Web-based Supplementary Materials

for

Warped Functional Analysis of Variance

by

Daniel Gervini and *Patrick A. Carter*

November 5, 2013

Contents

1 Web Appendix A: Estimation

1.1 Model specifications

Let $x_{ij}(t) = z_{ij}(w_{ij}^{-1}(t))$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J_i$. The $z_{ij}(t)$s follow a random-factor ANOVA model

$$z_{ij}(t) = \mu(t) + \alpha_i(t) + \beta_{ij}(t), \tag{1}$$

with the $\alpha_i(t)$s and the $\beta_{ij}(t)$s independent (among themselves and of each other) zero-mean random processes. The $\alpha_i(t)$s satisfy

$$\alpha_i(t) = \sum_{k=1}^{p} u_{ik}\phi_k(t), \tag{2}$$

with the $\phi_k(t)$s orthonormal, $\mathbf{u} = (u_1, \ldots, u_p)^T \sim N(\mathbf{0}, \mathbf{\Gamma})$ with $\mathbf{\Gamma} = \text{diag}(\gamma_1, \ldots, \gamma_p)$, $\gamma_1 \geq \cdots \geq \gamma_p > 0$. The $\beta_{ij}(t)$s satisfy

$$\beta_{ij}(t) = \sum_{k=1}^{q} v_{ijk} \psi_k(t), \tag{3}$$

with the $\psi_k(t)$s orthonormal and $\mathbf{v} = (v_1, \ldots, v_q)^T \sim N(\mathbf{0}, \mathbf{\Lambda})$ with $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_q)$, $\lambda_1 \geq \cdots \geq \lambda_q > 0$.

Let $\mathbf{b}(t) = (b_1(t), \ldots, b_s(t))^T$ be a B-spline basis in $L^2([a, b])$. Then

$$\mu(t) = \mathbf{b}(t)^T \mathbf{m}, \;\; \phi_k(t) = \mathbf{b}(t)^T \mathbf{c}_k, \;\; \psi_k(t) = \mathbf{b}(t)^T \mathbf{d}_k.$$

Let $\mathbf{J} = \int_a^b \mathbf{b}(t)\mathbf{b}(t)^T dt \in \mathbb{R}^{s \times s}$, $\mathbf{C} = [\mathbf{c}_1, \ldots, \mathbf{c}_p] \in \mathbb{R}^{s \times p}$ and $\mathbf{D} = [\mathbf{d}_1, \ldots, \mathbf{d}_q] \in \mathbb{R}^{s \times q}$. Then the orthogonality condition on the ϕ_ks and the ψ_ks translates into

$$\mathbf{C}^T \mathbf{J} \mathbf{C} = \mathbf{I}_p, \;\; \mathbf{D}^T \mathbf{J} \mathbf{D} = \mathbf{I}_q.$$

The warping functions $w_{ij}(t)$ are interpolating cubic Hermite splines with mean knots $\boldsymbol{\tau}_0 \in \mathbb{R}^r$ and individual knots $\boldsymbol{\tau}_{ij} \in \mathbb{R}^r$; let $\boldsymbol{\theta}_0$ and $\boldsymbol{\theta}_{ij}$ be the Jupp transforms of $\boldsymbol{\tau}_0$ and $\boldsymbol{\tau}_{ij}$, respectively. Then

$$\boldsymbol{\theta}_{ij} = \boldsymbol{\theta}_0 + \boldsymbol{\eta}_i + \boldsymbol{\xi}_{ij}$$

with the $\boldsymbol{\eta}_i$s and the $\boldsymbol{\xi}_{ij}$s independent among themselves and of each other, and also of the amplitude factors $\alpha_i(t)$ and $\beta_{ij}(t)$. Assume

$$\boldsymbol{\eta}_i \sim N(\mathbf{0}, \mathbf{\Sigma}), \;\; \boldsymbol{\xi}_{ij} \sim N(\mathbf{0}, \mathbf{\Omega}).$$

The raw data is $\{\mathbf{y}_{ij}\}$ with

$$\mathbf{y}_{ij} = x_{ij}(\mathbf{t}_{ij}) + \boldsymbol{\varepsilon}_{ij}, \tag{4}$$

where $\mathbf{t}_{ij} = (t_{ij,1}, \ldots, t_{ij,\nu_{ij}})^T \in \mathbb{R}^{\nu_{ij}}$ is the time grid and $\boldsymbol{\varepsilon}_{ij} \sim N(0, \sigma^2 \mathbf{I}_{\nu_{ij}})$ are i.i.d. measurement error, independent of the $x_{ij}(t)$s. Given $\boldsymbol{\theta}_{ij}$, the warped time grid is $\mathbf{t}^*_{ij}(\boldsymbol{\theta}_{ij}) = w^{-1}_{\boldsymbol{\theta}_{ij}}(\mathbf{t}_{ij})$, and the corresponding warped B-spline basis matrix is

$$\mathbf{B}^*_{ij}(\boldsymbol{\theta}_{ij}) = \left[b_1(\mathbf{t}^*_{ij}(\boldsymbol{\theta}_{ij})), \ldots, b_s(\mathbf{t}^*_{ij}(\boldsymbol{\theta}_{ij})) \right].$$

Then

$$\mathbf{y}_{ij} | (\mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) \sim N\left(\mathbf{B}^*_{ij}(\boldsymbol{\theta}_{ij}) \left(\mathbf{m} + \mathbf{C}\mathbf{u}_i + \mathbf{D}\mathbf{v}_{ij} \right), \sigma^2 \mathbf{I}_{\nu_{ij}} \right)$$

and they are conditionally independent given $(\mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij})$.

1.2 Penalized MLE via EM algorithm

Since the data is independent between groups, the log-likelihood function is

$$\ell = \sum_{i=1}^{I} \log f(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}).$$

We penalize the total warping variance $\mathrm{tr}(\boldsymbol{\Sigma} + \boldsymbol{\Omega})$, so the parameter estimators are defined as the maximizers of

$$\ell_\lambda = \ell - \frac{1}{2}\lambda \mathrm{tr}(\boldsymbol{\Sigma} + \boldsymbol{\Omega}),$$

for some penalization parameter $\lambda > 0$.

We compute the MLEs using the EM algorithm, treating the unobservable random effects as missing data. Let the "complete data" penalized likelihood be

$$\ell_\lambda^* = \sum_{i=1}^{I} \log f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i}, \mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) - \frac{1}{2}\lambda \mathrm{tr}(\boldsymbol{\Sigma} + \boldsymbol{\Omega}).$$

We have

$$\begin{aligned}
&\sum_{i=1}^{I} \log f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i}, \mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) \\
= \;& \sum_{i=1}^{I} \log f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i} | \mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) + \sum_{i=1}^{I} \log f(\mathbf{u}_i) \\
&+ \sum_{i=1}^{I} \log f(\{\mathbf{v}_{ij}\}_{j=1}^{J_i}) + \sum_{i=1}^{I} \log f(\boldsymbol{\eta}_i) + \sum_{i=1}^{I} \log f(\{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) \\
= \;& \sum_{i=1}^{I} \sum_{j=1}^{J_i} \log f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) + \sum_{i=1}^{I} \log f(\mathbf{u}_i) \\
&+ \sum_{i=1}^{I} \sum_{j=1}^{J_i} \log f(\mathbf{v}_{ij}) + \sum_{i=1}^{I} \log f(\boldsymbol{\eta}_i) + \sum_{i=1}^{I} \sum_{i=1}^{J_i} \log f(\boldsymbol{\xi}_{ij})
\end{aligned}$$

where

$$\sum_{i=1}^{I} \sum_{j=1}^{J_i} \log f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) \propto$$

$$-\frac{m}{2}\log\sigma^2-\sum_{i=1}^{I}\sum_{j=1}^{J_i}\frac{\|\mathbf{y}_{ij}-\mathbf{B}^*_{ij}(\boldsymbol{\theta}_{ij})(\mathbf{m}+\mathbf{C}\mathbf{u}_i+\mathbf{D}\mathbf{v}_{ij})\|^2}{2\sigma^2},$$

$$\begin{aligned}\sum_{i=1}^{I}\log f(\mathbf{u}_i) &\propto -\frac{I}{2}\log\det\boldsymbol{\Gamma}-\frac{1}{2}\sum_{i=1}^{I}\mathbf{u}_i^T\boldsymbol{\Gamma}^{-1}\mathbf{u}_i\\ &= -\frac{I}{2}\sum_{k=1}^{p}\log\gamma_k-\frac{1}{2}\sum_{k=1}^{p}\sum_{i=1}^{I}\frac{u_{ik}^2}{\gamma_k},\end{aligned}$$

$$\begin{aligned}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\log f(\mathbf{v}_{ij}) &\propto -\frac{n}{2}\log\det\boldsymbol{\Lambda}-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{v}_{ij}^T\boldsymbol{\Lambda}^{-1}\mathbf{v}_{ij}\\ &= -\frac{n}{2}\sum_{k=1}^{q}\log\lambda_k-\frac{1}{2}\sum_{k=1}^{q}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\frac{v_{ijk}^2}{\lambda_k},\end{aligned}$$

$$\sum_{i=1}^{I}\log f(\boldsymbol{\eta}_i)\propto -\frac{I}{2}\log\det\boldsymbol{\Sigma}-\frac{1}{2}\sum_{i=1}^{I}\boldsymbol{\eta}_i^T\boldsymbol{\Sigma}^{-1}\boldsymbol{\eta}_i,$$

$$\sum_{i=1}^{I}\sum_{i=1}^{J_i}\log f(\boldsymbol{\xi}_{ij})\propto -\frac{n}{2}\log\det\boldsymbol{\Omega}-\frac{1}{2}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\boldsymbol{\xi}_{ij}^T\boldsymbol{\Omega}^{-1}\boldsymbol{\xi}_{ij},$$

with $m=\sum_{i=1}^{I}\sum_{j=1}^{J_i}\nu_{ij}$ and $n=\sum_{i=1}^{I}J_i$.

The EM updates on each iteration are defined as the maximizers of $Q=\mathrm{E}(\ell^*_\lambda|\mathbf{y}_{11},\ldots,\mathbf{y}_{IJ_I})$, where the expectation is computed with the parameter estimates of the current iteration. From the above expressions, and the fact that the $\mathbf{y}_{ij}$s and the random effects are independent for different is, the following equations follow:

$$\hat{\gamma}_k=\frac{1}{I}\sum_{i=1}^{I}\mathrm{E}(u_{ik}^2|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}),\quad k=1,\ldots,p, \tag{5}$$

$$\hat{\lambda}_k=\frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{E}(v_{ijk}^2|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}),\quad k=1,\ldots,q, \tag{6}$$

$$\hat{\sigma}^2=\frac{1}{m}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{E}\{\|\mathbf{y}_{ij}-\mathbf{B}^*_{ij}(\boldsymbol{\theta}_{ij})(\hat{\mathbf{m}}+\hat{\mathbf{C}}\mathbf{u}_i+\hat{\mathbf{D}}\mathbf{v}_{ij})\|^2|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\},$$

and

$$\begin{aligned}\hat{\mathbf{m}} &= \left[\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\}\right]^{-1}\times \\ &\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{E}[\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\{\mathbf{y}_{ij}-\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})\hat{\mathbf{C}}\mathbf{u}_i-\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})\hat{\mathbf{D}}\mathbf{v}_{ij}\}|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}].\end{aligned}$$

The rest of the estimating equations are more complicated, due to the penalization term in the cases of $\mathbf{\Sigma}$ and $\mathbf{\Omega}$, and due to the orthogonality constraints in the cases of $\mathbf{C}$ and $\mathbf{D}$. Using differentials to compute the derivatives (Magnus and Neudecker, 1999), for $\mathbf{\Sigma}$ we have

$$\begin{aligned}\mathrm{d}_{\mathbf{\Sigma}}Q &= -\frac{I}{2}\mathrm{d}_{\mathbf{\Sigma}}\left(\log\det\mathbf{\Sigma}\right)-\frac{1}{2}\sum_{i=1}^{I}\mathrm{E}\left\{\boldsymbol{\eta}_i^T\mathrm{d}_{\mathbf{\Sigma}}\left(\mathbf{\Sigma}^{-1}\right)\boldsymbol{\eta}_i|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\right\}-\frac{1}{2}\lambda\mathrm{tr}(\mathrm{d}\mathbf{\Sigma}) \\ &= -\frac{I}{2}\mathrm{tr}\left(\mathbf{\Sigma}^{-1}\mathrm{d}\mathbf{\Sigma}\right)+\frac{1}{2}\sum_{i=1}^{I}\mathrm{E}\left\{\boldsymbol{\eta}_i^T\mathbf{\Sigma}^{-1}\left(\mathrm{d}\mathbf{\Sigma}\right)\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\right\}-\frac{1}{2}\lambda\mathrm{tr}(\mathrm{d}\mathbf{\Sigma}) \\ &= \mathrm{tr}\left[\left\{-\frac{I}{2}\mathbf{\Sigma}^{-1}+\frac{1}{2}\sum_{i=1}^{I}\mathrm{E}(\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\mathbf{\Sigma}^{-1}|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i})-\frac{\lambda}{2}\mathbf{I}_r\right\}\mathrm{d}\mathbf{\Sigma}\right]\end{aligned}$$

so

$$\frac{\partial Q}{\partial\mathbf{\Sigma}} = -\frac{I}{2}\mathbf{\Sigma}^{-1}+\frac{I}{2}\mathbf{\Sigma}^{-1}\mathbf{S}_{\boldsymbol{\eta\eta}}\mathbf{\Sigma}^{-1}-\frac{\lambda}{2}\mathbf{I}_r$$

where

$$\mathbf{S}_{\boldsymbol{\eta\eta}} = \frac{1}{I}\sum_{i=1}^{I}\mathrm{E}(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}).$$

Then the equation $\partial Q/\partial\mathbf{\Sigma}=\mathbf{O}$ is equivalent to

$$\frac{\lambda}{I}\hat{\mathbf{\Sigma}}^2+\hat{\mathbf{\Sigma}}-\mathbf{S}_{\boldsymbol{\eta\eta}}=\mathbf{O}, \tag{7}$$

whose explicit solution is

$$\hat{\mathbf{\Sigma}} = -\frac{1}{2(\lambda/I)}\mathbf{I}_r+\frac{1}{2(\lambda/I)}\left\{\mathbf{I}_r+4\left(\frac{\lambda}{I}\right)\mathbf{S}_{\boldsymbol{\eta\eta}}\right\}^{1/2}.$$

(This expression is easy to interpret in terms of the spectral decomposition of $\mathbf{S}_{\boldsymbol{\eta\eta}}$: if $\mathbf{S}_{\boldsymbol{\eta\eta}}=\mathbf{U}\mathbf{\Delta}\mathbf{U}^T$ with $\mathbf{\Delta}=\mathrm{diag}(\delta_1,\ldots,\delta_r)$ the eigenvalues and $\mathbf{U}\in\mathbb{R}^{r\times r}$ the eigenvectors, then $\hat{\boldsymbol{\Sigma}} = \mathbf{U}\boldsymbol{\Delta}_\lambda\mathbf{U}^T$ with $\boldsymbol{\Delta}_\lambda = \mathrm{diag}(\delta_{1,\lambda}, \ldots, \delta_{r,\lambda})$ and

$$\delta_{k,\lambda} = \frac{-1 + \{1 + 4(\lambda/I)\delta_k\}^{1/2}}{2(\lambda/I)}, \quad k = 1, \ldots, r.$$

Since $\delta_{k,\lambda} \to \delta_k$ when $\lambda \to 0$ and $\delta_{k,\lambda} \to 0$ when $\lambda \to +\infty$, the effect of the penalization term is to shrink the eigenvalues of $\mathbf{S}_{\boldsymbol{\eta}\boldsymbol{\eta}}$ towards zero, therefore controlling the warping flexibility by reducing the warping variance.) Similarly, for $\hat{\boldsymbol{\Omega}}$ we have

$$\frac{\partial Q}{\partial \boldsymbol{\Omega}} = -\frac{n}{2}\boldsymbol{\Omega}^{-1} + \frac{n}{2}\boldsymbol{\Omega}^{-1}\mathbf{S}_{\boldsymbol{\xi}\boldsymbol{\xi}}\boldsymbol{\Omega}^{-1} - \frac{\lambda}{2}\mathbf{I}_r$$

with

$$\mathbf{S}_{\boldsymbol{\xi}\boldsymbol{\xi}} = \frac{1}{n}\sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}),$$

so

$$\frac{\lambda}{n}\hat{\boldsymbol{\Omega}}^2 + \hat{\boldsymbol{\Omega}} - \mathbf{S}_{\boldsymbol{\xi}\boldsymbol{\xi}} = \mathbf{O} \tag{8}$$

and then

$$\hat{\boldsymbol{\Omega}} = -\frac{1}{2(\lambda/n)}\mathbf{I}_r + \frac{1}{2(\lambda/n)}\left\{\mathbf{I}_r + 4\left(\frac{\lambda}{n}\right)\mathbf{S}_{\boldsymbol{\xi}\boldsymbol{\xi}}\right\}^{1/2}.$$

To find $\hat{\mathbf{C}}$ and $\hat{\mathbf{D}}$, we maximize Q one column of $\mathbf{C}$ and $\mathbf{D}$ at a time. Concretely: suppose we want to update $\hat{\mathbf{c}}_k$ for a given k, with the other $\hat{\mathbf{c}}_j$s and $\hat{\mathbf{D}}$ kept fixed. Let

$$\mathbf{r}_{ij}^{(k)} = \mathbf{y}_{ij} - \mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})(\hat{\mathbf{m}} + \sum_{l \neq k} \hat{\mathbf{c}}_l u_{il} + \hat{\mathbf{D}}\mathbf{v}_{ij}).$$

Then

$$\hat{\mathbf{c}}_k = \arg\min_{\mathbf{c}_k} \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\|\mathbf{r}_{ij}^{(k)} - \mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})\mathbf{c}_k u_{ik}\|^2 | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}\}$$

subject to $\mathbf{c}_k^T\mathbf{J}\mathbf{c}_k = 1$ and $\hat{\mathbf{c}}_l^T\mathbf{J}\mathbf{c}_k = 0$ for all $l < k$. If we define

$$\begin{aligned}
\mathbf{A}_k &= \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})u_{ik}^2 | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}\}, \\
\mathbf{b}_k &= \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{r}_{ij}^{(k)} u_{ik} | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}\},
\end{aligned}$$

then $\hat{\mathbf{c}}_k$ is the solution to the constrained quadratic minimization problem

$$\begin{aligned}\text{minimize:} \quad & \mathbf{c}^T\mathbf{A}_k\mathbf{c} - 2\mathbf{b}_k^T\mathbf{c} \\ \text{subject to:} \quad & \hat{\mathbf{c}}_l^T\mathbf{J}\mathbf{c} = 0,\ l < k, \\ & \mathbf{c}^T\mathbf{J}\mathbf{c} = 1.\end{aligned}$$

The solution is very easy to find numerically. Note that

$$\begin{aligned}\mathbf{b}_k \;=\; & \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{r}_{ij}u_{ik}|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\} + \\ & \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})u_{ik}^2|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\}\mathbf{c}_k,\end{aligned}$$

so it's not necessary to recompute expectations for each $\mathbf{b}_k$. To update $\hat{\mathbf{D}}$ we proceed in a similar way, using

$$\mathbf{r}_{ij}^{(k)} = \mathbf{y}_{ij} - \mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})(\hat{\mathbf{m}} + \hat{\mathbf{C}}\mathbf{u}_i + \sum_{l\neq k}\hat{\mathbf{d}}_l v_{ijl}),$$

$$\mathbf{A}_k = \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})v_{ijk}^2|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\},$$

and

$$\mathbf{b}_k = \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{r}_{ij}^{(k)}v_{ijk}|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\}$$

instead, where now $\mathbf{b}_k$ satisfies

$$\begin{aligned}\mathbf{b}_k \;=\; & \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{r}_{ij}v_{ijk}|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\} + \\ & \sum_{i=1}^{I}\sum_{j=1}^{J_i} \mathrm{E}\{\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T\mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})v_{ijk}^2|\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}\}\mathbf{d}_k.\end{aligned}$$

1.3 Computing expectations by Monte Carlo integration

First off we need to compute the joint density $f(\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i})$ for each i. We have

$$f(\mathbf{y}_{i1},\ldots,\mathbf{y}_{iJ_i}) =$$

$$= \int \cdots \int f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i}, \mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) \mathrm{d}(\mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i})$$

$$\begin{aligned} = \quad & \int \cdots \int f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i} | \mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) f(\mathbf{u}_i) f(\{\mathbf{v}_{ij}\}_{j=1}^{J_i}) \times \\ & f(\boldsymbol{\eta}_i) f(\{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) \mathrm{d}(\mathbf{u}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \boldsymbol{\eta}_i, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) \end{aligned}$$

$$\begin{aligned} = \quad & \int \cdots \int \left\{ \prod_{j=1}^{J_i} f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) \right\} f(\mathbf{u}_i) \mathrm{d}\mathbf{u}_i \times \\ & \left\{ \prod_{j=1}^{J_i} f(\mathbf{v}_{ij}) \mathrm{d}\mathbf{v}_{ij} \right\} f(\boldsymbol{\eta}_i) \mathrm{d}\boldsymbol{\eta}_i \left\{ \prod_{j=1}^{J_i} f(\boldsymbol{\xi}_{ij}) \mathrm{d}\boldsymbol{\xi}_{ij} \right\} \end{aligned}$$

$$= \int \left\{ \prod_{j=1}^{J_i} \int f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) f(\mathbf{v}_{ij}) f(\boldsymbol{\xi}_{ij}) \mathrm{d}\mathbf{v}_{ij} \mathrm{d}\boldsymbol{\xi}_{ij} \right\} f(\mathbf{u}_i) f(\boldsymbol{\eta}_i) \mathrm{d}\mathbf{u}_i \mathrm{d}\boldsymbol{\eta}_i.$$

Then we take independent random samples $\{(\mathbf{u}_l, \boldsymbol{\eta}_l)\}_{l=1}^{N}$, $\{(\mathbf{v}_m, \boldsymbol{\xi}_m)\}_{m=1}^{M}$ and define

$$\hat{f}(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}) := \frac{1}{N} \sum_{l=1}^{N} \left[\prod_{j=1}^{J_i} \left\{ \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{y}_{ij} | \mathbf{u}_l, \mathbf{v}_m, \boldsymbol{\eta}_l, \boldsymbol{\xi}_m) \right\} \right].$$

All expectations that need to be computed have the form $\mathrm{E}(g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i})$ for some g, such as $g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) = \boldsymbol{\xi}_{ij} \boldsymbol{\xi}_{ij}^T$ or $g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) = \mathbf{B}_{ij}^*(\boldsymbol{\theta}_{ij})^T \mathbf{r}_{ij} \mathbf{v}_{ij}^T$. Note that

$$\mathrm{E}\{g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}\} =$$

$$\begin{aligned} = \quad & \int g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) f(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij} | \mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}) \mathrm{d}\mathbf{u}_i \mathrm{d}\boldsymbol{\eta}_i \mathrm{d}\mathbf{v}_{ij} \mathrm{d}\boldsymbol{\xi}_{ij} \\ = \quad & \frac{\int g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) f(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}, \mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) \mathrm{d}\mathbf{u}_i \mathrm{d}\boldsymbol{\eta}_i \mathrm{d}\mathbf{v}_{ij} \mathrm{d}\boldsymbol{\xi}_{ij}}{f(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i})}. \end{aligned}$$

The denominator $f(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i})$ was computed above. For the numerator, we have

$$f(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}, \mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) =$$

$$= \int \cdots \int f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i}, \mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}, \{\mathbf{v}_{ij'}\}_{j' \neq j}, \{\boldsymbol{\xi}_{ij'}\}_{j' \neq j}) \mathrm{d}(\{\mathbf{v}_{ij'}\}_{j' \neq j}, \{\boldsymbol{\xi}_{ij'}\}_{j' \neq j})$$

$$= \int \cdots \int f(\{\mathbf{y}_{ij}\}_{j=1}^{J_i} | \mathbf{u}_i, \boldsymbol{\eta}_i, \{\mathbf{v}_{ij}\}_{j=1}^{J_i}, \{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) f(\mathbf{u}_i) f(\boldsymbol{\eta}_i) \times f(\{\mathbf{v}_{ij}\}_{j=1}^{J_i}) f(\{\boldsymbol{\xi}_{ij}\}_{j=1}^{J_i}) \mathrm{d}(\{\mathbf{v}_{ij'}\}_{j' \neq j}, \{\boldsymbol{\xi}_{ij'}\}_{j' \neq j})$$

$$= \int \cdots \int \left\{ \prod_{j=1}^{J_i} f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) \right\} f(\mathbf{u}_i) f(\boldsymbol{\eta}_i) \times f(\mathbf{v}_{ij}) \left\{ \prod_{j' \neq j} f(\mathbf{v}_{ij'}) \mathrm{d}\mathbf{v}_{ij'} \right\} f(\boldsymbol{\xi}_{ij}) \left\{ \prod_{j' \neq j} f(\boldsymbol{\xi}_{ij'}) \mathrm{d}\boldsymbol{\xi}_{ij'} \right\}$$

$$= f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) f(\mathbf{u}_i) f(\boldsymbol{\eta}_i) f(\mathbf{v}_{ij}) f(\boldsymbol{\xi}_{ij}) \times \left\{ \prod_{j' \neq j} \int f(\mathbf{y}_{ij'} | \mathbf{u}_i, \mathbf{v}_{ij'}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij'}) f(\mathbf{v}_{ij'}) f(\boldsymbol{\xi}_{ij'}) \mathrm{d}\mathbf{v}_{ij'} \mathrm{d}\boldsymbol{\xi}_{ij'} \right\},$$

so

$$\int g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) f(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i}, \mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) \mathrm{d}\mathbf{u}_i \mathrm{d}\boldsymbol{\eta}_i \mathrm{d}\mathbf{v}_{ij} \mathrm{d}\boldsymbol{\xi}_{ij} =$$

$$= \int \left\{ \int g(\mathbf{u}_i, \boldsymbol{\eta}_i, \mathbf{v}_{ij}, \boldsymbol{\xi}_{ij}) f(\mathbf{y}_{ij} | \mathbf{u}_i, \mathbf{v}_{ij}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij}) f(\mathbf{v}_{ij}) f(\boldsymbol{\xi}_{ij}) \mathrm{d}\mathbf{v}_{ij} \mathrm{d}\boldsymbol{\xi}_{ij} \right\} \times \left\{ \prod_{j' \neq j} \int f(\mathbf{y}_{ij'} | \mathbf{u}_i, \mathbf{v}_{ij'}, \boldsymbol{\eta}_i, \boldsymbol{\xi}_{ij'}) f(\mathbf{v}_{ij'}) f(\boldsymbol{\xi}_{ij'}) \mathrm{d}\mathbf{v}_{ij'} \mathrm{d}\boldsymbol{\xi}_{ij'} \right\} f(\mathbf{u}_i) f(\boldsymbol{\eta}_i) \mathrm{d}\mathbf{u}_i \mathrm{d}\boldsymbol{\eta}_i.$$

This is approximated by

$$\frac{1}{N} \sum_{l=1}^{N} \left\{ \frac{1}{M} \sum_{m=1}^{M} g(\mathbf{u}_l, \boldsymbol{\eta}_l, \mathbf{v}_m, \boldsymbol{\xi}_m) f(\mathbf{y}_{ij} | \mathbf{u}_l, \mathbf{v}_m, \boldsymbol{\eta}_l, \boldsymbol{\xi}_m) \right\} \left[\prod_{j' \neq j} \left\{ \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{y}_{ij'} | \mathbf{u}_l, \mathbf{v}_m, \boldsymbol{\eta}_l, \boldsymbol{\xi}_m) \right\} \right].$$

Algorithmically, this is computed as follows:

For $i = 1, \ldots, I$

  For $j = 1, \ldots, J_i$

    $\hat{e}_{ij} = 0$

    For $l = 1, \ldots, N$

      For $j' = 1, \ldots, J_i$

        $\hat{f}_{j'} = \underset{1 \leq m \leq M}{\text{ave}} \{ f(\mathbf{y}_{ij'} | \mathbf{u}_l, \mathbf{v}_m, \boldsymbol{\eta}_l, \boldsymbol{\xi}_m) \}$

      End

      $\hat{g} = \underset{1 \leq m \leq M}{\text{ave}} \{ g(\mathbf{u}_l, \boldsymbol{\eta}_l, \mathbf{v}_m, \boldsymbol{\xi}_m) f(\mathbf{y}_{ij} | \mathbf{u}_l, \mathbf{v}_m, \boldsymbol{\eta}_l, \boldsymbol{\xi}_m) \}$

$$\hat{e}_{ij} = \hat{e}_{ij} + \hat{g} * \operatorname*{prod}_{j' \neq j} \hat{f}_{j'}/N$$

End

$\hat{e}_{ij} = \hat{e}_{ij}/\hat{f}(\mathbf{y}_{i1}, \ldots, \mathbf{y}_{iJ_i})$

End

End

1.4 Classic one-way ANOVA model (no warping)

For the no-warping model the $\mathbf{y}_{ij}$s can be written as

$$\mathbf{y}_{ij} = \boldsymbol{\mu}_{ij} + \boldsymbol{\Phi}_{ij}\mathbf{u}_i + \boldsymbol{\Psi}_{ij}\mathbf{v}_{ij} + \boldsymbol{\varepsilon}_{ij}$$

where $\boldsymbol{\mu}_{ij} = \mathbf{B}_{ij}\mathbf{m}$, $\boldsymbol{\Phi}_{ij} = \mathbf{B}_{ij}\mathbf{C}$ and $\boldsymbol{\Psi}_{ij} = \mathbf{B}_{ij}\mathbf{D}$. The random effects $\mathbf{u}_i$, $\mathbf{v}_{ij}$ and $\boldsymbol{\varepsilon}_{ij}$ satisfy the same distributional assumptions as before. Then

$$\boldsymbol{\Sigma}_{ij} := \operatorname{var}(\mathbf{y}_{ij}) = \boldsymbol{\Phi}_{ij}\boldsymbol{\Gamma}\boldsymbol{\Phi}_{ij}^T + \mathbf{A}_{ij}, \text{ where } \mathbf{A}_{ij} = \boldsymbol{\Psi}_{ij}\boldsymbol{\Lambda}\boldsymbol{\Psi}_{ij}^T + \sigma^2\mathbf{I}_{\nu_{ij}}.$$

The stacked data for group i will satisfy

$$\begin{bmatrix} \mathbf{y}_{i1} \\ \vdots \\ \mathbf{y}_{iJ_i} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\mu}_{i1} \\ \vdots \\ \boldsymbol{\mu}_{iJ_i} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Phi}_{i1} \\ \vdots \\ \boldsymbol{\Phi}_{iJ_i} \end{bmatrix} \mathbf{u}_i + \begin{bmatrix} \boldsymbol{\Psi}_{i1} & \cdots & \mathbf{O} \\ \vdots & & \vdots \\ \mathbf{O} & \cdots & \boldsymbol{\Psi}_{iJ_i} \end{bmatrix} \times \begin{bmatrix} \mathbf{v}_{i1} \\ \vdots \\ \mathbf{v}_{iJ_i} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_{i1} \\ \vdots \\ \boldsymbol{\varepsilon}_{iJ_i} \end{bmatrix}.$$

Then, if we denote by $\mathbf{y}_{i\cdot}$ the stacked $\mathbf{y}_{ij}$s and similarly define $\boldsymbol{\mu}_{i\cdot}$ and $\boldsymbol{\Phi}_{i\cdot}$, we have

$$\boldsymbol{\Sigma}_{i\cdot} := \operatorname{var}(\mathbf{y}_{i\cdot}) = \boldsymbol{\Phi}_{i\cdot}\boldsymbol{\Gamma}\boldsymbol{\Phi}_{i\cdot}^T + \mathbf{A}_{i\cdot}$$

with

$$\mathbf{A}_{i\cdot} = \begin{bmatrix} \mathbf{A}_{i1} & \cdots & \mathbf{O} \\ \vdots & & \vdots \\ \mathbf{O} & \cdots & \mathbf{A}_{iJ_i} \end{bmatrix}.$$

Also,

$$\operatorname{cov}(\mathbf{y}_{i\cdot}, \mathbf{u}_i) = \boldsymbol{\Phi}_{i\cdot}\boldsymbol{\Gamma},$$

$$\text{cov}(\mathbf{y}_{i\cdot}, \mathbf{v}_{ij}) = \begin{bmatrix} \mathbf{O} \\ \vdots \\ \mathbf{\Psi}_{ij}\mathbf{\Lambda} \\ \vdots \\ \mathbf{O} \end{bmatrix}, \; \text{cov}(\mathbf{y}_{i\cdot}, \boldsymbol{\varepsilon}_{ij}) = \begin{bmatrix} \mathbf{O} \\ \vdots \\ \sigma^2 \mathbf{I}_{\nu_{ij}} \\ \vdots \\ \mathbf{O} \end{bmatrix},$$

$$\text{cov}\left(\mathbf{y}_{i\cdot}, \begin{bmatrix} \mathbf{u}_i \\ \boldsymbol{\varepsilon}_{ij} \end{bmatrix}\right) = \begin{bmatrix} & \mathbf{O} \\ & \vdots \\ \mathbf{\Phi}_{i\cdot}\mathbf{\Gamma} & \sigma^2 \mathbf{I}_{\nu_{ij}} \\ & \vdots \\ & \mathbf{O} \end{bmatrix}, \; \text{cov}\left(\mathbf{y}_{i\cdot}, \begin{bmatrix} \mathbf{v}_{ij} \\ \boldsymbol{\varepsilon}_{ij} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{O} & \mathbf{O} \\ \vdots & \vdots \\ \mathbf{\Psi}_{ij}\mathbf{\Lambda} & \sigma^2 \mathbf{I}_{\nu_{ij}} \\ \vdots & \vdots \\ \mathbf{O} & \mathbf{O} \end{bmatrix}.$$

Then

$$\begin{aligned} \text{E}\left(\mathbf{u}_i | \mathbf{y}_{i\cdot}\right) &= -\mathbf{\Gamma}\mathbf{\Phi}_{i\cdot}^T \mathbf{\Sigma}_{i\cdot}^{-1}(\mathbf{y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}), \\ \text{var}\left(\mathbf{u}_i | \mathbf{y}_{i\cdot}\right) &= \mathbf{\Gamma} - \mathbf{\Gamma}\mathbf{\Phi}_{i\cdot}^T \mathbf{\Sigma}_{i\cdot}^{-1} \mathbf{\Phi}_{i\cdot}\mathbf{\Gamma}, \end{aligned}$$

$$\begin{aligned} \text{E}\left(\mathbf{v}_{ij} | \mathbf{y}_{i\cdot}\right) &= -\mathbf{\Lambda}\mathbf{\Psi}_{ij}^T [\mathbf{\Sigma}_{i\cdot}^{-1}(\mathbf{y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot})]_{\text{block } j}, \\ \text{var}\left(\mathbf{v}_{ij} | \mathbf{y}_{i\cdot}\right) &= \mathbf{\Lambda} - \mathbf{\Lambda}\mathbf{\Psi}_{ij}^T [\mathbf{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,j)} \mathbf{\Psi}_{ij}\mathbf{\Lambda}, \end{aligned}$$

and

$$\begin{aligned} \text{E}\left(\boldsymbol{\varepsilon}_{ij} | \mathbf{y}_{i\cdot}\right) &= -\sigma^2 [\mathbf{\Sigma}_{i\cdot}^{-1}(\mathbf{y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot})]_{\text{block } j}, \\ \text{var}\left(\boldsymbol{\varepsilon}_{ij} | \mathbf{y}_{i\cdot}\right) &= \sigma^2 \mathbf{I}_{\nu_{ij}} - \sigma^4 [\mathbf{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,j)}. \end{aligned}$$

Also,

$$\text{var}\left(\begin{bmatrix} \mathbf{u}_i \\ \boldsymbol{\varepsilon}_{ij} \end{bmatrix} \middle| \mathbf{y}_{i\cdot}\right) = \begin{bmatrix} \mathbf{\Gamma} & \mathbf{O} \\ \mathbf{O} & \sigma^2 \mathbf{I}_{\nu_{ij}} \end{bmatrix} - \begin{bmatrix} \mathbf{\Gamma}\mathbf{\Phi}_{i\cdot}^T \mathbf{\Sigma}_{i\cdot}^{-1} \mathbf{\Phi}_{i\cdot}\mathbf{\Gamma} & \sigma^2 \mathbf{\Gamma}\mathbf{\Phi}_{i\cdot}^T [\mathbf{\Sigma}_{i\cdot}^{-1}]_{\text{block } (\cdot,j)} \\ \sigma^2 [\mathbf{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,\cdot)} \mathbf{\Phi}_{i\cdot}\mathbf{\Gamma} & \sigma^4 [\mathbf{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,j)} \end{bmatrix}$$

and

$$\text{var}\left(\begin{bmatrix} \mathbf{v}_{ij} \\ \boldsymbol{\varepsilon}_{ij} \end{bmatrix} \middle| \mathbf{y}_{i\cdot}\right) = \begin{bmatrix} \mathbf{\Lambda} & \mathbf{O} \\ \mathbf{O} & \sigma^2 \mathbf{I}_{\nu_{ij}} \end{bmatrix} - \begin{bmatrix} \mathbf{\Lambda}\mathbf{\Psi}_{ij}^T \\ \sigma^2 \mathbf{I}_{\nu_{ij}} \end{bmatrix} [\mathbf{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,j)} \begin{bmatrix} \mathbf{\Psi}_{ij}\mathbf{\Lambda} & \sigma^2 \mathbf{I}_{\nu_{ij}} \end{bmatrix},$$

so

$$\begin{aligned}\mathrm{cov}(\boldsymbol{\varepsilon}_{ij}, \mathbf{u}_i | \mathbf{y}_{i\cdot}) &= -\sigma^2 [\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block }(j,\cdot)} \boldsymbol{\Phi}_{i\cdot} \boldsymbol{\Gamma}, \\ \mathrm{cov}(\boldsymbol{\varepsilon}_{ij}, \mathbf{v}_{ij} | \mathbf{y}_{i\cdot}) &= -\sigma^2 [\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block }(j,j)} \boldsymbol{\Psi}_{ij} \boldsymbol{\Lambda}\end{aligned}$$

Note that

$$\boldsymbol{\Sigma}_{i\cdot}^{-1} = \mathbf{A}_{i\cdot}^{-1} - \mathbf{A}_{i\cdot}^{-1} \boldsymbol{\Phi}_{i\cdot} \left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Phi}_{i\cdot}^T \mathbf{A}_{i\cdot}^{-1} \boldsymbol{\Phi}_{i\cdot} \right)^{-1} \boldsymbol{\Phi}_{i\cdot}^T \mathbf{A}_{i\cdot}^{-1}$$

with

$$\mathbf{A}_{i\cdot}^{-1} = \begin{bmatrix} \mathbf{A}_{i1}^{-1} & \cdots & \mathbf{O} \\ \vdots & & \vdots \\ \mathbf{O} & \cdots & \mathbf{A}_{iJ_i}^{-1} \end{bmatrix}$$

and

$$\mathbf{A}_{ij}^{-1} = \frac{1}{\sigma^2} \mathbf{I}_{\nu_{ij}} - \frac{1}{\sigma^4} \boldsymbol{\Psi}_{ij} \left(\boldsymbol{\Lambda}^{-1} + \frac{1}{\sigma^2} \boldsymbol{\Psi}_{ij}^T \boldsymbol{\Psi}_{ij} \right)^{-1} \boldsymbol{\Psi}_{ij}^T.$$

Then

$$\mathbf{A}_{i\cdot}^{-1} \boldsymbol{\Phi}_{i\cdot} = \begin{bmatrix} \mathbf{A}_{i1}^{-1} \boldsymbol{\Phi}_{i1} \\ \vdots \\ \mathbf{A}_{iJ_i}^{-1} \boldsymbol{\Phi}_{iJ_i} \end{bmatrix},$$

so

$$\begin{aligned}\boldsymbol{\Omega}_i &:= \boldsymbol{\Phi}_{i\cdot}^T \mathbf{A}_{i\cdot}^{-1} \boldsymbol{\Phi}_{i\cdot} = \sum_{j=1}^{J_i} \boldsymbol{\Phi}_{ij}^T \mathbf{A}_{ij}^{-1} \boldsymbol{\Phi}_{ij}, \\ \boldsymbol{\beta}_i &:= \boldsymbol{\Phi}_{i\cdot}^T \mathbf{A}_{i\cdot}^{-1} (\mathbf{y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}) = \sum_{j=1}^{J_i} \boldsymbol{\Phi}_{ij}^T \mathbf{A}_{ij}^{-1} (\mathbf{y}_{ij} - \boldsymbol{\mu}_{ij}),\end{aligned}$$

and

$$\boldsymbol{\Phi}_{i\cdot}^T \boldsymbol{\Sigma}_{i\cdot}^{-1} \boldsymbol{\Phi}_{i\cdot} = \boldsymbol{\Omega}_i - \boldsymbol{\Omega}_i \left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i \right)^{-1} \boldsymbol{\Omega}_i,$$

$$\boldsymbol{\Phi}_{i\cdot}^T \boldsymbol{\Sigma}_{i\cdot}^{-1} (\mathbf{y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot}) = \left\{ \mathbf{I} - \boldsymbol{\Omega}_i \left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i \right)^{-1} \right\} \boldsymbol{\beta}_i.$$

Also

$$\begin{aligned}[\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block }(j,j)} &= \mathbf{A}_{ij}^{-1} - \mathbf{A}_{ij}^{-1} \boldsymbol{\Phi}_{ij} \left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i \right)^{-1} \boldsymbol{\Phi}_{ij}^T \mathbf{A}_{ij}^{-1}, \\ [\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block }(j,k)} &= -\mathbf{A}_{ij}^{-1} \boldsymbol{\Phi}_{ij} \left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i \right)^{-1} \boldsymbol{\Phi}_{ik}^T \mathbf{A}_{ik}^{-1},\end{aligned}$$

so

$$\boldsymbol{\gamma}_{ij} := [\boldsymbol{\Sigma}_{i\cdot}^{-1}(\mathbf{y}_{i\cdot} - \boldsymbol{\mu}_{i\cdot})]_{\text{block } j} = \mathbf{A}_{ij}^{-1}(\mathbf{y}_{ij} - \boldsymbol{\mu}_{ij}) - \mathbf{A}_{ij}^{-1}\boldsymbol{\Phi}_{ij}\left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i\right)^{-1}\boldsymbol{\beta}_i,$$

$$[\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,\cdot)}\boldsymbol{\Phi}_{i\cdot} = \mathbf{A}_{ij}^{-1}\boldsymbol{\Phi}_{ij} - \mathbf{A}_{ij}^{-1}\boldsymbol{\Phi}_{ij}\left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i\right)^{-1}\boldsymbol{\Omega}_{i}.$$

With these conditional expectations and variances we easily get explicit EM updating equations, which are basically the same as above only that without the $\boldsymbol{\theta}$s:

$$\hat{\gamma}_k = \frac{1}{I}\sum_{i=1}^{I}\mathrm{E}(u_{ik}^2|\mathbf{y}_{i\cdot}),\quad k = 1,\ldots,p,$$

$$\hat{\lambda}_k = \frac{1}{\sum_{i=1}^{I} J_i}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{E}(v_{ijk}^2|\mathbf{y}_{i\cdot}),\quad k = 1,\ldots,q,$$

$$\hat{\sigma}^2 = \frac{1}{\sum_{i=1}^{I}\sum_{j=1}^{J_i}\nu_{ij}}\sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathrm{E}(\|\boldsymbol{\varepsilon}_{ij}\|^2|\mathbf{y}_{i\cdot}),$$

for which we use the expressions

$$\begin{aligned}\mathrm{E}(\mathbf{u}_i\mathbf{u}_i^T|\mathbf{y}_{i\cdot}) &= \boldsymbol{\Gamma} - \boldsymbol{\Gamma}\left\{\boldsymbol{\Omega}_i - \boldsymbol{\Omega}_i\left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i\right)^{-1}\boldsymbol{\Omega}_i\right\}\boldsymbol{\Gamma} + \\ &\quad \boldsymbol{\Gamma}\left\{\mathbf{I} - \boldsymbol{\Omega}_i\left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i\right)^{-1}\right\}\boldsymbol{\beta}_i\boldsymbol{\beta}_i^T\left\{\mathbf{I} - \boldsymbol{\Omega}_i\left(\boldsymbol{\Gamma}^{-1} + \boldsymbol{\Omega}_i\right)^{-1}\right\}\boldsymbol{\Gamma},\end{aligned}$$

$$\mathrm{E}(\mathbf{v}_{ij}\mathbf{v}_{ij}^T|\mathbf{y}_{i\cdot}) = \boldsymbol{\Lambda} - \boldsymbol{\Lambda}\boldsymbol{\Psi}_{ij}^T[\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,j)}\boldsymbol{\Psi}_{ij}\boldsymbol{\Lambda} + \boldsymbol{\Lambda}\boldsymbol{\Psi}_{ij}^T\boldsymbol{\gamma}_{ij}\boldsymbol{\gamma}_{ij}^T\boldsymbol{\Psi}_{ij}\boldsymbol{\Lambda},$$

and

$$\mathrm{E}(\|\boldsymbol{\varepsilon}_{ij}\|^2|\mathbf{y}_{i\cdot}) = \sigma^2\nu_{ij} - \sigma^4\mathrm{tr}\left([\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block } (j,j)}\right) + \sigma^4\|\boldsymbol{\gamma}_{ij}\|^2.$$

For the $\hat{\mathbf{c}}_k$s and the $\hat{\mathbf{d}}_k$s we proceed similarly. For $\hat{\mathbf{c}}_k$ we define

$$\begin{aligned}\mathbf{r}_{ij}^{(k)} &= \mathbf{y}_{ij} - \mathbf{B}_{ij}(\mathbf{m} + \sum_{l\neq k}\mathbf{c}_l u_{il} + \mathbf{D}\mathbf{v}_{ij}) \\ &= \boldsymbol{\varepsilon}_{ij} + \mathbf{B}_{ij}\mathbf{c}_k u_{ik}\end{aligned}$$

so

$$\begin{aligned}\mathbf{b}_k &= \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{B}_{ij}^T\mathrm{E}(\mathbf{r}_{ij}^{(k)}u_{ik}|\mathbf{y}_{i\cdot}) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{B}_{ij}^T\mathrm{E}(\boldsymbol{\varepsilon}_{ij}u_{ik}|\mathbf{y}_{i\cdot}) + \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{B}_{ij}^T\mathbf{B}_{ij}\mathbf{c}_k\mathrm{E}(u_{ik}^2|\mathbf{y}_{i\cdot}),\end{aligned}$$

for which we use

$$\begin{aligned}\mathrm{E}(\boldsymbol{\varepsilon}_{ij}\mathbf{u}_i^T|\mathbf{y}_{i\cdot}) &= -\sigma^2\left\{\mathbf{A}_{ij}^{-1}\boldsymbol{\Phi}_{ij} - \mathbf{A}_{ij}^{-1}\boldsymbol{\Phi}_{ij}\left(\boldsymbol{\Gamma}^{-1}+\boldsymbol{\Omega}_i\right)^{-1}\boldsymbol{\Omega}_i\right\}\boldsymbol{\Gamma} \\ &\quad +\sigma^2\boldsymbol{\gamma}_{ij}\boldsymbol{\beta}_i^T\left\{\mathbf{I}-\left(\boldsymbol{\Gamma}^{-1}+\boldsymbol{\Omega}_i\right)^{-1}\boldsymbol{\Omega}_i\right\}\boldsymbol{\Gamma}.\end{aligned}$$

For $\hat{\mathbf{d}}_k$ we define

$$\begin{aligned}\mathbf{r}_{ij}^{(k)} &= \mathbf{y}_{ij} - \mathbf{B}_{ij}(\mathbf{m}+\mathbf{C}\mathbf{u}_i + \sum_{l\neq k}\mathbf{d}_l v_{ijl}) \\ &= \boldsymbol{\varepsilon}_{ij} + \mathbf{B}_{ij}\mathbf{d}_k v_{ijk}\end{aligned}$$

so

$$\begin{aligned}\mathbf{b}_k &= \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{B}_{ij}^T\mathrm{E}(\mathbf{r}_{ij}^{(k)}v_{ijk}|\mathbf{y}_{i\cdot}) \\ &= \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{B}_{ij}^T\mathrm{E}(\boldsymbol{\varepsilon}_{ij}v_{ijk}|\mathbf{y}_{i\cdot}) + \sum_{i=1}^{I}\sum_{j=1}^{J_i}\mathbf{B}_{ij}^T\mathbf{B}_{ij}\mathbf{d}_k\mathrm{E}(v_{ijk}^2|\mathbf{y}_{i\cdot}),\end{aligned}$$

for which we use

$$\mathrm{E}(\boldsymbol{\varepsilon}_{ij}\mathbf{v}_{ij}^T|\mathbf{y}_{i\cdot}) = -\sigma^2[\boldsymbol{\Sigma}_{i\cdot}^{-1}]_{\text{block }(j,j)}\boldsymbol{\Psi}_{ij}\boldsymbol{\Lambda} + \sigma^2\boldsymbol{\gamma}_{ij}\boldsymbol{\gamma}_{ij}^T\boldsymbol{\Psi}_{ij}\boldsymbol{\Lambda}.$$

2 Web Appendix B: Asymptotics

2.1 Likelihood derivatives and Fisher Information Matrices

We have

$$f(\mathbf{y}_{i\cdot}) = \iint\left\{\prod_{j=1}^{J_i}\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})f(\mathbf{v}_{ij})f(\boldsymbol{\xi}_{ij})\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\}f(\mathbf{u}_i)f(\boldsymbol{\eta}_i)\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i. \tag{9}$$

Then

$$\frac{\partial\log f(\mathbf{y}_{i\cdot})}{\partial\gamma_k} = \frac{1}{f(\mathbf{y}_{i\cdot})}\frac{\partial f(\mathbf{y}_{i\cdot})}{\partial\gamma_k}$$

and

$$\frac{\partial f(\mathbf{y}_{i\cdot})}{\partial\gamma_k} = \iint\left\{\prod_{j=1}^{J_i}\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})f(\mathbf{v}_{ij})f(\boldsymbol{\xi}_{ij})\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\}\frac{\partial f(\mathbf{u}_i)}{\partial\gamma_k}f(\boldsymbol{\eta}_i)\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i.$$

Since

$$\log f(\mathbf{u}_i) \propto -\frac{1}{2}\sum_{k=1}^{p}\log\gamma_k - \frac{1}{2}\sum_{k=1}^{p}\frac{u_{ik}^2}{\gamma_k},$$

we have

$$\begin{aligned}\frac{\partial f(\mathbf{u}_i)}{\partial\gamma_k} &= f(\mathbf{u}_i)\frac{\partial\log f(\mathbf{u}_i)}{\partial\gamma_k}\\ &= f(\mathbf{u}_i)\left(-\frac{1}{2\gamma_k}+\frac{u_{ik}^2}{2\gamma_k^2}\right)\end{aligned}$$

and then

$$\frac{\partial\log f(\mathbf{y}_{i\cdot})}{\partial\gamma_k} =$$

$$\begin{aligned}&= \frac{1}{f(\mathbf{y}_{i\cdot})}\iint\left\{\prod_{j=1}^{J_i}\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})f(\mathbf{v}_{ij})f(\boldsymbol{\xi}_{ij})\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\}\\ &\quad\times f(\mathbf{u}_i)\left(-\frac{1}{2\gamma_k}+\frac{u_{ik}^2}{2\gamma_k^2}\right)f(\boldsymbol{\eta}_i)\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i\end{aligned}$$

$$\begin{aligned}&= \mathrm{E}\left\{\left(-\frac{1}{2\gamma_k}+\frac{u_{ik}^2}{2\gamma_k^2}\right)\Big|\,\mathbf{y}_{i\cdot}\right\}\\ &= -\frac{1}{2\gamma_k}+\frac{\widehat{u_{ik}^2}}{2\gamma_k^2}.\end{aligned}$$

Similarly,

$$\frac{\partial f(\mathbf{y}_{i\cdot})}{\partial\lambda_k} =$$

$$\begin{aligned}&\iint\Big[\sum_{j=1}^{J_i}\left\{\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})\frac{\partial f(\mathbf{v}_{ij})}{\partial\lambda_k}f(\boldsymbol{\xi}_{ij})\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\}\\ &\times\left\{\prod_{j'\neq j}\iint f(\mathbf{y}_{ij'}|\mathbf{u}_i,\mathbf{v}_{ij'},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij'})f(\mathbf{v}_{ij'})f(\boldsymbol{\xi}_{ij'})\mathrm{d}\mathbf{v}_{ij'}\mathrm{d}\boldsymbol{\xi}_{ij'}\right\}\Big]f(\mathbf{u}_i)f(\boldsymbol{\eta}_i)\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i\end{aligned}$$

and

$$\log f(\mathbf{v}_{ij}) \propto -\frac{1}{2}\sum_{k=1}^{q}\log\lambda_k - \frac{1}{2}\sum_{k=1}^{q}\frac{v_{ijk}^2}{\lambda_k},$$

so

$$\frac{\partial f(\mathbf{v}_{ij})}{\partial\lambda_k} = f(\mathbf{v}_{ij})\left(-\frac{1}{2\lambda_k}+\frac{v_{ijk}^2}{2\lambda_k^2}\right)$$

and then

$$\begin{aligned}\frac{\partial \log f(\mathbf{y}_{i\cdot})}{\partial \lambda_k} &= \sum_{j=1}^{J_i} \mathrm{E}\left\{ \left(-\frac{1}{2\lambda_k} + \frac{v_{ijk}^2}{2\lambda_k^2} \right) \middle| \mathbf{y}_{i\cdot} \right\} \\ &= -\frac{J_i}{2\lambda_k} + \frac{\sum_{j=1}^{J_i} \widehat{v_{ijk}^2}}{2\lambda_k^2}.\end{aligned}$$

The elements of **F** then come down to:

$$\begin{aligned}F_{kl} &= \mathrm{E}\left\{ \left(-\frac{1}{2\gamma_k} + \frac{\widehat{u_{ik}^2}}{2\gamma_k^2} \right) \left(-\frac{1}{2\gamma_l} + \frac{\widehat{u_{il}^2}}{2\gamma_l^2} \right) \right\} \\ &= \frac{1}{4\gamma_k\gamma_l} - \frac{\mathrm{E}(\widehat{u_{ik}^2})}{4\gamma_k^2\gamma_l} - \frac{\mathrm{E}(\widehat{u_{il}^2})}{4\gamma_k\gamma_l^2} + \frac{\mathrm{E}(\widehat{u_{ik}^2}\widehat{u_{il}^2})}{4\gamma_k^2\gamma_l^2} \\ &= -\frac{1}{4\gamma_k\gamma_l} + \frac{\mathrm{E}(\widehat{u_{ik}^2}\widehat{u_{il}^2})}{4\gamma_k^2\gamma_l^2}\end{aligned}$$

(since $\mathrm{E}\{\mathrm{E}(\widehat{u_{ik}^2})\} = \mathrm{E}(u_{ik}^2) = \lambda_k$) for $k = 1, \ldots, p$ and $l = 1, \ldots, p$;

$$\begin{aligned}F_{k,p+l} &= \mathrm{E}\left\{ \left(-\frac{1}{2\gamma_k} + \frac{\widehat{u_{ik}^2}}{2\gamma_k^2} \right) \left(-\frac{J_i}{2\lambda_l} + \frac{\sum_{j=1}^{J_i} \widehat{v_{ijl}^2}}{2\lambda_l^2} \right) \right\} \\ &= \frac{J_i}{4\gamma_k\lambda_l} - \frac{J_i\mathrm{E}(\widehat{u_{ik}^2})}{4\gamma_k^2\lambda_l} - \frac{\sum_{j=1}^{J_i} \mathrm{E}(\widehat{v_{ijl}^2})}{4\gamma_k\lambda_l^2} + \frac{\mathrm{E}(\widehat{u_{ik}^2}\sum_{j=1}^{J_i} \widehat{v_{ijl}^2})}{4\gamma_k^2\lambda_l^2} \\ &= -\frac{J_i}{4\gamma_k\lambda_l} + \frac{\mathrm{E}(\widehat{u_{ik}^2}\sum_{j=1}^{J_i} \widehat{v_{ijl}^2})}{4\gamma_k^2\lambda_l^2}\end{aligned}$$

for $k = 1, \ldots, p$ and $l = 1, \ldots, q$; and

$$\begin{aligned}F_{p+k,p+l} &= \mathrm{E}\left\{ \left(-\frac{J_i}{2\lambda_k} + \frac{\sum_{j=1}^{J_i} \widehat{v_{ijk}^2}}{2\lambda_k^2} \right) \left(-\frac{J_i}{2\lambda_l} + \frac{\sum_{j=1}^{J_i} \widehat{v_{ijl}^2}}{2\lambda_l^2} \right) \right\} \\ &= \frac{J_i^2}{4\lambda_k\lambda_l} - \frac{J_i\mathrm{E}(\sum_{j=1}^{J_i} \widehat{v_{ijk}^2})}{4\lambda_k^2\lambda_l} - \frac{J_i\mathrm{E}(\sum_{j=1}^{J_i} \widehat{v_{ijl}^2})}{4\lambda_k\lambda_l^2} + \frac{\mathrm{E}(\sum_{j=1}^{J_i} \widehat{v_{ijk}^2}\sum_{j=1}^{J_i} \widehat{v_{ijl}^2})}{4\lambda_k^2\lambda_l^2} \\ &= -\frac{J_i^2}{4\lambda_k\lambda_l} + \frac{\mathrm{E}(\sum_{j=1}^{J_i} \widehat{v_{ijk}^2}\sum_{j=1}^{J_i} \widehat{v_{ijl}^2})}{4\lambda_k^2\lambda_l^2}\end{aligned}$$

for $k = 1, \ldots, q$ and $l = 1, \ldots, q$. Once the parameters are replaced by their estimators and the expectations are replaced by averages over i, in view of (5) and (6) we obtain

$$\hat{F}_{kl} = -\frac{1}{4\hat{\gamma}_k\hat{\gamma}_l} + \frac{\text{ave}_i(\widehat{u_{ik}^2}\widehat{u_{il}^2})}{4\hat{\gamma}_k^2\hat{\gamma}_l^2}$$

for $k = 1, \ldots, p$ and $l = 1, \ldots, p$;

$$\hat{F}_{k,p+l} = -\frac{\text{ave}_i(J_i)}{4\hat{\gamma}_k\hat{\lambda}_l} + \frac{\text{ave}_i(\widehat{u_{ik}^2}\sum_{j=1}^{J_i}\widehat{v_{ijl}^2})}{4\hat{\gamma}_k^2\hat{\lambda}_l^2}$$

for $k = 1, \ldots, p$ and $l = 1, \ldots, q$; and

$$\hat{F}_{p+k,p+l} = -\frac{\text{ave}_i(J_i^2)}{4\hat{\lambda}_k\hat{\lambda}_l} + \frac{\text{ave}_i(\sum_{j=1}^{J_i}\widehat{v_{ijk}^2}\sum_{j=1}^{J_i}\widehat{v_{ijl}^2})}{4\hat{\lambda}_k^2\hat{\lambda}_l^2}$$

for $k = 1, \ldots, q$ and $l = 1, \ldots, q$.

For the warping parameters, we have

$$\log f(\boldsymbol{\eta}_i) \propto -\frac{1}{2}\log\det\mathbf{\Sigma} - \frac{1}{2}\boldsymbol{\eta}_i^T\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i$$

and then

$$\begin{aligned}
\mathrm{d}_{\mathbf{\Sigma}}\log f(\boldsymbol{\eta}_i) &= -\frac{1}{2}\text{tr}\left(\mathbf{\Sigma}^{-1}\mathrm{d}\mathbf{\Sigma}\right) + \frac{1}{2}\boldsymbol{\eta}_i^T\mathbf{\Sigma}^{-1}\left(\mathrm{d}\mathbf{\Sigma}\right)\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i \\
&= \left\{-\frac{1}{2}\text{vec}\left(\mathbf{\Sigma}^{-1}\right)^T + \frac{1}{2}\left(\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i \otimes \mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i\right)^T\right\}\text{vec}(\mathrm{d}\mathbf{\Sigma}).
\end{aligned}$$

The partial derivative is

$$\frac{\partial\log f(\boldsymbol{\eta}_i)}{\partial\Sigma_{kk}} = \left\{-\frac{1}{2}\text{vec}\left(\mathbf{\Sigma}^{-1}\right)^T + \frac{1}{2}\left(\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i \otimes \mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i\right)^T\right\}\mathbf{e}_{kk},$$

with $\mathbf{e}_{kk}$ the $r^2 \times 1$ vector having a 1 in the place corresponding to Σ_{kk} in $\text{vec}(\mathbf{\Sigma})$ and 0 elsewhere. Since $\mathbf{e}_{kk} = \text{vec}(\mathbf{E}_{kk})$ with $\mathbf{E}_{kk}$ the $r \times r$ matrix with 1 in the (k, k) coordinate and 0 elsewhere, it follows that

$$\begin{aligned}
\frac{\partial\log f(\boldsymbol{\eta}_i)}{\partial\Sigma_{kk}} &= -\frac{1}{2}\text{tr}\left(\mathbf{\Sigma}^{-1}\mathbf{E}_{kk}\right) + \frac{1}{2}\boldsymbol{\eta}_i^T\mathbf{\Sigma}^{-1}\left(\mathbf{E}_{kk}\right)\mathbf{\Sigma}^{-1}\boldsymbol{\eta}_i \\
&= -\frac{1}{2}\left(\mathbf{\Sigma}^{-1}\right)_{kk} + \frac{1}{2}\boldsymbol{\eta}_i^T\left(\mathbf{\Sigma}^{-1}\right)_{\cdot k}\left(\mathbf{\Sigma}^{-1}\right)_{\cdot k}^T\boldsymbol{\eta}_i,
\end{aligned}$$

where $(\mathbf{\Sigma}^{-1})_{\cdot k}$ denotes the kth columns of $\mathbf{\Sigma}^{-1}$. Then, as before, we have

$$\frac{\partial \log f(\mathbf{y}_{i\cdot})}{\partial \Sigma_{kk}} = \frac{1}{f(\mathbf{y}_{i\cdot})}\frac{\partial f(\mathbf{y}_{i\cdot})}{\partial \Sigma_{kk}}$$

$$= \frac{1}{f(\mathbf{y}_{i\cdot})}\iint \left\{\prod_{j=1}^{J_i}\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})f(\mathbf{v}_{ij})f(\boldsymbol{\xi}_{ij})\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\} f(\mathbf{u}_i)\frac{\partial f(\boldsymbol{\eta}_i)}{\partial \Sigma_{kk}}\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i$$

$$= \frac{1}{f(\mathbf{y}_{i\cdot})}\iint \left\{\prod_{j=1}^{J_i}\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})f(\mathbf{v}_{ij})f(\boldsymbol{\xi}_{ij})\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\} f(\mathbf{u}_i)\frac{\partial \log f(\boldsymbol{\eta}_i)}{\partial \Sigma_{kk}}f(\boldsymbol{\eta}_i)\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i$$

$$= -\frac{1}{2}\left(\mathbf{\Sigma}^{-1}\right)_{kk} + \frac{1}{2}\left(\mathbf{\Sigma}^{-1}\right)_{\cdot k}^{T}\mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\mathbf{\Sigma}^{-1}\right)_{\cdot k}.$$

Note that, by (7), this implies

$$(\hat{\mathbf{\Sigma}}^{-1})_{kk} = (\hat{\mathbf{\Sigma}}^{-1})_{\cdot k}^{T}\hat{\mathbf{\Sigma}}(\hat{\mathbf{\Sigma}}^{-1})_{\cdot k},$$

which is natural since any matrix and its inverse must satisfy this relationship. Similarly,

$$\log f(\boldsymbol{\xi}_{ij}) \propto -\frac{1}{2}\log\det\mathbf{\Omega} - \frac{1}{2}\boldsymbol{\xi}_{ij}^T\mathbf{\Omega}^{-1}\boldsymbol{\xi}_{ij},$$

so

$$\mathrm{d}_{\mathbf{\Omega}}\log f(\boldsymbol{\xi}_{ij}) = -\frac{1}{2}\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathrm{d}\Omega\right) + \frac{1}{2}\boldsymbol{\xi}_{ij}^T\mathbf{\Omega}^{-1}\left(\mathrm{d}\Omega\right)\mathbf{\Omega}^{-1}\boldsymbol{\xi}_{ij}$$

and then

$$\begin{aligned}\frac{\partial \log f(\boldsymbol{\xi}_{ij})}{\partial \Omega_{kk}} &= -\frac{1}{2}\mathrm{tr}\left(\mathbf{\Omega}^{-1}\mathbf{E}_{kk}\right) + \frac{1}{2}\boldsymbol{\xi}_{ij}^T\mathbf{\Omega}^{-1}\left(\mathbf{E}_{kk}\right)\mathbf{\Omega}^{-1}\boldsymbol{\xi}_{ij} \\ &= -\frac{1}{2}\left(\mathbf{\Omega}^{-1}\right)_{kk} + \frac{1}{2}\boldsymbol{\xi}_{ij}^T\left(\mathbf{\Omega}^{-1}\right)_{\cdot k}\left(\mathbf{\Omega}^{-1}\right)_{\cdot k}^T\boldsymbol{\xi}_{ij}.\end{aligned}$$

Therefore

$$\frac{\partial \log f(\mathbf{y}_{i\cdot})}{\partial \Omega_{kk}} = \frac{1}{f(\mathbf{y}_{i\cdot})}\frac{\partial f(\mathbf{y}_{i\cdot})}{\partial \Omega_{kk}}$$

$$\begin{aligned}= \ &\iint \Big[\sum_{j=1}^{J_i}\left\{\iint f(\mathbf{y}_{ij}|\mathbf{u}_i,\mathbf{v}_{ij},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij})f(\mathbf{v}_{ij})\frac{\partial f(\boldsymbol{\xi}_{ij})}{\partial \Omega_{kk}}\mathrm{d}\mathbf{v}_{ij}\mathrm{d}\boldsymbol{\xi}_{ij}\right\} \\ &\times\left\{\prod_{j'\neq j}\iint f(\mathbf{y}_{ij'}|\mathbf{u}_i,\mathbf{v}_{ij'},\boldsymbol{\eta}_i,\boldsymbol{\xi}_{ij'})f(\mathbf{v}_{ij'})f(\boldsymbol{\xi}_{ij'})\mathrm{d}\mathbf{v}_{ij'}\mathrm{d}\boldsymbol{\xi}_{ij'}\right\}\Big] f(\mathbf{u}_i)f(\boldsymbol{\eta}_i)\mathrm{d}\mathbf{u}_i\mathrm{d}\boldsymbol{\eta}_i\end{aligned}$$

$$= \sum_{j=1}^{J_i} \mathrm{E}\left\{ \left(-\frac{1}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{kk} + \frac{1}{2}\boldsymbol{\xi}_{ij}^T \left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k} \left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^T \boldsymbol{\xi}_{ij} \right) \middle| \mathbf{y}_{i\cdot} \right\}$$

$$= -\frac{J_i}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{kk} + \frac{1}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^T \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right) \left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k},$$

and by (8) we again have

$$(\hat{\boldsymbol{\Omega}}^{-1})_{kk} = (\hat{\boldsymbol{\Omega}}^{-1})_{\cdot k}^T \hat{\boldsymbol{\Omega}} (\hat{\boldsymbol{\Omega}}^{-1})_{\cdot k}.$$

The elements of $\mathbf{G}$ for $k = 1, \ldots, r$ and $l = 1, \ldots, r$, keeping in mind that $\mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right) = \boldsymbol{\Sigma}$, then come down to

$$\begin{aligned} G_{kl} &= \mathrm{E}\Big[\Big\{ -\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk} + \frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k} \Big\} \\ &\quad \times \Big\{ -\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\right)_{ll} + \frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l} \Big\}\Big] \end{aligned}$$

$$\begin{aligned} &= \frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Sigma}^{-1}\right)_{ll} - \frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Sigma}^{-1}\right)_{ll} \\ &\quad -\frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l} \\ &\quad +\frac{1}{4}\mathrm{E}\left\{ \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l} \right\} \end{aligned}$$

$$\begin{aligned} &= -\frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Sigma}^{-1}\right)_{ll} \\ &\quad +\frac{1}{4}\mathrm{E}\left\{ \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot l} \right\}. \end{aligned}$$

Now,

$$\begin{aligned} \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k} &= \left\{\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \otimes \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T\right\} \mathrm{vec}\left\{\mathrm{E}\left(\boldsymbol{\eta}_i\boldsymbol{\eta}_i^T|\mathbf{y}_{i\cdot}\right)\right\} \\ &= \left\{\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \otimes \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T\right\} \mathrm{E}\left(\boldsymbol{\eta}_i \otimes \boldsymbol{\eta}_i|\mathbf{y}_{i\cdot}\right) \end{aligned}$$

and similarly with the other factor (but transposed), so

$$\begin{aligned} G_{kl} &= -\frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Sigma}^{-1}\right)_{ll} \\ &\quad +\frac{1}{4}\left\{\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T \otimes \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^T\right\} \mathrm{E}\left(\widehat{\boldsymbol{\eta}_i^{\otimes 2}}\widehat{\boldsymbol{\eta}_i^{\otimes 2}}^T\right)\left\{\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k} \otimes \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\right\} \end{aligned}$$

where $\widehat{\boldsymbol{\eta}_i^{\otimes 2}} = \mathrm{E}\left(\boldsymbol{\eta}_i \otimes \boldsymbol{\eta}_i | \mathbf{y}_{i\cdot}\right)$. The other elements, also for $k = 1, \ldots, r$ and $l = 1, \ldots, r$, are

$$\begin{aligned} G_{k,r+l} &= \mathrm{E}\Big[\Big\{-\frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk} + \frac{1}{2}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^{T} \mathrm{E}\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_i^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\Big\} \\ &\quad \times \Big\{-\frac{J_i}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} + \frac{1}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}\Big\}\Big] \end{aligned}$$

$$\begin{aligned} &= \frac{J_i}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} - \frac{J_i}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^{T} \mathrm{E}\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_i^T\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} \\ &\quad -\frac{1}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l} \\ &\quad +\frac{1}{4}\mathrm{E}\Big\{\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^{T} \mathrm{E}\left(\boldsymbol{\eta}_i \boldsymbol{\eta}_i^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}\Big\} \end{aligned}$$

$$\begin{aligned} &= -\frac{J_i}{4}\left(\boldsymbol{\Sigma}^{-1}\right)_{kk}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} \\ &\quad +\frac{1}{4}\left\{\left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^{T} \otimes \left(\boldsymbol{\Sigma}^{-1}\right)_{\cdot k}^{T}\right\} \mathrm{E}\Big(\widehat{\boldsymbol{\eta}_i^{\otimes 2}} \sum_{j=1}^{J_i} \widehat{\boldsymbol{\xi}_{ij}^{\otimes 2}}^{T}\Big)\left\{\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l} \otimes \left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}\right\}, \end{aligned}$$

where $\widehat{\boldsymbol{\xi}_{ij}^{\otimes 2}} = \mathrm{E}\left(\boldsymbol{\xi}_{ij} \otimes \boldsymbol{\xi}_{ij} | \mathbf{y}_{i\cdot}\right)$, and

$$\begin{aligned} G_{r+k,r+l} &= \mathrm{E}\Big[\Big\{-\frac{J_i}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{kk} + \frac{1}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}\Big\} \\ &\quad \times \Big\{-\frac{J_i}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} + \frac{1}{2}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}\Big\}\Big] \end{aligned}$$

$$\begin{aligned} &= \frac{J_i^2}{4}\left(\boldsymbol{\Omega}^{-1}\right)_{kk}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} - \frac{J_i}{4}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} \\ &\quad -\frac{J_i}{4}\left(\boldsymbol{\Omega}^{-1}\right)_{kk}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l} \\ &\quad +\frac{1}{4}\mathrm{E}\Big\{\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}^{T} \sum_{j=1}^{J_i} \mathrm{E}\left(\boldsymbol{\xi}_{ij}\boldsymbol{\xi}_{ij}^T | \mathbf{y}_{i\cdot}\right)\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}\Big\} \end{aligned}$$

$$
\begin{aligned}
= & -\frac{J_i^2}{4}\left(\boldsymbol{\Omega}^{-1}\right)_{kk}\left(\boldsymbol{\Omega}^{-1}\right)_{ll} \\
& +\frac{1}{4}\left\{\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^T \otimes\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot k}^T\right\} \mathrm{E}\Big(\sum_{j=1}^{J_i} \widehat{\boldsymbol{\xi}_{ij}^{\otimes 2}} \sum_{j=1}^{J_i} \widehat{\boldsymbol{\xi}_{ij}^{\otimes 2}}^T\Big)\left\{\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l} \otimes\left(\boldsymbol{\Omega}^{-1}\right)_{\cdot l}\right\} .
\end{aligned}
$$

As always, to obtain $\hat{\mathrm{G}}$ expectations are replaced by averages over $i = 1, \ldots, I$.

3 Web Appendix C: Cubic Hermite Splines

3.1 Interpolation on a single interval

3.1.1 Interpolation on the interval $[0, 1]$

Given values of the function to be interpolated, f_0 at $t = 0$ and f_1 at $t = 1$, and values of the derivatives at those points, d_0 at $t = 0$ and d_1 at $t = 1$, then

$$
f(t) = h_{00}(t)f_0 + h_{10}(t)d_0 + h_{01}(t)f_1 + h_{11}(t)d_1
$$

with

$$
\begin{aligned}
h_{00}(t) &= 2t^3 - 3t^2 + 1 = (1 + 2t)(1 - t)^2 \\
h_{10}(t) &= t^3 - 2t^2 + t = t(1 - t)^2 \\
h_{01}(t) &= -2t^3 + 3t^2 = t^2(3 - 2t) = h_{00}(1 - t) \\
h_{11}(t) &= t^3 - t^2 = t^2(t - 1) = -h_{10}(1 - t)
\end{aligned}
$$

satisfies

$$
\begin{aligned}
f(0) &= f_0, \;\; f(1) = f_1, \\
f'_+(0) &= d_0, \;\; f'_-(1) = d_1.
\end{aligned}
$$

Since $f(t)$ is a polynomial of degree 3 (i.e. has 4 free coefficients) and satisfies these four conditions, it's the *only* cubic polynomial that satisfies these conditions.

3.1.2 Interpolation on a general interval $[x_k, x_{k+1}]$

If we are now given values of the function f_k at $t = x_k$ and f_{k+1} at $t = x_{k+1}$, and values of the derivatives at those points, d_k at $t = x_k$ and d_{k+1} at $t = x_{k+1}$, then

$$f(t) = h_{00}\left(\frac{t - x_k}{s_k}\right) f_k + h_{10}\left(\frac{t - x_k}{s_k}\right) s_k d_k + h_{01}\left(\frac{t - x_k}{s_k}\right) f_{k+1} + h_{11}\left(\frac{t - x_k}{s_k}\right) s_k d_{k+1}$$

with $s_k = x_{k+1} - x_k$ and $h_{ij}(t)$ as before, satisfies

$$\begin{aligned} f(x_k) &= f_k, \;\; f(x_{k+1}) = f_{k+1}, \\ f'_+(x_k) &= d_k, \;\; f'_-(x_{k+1}) = d_{k+1}. \end{aligned}$$

Again, this cubic polynomial is unique, subject to these four conditions.

3.2 Interpolating a data set

Suppose we have p ponts in some interval $[a, b]$, $a < x_1 < x_2 < \cdots < x_p < b$, and corresponding values of the function, f_k, and the derivative, d_k, at each point x_k. Let's define $x_0 = a$, $x_{p+1} = b$ and f_0, f_{p+1}, d_0 and d_{p+1} accordingly. Then the piecewise cubic interpolant is going to be

$$\begin{aligned} f(t) &= \sum_{k=0}^{p} \Bigg\{ h_{00}\left(\frac{t - x_k}{s_k}\right) f_k + h_{10}\left(\frac{t - x_k}{s_k}\right) s_k d_k \\ &\quad + h_{01}\left(\frac{t - x_k}{s_k}\right) f_{k+1} + h_{11}\left(\frac{t - x_k}{s_k}\right) s_k d_{k+1} \Bigg\} \\ &= \sum_{k=0}^{p} h_{00}\left(\frac{t - x_k}{s_k}\right) f_k + \sum_{k=0}^{p} h_{10}\left(\frac{t - x_k}{s_k}\right) s_k d_k \\ &\quad + \sum_{k=1}^{p+1} h_{01}\left(\frac{t - x_{k-1}}{s_{k-1}}\right) f_k + \sum_{k=1}^{p+1} h_{11}\left(\frac{t - x_{k-1}}{s_{k-1}}\right) s_{k-1} d_k \\ &= h_{00}\left(\frac{t - x_0}{s_0}\right) f_0 + h_{10}\left(\frac{t - x_0}{s_0}\right) s_0 d_0 \\ &\quad + \sum_{k=1}^{p} \left\{ h_{00}\left(\frac{t - x_k}{s_k}\right) + h_{01}\left(\frac{t - x_{k-1}}{s_{k-1}}\right) \right\} f_k \\ &\quad + \sum_{k=1}^{p} \left\{ h_{10}\left(\frac{t - x_k}{s_k}\right) s_k + h_{11}\left(\frac{t - x_{k-1}}{s_{k-1}}\right) s_{k-1} \right\} d_k \\ &\quad + h_{01}\left(\frac{t - x_p}{s_p}\right) f_{p+1} + h_{11}\left(\frac{t - x_p}{s_p}\right) s_p d_{p+1} \end{aligned}$$

with $s_k = x_{k+1} - x_k$ as before, and we are defining $h_{ij}(t) = 0$ for $t \notin [0, 1]$.

Then we can introduce the basis functions

$$\phi_k(t) = \begin{cases} h_{00}\left(\frac{t-x_0}{s_0}\right) & \text{if } k = 0 \\ h_{00}\left(\frac{t-x_k}{s_k}\right) + h_{01}\left(\frac{t-x_{k-1}}{s_{k-1}}\right) & \text{if } k = 1, \ldots, p \\ h_{01}\left(\frac{t-x_p}{s_p}\right) & \text{if } k = p+1 \end{cases}$$

and

$$\psi_k(t) = \begin{cases} h_{10}\left(\frac{t-x_0}{s_0}\right) s_0 & \text{if } k = 0 \\ h_{10}\left(\frac{t-x_k}{s_k}\right) s_k + h_{11}\left(\frac{t-x_{k-1}}{s_{k-1}}\right) s_{k-1} & \text{if } k = 1, \ldots, p \\ h_{11}\left(\frac{t-x_p}{s_p}\right) s_p & \text{if } k = p+1, \end{cases}$$

and we have

$$\boxed{f(t) = \sum_{k=0}^{p+1} \phi_k(t) f_k + \sum_{k=0}^{p+1} \psi_k(t) d_k}$$

Since $h_{01}(t) = h_{00}(1-t)$ and $h_{11}(t) = -h_{10}(1-t)$,

$$\phi_k(t) = \begin{cases} h_{00}\left(\frac{t-x_0}{s_0}\right) & \text{if } k = 0 \\ h_{00}\left(\frac{t-x_k}{s_k}\right) + h_{00}\left(\frac{x_k-t}{s_{k-1}}\right) & \text{if } k = 1, \ldots, p \\ h_{00}\left(\frac{x_{p+1}-t}{s_p}\right) & \text{if } k = p+1 \end{cases}$$

and

$$\psi_k(t) = \begin{cases} h_{10}\left(\frac{t-x_0}{s_0}\right) s_0 & \text{if } k = 0 \\ h_{10}\left(\frac{t-x_k}{s_k}\right) s_k - h_{10}\left(\frac{x_k-t}{s_{k-1}}\right) s_{k-1} & \text{if } k = 1, \ldots, p \\ -h_{10}\left(\frac{x_{p+1}-t}{s_p}\right) s_p & \text{if } k = p+1. \end{cases}$$

For $k = 1, \ldots, p$ we have

$$\phi_k(t) = \begin{cases} 0 & \text{if } t < x_{k-1} \text{ or } t > x_{k+1} \\ h_{00}\left(\frac{x_k-t}{s_{k-1}}\right) & \text{if } x_{k-1} \le t \le x_k \\ h_{00}\left(\frac{t-x_k}{s_k}\right) & \text{if } x_k \le t \le x_{k+1} \end{cases}$$

and

$$\psi_k(t) = \begin{cases} 0 & \text{if } t < x_{k-1} \text{ or } t > x_{k+1} \\ -h_{10}\left(\frac{x_k-t}{s_{k-1}}\right) s_{k-1} & \text{if } x_{k-1} \le t \le x_k \\ h_{10}\left(\frac{t-x_k}{s_k}\right) s_k & \text{if } x_k \le t \le x_{k+1}. \end{cases}$$

Since $h_{00}(0) = 1$, $h_{00}(1) = 0$ and $h'_{00}(0) = h'_{00}(1) = 0$, the ϕ_ks are continuous and differentiable everywhere as functions of t, with $\phi_k(x_{k-1}) = \phi_k(x_{k+1}) = 0$, $\phi_k(x_k) = 1$ and $\phi'_k(x_{k-1}) = \phi'_k(x_k) = \phi'_k(x_{k+1}) = 0$. Similarly, since $h_{10}(0) = h_{10}(1) = 0$, $h'_{10}(0) = 1$ and $h'_{10}(1) = 0$, the ψ_ks are also continuous and differentiable everywhere as functions of t, with $\psi_k(x_{k-1}) = \psi_k(x_k) = \psi_k(x_{k+1}) = 0$, $\psi'_k(x_{k-1}) = \psi'_k(x_{k+1}) = 0$ and $\psi'_k(x_k) = 1$.

For the "border" basis functions we have

$$\phi_0(t) = \begin{cases} 0 & \text{if } t < a \text{ or } t > x_1 \\ h_{00}\left(\frac{t-a}{x_1-a}\right) & \text{if } a \le t \le x_1 \end{cases}$$

which is discontinuous only at $t = a$, with $\phi_0(a) = 1$, $\phi_0(x_1) = 0$, and $(\phi_0)'_+(a) = \phi'_0(x_1) = 0$;

$$\phi_{p+1}(t) = \begin{cases} 0 & \text{if } t < x_p \text{ or } t > b \\ h_{00}\left(\frac{b-t}{b-x_p}\right) & \text{if } x_p \le t \le b \end{cases}$$

which is discontinuous only at $t = b$, with $\phi_{p+1}(x_p) = 0$, $\phi_{p+1}(b) = 1$, and $\phi'_{p+1}(x_p) = \left(\phi_{p+1}\right)'_-(b) = 0$;

$$\psi_0(t) = \begin{cases} 0 & \text{if } t < a \text{ or } t > x_1 \\ h_{10}\left(\frac{t-a}{x_1-a}\right) s_0 & \text{if } a \le t \le x_1 \end{cases}$$

which is continuous everywhere, with $\psi_0(a) = \psi_0(x_1) = 0$, $(\psi_0)'_+(a) = 1$, and $\psi'_0(x_1) = 0$; and

$$\psi_{p+1}(t) = \begin{cases} 0 & \text{if } t < x_p \text{ or } t > b \\ -h_{10}\left(\frac{b-t}{b-x_p}\right) s_p & \text{if } x_p \le t \le b \end{cases}$$

which is continuous everywhere, with $\psi_{p+1}(x_p) = \psi_{p+1}(b) = 0$, $\psi'_{p+1}(x_p) = 0$, and $\left(\psi_{p+1}\right)'_-(b) = 1$.

3.3 Monotone interpolation

Given $a = x_0 < x_1 < \cdots < x_p < x_{p+1} = b$ and $f_0 < \cdots < f_{p+1}$, with d_ks unspecified, it's always possible to find d_ks such that the resulting $f(t)$ is strictly increasing (Fritsch and Carlson, 1980). Let

$$\Delta_k = \frac{f_{k+1} - f_k}{x_{k+1} - x_k}, \quad \alpha_k = \frac{d_k}{\Delta_k}, \quad \beta_k = \frac{d_{k+1}}{\Delta_k}.$$

Then $f(t)$ is monotone in $[x_k, x_{k+1}]$ if and only if:

1. $\alpha_k + \beta_k - 2 \leq 0$ and $\text{sign}(d_k) = \text{sign}(d_{k+1}) = \text{sign}(\Delta_k)$; or

2. $\alpha_k + \beta_k - 2 > 0$, $\text{sign}(d_k) = \text{sign}(d_{k+1}) = \text{sign}(\Delta_k)$, and:

   (a) $2\alpha_k + \beta_k - 3 \leq 0$, or

   (b) $\alpha_k + 2\beta_k - 3 \leq 0$, or

   (c) $\alpha_k - (2\alpha_k + \beta_k - 3)^2/\{3(\alpha_k + \beta_k - 2)\} \geq 0$.

The condition

$$\sqrt{\alpha_k^2 + \beta_k^2} \leq 3$$

implies either 1 or 2(a)–2(c) above, so it's sufficient to guarantee monotonicity. This motivates the following algorithm for constructing the d_ks:

1. Initialize the derivatives $\{d_k\}$ so that $\text{sign}(d_k) = \text{sign}(d_{k+1}) = \text{sign}(\Delta_k)$. For instance,

$$d_0 = \Delta_0, \;\; d_k = \frac{\Delta_{k-1} + \Delta_k}{2} \text{ for } k = 1, \ldots, p, \;\; d_{p+1} = \Delta_p.$$

2. For $k = 0, \ldots, p$:

   (a) If $\sqrt{\alpha_k^2 + \beta_k^2} \leq 3$ the interpolant will be monotone in $[x_k, x_{k+1}]$; go to next k.

   (b) If $\sqrt{\alpha_k^2 + \beta_k^2} > 3$, let $\tau_k = 3/\sqrt{\alpha_k^2 + \beta_k^2}$, $\alpha_k^* = \tau_k \alpha_k$, and $\beta_k^* = \tau_k \beta_k$; set

$$d_k = \alpha_k^* \Delta_k, \;\; d_{k+1} = \beta_k^* \Delta_k.$$

   The interpolant will be monotone in $[x_k, x_{k+1}]$; go to next k.

The algorithm may change the value of each d_k at most twice from its initial value: first when the interval $[x_{k-1}, x_k]$ is considered and again when the interval $[x_k, x_{k+1}]$ is considered. But since $0 \leq \alpha_k^* \leq \alpha_k$ and $0 \leq \beta_k^* \leq \beta_k$, the modification of d_k for $[x_k, x_{k+1}]$ will maintain the monotonicity condition on $[x_{k-1}, x_k]$; see comments on p. 241 of Fritsch and Carlson (1980).

4 Web Appendix D: Beetle Growth Data Analysis

Figure 1 shows Normal Probability plots of the estimated random effects $\{\hat{U}_i\}$ (amplitude main factor), $\{\hat{V}_{ij}\}$ (amplitude residual factor), $\{\hat{\eta}_i\}$ (Jupp-transformed warping knot main

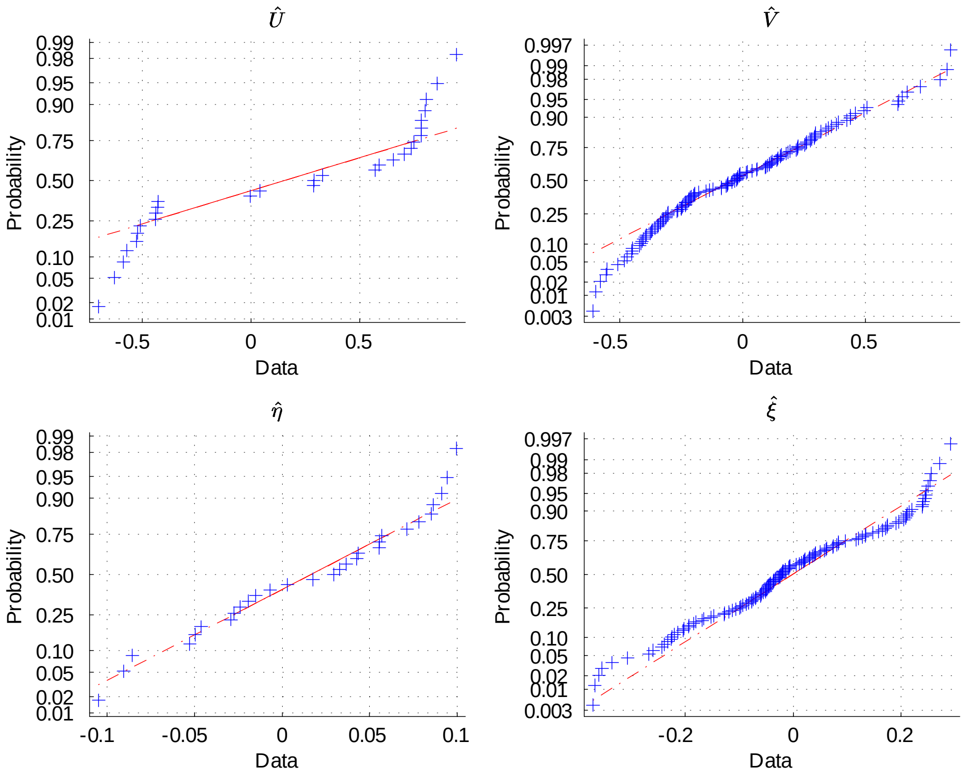

Figure 1: Flour Beetle Growth Example. Normal probability plots of estimated random effects for warped ANOVA fit.

factor), and $\{\hat{\xi}_{ij}\}$ (Jupp-transformed warping knot residual factor). The normality assumption is reasonable except perhaps for the $\hat{U}_i$s, which show lighter tails than expected for a Normal distribution, but no outliers or asymmetry. Figure 2 shows the grand mean $\hat{\mu}(t)$ alongside the 29 group means of the warped curves $\hat{z}_{ij}(t)$, confirming again that there are not outlying groups.

We also studied the residuals of the raw data with respect to the smooth estimated trajectories, $\hat{\varepsilon}_{ijk} = y_{ijk} - \hat{x}_{ij}(t_{ijk})$. Figure 3(a) shows a few isolated large residuals but does not reveal any individual outlying trajectory (i.e. no individuals stand out as consistently having large residuals). Similarly, the Normal probability plot in Figure 3(b) shows a few large residuals at the tails but the plot is generaly consistent with a Normal distribution.

To validate the asymptotic standard deviations for the variance ratios $\hat{h}_z$ and $\hat{h}_w$, we bootstrapped these estimators using residual-resampling bootstrap. The histograms based on 200 replications are shown in Figure 4. We see that the bootstrap distribution of $\hat{h}_z$ looks reasonably close to Normal but the truncated left tail for $\hat{h}_w$ is more noticeable,

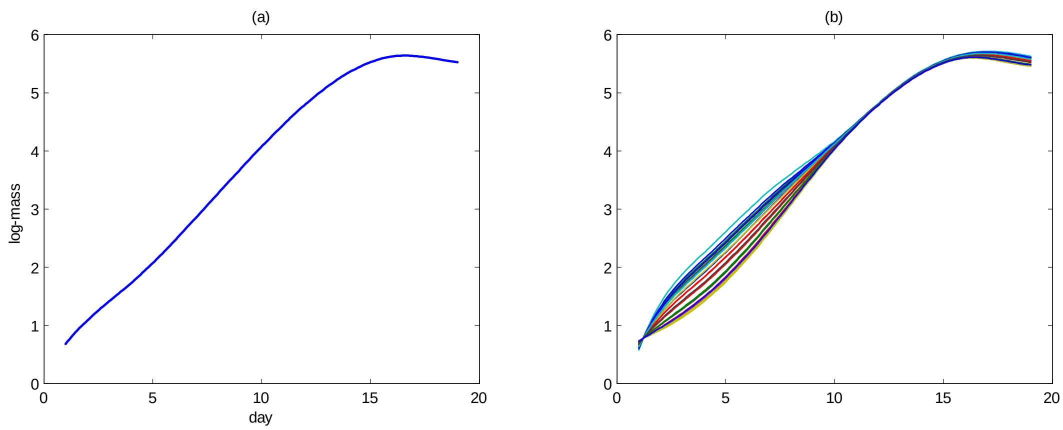

Figure 2: Flour Beetle Growth Example. (a) Grand mean $\hat{\mu}(t)$ and (b) the 29 group means $\bar{\hat{z}}_{i\cdot}(t)$ of the estimated warped curves.

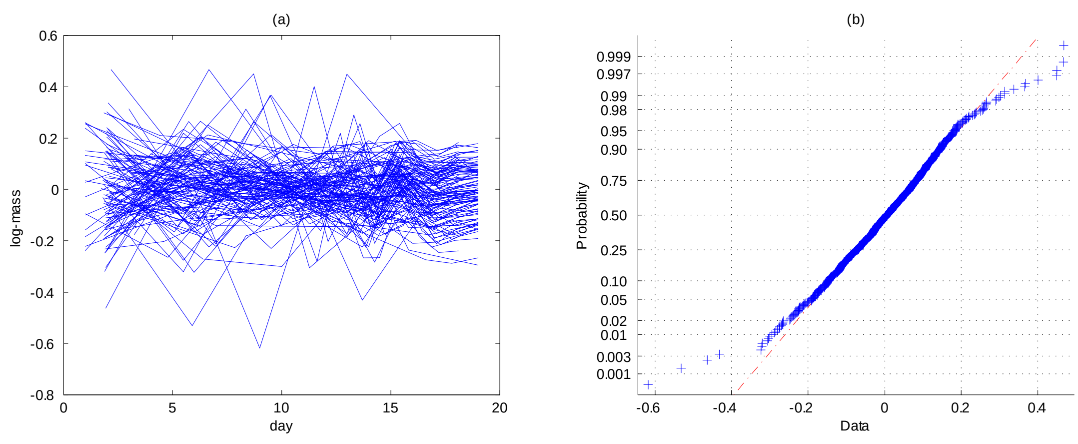

Figure 3: Flour Beetle Growth Example. (a) Residual trajectories $\hat{\varepsilon}_{ijk} = y_{ijk} - \hat{x}_{ij}(t_{ijk})$ and (b) Normal probability plot of the $\hat{\varepsilon}_{ijk}$s.

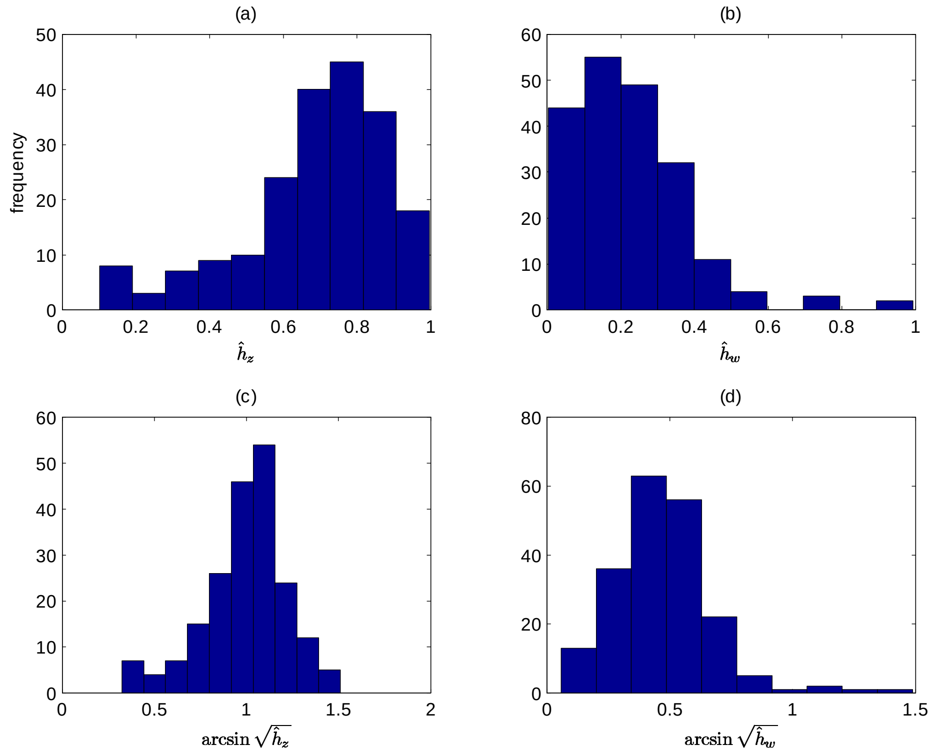

Figure 4: Flour Beetle Growth Example. Histograms of 200 bootstrap replications of (a) $\hat{h}_z$ and (b) $\hat{h}_w$, and their respective transformations (c) $\arcsin\sqrt{\hat{h}_z}$ and (d) $\arcsin\sqrt{\hat{h}_w}$.

whereas for the arcsin-root transformation the Normal approximation is accurate in both cases.

5 Web Appendix E: Model identifiability and uniqueness of warping functions

In this section we discuss the problem of identifiability of models of the form $x_i(t) = z_i\{w_i^{-1}(t)\}$ when the w_is belong to a given space of warping functions $\mathscr{W}$. We begin our discussion with a different but related issue: *when* is it possible to align curves in a meaningful way? Broadly speaking, curve alignment only makes sense when the curves have the same number of features (peaks and valleys). On the contrary, if for instance some curves in the sample have two peaks and some curves have three peaks, it is clearly not possible to align them in a meaningful way. In those cases it makes more sense to align the curves by clusters, as in Sangalli *et al.* (2010), but not to try to align the whole sample. So the first condition for a sensible curve alignment is that the curves have the same number of features. Once this is met, the existence and uniqueness of warping functions in $\mathscr{W}$ that perfectly align the curves is easy to establish. This discussion largely follows Kneip and Ramsay (2008), although we use a finite-dimensional $\mathscr{W}$.

Definition 1 *We say that the curves $\{x_1, \ldots, x_n\}$, with $x_i : [a, b] \to \mathbb{R}$ differentiable, have K features if*

$$\#\{t \in (a, b) : x_i'(t) = 0\} = K$$

for all $i = 1, \ldots, n$. In that case, for each curve $x_i(t)$ we define the vector of features $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{iK})$ as the zeros of $x_i'(t)$ taken in increasing order.

Definition 2 *We say that the curves $\{z_1, \ldots, z_n\}$ are aligned at $\boldsymbol{\zeta}_0 \in \mathbb{R}^K$ if they have K features and if $\boldsymbol{\zeta}_i = \boldsymbol{\zeta}_0$ for all $i = 1, \ldots, n$.*

Proposition 3 *If the curves $\{x_1, \ldots, x_n\}$ have K features and $\mathscr{W}$ is a K-dimensional linear space with basis $\{\psi_1, \ldots, \psi_K\}$ such that the vectors $\{\psi_1(\bar{\boldsymbol{\zeta}}), \ldots, \psi_K(\bar{\boldsymbol{\zeta}})\}$ are linearly independent, then there exist unique warping functions $\{w_1, \ldots, w_n\}$ in $\mathscr{W}$ such that $\{x_1 \circ w_1, \ldots, x_n \circ w_n\}$ are aligned at $\bar{\boldsymbol{\zeta}}$.*

Proof. The matrix $\boldsymbol{\Psi} = [\psi_1(\bar{\boldsymbol{\zeta}}), \ldots, \psi_K(\bar{\boldsymbol{\zeta}})] \in \mathbb{R}^{K \times K}$ is invertible by assumption. Let $w_i(t) = \sum_{j=1}^{K} \theta_{ij}\psi_j(t)$ with $\boldsymbol{\theta}_i = \boldsymbol{\Psi}^{-1}\boldsymbol{\zeta}_i$. Then the functions $\{x_1 \circ w_1, \ldots, x_n \circ w_n\}$ are aligned at $\bar{\boldsymbol{\zeta}}$, since

$$\begin{aligned}(x_i \circ w_i)'(\bar{\zeta}_{\cdot j}) &= x_i'(w_i(\bar{\zeta}_{\cdot j}))w_i'(\bar{\zeta}_{\cdot j}) \\ &= x_i'(\zeta_{ij})w_i'(\bar{\zeta}_{\cdot j}) \\ &= 0\end{aligned}$$

because the ζ_{ij}s are the zeros of $x_i'(t)$. Conversely, if $\{x_1 \circ \tilde{w}_1, \ldots, x_n \circ \tilde{w}_n\}$ are aligned at $\bar{\boldsymbol{\zeta}}$, then $(x_i \circ \tilde{w}_i)'(\bar{\zeta}_{\cdot j}) = 0$ for all i and j, which implies $x_i'(\tilde{w}_i(\bar{\zeta}_{\cdot j})) = 0$ for all i and j because $\tilde{w}_i'(\bar{\zeta}_{\cdot j}) > 0$. Then $\tilde{w}_i(\bar{\zeta}_{\cdot j}) = \zeta_{ij}$ for $j = 1, \ldots, K$, so if we write $\tilde{w}_i(t) = \sum_{j=1}^{K} \tilde{\theta}_{ij}\psi_j(t)$, we have $\boldsymbol{\Psi}\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\zeta}_i$. But $\boldsymbol{\Psi}$ is invertible, so $\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i$ and then $\tilde{w}_i = w_i$ for all i. ■

Note that $\{\psi_1, \ldots, \psi_K\}$ being linearly independent functions does not automatically imply that $\{\psi_1(\bar{\boldsymbol{\zeta}}), \ldots, \psi_K(\bar{\boldsymbol{\zeta}})\}$ are linearly independent vectors in $\mathbb{R}^K$. However, this will generally be the case if the family $\mathscr{W}$ is reasonably specified. Consider, for example, the case of monotone Hermite splines with knot vector $\boldsymbol{\tau}$. If $\boldsymbol{\tau} = \bar{\boldsymbol{\zeta}}$ then it is clear that $\{\psi_1(\bar{\boldsymbol{\zeta}}), \ldots, \psi_K(\bar{\boldsymbol{\zeta}})\}$ are independent, since $\psi_j(\boldsymbol{\tau}_0) = \tau_j\mathbf{e}_j$ by definition. But even if $\boldsymbol{\tau} \neq \bar{\boldsymbol{\zeta}}$, the vectors $\{\psi_1(\bar{\boldsymbol{\zeta}}), \ldots, \psi_K(\bar{\boldsymbol{\zeta}})\}$ will still be independent as long as $\boldsymbol{\tau}$ is reasonably close to $\bar{\boldsymbol{\zeta}}$. On the other hand, if $\boldsymbol{\tau}$ is grossly misspecified, this condition may fail:

for example, if $\bar{\boldsymbol{\zeta}} = (.2, .3)$ and we take $\boldsymbol{\tau} = (.6, .7)$, given that the Hermite basis functions have compact support between three consecutive knots (or two knots and a boundary point), we have $\psi_2(\bar{\boldsymbol{\zeta}}) = \mathbf{0}$ and then $\{\psi_1(\bar{\boldsymbol{\zeta}}), \psi_2(\bar{\boldsymbol{\zeta}})\}$ are clearly not independent. But ruling out grossly misspecified cases like this, the condition that $\{\psi_1(\bar{\boldsymbol{\zeta}}), \ldots, \psi_K(\bar{\boldsymbol{\zeta}})\}$ are linearly independent is, in general, easily met.

The next proposition shows that the model $x_i(t) = z_i\{v_i(t)\}$ is identifiable under similar conditions to those of Proposition 3, although more general, since the dimension of $\mathscr{W}$ is allowed to be smaller than the number of features K (because, for example, the model may warp some peaks but not others, as in the simulations.)

Proposition 4 *Let $\{x_1, \ldots, x_n\}$ be a sample of curves generated by the model $x_i(t) = z_i\{v_i(t)\}$, where $v_i(t) = w_i^{-1}(t)$, $w_i \in \mathscr{W}$, $\bar{w}(t) = t$ for all $t \in [a, b]$, the curves $\{z_1, \ldots, z_n\}$ are aligned at $\boldsymbol{\zeta}_0 \in \mathbb{R}^K$, and $\mathscr{W}$ is an r-dimensional linear space with $r \leq K$ and basis $\{\psi_1, \ldots, \psi_r\}$ such that $\{\psi_1(\boldsymbol{\zeta}_0), \ldots, \psi_r(\boldsymbol{\zeta}_0)\}$ are linearly independent in $\mathbb{R}^K$. Then, if $x_i(t) = \tilde{z}_i\{\tilde{v}_i(t)\}$ for $\tilde{v}_i(t) = \tilde{w}_i^{-1}(t)$ with $\tilde{w}_i \in \mathscr{W}$, $\overline{\tilde{w}}(t) = t$ for all $t \in [a, b]$, and $\{\tilde{z}_1, \ldots, \tilde{z}_n\}$ aligned at some $\tilde{\boldsymbol{\zeta}}_0 \in \mathbb{R}^{\tilde{K}}$, we have $\tilde{w}_i = w_i$ and $\tilde{z}_i = z_i$ for all i (and in particular, $\tilde{K} = K$ and $\tilde{\boldsymbol{\zeta}}_0 = \boldsymbol{\zeta}_0$).*

Proof. By construction, $x_i(t)$ has K features given by $\boldsymbol{\zeta}_i = w_i(\boldsymbol{\zeta}_0)$, and $w_i(t) = \sum_{j=1}^{r} \theta_{ij}\psi_j(t)$ for some $\boldsymbol{\theta}_i \in \mathbb{R}^r$ that satisfies $\boldsymbol{\Psi}\boldsymbol{\theta}_i = \boldsymbol{\zeta}_i$, with $\boldsymbol{\Psi} = [\psi_1(\boldsymbol{\zeta}_0), \ldots, \psi_r(\boldsymbol{\zeta}_0)] \in \mathbb{R}^{K \times r}$ a full-rank matrix. Also, since $\bar{w}(t) \equiv t$, we have $\bar{\boldsymbol{\zeta}} = \boldsymbol{\zeta}_0$. Now suppose $x_i(t) = \tilde{z}_i\{\tilde{v}_i(t)\}$ with the $\tilde{z}_i(t)$s aligned at some $\tilde{\boldsymbol{\zeta}}_0$. Since $x_i'(t) = 0$ if and only if $\tilde{z}_i'\{\tilde{v}_i(t)\} = 0$, and we already know that $x_i'(t) = 0$ if and only if $t = \zeta_{ij}$ for some $j = 1, \ldots, K$, it follows that $\tilde{v}_i(\zeta_{ij})$ must be equal to one of the coordinates of $\tilde{\boldsymbol{\zeta}}_0$. But the $\tilde{v}_i$s are strictly increasing, so $\tilde{K}$ must be equal to K, and then $\tilde{v}_i(\boldsymbol{\zeta}_i) = \tilde{\boldsymbol{\zeta}}_0$, so $\tilde{w}_i(\tilde{\boldsymbol{\zeta}}_0) = \boldsymbol{\zeta}_i$ for all i. Taking averages we get $\tilde{\boldsymbol{\zeta}}_0 = \bar{\boldsymbol{\zeta}}$, since $\overline{\tilde{w}}(t) = t$ for all t; so $\tilde{\boldsymbol{\zeta}}_0 = \boldsymbol{\zeta}_0$. Then $\tilde{w}_i(\boldsymbol{\zeta}_0) = \boldsymbol{\zeta}_i$, that is, there exists a coefficient vector $\tilde{\boldsymbol{\theta}}_i \in \mathbb{R}^r$ such that $\boldsymbol{\Psi}\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\zeta}_i$; but $\boldsymbol{\Psi}$ is a full-rank matrix, so $\tilde{\boldsymbol{\theta}}_i = \boldsymbol{\theta}_i$ and then $\tilde{w}_i = w_i$ for all i. Since $\tilde{z}_i(t) = x_i\{\tilde{w}_i(t)\}$ and $\tilde{w}_i = w_i$, it follows that $\tilde{z}_i = z_i$ for all i. ■

6 References

Fritsch, F.N. and Carlson, R.E. (1980). Monotone piecewise cubic interpolation. *SIAM J. Numer. Anal.* **17** 238–246.

Kneip, A., and Ramsay, J. O. (2008). Combining registration and fitting for functional models. *Journal of the American Statistical Association* **103** 1155–1165.

Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics (Second Edition)*. Wiley, New York.

Sangalli, L.M., Secchi, P., Vantini, S., and Vitelli, V. (2010). k-mean alignment for curve clustering. *Computational Statistics and Data Analysis* **54** 1219–1233.