Biostatistics (2024), **0**, 0, *pp.* 1–**??**
doi:10.1093/biostatistics/BCAP˙Supp˙Biostat˙revised

Supplementary Materials for "A Bayesian estimation of covariate assisted principal regression for brain functional connectivity"

HYUNG G. PARK*

Division of Biostatistics, Department of Population Health, New York University School of Medicine,

180 Madison Ave., New York, USA

parkh15@nyu.edu

S1. Relationship between the parameters in tangent space (2.7) and in generative model (2.2)

In this section, we examine the relationship between the parameters $\tilde{\boldsymbol{B}}$ in tangent space (2.7) and $\boldsymbol{B}$ in generative model (2.2). The proposed covariance model is

$$\boldsymbol{Y}_{it} = \boldsymbol{\Gamma}\boldsymbol{\Psi}_i^{\frac{1}{2}}\boldsymbol{s}_{it} + \boldsymbol{L}_i\boldsymbol{\epsilon}_{it} \tag{S.1}$$

with latent factors $\boldsymbol{s}_{it} \sim N(\boldsymbol{0}, \boldsymbol{I}_d)$ and $\boldsymbol{\epsilon}_{it} \sim N(\boldsymbol{0}, \boldsymbol{I}_{p-d})$, of dimensions d and $p-d$, where we parametrize $\boldsymbol{\Psi}_i^{\frac{1}{2}} = \exp(\mathrm{Diag}((\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i)/2)$.

Let us write $\boldsymbol{L}_i\boldsymbol{L}_i^\top = \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top$ by its eigen-decomposition with a normalized eigenvector matrix $\tilde{\boldsymbol{L}}_i \in \mathbb{R}^{p\times(p-d)}$ and the diagonal matrix of eigenvalues $\tilde{\boldsymbol{\Xi}}_i \in \mathbb{R}^{(p-d)\times(p-d)}$. Then we have

*To whom correspondence should be addressed.

$\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma}\boldsymbol{\Psi}_i\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top$, and correspondingly the Euclidean mean, $\bar{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\Sigma}_i$, is

$$\begin{aligned}\bar{\boldsymbol{\Sigma}} &= \frac{1}{n}\sum_{i=1}^n (\boldsymbol{\Gamma}\boldsymbol{\Psi}_i\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top) \\ &= \boldsymbol{\Gamma}\bar{\boldsymbol{\Psi}}\boldsymbol{\Gamma}^\top + \frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top,\end{aligned}$$

where the second term, $\frac{1}{n}\sum_{i=1}^n \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top$, can be further expressed through an eigen-decomposition of the form $\boldsymbol{L}\boldsymbol{\Xi}\boldsymbol{L}^\top$, for some population level $\boldsymbol{L} \in \mathbb{R}^{p\times(p-d)}$ and a diagonal matrix $\boldsymbol{\Xi} \in \mathbb{R}^{(p-d)\times(p-d)}$. This implies

$$\bar{\boldsymbol{\Sigma}} = \boldsymbol{\Gamma}\bar{\boldsymbol{\Psi}}\boldsymbol{\Gamma}^\top + \boldsymbol{L}\boldsymbol{\Xi}\boldsymbol{L}^\top = \boldsymbol{\Gamma}_{\text{aug}}\bar{\boldsymbol{\Psi}}_{\text{aug}}\boldsymbol{\Gamma}_{\text{aug}}^\top$$

where $\boldsymbol{\Gamma}_{\text{aug}} = [\boldsymbol{\Gamma}, \boldsymbol{L}]$ is a $p \times p$ orthonormal matrix and $\bar{\boldsymbol{\Psi}}_{\text{aug}}$ is a $p \times p$ diagonal matrix. Accordingly, we can write $\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}} = \boldsymbol{\Gamma}_{\text{aug}}(\bar{\boldsymbol{\Psi}}_{\text{aug}})^{-\frac{1}{2}}\boldsymbol{\Gamma}_{\text{aug}}^\top$, where $(\bar{\boldsymbol{\Psi}}_{\text{aug}})^{-\frac{1}{2}}$ is a $p \times p$ diagonal matrix. Thus, the $d \times d$ matrix $\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}_{\text{aug}}(\bar{\boldsymbol{\Psi}}_{\text{aug}})^{-\frac{1}{2}}\boldsymbol{\Gamma}_{\text{aug}}^\top\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^\top[\boldsymbol{\Gamma}, \boldsymbol{L}](\bar{\boldsymbol{\Psi}}_{\text{aug}})^{-\frac{1}{2}}[\boldsymbol{\Gamma}, \boldsymbol{L}]^\top\boldsymbol{\Gamma} = \boldsymbol{\Gamma}^\top\boldsymbol{\Gamma}\bar{\boldsymbol{\Psi}}^{-\frac{1}{2}}\boldsymbol{\Gamma}^\top\boldsymbol{\Gamma} = \bar{\boldsymbol{\Psi}}^{-\frac{1}{2}}$ is a diagonal matrix. The model (S.1) implies that

$$\begin{aligned}\text{var}(\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{Y}_{it}) &= \boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Sigma}_i\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma} \\ &= \boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}(\boldsymbol{\Gamma}\exp(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i)\boldsymbol{\Gamma}^\top + \boldsymbol{L}_i\boldsymbol{L}_i^\top)\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma} \\ &= \boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma}\exp(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i))\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma} \\ &= \exp(\text{Diag}(\log(\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma})))\exp(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i))\exp(\text{Diag}(\log(\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma}))) \\ &= \exp(\text{Diag}(\log(\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Gamma})) + \text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i)) \\ &= \exp(\text{Diag}(\log(\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-1}\boldsymbol{\Gamma}) + \boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i))\end{aligned} \tag{S.2}$$

where the third and fourth equalities follow from the diagonality of $\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{\Gamma} \in \mathbb{R}^{d\times d}$ and that any two diagonal matrices can commute. Under tangent-space parametrization (2.7) of the main manuscript, we use the following log-linear model for the variance of dimension-reduced signals,

$$\text{var}(\boldsymbol{\Gamma}^\top\bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}}\boldsymbol{Y}_{it}) = \exp(\text{Diag}(\tilde{\boldsymbol{B}}\boldsymbol{x}_i + \tilde{\boldsymbol{z}}_i)). \tag{S.3}$$

However, the terms on the right-hand side of (S.3) do not directly correspond to those in (S.2) of the original data space. This is because a tangent space is a local linear approximation to Sym_d^+

where the variance objects $\text{var}(\boldsymbol{\Gamma}^\top \bar{\boldsymbol{\Sigma}}^{-\frac{1}{2}} \boldsymbol{Y}_{it})$ $(i = 1, \ldots, n)$ are located, and the "true values" of these parametric terms on the right-hand side of (S.3) are affected by this local linearity assumption. However, for $\boldsymbol{\Psi}_i \approx \bar{\boldsymbol{\Psi}}$ (and thus $\phi_{\bar{\boldsymbol{\Psi}}}(\boldsymbol{\Psi}_i) \approx \boldsymbol{I}_d$), we can approximately identify the relationship between $\boldsymbol{B}$ and $\tilde{\boldsymbol{B}}$ by matching the parametrization in (S.2) and that in (S.3). For prior distributions on $\boldsymbol{z}_i$ and $\tilde{\boldsymbol{z}}_i$ centered at $\mathbf{0}$ (which serve as an identifiability condition), due to the fact that the term $\log(\boldsymbol{\Gamma}^\top \bar{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Gamma})$ in (S.2) is just a constant with respect to i, only the intercept column differs between $\boldsymbol{B}$ and $\tilde{\boldsymbol{B}}$. Specifically, the intercept vector $\boldsymbol{\beta}_0 \in \mathbb{R}^d$ in $\boldsymbol{B}$ under parametrization (S.2) differs by the diagonal elements of $\log(\boldsymbol{\Gamma}^\top \bar{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Gamma})$, i.e.,

$$\boldsymbol{\beta}_0 = \tilde{\boldsymbol{\beta}}_0 - \text{diag}(\log(\boldsymbol{\Gamma}^\top \bar{\boldsymbol{\Sigma}}^{-1} \boldsymbol{\Gamma})),$$

in which $\tilde{\boldsymbol{\beta}}_0 \in \mathbb{R}^d$ is the intercept column in $\tilde{\boldsymbol{B}}$ under the tangent-space parametrization (S.3).

S2. Performance of the expected deviance criterion in determining the number of components

In the main manuscript, we considered a WAIC-based criterion to select the number of projection components d. Specifically, we proposed to pick the dimensionally d of the outcome projection space in a way that maximizes the expected deviance, by estimating the expected log posterior ratio, $E\left[\log \frac{p(\boldsymbol{\Gamma}^\top \boldsymbol{Y}_{it} | \check{\boldsymbol{\Psi}}_i)}{p(\boldsymbol{\Gamma}^\top \boldsymbol{Y}_{it} | \bar{\boldsymbol{\Psi}})}\right] = E\left[\log p(\boldsymbol{\Gamma}^\top \boldsymbol{Y}_{it} | \check{\boldsymbol{\Psi}}_i)\right] - E\left[\log p(\boldsymbol{\Gamma}^\top \boldsymbol{Y}_{it} | \bar{\boldsymbol{\Psi}})\right]$, which is approximated by the WAIC-based formula presented in (2.15) of the main manuscript.

In this section, we report the results from a simulated experiment to demonstrate the performance of this expected deviance criterion. The data were generated following the simulation scenarios in Section 3.1 of the main manuscript. We vary $n \in \{100, 200, 300\}$, $T = T_i \in \{10, 20\}$ and $p \in \{10, 20\}$, with $d = 2$ number of "correct" components. We consider the both cases where the models are correctly specified, and where the models are misspecified in which there are no common eigenvectors $\boldsymbol{\Gamma}$ related to covariates across subjects.

For each scenario, we ran 50 simulation replications and reported the proportion (computed based on the 50 simulation runs) that the estimated number of components is the same as the

"truth" ($d = 2$). For each simulation run with different $d \in \{1, 2, 3, 4\}$ values, we computed the criterion (2.15) using the posterior MCMC samples, and selected d that minimizes the criterion.

In Table S.1 below, we display the proportions of the cases where the correct number of components are identified out of 50 simulation runs. In all scenarios, the results consistently show the proportions very close 1 ($\geqslant 0.86$), suggesting the reliability of this expected deviance criterion in choosing the number of components.

		Correctly specified		Model misspecified	
		$p = 10$	$p = 20$	$p = 10$	$p = 20$
$n = 100$	$T = 10$	0.94	0.94	0.92	0.98
	$T = 20$	0.86	0.98	0.94	0.94
$n = 200$	$T = 10$	1.00	1.00	0.94	0.98
	$T = 20$	1.00	1.00	1.00	1.00
$n = 300$	$T = 10$	1.00	1.00	1.00	0.98
	$T = 20$	0.96	1.00	0.98	0.98

Table S.1. The proportion of the cases for identifying the correct number of components out of 50 simulation replications, for different simulation scenarios with $n \in \{100, 200, 300\}$, $T \in \{10, 20\}$ and $p \in \{10, 20\}$, under both correctly specified and misspecified model conditions.

Alternatively, the number of components d can be determined based on a metric that measures the level of "deviation from diagonality" (DfD) of the sample version of the dimension-reduced covariance $\boldsymbol{\Psi}_i$, as in Zhao *and others* (2021). If this sample version, $\boldsymbol{\Gamma}^\top \widehat{\boldsymbol{\Sigma}}_i \boldsymbol{\Gamma} = \widehat{\boldsymbol{\Psi}}_i$, where $\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{T_i} \sum_{t=1}^{T_i} \boldsymbol{Y}_{it} \boldsymbol{Y}_{it}^\top$, deviates from the diagonality assumed in the model, then the following quantity $\log \mathrm{DfD}(\{\widehat{\boldsymbol{\Psi}}_i\}_{i=1}^n) := \frac{1}{\sum_{i=1}^n T_i} \sum_{i=1}^n T_i \left(\log|\mathrm{diag}(\widehat{\boldsymbol{\Psi}}_i)| - \log|\widehat{\boldsymbol{\Psi}}_i| \right)$, tends to deviate from 0, where $\mathrm{diag}(\widehat{\boldsymbol{\Psi}}_i)$ is a $d \times d$ diagonal matrix with its diagonal elements given by the diagonal elements of $\widehat{\boldsymbol{\Psi}}_i$, and the determinants $|\mathrm{diag}(\widehat{\boldsymbol{\Psi}}_i)|$ and $|\widehat{\boldsymbol{\Psi}}_i|$ correspond to the determinants of the $d \times d$ matrices $\mathrm{diag}(\widehat{\boldsymbol{\Psi}}_i)$ and $\widehat{\boldsymbol{\Psi}}_i$, respectively. From Hadamard's inequality, $\log \mathrm{DfD}(\{\widehat{\boldsymbol{\Psi}}_i\}_{i=1}^n) \geqslant 0$ (Zhao *and others*, 2021; Flury and Gautschi, 1986), and the equality occurs if the matrix $\widehat{\boldsymbol{\Psi}}_i$ is perfectly diagonal. The requirement for $\widehat{\boldsymbol{\Psi}}_i$ to be diagonal may become more stringent as the dimensionality d increases (Zhao *and others*, 2021), and thus, with a too large number of components d, the quantity $\log \mathrm{DfD}(\{\widehat{\boldsymbol{\Psi}}_i\}_{i=1}^n)$ tends to deviate from 0. We can utilize posterior MCMC samples

to compute the posterior mean, $E[\log \mathrm{DfD}(\{\widehat{\boldsymbol{\Psi}}_i\}_{i=1}^n)|\mathcal{D}]$, by using the HMC sampler outputs on $\widehat{\boldsymbol{\Psi}}_i^{(s)} = \boldsymbol{\Gamma}^{(s)\top}\widehat{\boldsymbol{\Sigma}}_i\boldsymbol{\Gamma}^{(s)}$ for each subject i $(i = 1, \ldots, n)$, where $\boldsymbol{\Gamma}^{(s)}$ is the sth posterior sampler output for $\boldsymbol{\Gamma}$ from the Markov chain. Heuristically, we can increase the projection dimension from 1 to some reasonable number, $\check{d}(\leqslant p)$, and choose a suitable number right before a large jump in the metric, or identify an appropriate number by finding the largest d such that the metric $E[\log \mathrm{DfD}(\{\widehat{\boldsymbol{\Psi}}_i\}_{i=1}^n)|\mathcal{D}]$ is less than some threshold, i.e., choose the largest d such that $E[\log \mathrm{DfD}(\{\widehat{\boldsymbol{\Psi}}_i\}_{i=1}^n)|\mathcal{D}] \leqslant c$ for some cutoff value c. However, the choice of c is less objective compared to the WAIC criterion discussed in the main manuscript, and we did not pursue this approach to optimize the rank of the model.

S3. Simulation illustration: $p = 10$ cases

The model parameter estimation performance for $p = 10$ case is provided in Figure S.1.

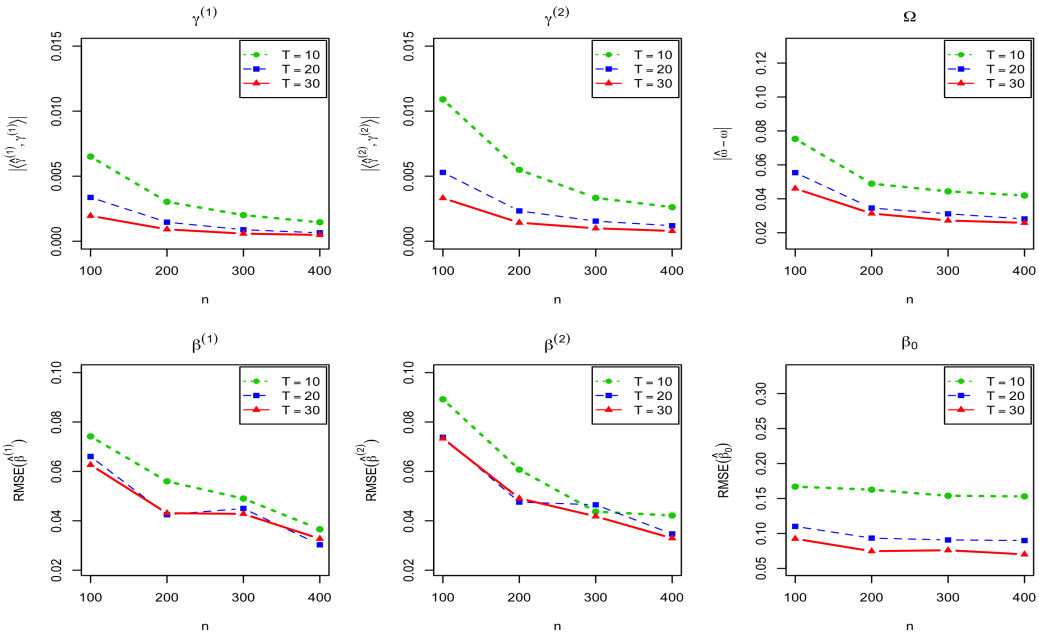

Fig. S.1. The model parameter estimation performance for $p = 10$ case, for the loading coefficient vectors $\boldsymbol{\gamma}^{(k)}$ $(k = 1, 2)$, the elements of the random effect covariance matrix $\boldsymbol{\Omega}$, the regression coefficients $\boldsymbol{\beta}^{(k)}$ $(k = 1, 2)$, and the intercept $\boldsymbol{\beta}_0$, averaged across 50 simulation replications, with varying $n \in \{100, 200, 300, 400\}$ and $T \in \{10, 20, 30\}$.

S4. Simulation illustration: When the model is misspecified

In this section, we report the estimation performance for $p = 10$ in terms of the absolute cosine similarity and the RMSE, when the model is misspecified: 1) when no random effect component $\boldsymbol{z}_i$ is considered in the model estimation, which is reported in Figure S.2; and 2) when there are no common "signal" (covariate relevant) eigenvectors across subjects, which is reported in Figure S.3. In addition, we report these misspecified cases' 95% credible interval coverage proportions in Table S.2.

When no random effect is used (Figure S.2), while the estimation performance results in terms of the absolute cosine similarity and the RMSE are comparable with those of the correctly-specified cases, the actual coverage proportions of the 95% credible intervals tend to deviate from the nominal level (the left columns in Table S.2), particularly for the regression coefficients $\boldsymbol{\beta}^{(k)}$, underestimating the uncertainties. When there is no common "signal" eigenvectors across subjects (Figure S.3), there was an added bias in the estimation of $\boldsymbol{\beta}^{(k)}$, and the coverage tended to be smaller than the nominal level (the right columns in Table S.2).

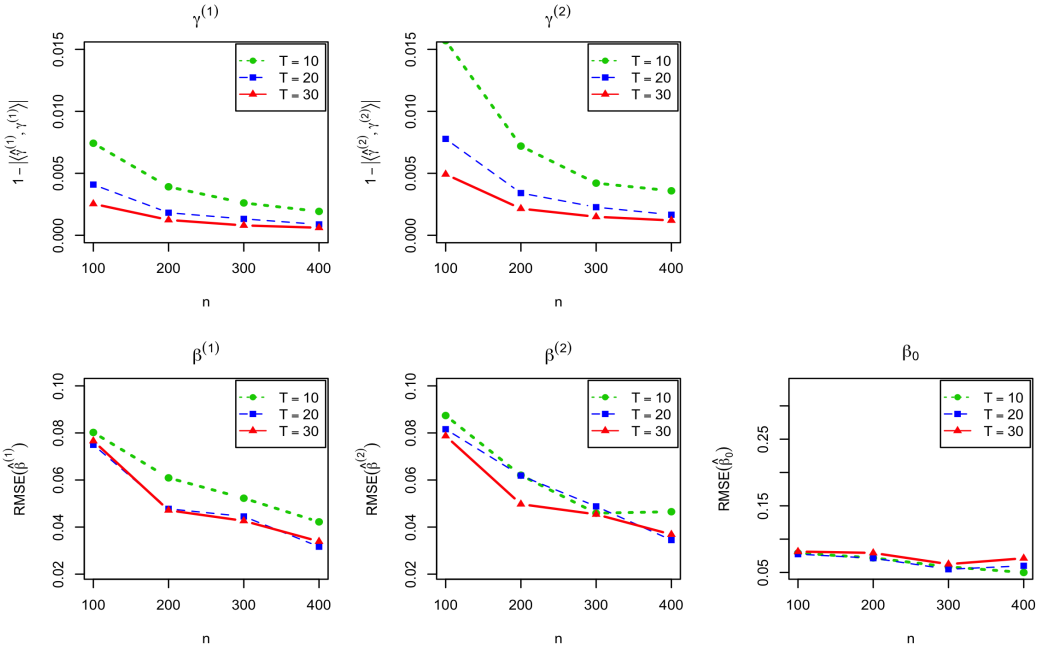

Fig. S.2. The cases where no random effect component $\boldsymbol{z}_i$ is considered in the model estimation. Model parameter estimation performance for the loading coefficient vectors $\boldsymbol{\gamma}^{(k)}$ $(k = 1, 2)$, the regression coefficients $\boldsymbol{\beta}^{(k)}$ $(k = 1, 2)$, and the intercept $\boldsymbol{\beta}_0$, averaged across 50 simulation replications, with varying $n \in \{100, 200, 300, 400\}$ and $T \in \{10, 20, 30\}$.

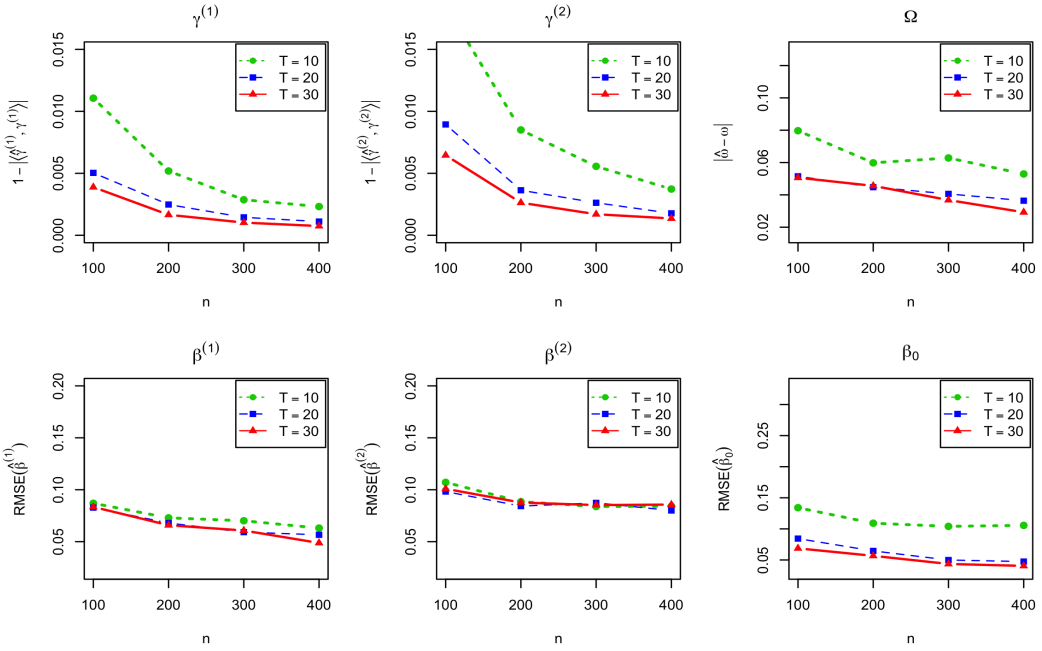

Fig. S.3. The cases where there are no common "signal" eigenvectors across subjects. Model parameter estimation performance for the loading coefficient vectors $\boldsymbol{\gamma}^{(k)}$ $(k = 1, 2)$, the elements of the random effect covariance matrix $\boldsymbol{\Omega}$, the regression coefficients $\boldsymbol{\beta}^{(k)}$ $(k = 1, 2)$, and the intercept $\boldsymbol{\beta}_0$, averaged across 50 simulation replications, with varying $n \in \{100, 200, 300, 400\}$ and $T \in \{10, 20, 30\}$.

		No random effect $\boldsymbol{z}_i$ in the model					No common eigenvectors $\boldsymbol{\Gamma}$				
n	T	$\boldsymbol{\gamma}^{(1)}$	$\boldsymbol{\gamma}^{(2)}$	$\boldsymbol{\beta}^{(1)}$	$\boldsymbol{\beta}^{(2)}$	$\boldsymbol{\Omega}$	$\boldsymbol{\gamma}^{(1)}$	$\boldsymbol{\gamma}^{(2)}$	$\boldsymbol{\beta}^{(1)}$	$\boldsymbol{\beta}^{(2)}$	$\boldsymbol{\Omega}$
100	10	0.90	0.90	0.80	0.76	NA	0.81	0.88	0.88	0.83	0.91
	20	0.86	0.86	0.68	0.60	NA	0.83	0.85	0.87	0.78	0.94
	30	0.89	0.88	0.59	0.57	NA	0.75	0.80	0.84	0.69	0.87
200	10	0.89	0.85	0.76	0.70	NA	0.85	0.88	0.86	0.74	0.83
	20	0.92	0.91	0.70	0.60	NA	0.81	0.85	0.82	0.67	0.77
	30	0.90	0.92	0.62	0.57	NA	0.77	0.81	0.81	0.62	0.75
300	10	0.88	0.92	0.73	0.78	NA	0.89	0.87	0.84	0.64	0.67
	20	0.85	0.89	0.62	0.60	NA	0.85	0.90	0.82	0.54	0.78
	30	0.90	0.93	0.58	0.54	NA	0.79	0.83	0.72	0.52	0.75
400	10	0.91	0.93	0.78	0.74	NA	0.86	0.91	0.78	0.54	0.71
	20	0.93	0.92	0.71	0.68	NA	0.86	0.88	0.76	0.43	0.74
	30	0.92	0.93	0.60	0.56	NA	0.84	0.87	0.78	0.36	0.81

Table S.2. The proportion of time that 95% posterior credible intervals contain the true values of the projection loading vectors $\boldsymbol{\gamma}^{(k)}$ $(k = 1, 2)$, regression coefficients $\boldsymbol{\beta}^{(k)}$ $(k = 1, 2)$, and elements of $\boldsymbol{\Omega}$, averaged across 50 simulation replications, with varying $n \in \{100, 200, 300, 400\}$ and $T \in \{10, 20, 30\}$. Coverage computed for each entry, then averaged within components ($\boldsymbol{\gamma}^{(k)}$, $\boldsymbol{\beta}^{(k)}$ and $\boldsymbol{\Omega}$) and across the simulation replications (rounded to two significant digits).

S5. Application: Estimated model parameters

The estimated model parameters (posterior median and 95% credible intervals) estimated on the first session of the HCP dataset are displayed in Figures S.4, S.5 and S.6 for $\mathbf{\Gamma}$, $\boldsymbol{B}$ and $\mathbf{\Omega}$, respectively.

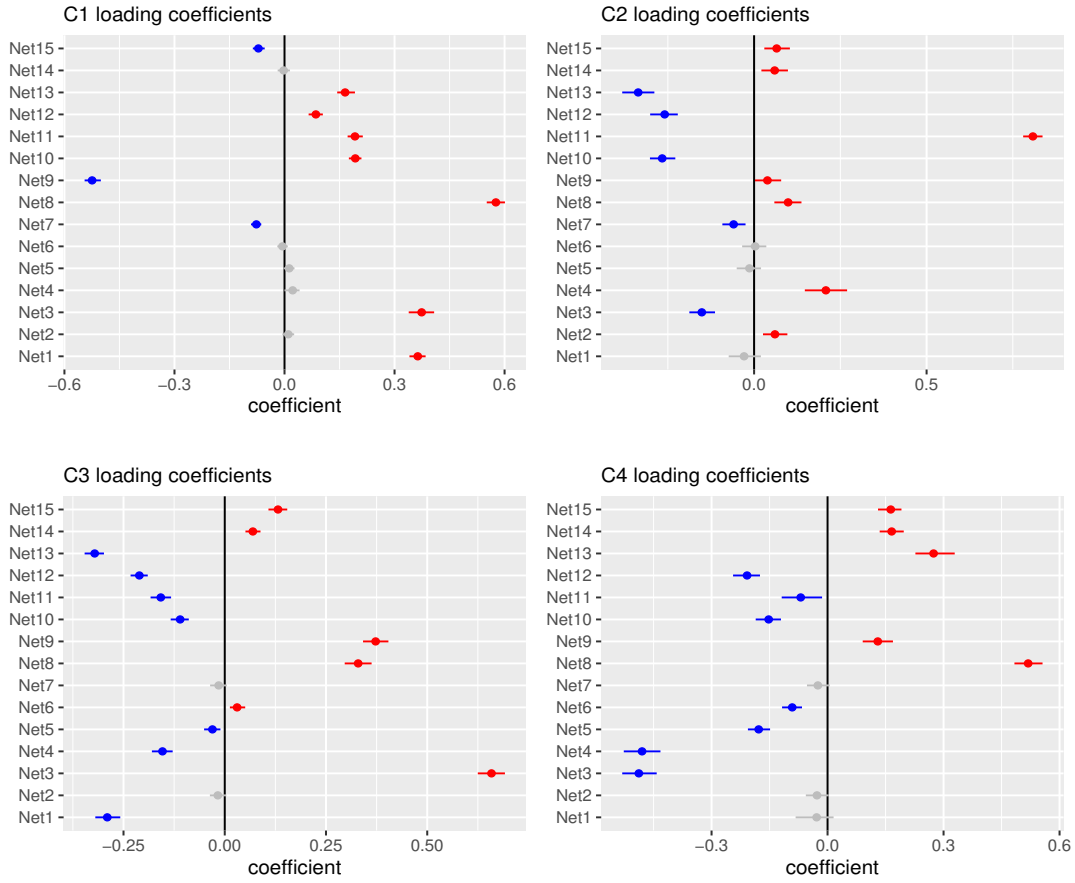

Fig. S.4. Credible intervals of the projection loading coefficients $\boldsymbol{\gamma}^{(1)}$, $\boldsymbol{\gamma}^{(2)}$, $\boldsymbol{\gamma}^{(3)}$ and $\boldsymbol{\gamma}^{(4)}$, i.e., those from $\mathbf{\Gamma} = [\boldsymbol{\gamma}^{(1)}; \boldsymbol{\gamma}^{(2)}; \boldsymbol{\gamma}^{(3)}; \boldsymbol{\gamma}^{(4)}] \in \mathbb{R}^{15\times 4}$ associated with the four outcome components C_1, C_2, C_3 and C_4, identified to be associated with covariates.

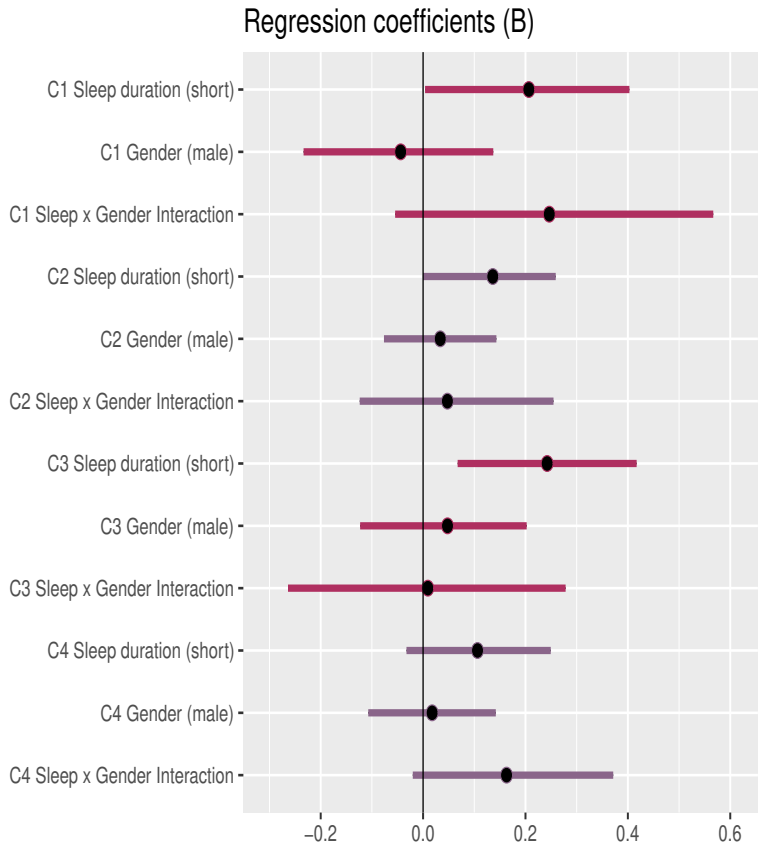

Fig. S.5. The posterior medians and 95% credible intervals of the regression coefficients $\boldsymbol{\beta}^{(1)}$, $\boldsymbol{\beta}^{(2)}$, $\boldsymbol{\beta}^{(3)}$ and $\boldsymbol{\beta}^{(4)}$, associated with the four estimated components C_1, C_2, C_3 and C_4.

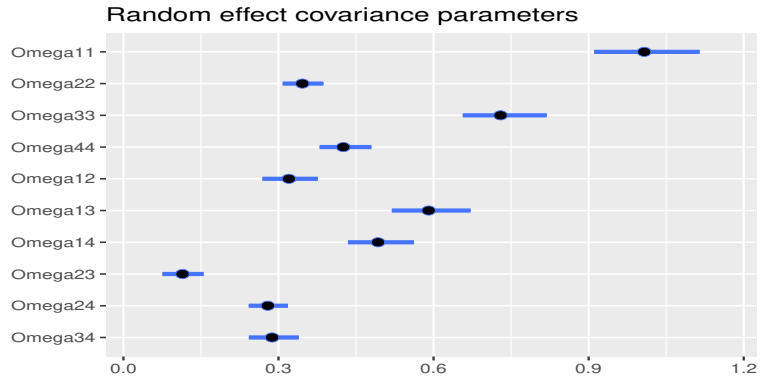

Fig. S.6. The posterior medians and 95% credible intervals of the random effect covariance components ($\boldsymbol{\Omega} \in \mathbb{R}^{4\times 4}$) (the 4 diagonal and 6 lower triangular elements) of the estimated projected signals $\boldsymbol{\Gamma}^\top \boldsymbol{Y}_{it} \in \mathbb{R}^4$, that model the residual covariances of the estimated projected signals not captured by the covariates $\boldsymbol{x}_i$.

S6. Application: Estimated model parameters from the four sessions of HCP

Besides focusing on the first session of the HCP data, in Table S.3 below, we report the models separately fitted for each of the 4 sessions of HCP. We report the posterior means (and 95% credible intervals) for the SleepDuration, Gender and their interaction regression coefficients ($\boldsymbol{\beta}^{(1)} = (\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)})^\top$, $\boldsymbol{\beta}^{(2)} = (\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)})^\top$, $\boldsymbol{\beta}^{(3)} = (\beta_1^{(3)}, \beta_2^{(3)}, \beta_3^{(3)})^\top$ and $\boldsymbol{\beta}^{(4)} = (\beta_1^{(4)}, \beta_2^{(4)}, \beta_3^{(4)})^\top$) and the random effect variances ($\omega_{11}, \omega_{22}, \omega_{33}$ and ω_{44}), associated with the 4 projection components (C_1, C_2, C_3 and C_4), where the models are separately fitted for each of the 4 sessions of HCP. The estimated model parameters in Table S.3 suggest that these regression parameters exhibits a high level of consistency across all four scanning sessions.

For each estimated (posterior mean) projection loading coefficient $\hat{\boldsymbol{\gamma}}^{(k)}$ ($k = 1, \ldots, 4$), we compute $\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(k)\top}\boldsymbol{Y}_{it})) = \log(\hat{\boldsymbol{\gamma}}^{(k)\top}\widehat{\boldsymbol{\Sigma}}_i\hat{\boldsymbol{\gamma}}^{(k)})$ ($i = 1, \ldots, n$), where $\widehat{\boldsymbol{\Sigma}}_i = \frac{1}{T_i}\sum_{t=1}^{T_i}\boldsymbol{Y}_{it}\boldsymbol{Y}_{it}^\top$. In Table S.3, we report the sample mean of the log-variances, i.e., $\frac{1}{n}\sum_{i=1}^{n}\log(\hat{\boldsymbol{\gamma}}^{(k)\top}\widehat{\boldsymbol{\Sigma}}_i\hat{\boldsymbol{\gamma}}^{(k)})$ (and the sample standard deviation), computed for each of the 4 projection components (C_1, C_2, C_3 and C_4) and for each of the 4 sessions.

Furthermore, for each (the kth) projection component (C_1, C_2, C_3 and C_4), we compute the intra-cluster correlation (ICC) coefficient value of the log-variance $\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(k)\top}\boldsymbol{Y}_{it}))$ ($i = 1, \ldots, n$) across the 4 sessions. This ICC quantifies the fraction of the variability in log-variance attributed to between-subject differences, relative to the combined variability from both between-subject and between-session differences.

In Table S.3, the ICC values are 0.84, 0.72, 0.84 and 0.83, for the 4 identified network components, indicating that the functional connectivity associated with the 4 identified components exhibits a relatively high level of consistency across all four scanning sessions.

		Session 1	Session 2	Session 3	Session 4	ICC
C_1	Mean (SD) of $\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(1)\top}\boldsymbol{Y}_{it}))$	8.70 (0.98)	8.41 (0.99)	8.48 (0.99)	8.74 (1.01)	0.844
	SleepDuration $\beta_1^{(1)}$	0.20 (0.00 0.40)	0.27 (0.05 0.50)	0.32 (0.11 0.55)	0.21 (-0.04 0.43)	
	Gender $\beta_2^{(1)}$	-0.04 (-0.23 0.14)	0.27 (0.08 0.46)	0.23 (0.05 0.42)	0.00 (-0.19 0.19)	
	Interaction $\beta_3^{(1)}$	0.25 (-0.05 0.57)	0.11 (-0.22 0.42)	0.15 (-0.21 0.43)	0.27 (-0.06 0.62)	
	Random-effect variance ω_{11}	1.01 (0.91 1.11)	1.00 (0.89 1.12)	0.98 (0.87 1.09)	1.04 (0.94 1.16)	
C_2	Mean (SD) of $\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(2)\top}\boldsymbol{Y}_{it}))$	7.11 (0.60)	7.49 (0.59)	7.49 (0.64)	7.33 (0.61)	0.721
	SleepDuration $\beta_1^{(2)}$	0.13 (0.00 0.26)	0.15 (0.03 0.28)	0.17 (0.04 0.31)	0.24 (0.10 0.36)	
	Gender $\beta_2^{(2)}$	0.03 (-0.08 0.14)	0.13 (0.02 0.23)	0.06 (-0.05 0.18)	0.08 (-0.05 0.21)	
	Interaction $\beta_3^{(2)}$	0.06 (-0.12 0.25)	0.00 (-0.18 0.19)	0.08 (-0.13 0.27)	-0.14 (-0.34 0.05)	
	Random-effect variance ω_{22}	0.35 (0.31 0.39)	0.31 (0.27 0.35)	0.35 (0.31 0.39)	0.35 (0.31 0.40)	
C_3	Mean (SD) of $\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(3)\top}\boldsymbol{Y}_{it}))$	8.85 (0.86)	8.58 (0.89)	8.59 (0.86)	8.76 (0.81)	0.843
	SleepDuration $\beta_1^{(3)}$	0.24 (0.07 0.42)	0.23 (0.04 0.42)	0.23 (0.03 0.44)	0.10 (-0.10 0.28)	
	Gender $\beta_2^{(3)}$	0.04 (-0.12 0.20)	0.15 (-0.01 0.31)	0.19 (0.04 0.35)	0.06 (-0.10 0.22)	
	Interaction $\beta_3^{(3)}$	0.01 (-0.26 0.28)	0.03 (-0.26 0.32)	0.03 (-0.27 0.31)	0.13 (-0.12 0.40)	
	Random-effect variance ω_{33}	0.73 (0.66 0.82)	0.77 (0.68 0.87)	0.72 (0.65 0.81)	0.63 (0.56 0.70)	
C_4	Mean (SD) of $\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(4)\top}\boldsymbol{Y}_{it}))$	8.28 (0.70)	8.24 (0.66)	8.33 (0.70)	8.26 (0.74)	0.833
	SleepDuration $\hat{\beta}_1^{(4)}$	0.11 (-0.03 0.25)	0.19 (0.06 0.33)	0.18 (0.04 0.34)	0.23 (0.07 0.39)	
	Gender $\beta_2^{(4)}$	0.02 (-0.11 0.14)	0.14 (0.02 0.25)	0.13 (-0.01 0.26)	0.09 (-0.05 0.23)	
	Interaction $\beta_3^{(4)}$	0.17 (-0.02 0.37)	-0.05 (-0.25 0.16)	0.05 (-0.18 0.26)	0.05 (-0.18 0.27)	
	Random-effect variance ω_{44}	0.43 (0.38 0.48)	0.37 (0.32 0.41)	0.40 (0.36 0.45)	0.50 (0.45 0.57)	

Table S.3. The posterior means (and 95% credible intervals) of the regression coefficients and the random effect variance components (ω_{kk}), along with the log-variance of the projected signals ($\log(\text{var}(\hat{\boldsymbol{\gamma}}^{(k)\top}\boldsymbol{Y}_{it}))$), for each of the 4 estimated projection components and for each of the 4 sessions of HCP.

S7. Description of the Contrast Map $\boldsymbol{\Gamma}(Diag(\boldsymbol{B\delta}))\boldsymbol{\Gamma}^\top$

We can think of a "contrast" map as the changes in the log covariance due to alterations in certain variables (i.e., covariates). To understand the impact of these changes on the covariance, we first focus on the "variance ratio" (VR). This helps us grasp how the variance of the signals is expected to change due to the changes in the covariates represented by $\boldsymbol{\delta}$.

Estimating the probabilistic model (S.1) allows us to identify eigenvectors $\boldsymbol{\Gamma}$ relevant to covariates, simplifying our analysis of the covariates-covariance association, particularly when the global log transformation is involved in modeling for tangent-space mapping. Having identified relevant eigenvectors $\boldsymbol{\Gamma}$ that align with the covariates' impact directions, the global log transformation maintains their orientation associated with covariate effects, thus preserving the interpretability of pairwise connectivity differences in the log contrast matrix given a $\boldsymbol{\delta}$-change in covariates, upon regression modeling in tangent space.

More specifically, each pair of the off diagonal elements of the log contrast matrix $\boldsymbol{\Gamma}\left(\text{Diag}(\boldsymbol{B}\boldsymbol{\delta})\right)\boldsymbol{\Gamma}^\top$ (see below) maintain its interpretability as the corresponding pairwise connectivity difference due to the $\boldsymbol{\delta}$ contrast in the covariates (but on the log scale). If we assume covariate dependent eigenvectors, then the global log transformation changes the covariate's impact directions, leading to an interpretational challenge.

Under model (2.1) of the main manuscript, we have a covariance decomposition

$$\boldsymbol{\Sigma}_i = \boldsymbol{\Gamma}\exp(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i))\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top,$$

where $\boldsymbol{L}_i\boldsymbol{L}_i^\top$ is eigen-decomposed into $\tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top$. Now we consider the impact of a $\boldsymbol{\delta}$-change in covariates on the covariance on the log scale,

$$\begin{aligned}
&\log\left(\boldsymbol{\Sigma}_{i,\boldsymbol{\delta}}\right) - \log\left(\boldsymbol{\Sigma}_i\right)\\
&= \log\left(\boldsymbol{\Gamma}\exp(\text{Diag}(\boldsymbol{B}(\boldsymbol{x}_i+\boldsymbol{\delta}) + \boldsymbol{z}_i))\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top\right) - \log\left(\boldsymbol{\Gamma}\exp(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i))\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\tilde{\boldsymbol{\Xi}}_i\tilde{\boldsymbol{L}}_i^\top\right)\\
&= \boldsymbol{\Gamma}\log\left(\exp(\text{Diag}(\boldsymbol{B}(\boldsymbol{x}_i+\boldsymbol{\delta}) + \boldsymbol{z}_i))\right)\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\log(\tilde{\boldsymbol{\Xi}}_i)\tilde{\boldsymbol{L}}_i^\top - \left(\boldsymbol{\Gamma}\log\left(\exp(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i))\right)\boldsymbol{\Gamma}^\top + \tilde{\boldsymbol{L}}_i\log(\tilde{\boldsymbol{\Xi}}_i)\tilde{\boldsymbol{L}}_i^\top\right)\\
&= \boldsymbol{\Gamma}\left(\text{Diag}(\boldsymbol{B}(\boldsymbol{x}_i+\boldsymbol{\delta}) + \boldsymbol{z}_i)\right)\boldsymbol{\Gamma}^\top - \boldsymbol{\Gamma}\left(\text{Diag}(\boldsymbol{B}\boldsymbol{x}_i + \boldsymbol{z}_i)\right)\boldsymbol{\Gamma}^\top\\
&= \boldsymbol{\Gamma}\left(\text{Diag}(\boldsymbol{B}\boldsymbol{\delta})\right)\boldsymbol{\Gamma}^\top.
\end{aligned} \tag{S.4}$$

Note, the logarithm of SPD matrices $\boldsymbol{\Sigma}_{i,\boldsymbol{\delta}}$ and $\boldsymbol{\Sigma}_i$, are taken relative to a chosen reference point, which defines the tangent space, and the choice of tangent space may impact the results. However, when we subtract the logarithms of two matrices, the reference point no longer affects the result because the subtraction operation is performed in the tangent space itself. Therefore, the choice of the reference point for the Log map does not impact the interpretation of $\log\left(\boldsymbol{\Sigma}_{i,\boldsymbol{\delta}}\right) - \log\left(\boldsymbol{\Sigma}_i\right)$ in (S.4)

S8. Application: Similarity between the components identified by CAP and Bayesian CAP

In this section, we display the cosine similarity (similarity between -1 and 1, with 0 indicating orthogonal) of the estimated projection directions from CAP (Zhao *and others*, 2021) (in their

first four leading components) and those from the proposed method. Table S.4 below shows positive association for each projection direction with the similarity at least 0.4. In Table S.4, we also report the CAP regression coefficients (with 95% bootstrap confidence intervals) for each projected outcome component.

		Similarity with BayesianCAP				CAP coefficient (95% bootstrap CI)		
		C_1	C_2	C_3	C_4	SleepDuration	Gender	Interaction
CAP	Component 1	0.34	0.70	-0.11	0.50	0.18 (-0.02 0.37)	-0.02 (-0.19 0.14)	0.26 (-0.03 0.56)
	Component 2	0.71	-0.50	-0.25	0.43	0.16 (-0.05 0.36)	-0.07 (-0.25 0.12)	0.23 (-0.07 0.52)
	Component 3	0.17	-0.05	0.42	-0.24	0.23 (0.09 0.36)	0.14 (0.03 0.26)	0.02 (-0.17 0.20)
	Component 4	0.47	0.02	0.41	-0.04	0.16 (-0.02 0.34)	0.01 (-0.15 0.15)	0.19 (-0.04 0.43)

Table S.4. Cosine similarity between the components identified by CAP and Bayesian CAP (in their first four leading components), and CAP regression coefficients (and 95% bootstrap confidence intervals) for each of the CAP projection components.

S9. Interpretation of the projection loading coefficients

In this section, we discuss how the projection loading coefficients $\boldsymbol{\gamma}^{(k)}$ can be interpreted in relation to the corresponding regression coefficient $\boldsymbol{\beta}^{(k)}$, as primarily considered in Zhao *and others* (2021).

For subject i with subject-specific pairwise correlation matrix $\boldsymbol{\Sigma}_i = [\sigma_{jj',i}] \in \mathrm{Sym}_p^+$, the log-variance of the (kth) projection component is

$$\log\left(\boldsymbol{\gamma}^{(k)\top}\boldsymbol{\Sigma}_i\boldsymbol{\gamma}^{(k)}\right) = \log\left(1 + \sum_{j\neq j'}^{p} \gamma_j^{(k)}\gamma_{j'}^{(k)}\sigma_{jj',i}\right) = \beta_0^{(k)} + x_i\beta_1^{(k)} + z_i^{(k)} \tag{S.5}$$

where the kth loading direction is $\boldsymbol{\gamma}^{(k)} = (\gamma_1^{(k)}, \ldots, \gamma_j^{(k)}, \ldots, \gamma_{j'}^{(k)}, \ldots, \gamma_p^{(k)})^\top \in \mathbb{R}^p$ (subject to $\|\boldsymbol{\gamma}^{(k)}\| = 1$). Model (S.5) indicates that, for any region pair (j, j'), the signs of the corresponding loading coefficient product $\gamma_j^{(k)}\gamma_{j'}^{(k)}$ and of the regression coefficient $\beta_1^{(k)}$ determine the direction of association between the covariate x_i and the pairwise correlation $\sigma_{jj',i}$. In (S.5), for the simplicity in illustration, $\boldsymbol{\Sigma}_i$ is assumed to be a correlation matrix, however, the same interpretation in terms of the direction of association is applicable to any covariance matrix $\boldsymbol{\Sigma}_i \in \mathrm{Sym}_p^+$.

First let us consider the case $\beta_1^{(k)} > 0$; if two brain regions (j and j') have the same loading

signs, the model suggests a positive association between the correlation $\sigma_{jj',i}$ and the covariate x_i, holding all the other components constant; on the other hand, if the brain regions have opposite loading signs, the model suggests a negative association between the correlation $\sigma_{jj',i}$ and the covariate x_i, holding all the other components constant.

Now let us consider the case $\beta_1^{(k)} < 0$; if two brain regions (j and j') have the same loading signs, the model suggests a negative association between the correlation $\sigma_{jj',i}$ and the covariate x_i, holding all the other components constant; on the other hand, if the brain regions have opposite loading signs, the model suggests a positive association between the correlation $\sigma_{jj',i}$ and the covariate x_i, holding all the other components constant.

S10. Application: Results from element-wise regression

One conventional method for analyzing group ICA data involves initially computing subject-level Pearson correlations between the ICs, which are then Fisher z-transformed. This process is performed on ($p(p-1)/2 =$) 105 pairs of correlations (calculated from 15 ICs), while the element-wise log transformation was conducted on the $p = 15$ diagonal elements. A total of 120 element-wise linear regressions were then conducted on SleepDuration, Gender and their interaction, and p-values are corrected for multiplicity using the Benjamini-Hochberg (BH) (Benjamini and Hochberg, 1995) procedure to control the false discovery rate (FDR) at 0.05. The patterns of the connectivity differences from this mass-univariate approach are presented in Figure S.7 below. The patterns are similar with the results from Bayesian CAP in Figure 4 of the main manuscript. However, compared to Bayesian CAP, far fewer statistically significant elements were identified (13 vs. 77, out of 480 elements).

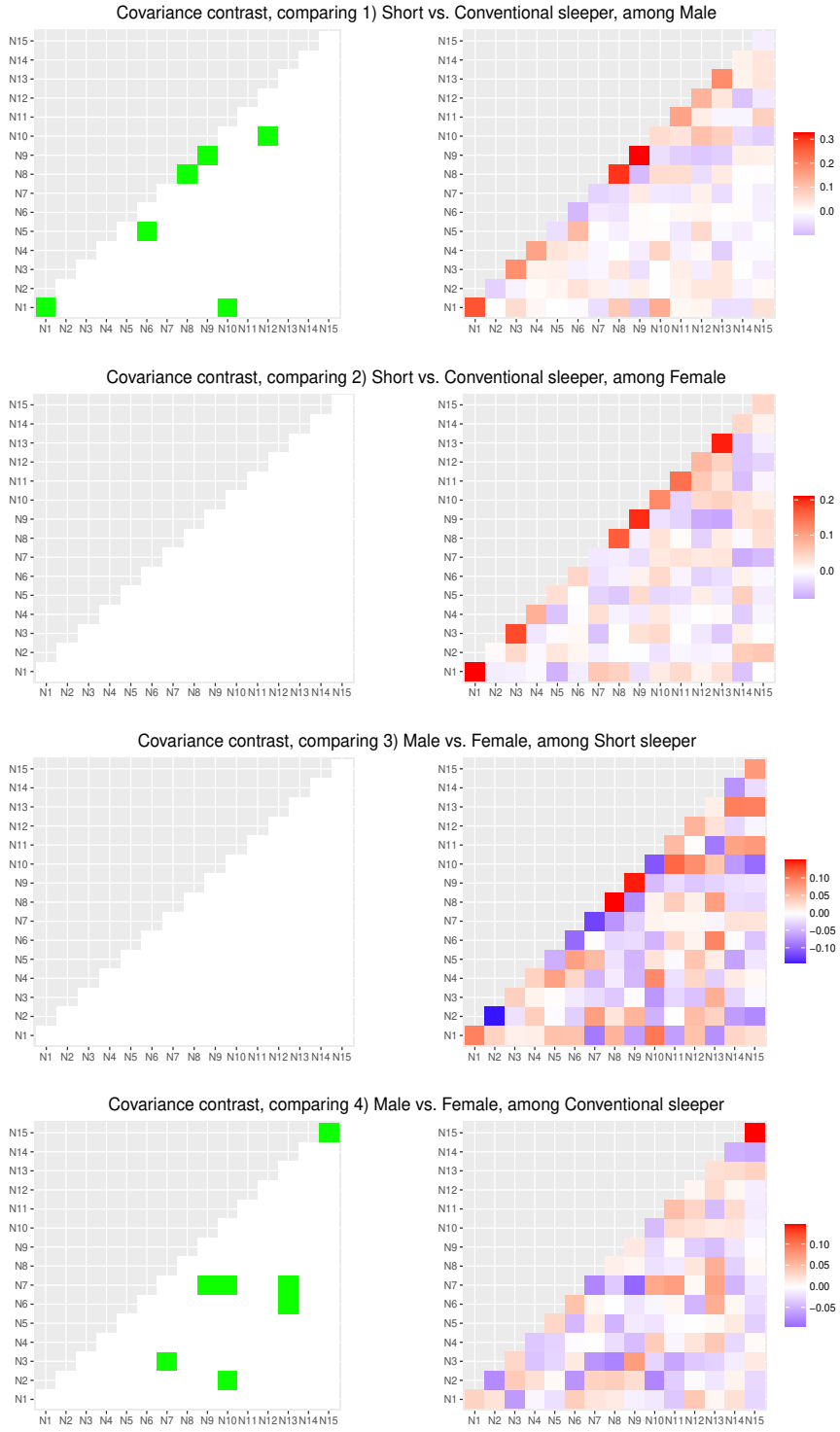

Fig. S.7. Elementwise regression. The statistical significance map (the left column panels) and the point estimate of the connectivity elements from mass-univariate regression with the log-transformed variance (diagonal) elements and Fisher z-transformed correlation (off-diagonal) elements, for each of the four covariate contrasts $\boldsymbol{\delta}$, derived from the SleepDuration $\times$ Gender interaction.

References

Benjamini, Y. and Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. **57**, 289–300.

Flury, B. and Gautschi, W. (1986). An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *SIAM Journal on Scientific and Statistical Computing* **7**, 169–184.

Zhao, Y., Wang, B., Mostofsky, S. H., Caffo, B.S. and Luo, X. (2021). Covariate assisted principal regression for covariance matrix outcomes. *Biostatistics* **22**, 629–645.

[*Received January 1, ****; revised January 1, ****; accepted for publication January 1, ****]