables from $\mathrm{N}(0,1)$, shifted them away from zero by $1/2$, and set them to be f^{tr}_{jk} for OTUs 51-100 and $f^{\mathrm{tr}}_{jk} = 0$ for the remaining OTUs. And we have $f^{\mathrm{tr}}_{kp} \overset{iid}{\sim} \mathrm{Unif}(-1,1)$ and $\sigma^{2,\mathrm{tr}} = 0.5^2$. Under this design, interactions among OTUs 1-50 do not change with covariates, while among OTUs 51-100 change with the covariates. Furthermore, OTUs 26-50 do not interact with the other OTUs. The covariance matrix corresponding to $\Sigma^{\mathrm{tr}}(\boldsymbol{x}_i)$ is illustrated in the lower triangle of Fig 3(b) and (c) for two selected samples. Excess zeros and large inter-subject variability are commonly observed in microbiome data, and we reflect those in the simulated dataset. For the normalized abundance level, we first set $\xi^{\alpha,\mathrm{tr}}_{j1} = -5$, $\xi^{\alpha,\mathrm{tr}}_{j2} \sim \mathrm{N}(2.5, 0.5)$ and $\xi^{\alpha,\mathrm{tr}}_{j3} \sim \mathrm{N}(5, 0.5)$ and simulated $\boldsymbol{\psi}^{\mathrm{tr}}_j = (\psi^{\mathrm{tr}}_{j1}, \psi^{\mathrm{tr}}_{j2}, \psi^{\mathrm{tr}}_{j3}) \sim \mathrm{Dir}(30, 40, 30)$. The three values, $\xi^{\alpha,\mathrm{tr}}_{jl}$, $l = 1$, 2 and 3, represent zero, small and large counts, respectively. We then let $\alpha^{\mathrm{tr}}_{sj} = \xi^{\alpha,\mathrm{tr}}_{jl}$ with probability ψ^{tr}_{jl} for $s \in \{1, \ldots, S\}$. We next simulated size factors $r^{\mathrm{tr}}_i \overset{iid}{\sim} \mathrm{Unif}(0,2)$ and regression coefficients $\beta^{\mathrm{tr}}_{jp} \overset{iid}{\sim} \mathrm{N}(0,1)$. We let $\boldsymbol{\mu}^{\mathrm{tr}}(\boldsymbol{x}_i) = r^{\mathrm{tr}}_i \mathbf{1}_J + \boldsymbol{\alpha}^{\mathrm{tr}}_{s_i} + \boldsymbol{\beta}' \boldsymbol{x}_i$. Under this setup, approximately 30.76% of Y_{ij}'s are 0. We specified the hyper-parameters values similar to Simulation 1 with $K = 7$. We ran MCMC for 160,000 iterations and

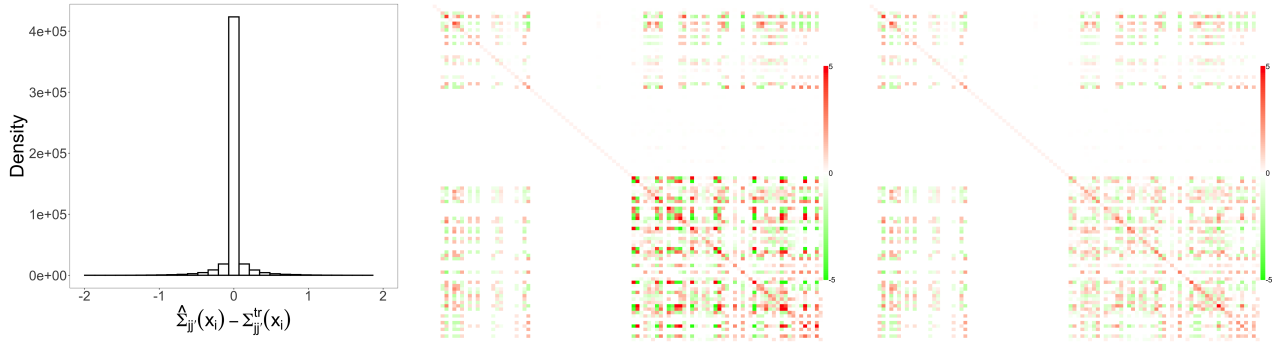
discarded the first half for burn-in. It took 23 hours on an Apple M1 chip laptop.

Fig 3(a) plots the differences $\hat{\Sigma}_{jj'}(\boldsymbol{x}_i) - \Sigma^{\text{tr}}_{jj'}(\boldsymbol{x}_i)$ of all samples, $j \leq j'$. *S, $j \leq j'$??* It shows that the extended model captures covariate varying feature interactions reasonably well even when excess zeros are present and counts greatly vary across subjects. Fig 3(b) and (c) compare the posterior median estimates of $\Sigma(\boldsymbol{x})$ to its truth. Their $\boldsymbol{x}$'s have the same value of the continuous covariate, $x^c = -1.23$, but the discrete covariate x_d takes the value of 0 and 1, respectively. Fig 4 presents the posterior estimates of covariance for some selected pairs of OTUs by varying the values of x^c and x^d. The solid and dashed lines are the truth and posterior estimates, and the red and blue colors correspond to $x^d = 0$ and 1, respectively. The shaded area represents point-wise 95% posterior credible interval estimates. The crosses at the bottom of the plots are the observed values of x^c. Fig 3 and Fig 4 show that our model identifies truly inactive OTUs and estimates the baseline covariance successfully, and the OTU interrelationship structure varying with the covariates is also reasonably well captured even when the sample size is smaller than the number of OTUs.

Supp. Fig 2 compares posterior median estimates of sample size factor r_i and the mean abundance μ_{ij} to their truth. In the figure, the library size factor and mean abundances are well estimated, serving as a reliable foundation for estimating the parameters of primary interest, such as $\Sigma(\boldsymbol{x}_i)$.

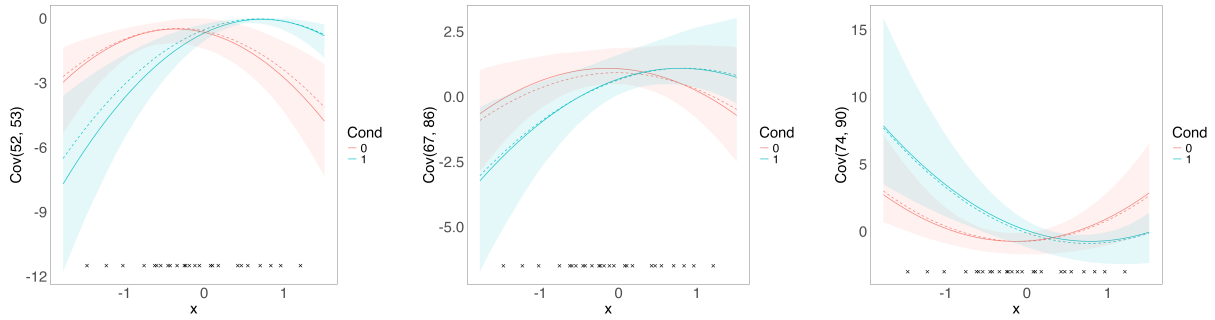
4.3 Simulation 3

For Simulation 3, we let the covariance matrix arbitrarily vary with a binary covariate, $x^d \in \{0, 1\}$ that represents two experimental conditions. Specifically, arbitrary covariance matrices were generated using the vine method in Lewandowski et al. (2009), separately for each value of x^d, as follows; We simulated partial correlations from linearly transformed Be$(1, 1)$ distribution over the interval of $(-1, 1)$. To encourage sparsity in $\Sigma^{\text{tr}}(\boldsymbol{x})$, we



(a) $\hat{\Sigma}_{jj'}(\boldsymbol{x}) - \Sigma^{\text{tr}}_{jj'}(\boldsymbol{x})$ (b) $\hat{\Sigma}(\boldsymbol{x}_2)$ vs $\Sigma^{\text{tr}}(\boldsymbol{x}_2)$ (c) $\hat{\Sigma}(\boldsymbol{x}_{27})$ vs $\Sigma^{\text{tr}}(\boldsymbol{x}_{27})$

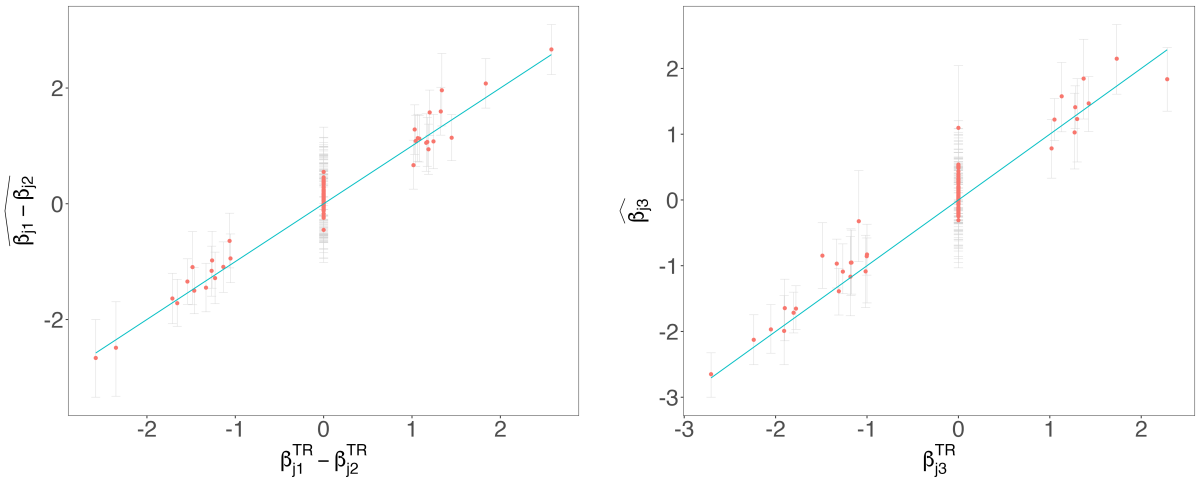
Figure 3: [Simulation 2] Panel (a) has a histogram of differences between $\hat{\Sigma}_{jj'}(\boldsymbol{x}_i)$ and $\Sigma^{\text{tr}}_{jj'}(\boldsymbol{x}_i)$ of all samples. In (b), the lower left and upper right triangles of the heatmap illustrate true values $\Sigma^{\text{tr}}_{jj'}$ and their posterior estimates of correlations $\hat{\Sigma}_{jj'}$, respectively. Two samples, samples 2 and 27, from subject 2, are arbitrarily chosen for illustration. Their covariates are $\boldsymbol{x}_2 = (1, -1.23), \boldsymbol{x}_{27} = (0, -1.23)$.



(a) OTUs 52 & 53 (b) OTUs 67 & 86 (c) OTUs 74 & 90

Figure 4: [Simulation 2] Scatter plots of $\Sigma_{jj'}(\boldsymbol{x})$ (dashed line) and $\Sigma^{\text{tr}}_{jj'}(\boldsymbol{x})$ (solid line) are plotted for three arbitrarily chosen OTU pairs, OTUs 52 and 53 in panel (a), OTUs 67 and 86 in panel (b), and OTUs 74 and 90 in panel (c). Crosses are observed values of the continuous covariate x_c. The red and blue colors are for $x^d = 0$ and 1, respectively. The shades represent pointwise 95% posterior credible interval estimates.

set the partial correlations below 0.8 to 0 and generated a correlation matrix, $\rho^{\text{tr}}(x^d)$ using their recursive formula. We then sampled independently from Unif$(1, 1.5)$ and let $\Sigma^{\text{tr}}_{jj'}(x^d) = \sigma^{2,\text{tr}}_j(x^d)\sigma^{2,\text{tr}}_{j'}(x^d)\rho^{\text{tr}}_{jj'}(x^d)$. $\Sigma^{\text{tr}}(0)$ and $\Sigma^{\text{tr}}(1)$ are shown in the lower triangle of Fig 6(b) and (c), respectively. We kept the rest of the simulation setup the same as in Simulation 1. We also used the same fixed hyperparameter values as in Simulation 2 except $K = 25$. We examined the prior sensitivity analysis by varying the value of K. The covariate-varying covariance matrix estimates remain largely unchanged for large enough



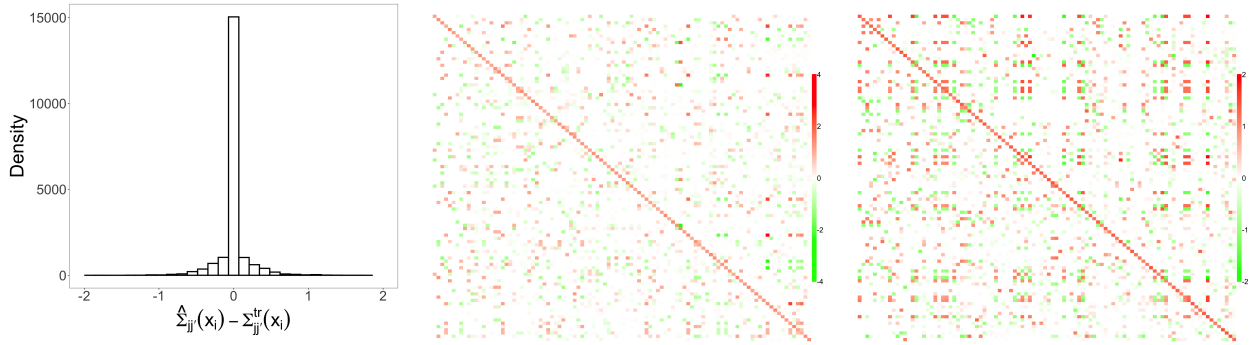
(a) $\widehat{\beta_{j1}-\beta_{j2}}$ versus $\beta_{j1}^{\text{tr}}-\beta_{j2}^{\text{tr}}$ (b) $\hat{\beta}_{j3}$ versus β_{j3}^{tr}

Figure 5: [Simulation 2] The posterior estimates of the effects of the binary and continuous covariates on abundance are plotted in panels (a) and (b), respectively. The dots represent the posterior median estimates, while the vertical lines indicate their corresponding 95% credible interval estimates.

values of K. More detailed results are summarized in Supp.3. We used MCMC to draw posterior samples of the random parameters. Examination of the MCMC simulation using traceplots indicated no evidence of convergence or mixing problems.

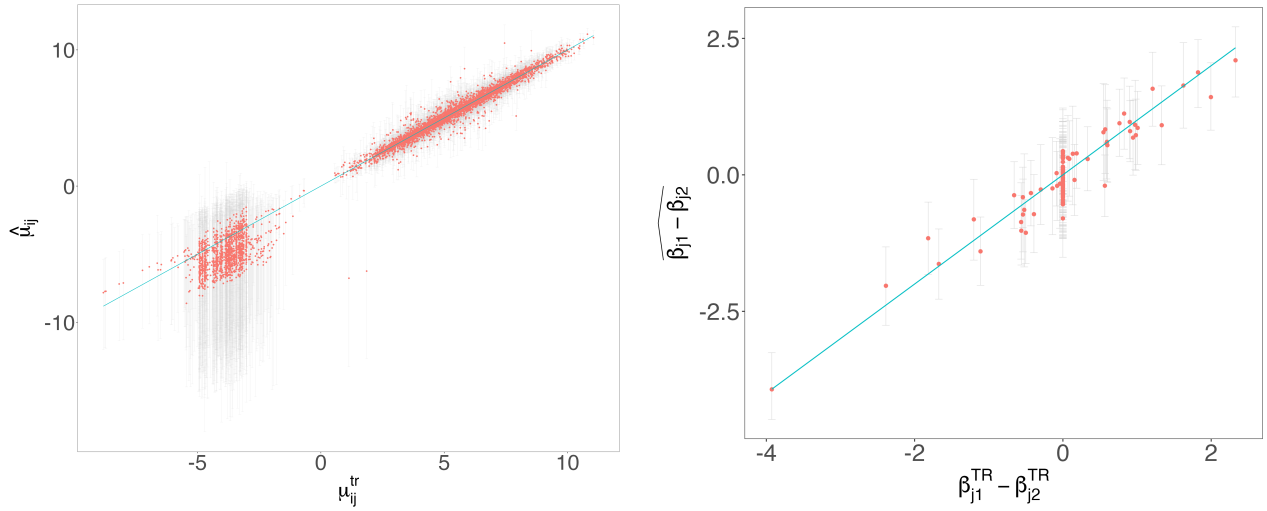
Fig 6(a) has a histogram of the differences $\hat{\Sigma}_{jj'}(x_d) - \Sigma_{jj'}(x^d)$, $j \leq j'$ for $x^d \in \{0,1\}$. Fig 6(b) and (c) compares the posterior estimates $\hat{\Sigma}(0)$ and $\hat{\Sigma}(1)$ under our model to their truth. Arbitrary $\Sigma^{\text{tr}}(x^d)$ are generated for each condition, while our model has a covariance regression with low rank assumption. However, the model approximates the truth structure in $\Sigma^{\text{tr}}(x^d)$ reasonably well.

Fig 7(a) shows a posterior median estimate of mean abundance μ_{ij} with their 95% interval estimate. Panel (b) of the figure compares the posterior estimates of the covariate effect $\beta_{j1} - \beta_{j2}$ on the mean abundance to the truth. The dots dots are point estimates, and the vertical lines 95% credible interval estimates. Our model effectively captures the covariate effects on the mean abundance.



(a) $\hat{\Sigma}_{jj'} - \Sigma^{\text{tr}}_{jj'}$ (b) $\hat{\Sigma}_0$ vs Σ^{tr}_0 (c) $\hat{\Sigma}_1$ vs Σ^{tr}_1

Figure 6: [Simulation 3] Panel (a) presents a histogram of the differences between $\hat{\Sigma}_{jj'}(x_d)$ and $\Sigma^{\text{tr}}_{jj'}(x_d)$, for $j \leq j'$ and $x_d \in \{0, 1\}$, where $\hat{\Sigma}_{jj'}(x_d)$ are the posterior median estimates of $\Sigma_{jj'}(x_d)$. Panels (b) and (c) compare $\hat{\Sigma}(x_d)$ to their true values for each x_d.



(a) $\hat{\mu}_{ij}$ vs μ^{tr}_{ij} (b) $\widehat{\beta_{j1} - \beta_{j2}}$ vs $\beta^{\text{tr}}_{j1} - \beta^{\text{tr}}_{j2}$

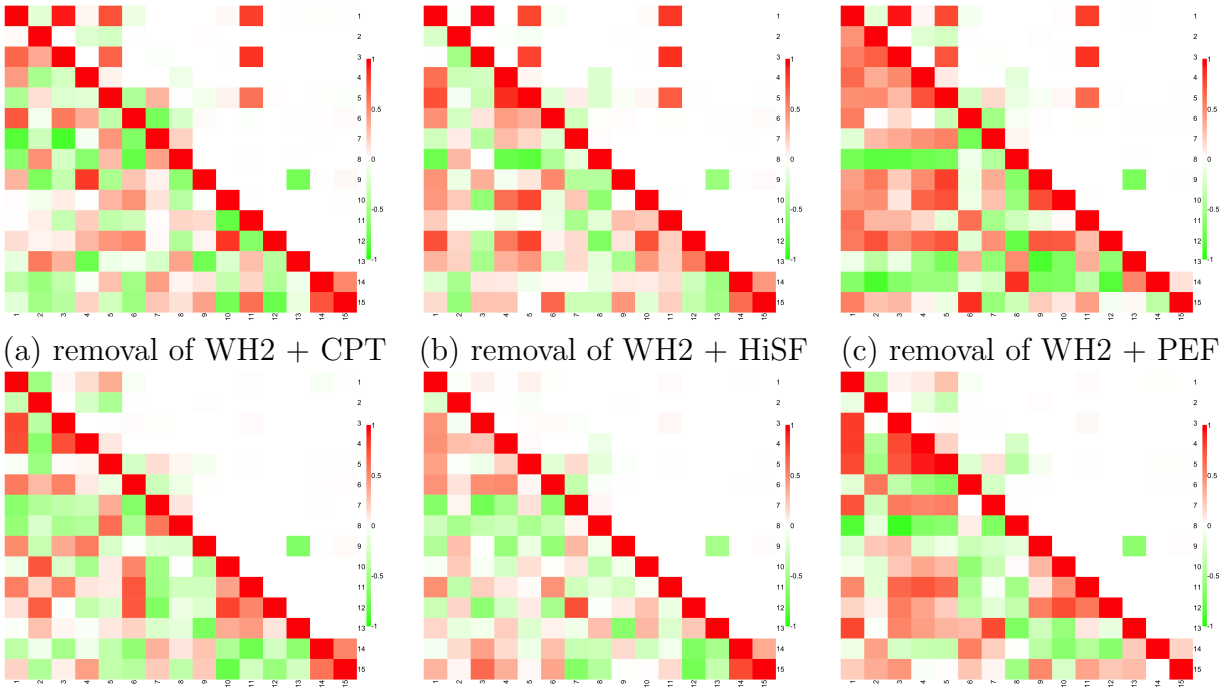
Figure 7: [Simulation 3] Panels (a) and (b) compare the posterior estimates of μ_{ij} and $\beta_{j1} - \beta_{j2}$ to their true values. The dots represent the posterior median estimates, while the vertical lines indicate the 95% credible interval estimates.

5 Mice Gut Microbiome Data Analysis

We applied our method to a subset of the mice gut microbiome data from Patnode et al. (2019). The experiment aimed to understand how individual species in a human gut microbial community interact with others and respond to dietary changes. Specifically, it examined the effects of a human diet representing the upper tertile of saturated fat and lower tertile of fruit and vegetable consumption (HiSF-LoFV) supplemented with differ-

ent fibers. The abundance of beneficial gut microbes may change differently with fiber supplements, and their competition with other microbes may be influenced by the type of supplement. Furthermore, the gut microbiome's membership was manipulated to assess how these changes affect microbial community dynamics. In particular, gnotobiotic mice colonized with a 15-member consortium of human-gut-derived bacterial strains were fed the HiSF-LoFV diet supplemented with various food-grade fibers. Each mouse received a different fiber-supplemented diet for a total of 4 weeks. The microbial composition of their fecal samples was measured over time using 16S rRNA gene sequencing. For our analysis, we focused on the microbial composition at day 12 post-colonization, comparing three diets: two different fiber supplements, 10% citrus pectin (CPT) and 10% pea pectin (PEF), and HiSF-LoFV as a control. This resulted in $N = 69$ samples, with mice assigned to the three diets: HiSF-LoFV, CPT, and PEF, respectively. In addition, one specific speices, *Bacteroides cellulosilyticus* (*B. cellulosilyticus*), was removed from half of the mice at the time of colonization. A list of the 15 strains in the defined bacterial consortium can be found in Supp. Tab. 1, and the number of samples for each of the six experimental conditions defined by diet and the presence/absence of B. *cellulosilyticus* is provided in Supp. Tab. 5.

The lower triangle of the heatmaps in Fig 8 (a)-(f) shows empirical correlation estimates, $\rho^{\text{em}}_{jj'}(\boldsymbol{x})$, computed using $\log(y_{ij} + 0.01)$ after normalization by the log of the total count sample size factor estimates for each of the six conditions. To fit our model, we first constructed covariate vectors for μ and Σ, assuming that diet types influence inter-species competition and abundances, while the removal of *B. cellulosilyticus* has indirect effects on the other species through the microbe-microbe interaction mechanisms (excluding the abundance of *B. cellulosilyticus* itself). The fixed hyperparameters were set similarly to those in Simulation 1, with $K = 8$. The MCMC simulation ran for 160,000 iterations, discarding the first half as burn-in and using the second half for inference. The computation



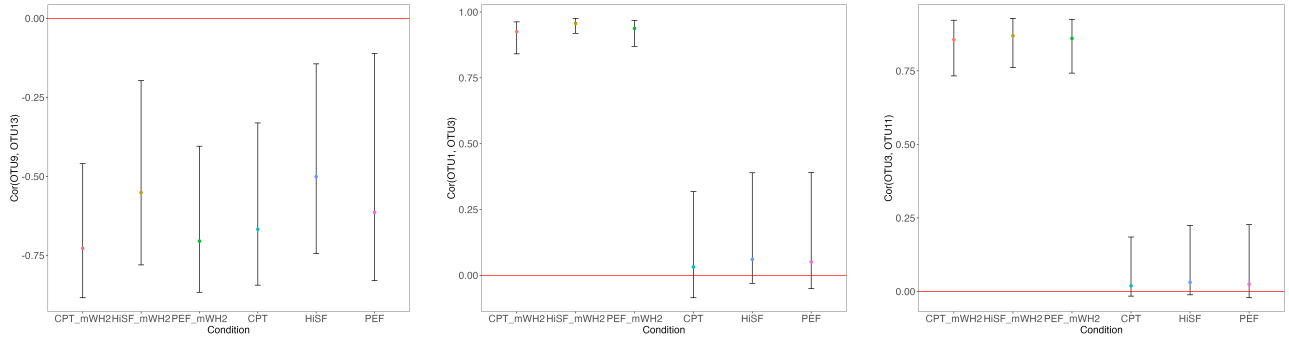
(a) removal of WH2 + CPT (b) removal of WH2 + HiSF (c) removal of WH2 + PEF

(d) CPT (e) HiSF (f) PEF

Figure 8: [Mice Data] The lower left and upper right triangles of the heatmap illustrate empirical correlation estimates $\rho^{\text{em}}jj'(\boldsymbol{x})$ and their posterior estimates $\hat{\rho}_{jj'}(\boldsymbol{x})$ under the six different experimental conditions, respectively.

took 22 minutes on an M1 Mac.

Fig 8(a)-(f) illustrates posterior mean estimates of correlation $\hat{\rho}_{jj'}(\boldsymbol{x})$ of the OTUs under the six different conditions for easier interpretation. Overall $\hat{\rho}_{jj'}(\boldsymbol{x})$'s tend smaller than their corresponding empirical estimate. The interaction mechanism vary noticeably by the removal of *B. cellulosilyticus*, but its change by the diet is minimal. Fig 9 illustrates posterior inference on $\rho_{jj'}(\boldsymbol{x})$ for some pairs of species, where the dots and vertical lines represent the posterior median estimates and 95% credible interval estimates, respectively. From panel (a), the 95% posterior credible interval estimate of $\rho_{jj'}(\boldsymbol{x})$ of OTUs 9 and 13, corresponding to species *Collinsella aerofaciens* and *Ruminococcaceae*, remains below zero under all conditions, indicating that the species' abundances are negatively associated. Panels (b) and (c) show the interval estimates for two pairs, OTUs 1 and 3 and OTUs 3 and 11, where OTUs 1, 3 and 11 correspond to species *B. ovatus*, *B. thetaiotaomicron* and *Odoribacter splanchnicus*. Their interactions are significantly positive in the absence

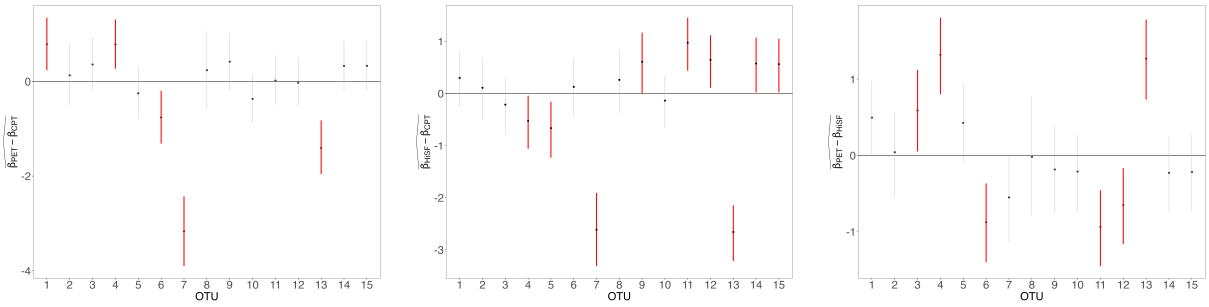


(a) OTUs 9 and 13 (b) OTUs 1 and 3 (c) OTUs 3 and 11

Figure 9: [Mice Data] Posterior inference on correlations $\rho_{jj'}(\boldsymbol{x})$ under the six experimental conditions is illustrated for three pairs. The dots denote posterior median estimates, and the vertical lines 95% credible interval estimates.

of *B. cellulosilyticus*, regardless of the diet but they have no significant interaction in the presence for all diet.

Fig 10 illustrates posterior inferences on the differential microbial abundances across diets. The dots and vertical lines represent posterior median estimates and 95% credible intervals, with intervals not containing zero highlighted in red. From panel (a), the diet supplemented with pea pectin is associated with statistically significant changes in abundance for five strains compared to the diet supplemented with citrus pectin. The effect estimates are positive for OTUs 1 and 4, and negative for OTUs 6, 7, and 13. OTUs 4, 6, and 7 correspond to *B. thetaiotaomicron*, and *B. finegoldii*, respectively. This implies that *B. ovatus* and *B. thetaiotaomicron* are more abundant with the pea pectin diet, while *B. finegoldii*, and *Ruminococcaceae* are less abundant. Panels (b) and (c) compare the control diet to the citrus pectin and pea pectin diets, respectively. From the figure, the abundance of *B. thetaiotaomicron* and *B. finegoldii* changes across all diet types. The shift from the control to the citrus pectin diet is associated with the most significant changes in microbial abundance.



(a) $\beta_{PET} - \beta_{CPT}$ (b) $\beta_{HiSF} - \beta_{CPT}$ (c) $\beta_{PET} - \beta_{HiSF}$

Figure 10: [Mice Data] Posterior estimates of diet effects on mean abundances are illustrated. Posterior mean estimates are denoted by dots, and 95% credible estimates with vertical lines. The intervals that do not contain zero are marked in red.

6 Conclusion

In this paper, we developed a Bayesian joint model of mean and covariance varying with covariates for high-dimensional multivariate count data. This method utilizes a covariate-dependent factor model for the covariance matrix and models the mean abundance using a flexible DP mixture. The model enables the assessment of covariate effects on mean and covariance in tandem. We place a Dir-Horseshoe prior on the covariate-dependent loading matrix to induce sparse feature interactions. The flexible mean mixture kernels handle the excess zeros and over-dispersion problems in the count data. The model is demonstrated through simulations and a real data example with categorical covariates.

Our methods can be further extended by relaxing the linear covariance regression to a more complicated regression formula, such as introducing the transformation of covariates $\sqrt{x}$ and $x^{\frac{3}{2}}$. Introducing different orders of covariates induces a higher order of covariance-covariate relationship. It is important to note that while this extension offers greater flexibility, the same higher order of covariates entering the mean regression would need more exploration (variables selection). With added parameters, more samples are needed to obtain reasonable inferences. A further research field is to study covariate-assisted interactions among temporal and/or longitudinal multivariate count data. In spatial and temporal heteroscedasticity, it's natural to consider the conditional dependence through

time or locations. Fieuws and Verbeke (2006) discussed a pairwise approach jointly modeling of multivariate longitudinal data using mixed models, providing a foundation for understanding covariance structures in such contexts. Adapting the factor loading matrix to a time or location-dependent one has the potential to enhance the inference of interaction structures in other domains.

References

Ahn, S. and Datta, S. (2024). Differential network connectivity analysis for microbiome data adjusted for clinical covariates using jackknife pseudo-values. *BMC bioinformatics*, 25(1):117.

Argelaguet, R., Velten, B., Arnol, D., Dietrich, S., Zenz, T., Marioni, J. C., Buettner, F., Huber, W., and Stegle, O. (2018). Multi-Omics Factor Analysis-a framework for unsupervised integration of multi-omics data sets. *Molecular Systems Biology*, 14:e8124.

Ban, Y., An, L., and Jiang, H. (2015). Investigating Microbial Co-Occurrence Patterns Based on Metagenomic Compositional Data. *Bioinformatics*, 31:3322–3329.

Bhattacharya, A. and Dunson, D. B. (2011). Sparse bayesian infinite factor models. *Biometrika*, pages 291–306.

Cai, T., Ma, Z., and Wu, Y. (2015). Optimal estimation and rank detection for sparse spiked covariance matrices. *Probability theory and related fields*, 161(3):781–815.

Canale, A. and Dunson, D. B. (2011). Bayesian kernel mixtures for counts. *Journal of the American Statistical Association*, 106(496):1528–1539.

Cao, Y., Lin, W., and Li, H. (2019). Large Covariance Estimation for Compositional Data via Composition-Adjusted Thresholding. *Journal of the American Statistical Association*, 114:759–772.

Carroll, R. J. and Ruppert, D. (1982). Robust Estimation in Heteroscedastic Linear Models. *The Annals of Statistics*, 10(2):429 – 441.

Chiu, T. Y., Leonard, T., and Tsui, K.-W. (1996). The matrix-logarithmic covariance model. *Journal of the American Statistical Association*, 91(433):198–210.

De Vito, R., Bellio, R., Trippa, L., and Parmigiani, G. (2019). Multi-study factor analysis. *Biometrics*, 75:337–346.

De Vito, R., Bellio, R., Trippa, L., and Parmigiani, G. (2021). Bayesian multistudy factor analysis for high-throughput biological data. *The annals of applied statistics*, 15(4):1723–1741.

Fan, J., Liao, Y., and Mincheva, M. (2013). Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 75(4):603–680.

Fang, H., Huang, C., Zhao, H., and Deng, M. (2015). Cclasso: correlation inference for compositional data through lasso. *Bioinformatics*, 31(19):3172–3180.

Fieuws, S. and Verbeke, G. (2006). Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics*, 62(2):424–431.

Fong, P. W., Li, W. K., and An, H.-Z. (2006). A simple multivariate arch model specified by random coefficients. *Computational statistics & data analysis*, 51(3):1779–1802.

Fox, E. B. and Dunson, D. B. (2015). Bayesian nonparametric covariance regression. *The Journal of Machine Learning Research*, 16(1):2501–2542.

Friedman, J. and Alm, E. J. (2012). Inferring correlation networks from genomic survey data. *PLOS Computational Biology*, 8:1–11.

Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441.

Green, P. J. and Hastie, D. I. (2009). Reversible jump mcmc. *Genetics*, 155(3):1391–1403.

Haario, H., Saksman, E., and Tamminen, J. (2001). An adaptive metropolis algorithm. *Bernoulli*, pages 223–242.

Hoff, P. D. and Niu, X. (2012). A Covariance Regression Model. *Statistica Sinica*, pages 729–753.

Hotelling, H. (1992). Relations between two sets of variates. In *Breakthroughs in statistics: methodology and distribution*, pages 162–190. Springer.

Ishwaran, H. and James, L. F. (2001). Gibbs Sampling Methods for Stick-breaking Priors. *Journal of the American Statistical Association*, 96:161–173.

Johnstone, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *The Annals of statistics*, 29(2):295–327.

Kurtz, Z. D., Müller, C. L., Miraldi, E. R., Littman, D. R., Blaser, M. J., and Bonneau, R. A. (2015). Sparse and compositionally robust inference of microbial ecological networks. *PLoS computational biology*, 11(5):e1004226.

Leonard, T. and Hsu, J. S. J. (1992). Bayesian Inference for a Covariance Matrix. *The Annals of Statistics*, 20(4):1669 – 1696.

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating Random Correlation Matrices Based on Vines and Extended Onion Method. *Journal of Multivariate Analysis*, 100(9):1989–2001.

Moran, K. R., Turner, E. L., Dunson, D., and Herring, A. H. (2021). Bayesian hierarchical

factor regression models to infer cause of death from verbal autopsy data. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 70(3):532–557.

Ni, Y., Stingo, F. C., and Baladandayuthapani, V. (2019). Bayesian graphical regression. *Journal of the American Statistical Association*, 114(525):184–197.

Niu, X. and Hoff, P. D. (2019). Joint mean and covariance modeling of multiple health outcome measures. *The annals of applied statistics*, 13(1):321.

Niu, Y., Ni, Y., Pati, D., and Mallick, B. K. (2023). Covariate-assisted bayesian graph learning for heterogeneous data. *Journal of the American Statistical Association*, pages 1–15.

Pati, D., Bhattacharya, A., Pillai, N. S., Dunson, D., et al. (2014). Posterior contraction in sparse bayesian factor models for massive covariance matrices. *Annals of Statistics*, 42(3):1102–1130.

Patnode, M. L., Beller, Z. W., Han, N. D., Cheng, J., Peters, S. L., Terrapon, N., Henrissat, B., Le Gall, S., Saulnier, L., Hayashi, D. K., et al. (2019). Interspecies competition impacts targeted manipulation of human gut bacteria by fiber-derived glycans. *Cell*, 179(1):59–73.

Pearson, K. (1901). Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690.

Pourahmadi, M. (2011). Covariance Estimation: The GLM and Regularization Perspectives. *Statistical Science*, 26(3):369 – 387.

Rummel, R. J. (1988). *Applied factor analysis*. Northwestern University Press.

Rutemiller, H. C. and Bowers, D. A. (1968). Estimation in a heteroscedastic regression model. *Journal of the American Statistical Association*, 63(322):552–557.

Sethuraman, J. (1994). A Constructive Definition of the Dirichlet Prior. *Statistica Sinica*, 4:639–650.

Shuler, K., Verbanic, S., Chen, I. A., and Lee, J. (2021). A bayesian nonparametric analysis for zero-inflated multivariate count data with application to microbiome study. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*.

Smyth, G. K. (1989). Generalized linear models with varying dispersion. *Journal of the Royal Statistical Society: Series B (Methodological)*, 51(1):47–60.

Xie, F., Xu, Y., Priebe, C. E., and Cape, J. (2018). Bayesian estimation of sparse spiked covariance matrices in high dimensions. *arXiv preprint arXiv:1808.07433*.

Zhang, S., Shen, Y., Chen, I. A., and and, J. L. (2025). Sparse bayesian group factor model for feature interactions in multiple count tables data. *Journal of the American Statistical Association*, 0(0):1–14.

Zhang, W. and Leng, C. (2012). A moving average cholesky factor model in covariance modelling for longitudinal data. *Biometrika*, 99(1):141–150.