JOINT MODELS FOR GRID POINT AND RESPONSE PROCESSES IN LONGITUDINAL AND FUNCTIONAL DATA

Department of Mathematical Sciences, University of Wisconsin–Milwaukee

Supplementary Material

This supplement contains detailed technical derivations, proofs of theorems, and additional simulation results.

0.1 Estimation

0.1.1 The model

We observe the data $\{(x_{i1}, y_{i1}), \ldots, (x_{im_i}, y_{im_i}) : i = 1, \ldots, n\}$, which we write in vector form as $\{(\mathbf{x}_i, m_i, \mathbf{y}_i) : i = 1, \ldots, n\}$. We assume

$$y_{ij} = g_i(x_{ij}) + \eta_{ij},$$

where the g_is are independent identically distributed realizations of a stochastic process $G : \mathscr{S} \to \mathbb{R}$ and the η_{ij}s are independent identically distributed $N(0, \sigma_\eta^2)$ random noise.

We assume G admits a finite Karhunen–Loève decomposition

$$G(x) = \nu(x) + \sum_{k=1}^{p_2} v_k \psi_k(x),$$

where the ψ_ks are orthonormal in $L^2(\mathscr{S})$ and the v_ks are uncorrelated zero-mean random variables independent of the η_{ij}s.

The grid points x_{ij}s are assumed to be realizations of a doubly-stochastic Poisson process with intensity $\Lambda(x)$. Therefore $(\mathbf{x}_i, m_i) \mid \Lambda = \lambda_i$ is a realization of a Poisson process with intensity $\lambda_i(x)$ for each i. We assume $\log \Lambda$ also admits a finite Karhunen–Loève decomposition

$$\log \Lambda(x) = \mu(x) + \sum_{k=1}^{p_1} u_k \phi_k(x),$$

where the ϕ_ks are orthonormal in $L^2(\mathscr{S})$ and the u_ks are uncorrelated zero-mean random variables, also independent of the η_{ij}s.

Let $\mathbf{u}$ and $\mathbf{v}$ be the vectors of the u_ks and v_ks, respectively, and $\boldsymbol{\sigma}_u^2$ and $\boldsymbol{\sigma}_v^2$ their variances. We assume the joint distribution of $(\mathbf{u}, \mathbf{v})$ is $N(\mathbf{0}, \boldsymbol{\Sigma})$ with

$$\boldsymbol{\Sigma} = \begin{pmatrix} \operatorname{diag}(\boldsymbol{\sigma}_u^2) & \boldsymbol{\Sigma}_{uv} \\ \boldsymbol{\Sigma}_{uv}^T & \operatorname{diag}(\boldsymbol{\sigma}_v^2) \end{pmatrix}.$$

For the functional parameters μ, ν, ϕ_ks and ψ_ks we assume semiparametric basis function models. Given a family of basis functions $\{\gamma_1(x), \ldots, \gamma_q(x)\}$, with $\gamma_l : \mathscr{S} \to \mathbb{R}$, let $\boldsymbol{\gamma}(x)$ be the vector of the $\gamma_l(x)$s. Then we have $\mu(x) = \mathbf{c}_0^T \boldsymbol{\gamma}(x)$, $\phi_k(x) = \mathbf{c}_k^T \boldsymbol{\gamma}(x)$, $\nu(x) = \mathbf{d}_0^T \boldsymbol{\gamma}(x)$ and $\psi_k(x) = \mathbf{d}_k^T \boldsymbol{\gamma}(x)$.

Let $\mathbf{C} = (\mathbf{c}_1, \ldots, \mathbf{c}_{p_1})$ and $\mathbf{D} = (\mathbf{d}_1, \ldots, \mathbf{d}_{p_2})$; then $\boldsymbol{\phi}(x) = \mathbf{C}^T\boldsymbol{\gamma}(x)$ and $\boldsymbol{\psi}(x) = \mathbf{D}^T\boldsymbol{\gamma}(x)$, where $\boldsymbol{\phi}(x)$ is the vector of the $\phi_k(x)$s and $\boldsymbol{\psi}(x)$ is the vector of the $\psi_k(x)$s. For a vector of observations $\mathbf{x}$ we define the matrices $\boldsymbol{\Gamma}(\mathbf{x}) = [\gamma_1(\mathbf{x}), \ldots, \gamma_q(\mathbf{x})]$, $\boldsymbol{\Phi}(\mathbf{x}) = [\phi_1(\mathbf{x}), \ldots, \phi_{p_1}(\mathbf{x})]$ and $\boldsymbol{\Psi}(\mathbf{x}) = [\psi_1(\mathbf{x}), \ldots, \psi_{p_2}(\mathbf{x})]$, where function evaluation of univariate functions at $\mathbf{x}$ is understood in a componentwise way. For data vectors $\mathbf{x}_i$ we will use the shorthands $\boldsymbol{\Gamma}_i$, $\boldsymbol{\Phi}_i$ and $\boldsymbol{\Psi}_i$, respectively.

0.1.2 Joint and conditional densities

Then the joint density of the observations and the latent variables is

$$f_{\boldsymbol{\theta}}(\mathbf{x}, m, \mathbf{y}, \mathbf{u}, \mathbf{v}) = f_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{x}, m, \mathbf{u}, \mathbf{v}) f_{\boldsymbol{\theta}}(\mathbf{x}, m \mid \mathbf{u}, \mathbf{v}) f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v}).$$

Since $f_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{x}, m, \mathbf{u}, \mathbf{v})$ does not explicitly depend on $\mathbf{u}$ and $f_{\boldsymbol{\theta}}(\mathbf{x}, m \mid \mathbf{u}, \mathbf{v})$ does not explicitly depend on $\mathbf{v}$, we have

$$f_{\boldsymbol{\theta}}(\mathbf{x}, m, \mathbf{y}, \mathbf{u}, \mathbf{v}) = f_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v}) f_{\boldsymbol{\theta}}(\mathbf{x}, m \mid \mathbf{u}) f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v}),$$

where

$$f_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v}) = \frac{1}{(2\pi\sigma_\eta^2)^{m/2}} \exp\left\{-\frac{1}{2\sigma_\eta^2}\|\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\|^2\right\}, \qquad (1)$$

$$f_{\boldsymbol{\theta}}(\mathbf{x}, m \mid \mathbf{u}) = \exp\left\{-\int \lambda_{\mathbf{u}}(t)dt\right\} \frac{1}{m!} \prod_{j=1}^{m} \lambda_{\mathbf{u}}(x_j), \qquad (2)$$

where $\lambda_{\mathbf{u}}(x) = \exp\{\mu(x) + \mathbf{u}^T\boldsymbol{\phi}(x)\}$, and

$$f_{\boldsymbol{\theta}}(\mathbf{u},\mathbf{v}) = \frac{1}{(2\pi)^{(p_1+p_2)/2}(\det\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{u}^T,\mathbf{v}^T)\boldsymbol{\Sigma}^{-1}(\mathbf{u}^T,\mathbf{v}^T)^T\right\}. \quad (3)$$

Each conditional density is a function of the following model parameters:

$$\begin{aligned} f_{\boldsymbol{\theta}}(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v}) &\longrightarrow \mathbf{d}_0, \mathbf{D}, \text{ and } \sigma_\eta^2 \\ f_{\boldsymbol{\theta}}(\mathbf{x}, m \mid \mathbf{u}) &\longrightarrow \mathbf{c}_0 \text{ and } \mathbf{C} \\ f_{\boldsymbol{\theta}}(\mathbf{u},\mathbf{v}) &\longrightarrow \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_v^2 \text{ and } \boldsymbol{\Sigma}_{uv} \end{aligned}$$

0.1.3 EM algorithm

The penalized maximum likelihood estimator $\hat{\boldsymbol{\theta}}$ is the maximizer of

$$\ell(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i) - \xi_1 P(\mu) - \xi_2 \sum_{k=1}^{p_1} P(\phi_k) - \xi_3 P(\nu) - \xi_4 \sum_{k=1}^{p_2} P(\psi_k)$$

subject to the constraints $\mathbf{c}_k^T\mathbf{J}\mathbf{c}_l = \mathbf{d}_k^T\mathbf{J}\mathbf{d}_l = \delta_{kl}$, $\sigma_\eta^2 > 0$ and $\boldsymbol{\Sigma}$ symmetric positive definite. The penalty functions are quadratic on the basis coefficients: if $f(x) = \mathbf{c}^T\boldsymbol{\gamma}(x)$ then $P(f) = \mathbf{c}^T\boldsymbol{\Omega}\mathbf{c}$ for $\boldsymbol{\Omega}$ that depends only on $\boldsymbol{\gamma}$. Explicitly: if the γ_ls are univariate (temporal processes) and $P(f) = \int (f'')^2$, then

$$\boldsymbol{\Omega} = \int \boldsymbol{\gamma}''(x)\boldsymbol{\gamma}''(x)^T dx.$$

If the γ_ls are bivariate (spatial processes) and $P(f) = \iint \{(\frac{\partial^2 f}{\partial t_1^2})^2 + 2(\frac{\partial^2 f}{\partial t_1 \partial t_2})^2 + (\frac{\partial^2 f}{\partial t_2^2})^2\}$, then

$$\boldsymbol{\Omega} = \mathbf{J}_{11} + 2\mathbf{J}_{12} + \mathbf{J}_{22}$$

with

$$\mathbf{J}_{ij} = \iint \left(\frac{\partial^2 \boldsymbol{\gamma}}{\partial t_i \partial t_j}\right) \left(\frac{\partial^2 \boldsymbol{\gamma}}{\partial t_i \partial t_j}\right)^T.$$

Then

$$\ell(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i) - \xi_1 \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0 - \xi_2 \operatorname{tr}(\mathbf{C}^T \boldsymbol{\Omega} \mathbf{C}) - \xi_3 \mathbf{d}_0^T \boldsymbol{\Omega} \mathbf{d}_0 - \xi_4 \operatorname{tr}(\mathbf{D}^T \boldsymbol{\Omega} \mathbf{D}).$$

The EM algorithm (Dempster et al., 1977) works iteratively as follows: given the current value of the estimator $\hat{\boldsymbol{\theta}}_{(k-1)}$, the updated value $\hat{\boldsymbol{\theta}}_{(k)}$ is defined as the maximizer of

$$\begin{aligned} Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i, \mathbf{y}_i, \mathbf{u}, \mathbf{v}) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\} \\ &\quad - \xi_1 \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0 - \xi_2 \operatorname{tr}(\mathbf{C}^T \boldsymbol{\Omega} \mathbf{C}) - \xi_3 \mathbf{d}_0^T \boldsymbol{\Omega} \mathbf{d}_0 - \xi_4 \operatorname{tr}(\mathbf{D}^T \boldsymbol{\Omega} \mathbf{D}) \end{aligned}$$

subject to the parameter constraints. Considering the factorization of the joint density and the dependence of each factor on the model parameters, we can write

$$Q(\boldsymbol{\theta} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) = Q_1(\mathbf{d}_0, \mathbf{D}, \sigma_\eta^2 \mid \hat{\boldsymbol{\theta}}_{(k-1)}) + Q_2(\mathbf{c}_0, \mathbf{C} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) + Q_3(\boldsymbol{\Sigma} \mid \hat{\boldsymbol{\theta}}_{(k-1)}),$$

where

$$\begin{aligned} Q_1(\mathbf{d}_0, \mathbf{D}, \sigma_\eta^2 \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\log f_{\boldsymbol{\theta}}(\mathbf{y}_i \mid \mathbf{x}_i, m_i, \mathbf{v}) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\} \\ &\quad - \xi_3 \mathbf{d}_0^T \boldsymbol{\Omega} \mathbf{d}_0 - \xi_4 \operatorname{tr}(\mathbf{D}^T \boldsymbol{\Omega} \mathbf{D}), \end{aligned}$$

$$\begin{aligned} Q_2(\mathbf{c}_0, \mathbf{C} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i \mid \mathbf{u}) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\} \\ &\quad - \xi_1 \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0 - \xi_2 \operatorname{tr}(\mathbf{C}^T \boldsymbol{\Omega} \mathbf{C}), \end{aligned}$$

and

$$Q_3(\boldsymbol{\Sigma} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\log f_{\boldsymbol{\theta}}(\mathbf{u},\mathbf{v}) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\}.$$

0.1.4 Estimating equations

For σ^2_η: Since

$$\log f_{\boldsymbol{\theta}}(\mathbf{y}_i \mid \mathbf{x}_i, m_i, \mathbf{v}) = -\frac{m_i}{2}\log(2\pi\sigma^2_\eta) - \frac{1}{2\sigma^2_\eta}\|\mathbf{y}_i - \nu(\mathbf{x}_i) - \boldsymbol{\Psi}(\mathbf{x}_i)\mathbf{v}\|^2$$

we have

$$\begin{aligned}
\frac{\partial}{\partial\sigma^2_\eta} Q_1(\mathbf{d}_0, \mathbf{D}, \sigma^2_\eta \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\frac{\partial}{\partial\sigma^2_\eta}\log f_{\boldsymbol{\theta}}(\mathbf{y}_i \mid \mathbf{x}_i, m_i, \mathbf{v}) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\} \\
&= -\frac{1}{n}\sum_{i=1}^{n}\frac{m_i}{2}\frac{1}{\sigma^2_\eta} \\
&\quad +\frac{1}{n}\sum_{i=1}^{n}\frac{1}{2(\sigma^2_\eta)^2} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\|\mathbf{y}_i - \nu(\mathbf{x}_i) - \boldsymbol{\Psi}(\mathbf{x}_i)\mathbf{v}\|^2 \mid \mathbf{x}_i, m_i, \mathbf{y}_i\}
\end{aligned}$$

and then

$$\hat{\sigma}^2_\eta = \frac{1}{\sum_{i=1}^n m_i}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\|\mathbf{y}_i - \hat{\nu}(\mathbf{x}_i) - \hat{\boldsymbol{\Psi}}(\mathbf{x}_i)\mathbf{v}\|^2 \mid \mathbf{x}_i, m_i, \mathbf{y}_i\}.$$

For $\mathbf{d}_0$: Since

$$\begin{aligned}
\mathsf{D}_{\mathbf{d}_0}\log f_{\boldsymbol{\theta}}(\mathbf{y}_i \mid \mathbf{x}_i, m_i, \mathbf{v}) &= \frac{1}{\sigma^2_\eta}\{\mathbf{y}_i - \nu(\mathbf{x}_i) - \boldsymbol{\Psi}(\mathbf{x}_i)\mathbf{v}\}^T\boldsymbol{\Gamma}(\mathbf{x}_i) \\
&= \frac{1}{\sigma^2_\eta}(\mathbf{y}_i - \boldsymbol{\Psi}_i\mathbf{v})^T\boldsymbol{\Gamma}_i - \frac{1}{\sigma^2_\eta}\mathbf{d}_0^T\boldsymbol{\Gamma}_i^T\boldsymbol{\Gamma}_i
\end{aligned}$$

and

$$\begin{aligned}
\mathsf{D}_{\mathbf{d}_0} Q_1(\mathbf{d}_0, \mathbf{D}, \sigma^2_\eta \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\{\mathsf{D}_{\mathbf{d}_0}\log f_{\boldsymbol{\theta}}(\mathbf{y}_i \mid \mathbf{x}_i, m_i, \mathbf{v}) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\} \\
&\quad -\xi_3 2\mathbf{d}_0^T\boldsymbol{\Omega},
\end{aligned}$$

we have

$$\mathsf{D}_{\mathbf{d}_0} Q_1(\mathbf{d}_0, \mathbf{D}, \sigma_\eta^2 \mid \hat{\boldsymbol{\theta}}_{k-1}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_\eta^2}(\mathbf{y}_i - \boldsymbol{\Psi}_i \hat{\mathbf{v}}_i)^T \boldsymbol{\Gamma}_i - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_\eta^2}\mathbf{d}_0^T \boldsymbol{\Gamma}_i^T \boldsymbol{\Gamma}_i - 2\xi_3 \mathbf{d}_0^T \boldsymbol{\Omega},$$

where $\hat{\mathbf{v}}_i = E_{\hat{\boldsymbol{\theta}}_{(k-1)}}(\mathbf{v} \mid \mathbf{x}_i, m_i, \mathbf{y}_i)$, so

$$\hat{\mathbf{d}}_0 = \left(\frac{1}{n}\sum_{i=1}^{n} \frac{1}{\hat{\sigma}_\eta^2} \boldsymbol{\Gamma}_i^T \boldsymbol{\Gamma}_i + 2\xi_3 \boldsymbol{\Omega} \right)^{-1} \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\hat{\sigma}_\eta^2} \boldsymbol{\Gamma}_i^T (\mathbf{y}_i - \boldsymbol{\Psi}_i \hat{\mathbf{v}}_i).$$

For $\mathbf{D}$: We can write $\boldsymbol{\Psi}_i \mathbf{v} = \boldsymbol{\Gamma}_i \mathbf{D} \mathbf{v} = (\mathbf{v}^T \otimes \boldsymbol{\Gamma}_i) \operatorname{vec} \mathbf{D}$, so

$$\mathsf{D}_{\operatorname{vec} \mathbf{D}} \log f_{\boldsymbol{\theta}}(\mathbf{y}_i \mid \mathbf{x}_i, m_i, \mathbf{v}) = \frac{1}{\sigma_\eta^2} \{\mathbf{y}_i - \nu(\mathbf{x}_i) - \boldsymbol{\Psi}(\mathbf{x}_i)\mathbf{v}\}^T (\mathbf{v}^T \otimes \boldsymbol{\Gamma}_i).$$

Also, $\operatorname{tr}(\mathbf{D}^T \boldsymbol{\Omega} \mathbf{D}) = \operatorname{vec} \mathbf{D}^T (\mathbf{I}_{p_2} \otimes \boldsymbol{\Omega}) \operatorname{vec} \mathbf{D}$, so

$$\begin{aligned} \mathsf{D}_{\operatorname{vec} \mathbf{D}} Q_1(\mathbf{d}_0, \mathbf{D}, \sigma_\eta^2 \mid \hat{\boldsymbol{\theta}}_{k-1}) &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_\eta^2} \{\mathbf{y}_i - \nu(\mathbf{x}_i)\}^T (\hat{\mathbf{v}}_i^T \otimes \boldsymbol{\Gamma}_i) \\ &\quad - \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_\eta^2} \operatorname{vec} \mathbf{D}^T (\widehat{\mathbf{v}_i \mathbf{v}_i^T} \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Gamma}_i) \\ &\quad - 2\xi_4 \operatorname{vec} \mathbf{D}^T (\mathbf{I}_{p_2} \otimes \boldsymbol{\Omega}). \end{aligned}$$

Then we can write

$$\nabla_{\operatorname{vec} \mathbf{D}} Q_1(\mathbf{d}_0, \mathbf{D}, \sigma_\eta^2 \mid \hat{\boldsymbol{\theta}}_{(k-1)}) = -\mathbf{Q} \operatorname{vec} \mathbf{D} + \mathbf{b}$$

with

$$\begin{aligned} \mathbf{Q} &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_\eta^2} (\widehat{\mathbf{v}_i \mathbf{v}_i^T} \otimes \boldsymbol{\Gamma}_i^T \boldsymbol{\Gamma}_i) + 2\xi_4 (\mathbf{I}_{p_2} \otimes \boldsymbol{\Omega}), \\ \mathbf{b} &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sigma_\eta^2} (\hat{\mathbf{v}}_i \otimes \boldsymbol{\Gamma}_i^T) \{\mathbf{y}_i - \nu(\mathbf{x}_i)\}, \end{aligned}$$

where $\widehat{\mathbf{v}_i\mathbf{v}_i^T} = E_{\hat{\boldsymbol{\theta}}_{(k-1)}}(\mathbf{v}\mathbf{v}^T \mid \mathbf{x}_i, m_i, \mathbf{y}_i)$. The orthogonality constraints can be expressed as $\mathbf{h}^D(\boldsymbol{\theta}) = \mathbf{0}$, where $\mathbf{h}^D(\boldsymbol{\theta})$ is a $p_2(p_2+1)/2$-dimensional vector with elements $h^D_{kl}(\boldsymbol{\theta}) = \mathbf{d}_k^T\mathbf{J}\mathbf{d}_l - \delta_{kl}$. We can write $\mathbf{d}_k^T\mathbf{J}\mathbf{d}_l = \mathbf{e}_k^T\mathbf{D}^T\mathbf{J}\mathbf{D}\mathbf{e}_l = \operatorname{tr}(\mathbf{D}^T\mathbf{J}\mathbf{D}\mathbf{e}_l\mathbf{e}_k^T)$, with $\mathbf{e}_k$ the kth canonical vector, so

$$h^D_{kl}(\boldsymbol{\theta}) = \operatorname{vec}\mathbf{D}^T(\mathbf{e}_k\mathbf{e}_l^T \otimes \mathbf{J})\operatorname{vec}\mathbf{D} - \delta_{kl}.$$

We can linearize the constraints by using the current value of $\mathbf{D}$ on the left, so

$$\mathbf{h}^D(\boldsymbol{\theta}) = \mathbf{A}\operatorname{vec}\mathbf{D} - \mathbf{f},$$

where $\mathbf{A}$ is the $p_2(p_2+1)/2 \times qp_2$ matrix with rows $\operatorname{vec}\hat{\mathbf{D}}^T_{(k-1)}(\mathbf{e}_k\mathbf{e}_l^T \otimes \mathbf{J})$ and $\mathbf{f}$ are the corresponding δ_{kl}s. The Lagrange condition for $\operatorname{vec}\hat{\mathbf{D}}_{(k)}$ to be a local maximizer under the constraints $\mathbf{h}^D(\boldsymbol{\theta}) = \mathbf{0}$ is then

$$\begin{aligned}\mathsf{D}_{\operatorname{vec}\mathbf{D}}Q_1(\mathbf{d}_0, \mathbf{D}, \sigma^2_\eta \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= \boldsymbol{\kappa}^T\mathsf{D}_{\operatorname{vec}\mathbf{D}}\mathbf{h}^D(\boldsymbol{\theta}) \\ &= \boldsymbol{\kappa}^T\mathbf{A},\end{aligned}$$

where $\boldsymbol{\kappa}$ is the $p_2(p_2+1)/2$-dimensional vector of Lagrange multipliers. Transposing both sides on the last equation we get

$$-\mathbf{Q}\operatorname{vec}\hat{\mathbf{D}}_{(k)} + \mathbf{b} = \mathbf{A}^T\boldsymbol{\kappa},$$

which together with the constraints can be written as a system

$$\begin{pmatrix} \mathbf{Q} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \operatorname{vec} \hat{\mathbf{D}}_{(k)} \\ \boldsymbol{\kappa} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{f} \end{pmatrix}.$$

Solving this linear system gives the updated $\operatorname{vec} \hat{\mathbf{D}}_{(k)}$.

For $\mathbf{c}_0$: Since

$$\begin{aligned} \log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i \mid \mathbf{u}) &= -\int \lambda_{\mathbf{u}}(t)dt - \log m_i! + \sum_{j=1}^{m_i} \log \lambda_{\mathbf{u}}(x_{ij}) \\ &= -\int \lambda_{\mathbf{u}}(t)dt - \log m_i! + \sum_{j=1}^{m_i} \{\boldsymbol{\gamma}(x_{ij})^T \mathbf{c}_0 + \mathbf{u}^T \boldsymbol{\phi}(x_{ij})\} \end{aligned}$$

we have

$$\mathsf{D}_{\mathbf{c}_0} \log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i \mid \mathbf{u}) = -\int \mathsf{D}_{\mathbf{c}_0} \lambda_{\mathbf{u}}(t)dt + \sum_{j=1}^{m_i} \boldsymbol{\gamma}(x_{ij})^T,$$

where

$$\begin{aligned} \mathsf{D}_{\mathbf{c}_0} \lambda_{\mathbf{u}}(t) &= \lambda_{\mathbf{u}}(t) \mathsf{D}_{\mathbf{c}_0} \log \lambda_{\mathbf{u}}(t) \\ &= \lambda_{\mathbf{u}}(t) \boldsymbol{\gamma}(t)^T. \end{aligned}$$

Then

$$\begin{aligned} \mathsf{D}_{\mathbf{c}_0} Q_2(\mathbf{c}_0, \mathbf{C} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) &= -\frac{1}{n} \sum_{i=1}^{n} \int \hat{\lambda}_{i,\mathbf{c}_0,\mathbf{C}}(t) \boldsymbol{\gamma}(t)^T dt + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{m_i} \boldsymbol{\gamma}(x_{ij})^T \\ &\quad -2\xi_1 \mathbf{c}_0^T \boldsymbol{\Omega}, \end{aligned}$$

where

$$\begin{aligned} \hat{\lambda}_{i,\mathbf{c}_0,\mathbf{C}}(t) &= E_{\hat{\boldsymbol{\theta}}_{(k-1)}} \{\lambda_{\mathbf{u}}(t) \mid \mathbf{x}_i, m_i, \mathbf{y}_i\} \\ &= E_{\hat{\boldsymbol{\theta}}_{(k-1)}} [\exp\{\boldsymbol{\gamma}(t)^T \mathbf{c}_0 + \boldsymbol{\gamma}(t)^T \mathbf{C}\mathbf{u}\} \mid \mathbf{x}_i, m_i, \mathbf{y}_i]. \end{aligned}$$

A Taylor expansion of $\lambda_{\mathbf{u}}(t)$ on the variable $\mathbf{c}_0$ at the current $\hat{\mathbf{c}}_{0(k-1)}$ gives

$$\begin{aligned}
\lambda_{\mathbf{u},\mathbf{c}_0,\mathbf{C}}(t) &\approx \lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\mathbf{C}}(t) + \mathsf{D}_{\mathbf{c}_0}\lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\mathbf{C}}(t)(\mathbf{c}_0 - \hat{\mathbf{c}}_{0(k-1)}) \\
&= \lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\mathbf{C}}(t) + \lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\mathbf{C}}(t)\boldsymbol{\gamma}(t)^T(\mathbf{c}_0 - \hat{\mathbf{c}}_{0(k-1)}) \\
&= \lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\mathbf{C}}(t)\{1 - \boldsymbol{\gamma}(t)^T\hat{\mathbf{c}}_{0(k-1)}\} + \lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\mathbf{C}}(t)\boldsymbol{\gamma}(t)^T\mathbf{c}_0,
\end{aligned}$$

so

$$\begin{aligned}
\mathsf{D}_{\mathbf{c}_0}Q_2(\hat{\mathbf{c}}_{0(k)}, \hat{\mathbf{C}}_{(k-1)} \mid \hat{\boldsymbol{\theta}}_{k-1}) \approx\ & -\frac{1}{n}\sum_{i=1}^{n}\int \hat{\lambda}_i(t)\{1 - \hat{\mu}_{(k-1)}(t)\}\boldsymbol{\gamma}(t)^T dt \\
& -\frac{1}{n}\sum_{i=1}^{n}\int \hat{\lambda}_i(t)\hat{\mathbf{c}}_{0(k)}^T\boldsymbol{\gamma}(t)\boldsymbol{\gamma}(t)^T dt \\
& +\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\boldsymbol{\gamma}(x_{ij})^T - 2\xi_1\hat{\mathbf{c}}_{0(k)}^T\boldsymbol{\Omega},
\end{aligned}$$

where $\hat{\lambda}_i(t) = \hat{\lambda}_{i,\hat{\mathbf{c}}_{0(k-1)},\hat{\mathbf{C}}_{(k-1)}}(t)$. Equating to zero and solving, we get

$$\begin{aligned}
\hat{\mathbf{c}}_0^{(k)} =\ & \left\{\frac{1}{n}\sum_{i=1}^{n}\int \hat{\lambda}_i(t)\boldsymbol{\gamma}(t)\boldsymbol{\gamma}(t)^T dt + 2\xi_1\boldsymbol{\Omega}\right\}^{-1} \times \\
& \left[-\frac{1}{n}\sum_{i=1}^{n}\int \hat{\lambda}_i(t)\{1 - \hat{\mu}(t)\}\boldsymbol{\gamma}(t)dt + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\boldsymbol{\gamma}(x_{ij})\right]
\end{aligned}$$

For $\mathbf{C}$: We can write

$$\begin{aligned}
\log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i \mid \mathbf{u}) &= -\int \lambda_{\mathbf{u}}(t)dt - \log m_i! + \sum_{j=1}^{m_i}\{\mu(x_{ij}) + \boldsymbol{\gamma}(x_{ij})^T\mathbf{C}\mathbf{u}\} \\
&= -\int \lambda_{\mathbf{u}}(t)dt - \log m_i! + \sum_{j=1}^{m_i}[\mu(x_{ij}) + \{\mathbf{u}^T \otimes \boldsymbol{\gamma}(x_{ij})^T\}\operatorname{vec}\mathbf{C}],
\end{aligned}$$

so

$$\mathsf{D}_{\operatorname{vec}\mathbf{C}}\log f_{\boldsymbol{\theta}}(\mathbf{x}_i, m_i \mid \mathbf{u}) = -\int \mathsf{D}_{\operatorname{vec}\mathbf{C}}\lambda_{\mathbf{u}}(t)dt + \sum_{j=1}^{m_i}\{\mathbf{u}^T \otimes \boldsymbol{\gamma}(x_{ij})^T\}.$$

As before,

$$\begin{aligned}\mathsf{D}_{\mathrm{vec}\,\mathbf{C}}\lambda_{\mathbf{u}}(t) &= \lambda_{\mathbf{u}}(t)\mathsf{D}_{\mathrm{vec}\,\mathbf{C}}\log\lambda_{\mathbf{u}}(t)\\ &= \lambda_{\mathbf{u}}(t)\{\mathbf{u}^T\otimes\boldsymbol{\gamma}(t)^T\},\end{aligned}$$

so

$$\begin{aligned}\mathsf{D}_{\mathrm{vec}\,\mathbf{C}}Q_2(\mathbf{c}_0,\mathbf{C}\mid\hat{\boldsymbol{\theta}}_{(k-1)}) &= -\frac{1}{n}\sum_{i=1}^n\int E_{\hat{\boldsymbol{\theta}}_{(k-1)}}[\lambda_{\mathbf{u}}(t)\{\mathbf{u}^T\otimes\boldsymbol{\gamma}(t)^T\}\mid\mathbf{x}_i,m_i,\mathbf{y}_i]dt\\ &\quad+\frac{1}{n}\sum_{i=1}^n\{\hat{\mathbf{u}}_i^T\otimes\sum_{j=1}^{m_i}\boldsymbol{\gamma}(x_{ij})^T\}-2\xi_2\,\mathrm{vec}\,\mathbf{C}^T(\mathbf{I}_{p_1}\otimes\boldsymbol{\Omega}),\end{aligned}$$

where $\hat{\mathbf{u}}_i = E_{\hat{\boldsymbol{\theta}}_{(k-1)}}(\mathbf{u}\mid\mathbf{x}_i,m_i,\mathbf{y}_i)$. Expanding $\lambda_{\mathbf{u}}(t)$ as before, but on the variable $\mathrm{vec}\,\mathbf{C}$, we get

$$\begin{aligned}\lambda_{\mathbf{u},\mathbf{c}_0,\mathbf{C}}(t) &\approx \lambda_{\mathbf{u},\mathbf{c}_0,\hat{\mathbf{C}}_{(k-1)}}(t)+\lambda_{\mathbf{u},\mathbf{c}_0,\hat{\mathbf{C}}_{(k-1)}}(t)\{\mathbf{u}^T\otimes\boldsymbol{\gamma}(t)^T\}(\mathrm{vec}\,\mathbf{C}-\mathrm{vec}\,\hat{\mathbf{C}}_{(k-1)})\\ &= \lambda_{\mathbf{u},\mathbf{c}_0,\hat{\mathbf{C}}_{(k-1)}}(t)\{1-\mathbf{u}^T\hat{\boldsymbol{\phi}}_{(k-1)}(t)\}+\lambda^{(k-1)}_{\mathbf{u},\mathbf{c}_0,\hat{\mathbf{C}}_{k-1}}(t)\{\mathbf{u}^T\otimes\boldsymbol{\gamma}(t)^T\}\,\mathrm{vec}\,\mathbf{C}.\end{aligned}$$

Then

$$\mathsf{D}_{\mathrm{vec}\,\mathbf{C}}Q_2(\hat{\mathbf{c}}_{0(k-1)},\hat{\mathbf{C}}_{(k)}\mid\hat{\boldsymbol{\theta}}_{(k-1)})\approx$$

$$\begin{aligned}&-\frac{1}{n}\sum_{i=1}^n\int E[\hat{\lambda}_{\mathbf{u}(k-1)}(t)\{1-\mathbf{u}^T\hat{\boldsymbol{\phi}}_{(k-1)}(t)\}\{\mathbf{u}^T\otimes\boldsymbol{\gamma}(t)^T\}\mid\mathbf{x}_i,m_i,\mathbf{y}_i]dt\\ &-\mathrm{vec}\,\hat{\mathbf{C}}^T_{(k)}\frac{1}{n}\sum_{i=1}^n\int E[\hat{\lambda}_{\mathbf{u}(k-1)}(t)\{\mathbf{u}\mathbf{u}^T\otimes\boldsymbol{\gamma}(t)\boldsymbol{\gamma}(t)^T\}\mid\mathbf{x}_i,m_i,\mathbf{y}_i]dt\\ &+\frac{1}{n}\sum_{i=1}^n\{\hat{\mathbf{u}}_i^T\otimes\sum_{j=1}^{m_i}\boldsymbol{\gamma}(x_{ij})^T\}-2\xi_2\,\mathrm{vec}\,\hat{\mathbf{C}}^T_{(k)}(\mathbf{I}_{p_1}\otimes\boldsymbol{\Omega}),\end{aligned}$$

where $\hat{\lambda}_{\mathbf{u}(k-1)}(t)=\lambda_{\mathbf{u},\hat{\mathbf{c}}_{0(k-1)},\hat{\mathbf{C}}_{(k-1)}}(t)$. As we did above for $\mathbf{D}$, this can be

expressed in linear form as

$$\mathsf{D}_{\operatorname{vec}\mathbf{C}} Q_2(\hat{\mathbf{c}}_{0(k-1)}, \hat{\mathbf{C}}_{(k)} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) \approx -\operatorname{vec}\hat{\mathbf{C}}_{(k)}^T \mathbf{Q} + \mathbf{b}^T$$

with

$$\mathbf{Q} = \frac{1}{n}\sum_{i=1}^{n} \int E[\hat{\lambda}_{\mathbf{u}(k-1)}(t)\{\mathbf{u}\mathbf{u}^T \otimes \boldsymbol{\gamma}(t)\boldsymbol{\gamma}(t)^T\} \mid \mathbf{x}_i, m_i, \mathbf{y}_i]dt + 2\xi_2(\mathbf{I}_{p_1} \otimes \boldsymbol{\Omega})$$

and

$$\begin{aligned}\mathbf{b} &= -\frac{1}{n}\sum_{i=1}^{n} \int E[\hat{\lambda}_{\mathbf{u}(k-1)}(t)\{1 - \mathbf{u}^T\hat{\boldsymbol{\phi}}_{(k-1)}(t)\}\{\mathbf{u} \otimes \boldsymbol{\gamma}(t)\} \mid \mathbf{x}_i, m_i, \mathbf{y}_i]dt \\ &\quad + \frac{1}{n}\sum_{i=1}^{n}\{\hat{\mathbf{u}}_i \otimes \sum_{j=1}^{m_i} \boldsymbol{\gamma}(x_{ij})\}.\end{aligned}$$

The orthogonality constraints are handled as before: $\mathbf{h}^C(\boldsymbol{\theta}) = \mathbf{A}\operatorname{vec}\mathbf{C} - \mathbf{f}$ with $\mathbf{A}$ the $p_1(p_1+1)/2 \times qp_1$ matrix with rows $\operatorname{vec}\hat{\mathbf{C}}_{(k-1)}^T(\mathbf{e}_k\mathbf{e}_l^T \otimes \mathbf{J})$ and $\mathbf{f}$ the corresponding δ_{kl}s. Then $\operatorname{vec}\hat{\mathbf{C}}_{(k)}$ is obtained by solving the system

$$\begin{pmatrix} \mathbf{Q} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \operatorname{vec}\hat{\mathbf{C}}_{(k)} \\ \boldsymbol{\kappa} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{f} \end{pmatrix}.$$

For $\boldsymbol{\Sigma}$: Let $\mathbf{w} = (\mathbf{u}^T, \mathbf{v}^T)^T$. Then

$$\begin{aligned}\log f_{\boldsymbol{\theta}}(\mathbf{u}, \mathbf{v}) &\propto -\frac{1}{2}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\mathbf{w} \\ &= -\frac{1}{2}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{w}\mathbf{w}^T),\end{aligned}$$

so

$$Q_3(\boldsymbol{\Sigma} \mid \hat{\boldsymbol{\theta}}_{(k-1)}) \propto -\frac{1}{2}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathbf{S})$$

where

$$\mathbf{S} = \frac{1}{n}\sum_{i=1}^{n} E_{\hat{\boldsymbol{\theta}}_{(k-1)}}\left(\mathbf{w}\mathbf{w}^T \mid \mathbf{x}_i, m_i, \mathbf{y}_i\right).$$

This $Q_3(\boldsymbol{\Sigma} \mid \hat{\boldsymbol{\theta}}_{(k-1)})$ is the classical log-likelihood function of a multivariate normal density, and it is well-known that the (unconstrained) maximizer is $\mathbf{S}$. However, $\mathbf{S}$ must be rotated to satisfy the constraints that the u_ks and the v_ks be uncorrelated, while maintaining the positive-definitness of the whole $\hat{\boldsymbol{\Sigma}}$. To this end we compute the spectral decompositions of the blocks $\mathbf{S}_{uu}$ and $\mathbf{S}_{vv}$,

$$\begin{aligned}\mathbf{U}_1\mathbf{L}_1\mathbf{U}_1^T &= \mathbf{S}_{uu},\\ \mathbf{U}_2\mathbf{L}_2\mathbf{U}_2^T &= \mathbf{S}_{vv},\end{aligned}$$

with the $\mathbf{U}$s orthogonal and the $\mathbf{L}$s diagonal, and let

$$\hat{\boldsymbol{\Sigma}}_{(k)} = \begin{pmatrix} \mathbf{U}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2^T \end{pmatrix} \mathbf{S} \begin{pmatrix} \mathbf{U}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{U}_2 \end{pmatrix}.$$

Then the blocks $\hat{\boldsymbol{\Sigma}}_{(k),uu}$ and $\hat{\boldsymbol{\Sigma}}_{(k),vv}$ are diagonal and equal to $\mathbf{L}_1$ and $\mathbf{L}_2$, respectively. Then $\hat{\boldsymbol{\sigma}}^2_{u(k)} = \operatorname{diag}\hat{\boldsymbol{\Sigma}}_{(k),uu}$, $\hat{\boldsymbol{\sigma}}^2_{v(k)} = \operatorname{diag}\hat{\boldsymbol{\Sigma}}_{(k),vv}$ and $\hat{\boldsymbol{\Sigma}}_{uv(k)} = \hat{\boldsymbol{\Sigma}}_{(k),uv}$. The respective component scores and coefficients must be rotated too, in order that $\hat{\boldsymbol{\Sigma}}_{(k)}$ be the covariance matrix of the $\hat{\mathbf{w}}_i$s and that the

values $\hat{\boldsymbol{\phi}}(x)^T\hat{\mathbf{u}}$ and $\hat{\boldsymbol{\psi}}(x)^T\hat{\mathbf{v}}$ be preserved:

$$\begin{aligned}
\hat{\mathbf{u}}_i &\longleftarrow \mathbf{U}_1^T\hat{\mathbf{u}}_i, \\
\hat{\mathbf{v}}_i &\longleftarrow \mathbf{U}_2^T\hat{\mathbf{v}}_i, \\
\hat{\mathbf{C}} &\longleftarrow \hat{\mathbf{C}}\mathbf{U}_1, \\
\hat{\mathbf{D}} &\longleftarrow \hat{\mathbf{D}}\mathbf{U}_2.
\end{aligned}$$

0.1.5 Algorithm initialization

As initial estimators for the EM algorithm we use the multiplicative component model of Gervini (2017) for the point process X, which gives us initial $\hat{\mathbf{c}}_0$, $\hat{\mathbf{C}}$, $\hat{\boldsymbol{\sigma}}_u^2$ and $\hat{\mathbf{u}}_i$s, and the reduced-rank principal component model of James et al. (2000) for the process Y, which gives us initial $\hat{\mathbf{d}}_0$, $\hat{\mathbf{D}}$, $\hat{\boldsymbol{\sigma}}_v^2$, $\hat{\sigma}_\eta^2$ and $\hat{\mathbf{v}}_i$s. As initial $\hat{\boldsymbol{\Sigma}}_{uv}$ we then use the cross-covariance matrix of the $\hat{\mathbf{u}}_i$s and the $\hat{\mathbf{v}}_i$s.

0.1.6 Laplace approximation of integrals

The marginal densities $f(\mathbf{x}, m, \mathbf{y})$ are computed by Laplace approximation. We have

$$\begin{aligned}
f(\mathbf{x}, m, \mathbf{y}) &= \iint f(\mathbf{x}, m, \mathbf{y} \mid \mathbf{w}) f(\mathbf{w}) d\mathbf{w} \\
&= \iint \exp g(\mathbf{w}) d\mathbf{w}
\end{aligned}$$

with

$$g(\mathbf{w}) = \log f(\mathbf{x}, m, \mathbf{y} \mid \mathbf{w}) + \log f(\mathbf{w}).$$

If $\hat{\mathbf{w}} = \arg\max g(\mathbf{w})$ then $g(\mathbf{w}) \approx g(\hat{\mathbf{w}}) + .5(\mathbf{w} - \hat{\mathbf{w}})^T \mathsf{H} g(\hat{\mathbf{w}})(\mathbf{w} - \hat{\mathbf{w}})$ and

$$f(\mathbf{x}, m, \mathbf{y}) \approx \exp\{g(\hat{\mathbf{w}})\}(2\pi)^{p/2} \det(\mathbf{S})^{1/2}$$

with $p = p_1 + p_2$ and

$$\mathbf{S} = \{-\mathsf{H} g(\hat{\mathbf{w}})\}^{-1}.$$

In effect, we are approximating

$$f(\mathbf{x}, m, \mathbf{y} \mid \mathbf{w}) f(\mathbf{w}) \approx \exp\{g(\hat{\mathbf{w}})\}(2\pi)^{p/2} \det(\mathbf{S})^{1/2} \varphi_{(\hat{\mathbf{w}}, \mathbf{S})}(\mathbf{w})$$

where $\varphi_{(\hat{\mathbf{w}}, \mathbf{S})}(\mathbf{w})$ denotes the pdf of a $N_p(\hat{\mathbf{w}}, \mathbf{S})$, so $\mathbf{W} \mid (\mathbf{x}, m, \mathbf{y}) \approx N_p(\hat{\mathbf{w}}, \mathbf{S})$. Then we can also approximate the moments:

$$\begin{aligned} E(\mathbf{W} \mid \mathbf{x}, m, \mathbf{y}) &\approx \hat{\mathbf{w}}, \\ E(\mathbf{W}\mathbf{W}^T \mid \mathbf{x}, m, \mathbf{y}) &\approx \mathbf{S} + \hat{\mathbf{w}}\hat{\mathbf{w}}^T. \end{aligned}$$

We find $\hat{\mathbf{w}}$ by (a few steps of) Newton–Raphson for each $(\mathbf{x}_i, m_i, \mathbf{y}_i)$. Since

$$\begin{aligned} g(\mathbf{w}) = & -\int \lambda_{\mathbf{u}}(t)dt + \sum_{j=1}^{m} \log \lambda_{\mathbf{u}}(x_j) - \log m! \\ & -\frac{m}{2} \log 2\pi\sigma_\eta^2 - \frac{1}{2\sigma_\eta^2} \|\mathbf{y} - \nu(\mathbf{x}) - \mathbf{\Psi}(\mathbf{x})\mathbf{v}\|^2 \\ & -\frac{p}{2} \log 2\pi - \frac{1}{2} \log\det \mathbf{\Sigma} - \frac{1}{2} \mathbf{w}^T \mathbf{\Sigma}^{-1} \mathbf{w} \end{aligned}$$

the derivatives with respect to $\mathbf{w} = (\mathbf{u}, \mathbf{v})$ are

$$\nabla g(\mathbf{w}) = \begin{bmatrix} -\int \lambda_{\mathbf{u}}(t)\boldsymbol{\phi}(t)dt + \sum_{j=1}^{m} \boldsymbol{\phi}(x_j) \\ \frac{1}{\sigma_\eta^2}\boldsymbol{\Psi}(\mathbf{x})^T\{\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\} \end{bmatrix} - \boldsymbol{\Sigma}^{-1}\mathbf{w}$$

and

$$\mathsf{H}g(\mathbf{w}) = \begin{bmatrix} -\int \lambda_{\mathbf{u}}(t)\boldsymbol{\phi}(t)\boldsymbol{\phi}(t)^T dt & \mathbf{O} \\ \mathbf{O} & -\frac{1}{\sigma_\eta^2}\boldsymbol{\Psi}(\mathbf{x})^T\boldsymbol{\Psi}(\mathbf{x}) \end{bmatrix} - \boldsymbol{\Sigma}^{-1}.$$

0.2 Asymptotics

0.2.1 Explicit Fisher's information matrix

Fisher's information matrix $\mathbf{F}_0 = E_{\boldsymbol{\theta}_0}\{\nabla \log f(\mathbf{x}, \mathbf{y}, m; \boldsymbol{\theta}_0)\nabla \log f(\mathbf{x}, \mathbf{y}, m; \boldsymbol{\theta}_0)^T\}$,

used in the asymptotic results below, is estimated by

$$\hat{\mathbf{F}} = \frac{1}{n}\sum_{i=1}^{n} \nabla \log f(\mathbf{x}_i, m_i, \mathbf{y}_i; \hat{\boldsymbol{\theta}})\nabla \log f(\mathbf{x}_i, m_i, \mathbf{y}_i; \hat{\boldsymbol{\theta}})^T.$$

Here we derive $\nabla \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta})$ by blocks of $\boldsymbol{\theta} = (\operatorname{vec} \boldsymbol{\Sigma}_{uv}, \mathbf{c}_0, \operatorname{vec} \mathbf{C}, \mathbf{d}_0, \operatorname{vec} \mathbf{D}, \boldsymbol{\sigma}_u^2, \boldsymbol{\sigma}_v^2, \sigma_\eta^2)$.

▶ For $\operatorname{vec} \boldsymbol{\Sigma}_{uv}$, since only $f(\mathbf{w})$ depends on $\boldsymbol{\Sigma}_{uv}$, we have

$$\begin{aligned} & \nabla_{\operatorname{vec} \boldsymbol{\Sigma}_{uv}} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) \\ = & \frac{1}{f(\mathbf{x}, m, \mathbf{y})} \iint f(\mathbf{x}, m, \mathbf{y} \mid \mathbf{w}) \nabla_{\operatorname{vec} \boldsymbol{\Sigma}_{uv}} f(\mathbf{w}) \, d\mathbf{w} \\ = & \iint \frac{\nabla_{\operatorname{vec} \boldsymbol{\Sigma}_{uv}} f(\mathbf{w})}{f(\mathbf{w})} \, \frac{f(\mathbf{x}, m, \mathbf{y}, \mathbf{w})}{f(\mathbf{x}, m, \mathbf{y})} \, d\mathbf{w} \\ = & \iint \nabla_{\operatorname{vec} \boldsymbol{\Sigma}_{uv}} \log f(\mathbf{w}) \; f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y}) \; d\mathbf{w}. \end{aligned}$$

Since

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

the differential with respect to $\boldsymbol{\Sigma}$ is

$$\mathsf{d}\log f(\mathbf{w}) = -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathsf{d}\boldsymbol{\Sigma}) + \frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}(\mathsf{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{w}.$$

Now, differentiating with respect to $\boldsymbol{\Sigma}_{uv}$,

$$\mathsf{d}\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{O} & \mathsf{d}\boldsymbol{\Sigma}_{uv} \\ \mathsf{d}\boldsymbol{\Sigma}_{uv}^T & \mathbf{O} \end{pmatrix}.$$

Then if we split $\boldsymbol{\Sigma}^{-1}$ into four blocks $\boldsymbol{\Sigma}_{11}^{-1}$, $\boldsymbol{\Sigma}_{12}^{-1}$, $\boldsymbol{\Sigma}_{21}^{-1}$ and $\boldsymbol{\Sigma}_{22}^{-1}$ commensurate with the four blocks of $\boldsymbol{\Sigma}$, and the vector $\boldsymbol{\Sigma}^{-1}\mathbf{w}$ into the first p_1 coordinates $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1$ and the last p_2 coordinates $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2$, we have

$$\boldsymbol{\Sigma}^{-1}\mathsf{d}\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{12}^{-1}\mathsf{d}\boldsymbol{\Sigma}_{uv}^T & \boldsymbol{\Sigma}_{11}^{-1}\mathsf{d}\boldsymbol{\Sigma}_{uv} \\ \boldsymbol{\Sigma}_{22}^{-1}\mathsf{d}\boldsymbol{\Sigma}_{uv}^T & \boldsymbol{\Sigma}_{21}^{-1}\mathsf{d}\boldsymbol{\Sigma}_{uv} \end{pmatrix}$$

and then

$$\begin{aligned} \operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathsf{d}\boldsymbol{\Sigma}) &= \operatorname{tr}(\boldsymbol{\Sigma}_{12}^{-1}\mathsf{d}\boldsymbol{\Sigma}_{uv}^T) + \operatorname{tr}(\boldsymbol{\Sigma}_{21}^{-1}\mathsf{d}\boldsymbol{\Sigma}_{uv}) \\ &= 2\operatorname{tr}(\mathsf{d}\boldsymbol{\Sigma}_{uv}^T\boldsymbol{\Sigma}_{12}^{-1}) \\ &= 2\operatorname{vec}(\mathsf{d}\boldsymbol{\Sigma}_{uv})^T\operatorname{vec}(\boldsymbol{\Sigma}_{12}^{-1}) \end{aligned}$$

and

$$\begin{aligned}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}(\mathsf{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{w} &= 2(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T(\mathsf{d}\boldsymbol{\Sigma}_{uv}^T)(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1 \\ &= 2\,\mathrm{tr}\{\mathsf{d}\boldsymbol{\Sigma}_{uv}^T(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T\} \\ &= 2\,\mathrm{vec}(\mathsf{d}\boldsymbol{\Sigma}_{uv})^T\,\mathrm{vec}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T\}.\end{aligned}$$

Then

$$\mathsf{d}\log f(\mathbf{w}) = -\,\mathrm{vec}(\mathsf{d}\boldsymbol{\Sigma}_{uv})^T\,\mathrm{vec}(\boldsymbol{\Sigma}_{12}^{-1}) + \mathrm{vec}(\mathsf{d}\boldsymbol{\Sigma}_{uv})^T\,\mathrm{vec}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T\},$$

which implies

$$\nabla_{\mathrm{vec}\,\boldsymbol{\Sigma}_{uv}} f(\mathbf{w}) = -\,\mathrm{vec}(\boldsymbol{\Sigma}_{12}^{-1}) + \mathrm{vec}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T\}$$

and then

$$\boxed{\nabla_{\mathrm{vec}\,\boldsymbol{\Sigma}_{uv}} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = -\,\mathrm{vec}(\boldsymbol{\Sigma}_{12}^{-1}) + \mathrm{vec}\,\mathbb{E}_{\boldsymbol{\theta}}\left\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T \mid \mathbf{x}, m, \mathbf{y}\right\}.}$$

The second term can be written more explicitly in terms of $\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{w}\mathbf{w}^T \mid \mathbf{x}, m, \mathbf{y})$: since $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1 = [\mathbf{I}_{p_1}, \mathbf{O}]\,\boldsymbol{\Sigma}^{-1}\mathbf{w}$ and $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2 = [\mathbf{O}, \mathbf{I}_{p_2}]\,\boldsymbol{\Sigma}^{-1}\mathbf{w}$, we have

$$\begin{aligned}(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T &= [\mathbf{I}_{p_1}, \mathbf{O}]\,\boldsymbol{\Sigma}^{-1}\mathbf{w}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{O}\\ \\ \mathbf{I}_{p_2}\end{bmatrix} \\ &= \left[\boldsymbol{\Sigma}_{11}^{-1}, \boldsymbol{\Sigma}_{12}^{-1}\right]\mathbf{w}\mathbf{w}^T\begin{bmatrix}\boldsymbol{\Sigma}_{12}^{-1}\\ \\ \boldsymbol{\Sigma}_{22}^{-1}\end{bmatrix}\end{aligned}$$

and then we take $\mathbb{E}_{\boldsymbol{\theta}}$.

▶ For $\mathbf{c}_0$, since only $f(\mathbf{x}, m \mid \mathbf{w})$ depends on $\mathbf{c}_0$, we have

$$\nabla_{\mathbf{c}_0} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = \iint \nabla_{\mathbf{c}_0} \log f(\mathbf{x}, m \mid \mathbf{w})\ f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y})\ d\mathbf{w}.$$

Here

$$\log f(\mathbf{x}, m \mid \mathbf{w}) = -\int \lambda_{\mathbf{u}}(t)dt + \sum_{j=1}^{m} \log \lambda_{\mathbf{u}}(x_j) - \log m!$$

with $\lambda_{\mathbf{u}}(t) = \exp\{\mathbf{c}_0^T \boldsymbol{\gamma}(t) + \mathbf{u}^T \boldsymbol{\phi}(t)\}$, so

$$\nabla_{\mathbf{c}_0} \log f(\mathbf{x}, m \mid \mathbf{w}) = -\int \lambda_{\mathbf{u}}(t)\boldsymbol{\gamma}(t)dt + \sum_{j=1}^{m} \boldsymbol{\gamma}(x_j)$$

and then

$$\boxed{\nabla_{\mathbf{c}_0} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = -\int \mathbb{E}_{\boldsymbol{\theta}}\{\lambda_{\mathbf{u}}(t) \mid \mathbf{x}, m, \mathbf{y}\}\boldsymbol{\gamma}(t)dt + \sum_{j=1}^{m} \boldsymbol{\gamma}(x_j).}$$

▶ For $\operatorname{vec} \mathbf{C}$, again only $f(\mathbf{x}, m \mid \mathbf{w})$ depends on $\operatorname{vec} \mathbf{C}$, so

$$\nabla_{\operatorname{vec} \mathbf{C}} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = \iint \nabla_{\operatorname{vec} \mathbf{C}} \log f(\mathbf{x}, m \mid \mathbf{w})\ f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y})\ d\mathbf{w}$$

as above. Since $\lambda_{\mathbf{u}}(t) = \exp\{\mu(t) + \boldsymbol{\gamma}(t)^T \mathbf{C}\mathbf{u}\} = \exp[\mu(t) + \{\mathbf{u}^T \otimes \boldsymbol{\gamma}(t)^T\} \operatorname{vec} \mathbf{C}]$,

we have

$$\nabla_{\operatorname{vec} \mathbf{C}} \log f(\mathbf{x}, m \mid \mathbf{w}) = -\int \lambda_{\mathbf{u}}(t)\{\mathbf{u} \otimes \boldsymbol{\gamma}(t)\}dt + \sum_{j=1}^{m} \{\mathbf{u} \otimes \boldsymbol{\gamma}(x_j)\}$$

and then

$$\boxed{\nabla_{\operatorname{vec} \mathbf{C}} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = -\int [\mathbb{E}_{\boldsymbol{\theta}}\{\lambda_{\mathbf{u}}(t)\mathbf{u} \mid \mathbf{x}, m, \mathbf{y}\} \otimes \boldsymbol{\gamma}(t)]dt + \sum_{j=1}^{m} \{\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{u} \mid \mathbf{x}, m, \mathbf{y}) \otimes \boldsymbol{\gamma}(x_j)\}.}$$

▶ For $\mathbf{d}_0$, since only $f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v})$ depends on this parameter, we have

$$\nabla_{\mathbf{d}_0} \log f(\mathbf{x}, \mathbf{y}, m; \boldsymbol{\theta}) = \iint \nabla_{\mathbf{d}_0} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) \; f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y}) \; d\mathbf{w}.$$

Since

$$\log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) \propto -\frac{1}{2\sigma_\eta^2} \|\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{x})\mathbf{d}_0 - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\|^2,$$

we have

$$\nabla_{\mathbf{d}_0} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) = \frac{1}{\sigma_\eta^2} \boldsymbol{\Gamma}(\mathbf{x})^T \{\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{x})\mathbf{d}_0 - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\}$$

and then

$$\boxed{\nabla_{\mathbf{d}_0} \log f(\mathbf{x}, \mathbf{y}, m; \boldsymbol{\theta}) = \frac{1}{\sigma_\eta^2} \boldsymbol{\Gamma}(\mathbf{x})^T \{\mathbf{y} - \boldsymbol{\Gamma}(\mathbf{x})\mathbf{d}_0 - \boldsymbol{\Psi}(\mathbf{x}) \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{v} \mid \mathbf{x}, m, \mathbf{y})\}.}$$

▶ For $\operatorname{vec} \mathbf{D}$, only $f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{v})$ depends on this parameter, so

$$\nabla_{\operatorname{vec} \mathbf{D}} \log f(\mathbf{x}, \mathbf{y}, m; \boldsymbol{\theta}) = \iint \nabla_{\operatorname{vec} \mathbf{D}} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) \; f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y}) \; d\mathbf{w}.$$

Since

$$\begin{aligned} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) &\propto -\frac{1}{2\sigma_\eta^2} \|\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Gamma}(\mathbf{x})\mathbf{D}\mathbf{v}\|^2 \\ &= -\frac{1}{2\sigma_\eta^2} \|\mathbf{y} - \nu(\mathbf{x}) - \{\mathbf{v}^T \otimes \boldsymbol{\Gamma}(\mathbf{x})\} \operatorname{vec} \mathbf{D}\|^2, \end{aligned}$$

we have

$$\begin{aligned} \nabla_{\operatorname{vec} \mathbf{D}} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) &= \frac{1}{\sigma_\eta^2} \{\mathbf{v} \otimes \boldsymbol{\Gamma}(\mathbf{x})^T\} \{\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\} \\ &= \frac{1}{\sigma_\eta^2} \operatorname{vec}[\boldsymbol{\Gamma}(\mathbf{x})^T \{\mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}\}\mathbf{v}^T] \end{aligned}$$

and then

$$\boxed{\nabla_{\mathrm{vec}\,\mathbf{D}} \log f(\mathbf{x}, \mathbf{y}, m; \boldsymbol{\theta}) = \frac{1}{\sigma_\eta^2} \mathrm{vec}(\boldsymbol{\Gamma}(\mathbf{x})^T [\{\mathbf{y} - \nu(\mathbf{x})\} \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{v}^T \mid \mathbf{x}, m, \mathbf{y}) - \boldsymbol{\Psi}(\mathbf{x}) \mathbb{E}_{\boldsymbol{\theta}}(\mathbf{v}\mathbf{v}^T \mid \mathbf{x}, m, \mathbf{y})]).}$$

▶ For $\boldsymbol{\sigma}_u^2$, we have only $f(\mathbf{w})$ depending on this parameter, so

$$\nabla_{\boldsymbol{\sigma}_u^2} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = \iint \nabla_{\boldsymbol{\sigma}_u^2} \log f(\mathbf{w})\ f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y})\ d\mathbf{w}.$$

As before,

$$\mathsf{d} \log f(\mathbf{w}) = -\frac{1}{2} \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \mathsf{d}\boldsymbol{\Sigma}) + \frac{1}{2} \mathbf{w}^T \boldsymbol{\Sigma}^{-1} (\mathsf{d}\boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \mathbf{w},$$

but now, differentiating with respect to $\boldsymbol{\sigma}_u^2$,

$$\mathsf{d}\boldsymbol{\Sigma} = \begin{pmatrix} \mathrm{diag}(\mathsf{d}\boldsymbol{\sigma}_u^2) & \mathbf{O} \\ \mathbf{O} & \mathbf{O} \end{pmatrix}.$$

Then

$$\begin{aligned} \mathrm{tr}(\boldsymbol{\Sigma}^{-1} \mathsf{d}\boldsymbol{\Sigma}) &= \mathrm{tr}\{\boldsymbol{\Sigma}_{11}^{-1} \mathrm{diag}(\mathsf{d}\boldsymbol{\sigma}_u^2)\} \\ &= \mathrm{diag}(\boldsymbol{\Sigma}_{11}^{-1})^T \mathsf{d}\boldsymbol{\sigma}_u^2 \end{aligned}$$

and

$$\begin{aligned} \mathbf{w}^T \boldsymbol{\Sigma}^{-1} (\mathsf{d}\boldsymbol{\Sigma}) \boldsymbol{\Sigma}^{-1} \mathbf{w} &= (\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^T \mathrm{diag}(\mathsf{d}\boldsymbol{\sigma}_u^2) (\boldsymbol{\Sigma}^{-1}\mathbf{w})_1 \\ &= \{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^{\odot 2}\}^T \mathsf{d}\boldsymbol{\sigma}_u^2, \end{aligned}$$

where $^{\odot 2}$ denotes element-wise squaring. Then

$$\mathsf{d} \log f(\mathbf{w}) = -\frac{1}{2} \mathrm{diag}(\boldsymbol{\Sigma}_{11}^{-1})^T \mathsf{d}\boldsymbol{\sigma}_u^2 + \frac{1}{2} \{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^{\odot 2}\}^T \mathsf{d}\boldsymbol{\sigma}_u^2,$$

so

$$\nabla_{\boldsymbol{\sigma}_u^2} \log f(\mathbf{w}) = -\frac{1}{2}\operatorname{diag}(\boldsymbol{\Sigma}_{11}^{-1}) + \frac{1}{2}(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^{\odot 2}$$

and then

$$\boxed{\nabla_{\boldsymbol{\sigma}_u^2} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = -\frac{1}{2}\operatorname{diag}(\boldsymbol{\Sigma}_{11}^{-1}) + \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^{\odot 2} \mid \mathbf{x}, m, \mathbf{y}\}.}$$

The second term can be written more explicitly in terms of $\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{w}\mathbf{w}^T \mid \mathbf{x}, m, \mathbf{y})$: since $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1 = [\mathbf{I}_{p_1}, \mathbf{O}]\, \boldsymbol{\Sigma}^{-1}\mathbf{w}$ and $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^{\odot 2} = \operatorname{diag}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^T\}$, we have

$$\begin{aligned}(\boldsymbol{\Sigma}^{-1}\mathbf{w})_1^{\odot 2} &= \operatorname{diag}\{[\mathbf{I}_{p_1}, \mathbf{O}]\, \boldsymbol{\Sigma}^{-1}\mathbf{w}\mathbf{w}^T\boldsymbol{\Sigma}^{-1} \begin{bmatrix} \mathbf{I}_{p_1} \\ \mathbf{O} \end{bmatrix}\} \\ &= \operatorname{diag}\{\left[\boldsymbol{\Sigma}_{11}^{-1}, \boldsymbol{\Sigma}_{12}^{-1}\right] \mathbf{w}\mathbf{w}^T \begin{bmatrix} \boldsymbol{\Sigma}_{11}^{-1} \\ \boldsymbol{\Sigma}_{21}^{-1} \end{bmatrix}\}\end{aligned}$$

and then we take $\mathbb{E}_{\boldsymbol{\theta}}$, which commutes with the diag operator.

▶ Similarly, for $\boldsymbol{\sigma}_v^2$ we have

$$\nabla_{\boldsymbol{\sigma}_v^2} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = \iint \nabla_{\boldsymbol{\sigma}_v^2} \log f(\mathbf{w})\ f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y})\ d\mathbf{w}.$$

Since

$$\mathsf{d} \log f(\mathbf{w}) = -\frac{1}{2}\operatorname{tr}(\boldsymbol{\Sigma}^{-1}\mathsf{d}\boldsymbol{\Sigma}) + \frac{1}{2}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}(\mathsf{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{w},$$

differentiating with respect to $\boldsymbol{\sigma}_v^2$ we get

$$\mathsf{d}\boldsymbol{\Sigma} = \begin{pmatrix} \mathbf{O} & \mathbf{O} \\ \mathbf{O} & \operatorname{diag}(\mathsf{d}\boldsymbol{\sigma}_v^2) \end{pmatrix}$$

and then

$$\begin{aligned}\mathrm{tr}(\boldsymbol{\Sigma}^{-1}\mathsf{d}\boldsymbol{\Sigma}) &= \mathrm{tr}\{\boldsymbol{\Sigma}_{22}^{-1}\,\mathrm{diag}(\mathsf{d}\boldsymbol{\sigma}_v^2)\}\\ &= \mathrm{diag}(\boldsymbol{\Sigma}_{22}^{-1})^T\mathsf{d}\boldsymbol{\sigma}_v^2\end{aligned}$$

and

$$\begin{aligned}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}(\mathsf{d}\boldsymbol{\Sigma})\boldsymbol{\Sigma}^{-1}\mathbf{w} &= (\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T\,\mathrm{diag}(\mathsf{d}\boldsymbol{\sigma}_v^2)(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2\\ &= \{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^{\odot 2}\}^T\mathsf{d}\boldsymbol{\sigma}_v^2.\end{aligned}$$

So, as before,

$$\nabla_{\boldsymbol{\sigma}_v^2}\log f(\mathbf{w}) = -\frac{1}{2}\,\mathrm{diag}(\boldsymbol{\Sigma}_{22}^{-1}) + \frac{1}{2}(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^{\odot 2}$$

and then

$$\boxed{\nabla_{\boldsymbol{\sigma}_v^2}\log f(\mathbf{x},m,\mathbf{y};\boldsymbol{\theta}) = -\frac{1}{2}\,\mathrm{diag}(\boldsymbol{\Sigma}_{22}^{-1}) + \frac{1}{2}\mathbb{E}_{\boldsymbol{\theta}}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^{\odot 2}\mid \mathbf{x},m,\mathbf{y}\}.}$$

Again, the second term can be written out in terms of $\mathbb{E}_{\boldsymbol{\theta}}(\mathbf{w}\mathbf{w}^T\mid \mathbf{x},m,\mathbf{y})$ using that $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2 = [\mathbf{O},\mathbf{I}_{p_2}]\,\boldsymbol{\Sigma}^{-1}\mathbf{w}$ and $(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^{\odot 2} = \mathrm{diag}\{(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^T\}$, so

$$\begin{aligned}(\boldsymbol{\Sigma}^{-1}\mathbf{w})_2^{\odot 2} &= \mathrm{diag}\{[\mathbf{O},\mathbf{I}_{p_2}]\,\boldsymbol{\Sigma}^{-1}\mathbf{w}\mathbf{w}^T\boldsymbol{\Sigma}^{-1}\begin{bmatrix}\mathbf{O}\\ \\ \mathbf{I}_{p_2}\end{bmatrix}\}\\ &= \mathrm{diag}\{\left[\boldsymbol{\Sigma}_{21}^{-1},\boldsymbol{\Sigma}_{22}^{-1}\right]\mathbf{w}\mathbf{w}^T\begin{bmatrix}\boldsymbol{\Sigma}_{12}^{-1}\\ \\ \boldsymbol{\Sigma}_{22}^{-1}\end{bmatrix}\}.\end{aligned}$$

Then we take $\mathbb{E}_{\boldsymbol{\theta}}$, which commutes with the diag operator.

► Finally, for σ_η^2, which is only present in $f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w})$, we have

$$\frac{\partial}{\partial \sigma_\eta^2} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = \iint \frac{\partial}{\partial \sigma_\eta^2} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) \; f(\mathbf{w} \mid \mathbf{x}, m, \mathbf{y}) \; d\mathbf{w},$$

where

$$\log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) = -\frac{m}{2} \log 2\pi\sigma_\eta^2 - \frac{1}{2\sigma_\eta^2} \|\boldsymbol{\eta}\|^2$$

with $\boldsymbol{\eta} = \mathbf{y} - \nu(\mathbf{x}) - \boldsymbol{\Psi}(\mathbf{x})\mathbf{v}$. Then

$$\frac{\partial}{\partial \sigma_\eta^2} \log f(\mathbf{y} \mid \mathbf{x}, m, \mathbf{w}) = -\frac{m}{2\sigma_\eta^2} + \frac{1}{2(\sigma_\eta^2)^2} \|\boldsymbol{\eta}\|^2$$

and consequently

$$\boxed{\frac{\partial}{\partial \sigma_\eta^2} \log f(\mathbf{x}, m, \mathbf{y}; \boldsymbol{\theta}) = -\frac{m}{2\sigma_\eta^2} + \frac{1}{2(\sigma_\eta^2)^2} \mathbb{E}_{\boldsymbol{\theta}}(\|\boldsymbol{\eta}\|^2 \mid \mathbf{x}, m, \mathbf{y}).}$$

0.2.2 Consistency

The consistency proof follows the usual steps for maximum likelihood estimators and M-estimators; see e.g. Pollard (1984) and Van der Vaart (2000). First we show that the asymptotic objective function has a unique maximum at the true parameter $\boldsymbol{\theta}_0$, then that $\{\hat{\boldsymbol{\theta}}_n\}$ is bounded in probability, and finally, via the Argmax Theorem, that $\hat{\boldsymbol{\theta}}_n$ converges to $\boldsymbol{\theta}_0$ in probability. In the following we use $Z = (\mathbf{x}, m, \mathbf{y})$ to simplify the notation. Also, we define $\boldsymbol{\xi}_n = (\xi_{1n}, \xi_{2n}, \xi_{3n}, \xi_{4n})^T$ and $\mathbf{P}(\boldsymbol{\theta}) = (P(\mu), \sum_{k=1}^{p_1} P(\phi_k), P(\nu), \sum_{k=1}^{p_2} P(\psi_k))^T$.

Lemma 1. *Under assumption A2, the function* $M(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}\{\log f(Z;\boldsymbol{\theta})\}$ *has a unique maximum at* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Proof. This is a consequence of Jensen's Inequality and model identifiability:

$$E_{\boldsymbol{\theta}_0}\left\{\log \frac{f(Z;\boldsymbol{\theta})}{f(Z;\boldsymbol{\theta}_0)}\right\} \leq \log E_{\boldsymbol{\theta}_0}\left\{\frac{f(Z;\boldsymbol{\theta})}{f(Z;\boldsymbol{\theta}_0)}\right\} = 0 \tag{1}$$

because

$$E_{\boldsymbol{\theta}_0}\left\{\frac{f(Z;\boldsymbol{\theta})}{f(Z;\boldsymbol{\theta}_0)}\right\} = 1$$

for all $\boldsymbol{\theta}$. Moreover, inequality (1) is strict unless $P_{\boldsymbol{\theta}_0}\{f(Z;\boldsymbol{\theta})/f(Z;\boldsymbol{\theta}_0) = 1\} = 1$, which happens only if $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ by identifiability. Then $E_{\boldsymbol{\theta}_0}\{\log f(Z;\boldsymbol{\theta})\} < E_{\boldsymbol{\theta}_0}\{\log f(Z;\boldsymbol{\theta}_0)\}$ for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. □

Lemma 2. *Under assumptions A1 and A3,* $\|\hat{\boldsymbol{\theta}}_n\| = O_P(1)$.

Proof. Let

$$M_n(\boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \log f(Z_i;\boldsymbol{\theta}).$$

By definition, $\hat{\boldsymbol{\theta}}_n$ maximizes

$$\ell_n(\boldsymbol{\theta}) = M_n(\boldsymbol{\theta}) - \boldsymbol{\xi}_n^T \mathbf{P}(\boldsymbol{\theta}),$$

so we have

$$M_n(\hat{\boldsymbol{\theta}}_n) - M_n(\boldsymbol{\theta}_0) \geq \boldsymbol{\xi}_n^T\{\mathbf{P}(\hat{\boldsymbol{\theta}}_n) - \mathbf{P}(\boldsymbol{\theta}_0)\}.$$

Since $\mathbf{P}(\boldsymbol{\theta}) \geq 0$ for all $\boldsymbol{\theta}$, this implies

$$\frac{1}{n}\sum_{i=1}^{n} \log \frac{f(Z_i;\hat{\boldsymbol{\theta}}_n)}{f(Z_i;\boldsymbol{\theta}_0)} \geq -\boldsymbol{\xi}_n^T \mathbf{P}(\boldsymbol{\theta}_0), \tag{2}$$

with the right-hand side going to zero as $n \to \infty$. As in Van der Vaart (2000, p. 63), consider the surrogate functions

$$g(z;\boldsymbol{\theta}) = \log\left\{\frac{f(z;\boldsymbol{\theta}) + f(z;\boldsymbol{\theta}_0)}{2f(z;\boldsymbol{\theta}_0)}\right\}$$

which satisfy

$$\log(\frac{1}{2}) \leq g(z;\boldsymbol{\theta}) \leq \log\left\{\frac{c(z) + f(z;\boldsymbol{\theta}_0)}{2f(z;\boldsymbol{\theta}_0)}\right\}$$

where $c(z) \geq f(z;\boldsymbol{\theta})$ for all $\boldsymbol{\theta}$. By concavity of the logarithm,

$$g(z;\boldsymbol{\theta}) \geq \frac{1}{2}\log\frac{f(z;\boldsymbol{\theta})}{f(z;\boldsymbol{\theta}_0)} + \frac{1}{2}\log(1) = \frac{1}{2}\log\frac{f(z;\boldsymbol{\theta})}{f(z;\boldsymbol{\theta}_0)},$$

so (2) implies

$$\frac{1}{n}\sum_{i=1}^{n} g(Z_i;\hat{\boldsymbol{\theta}}_n) \geq \frac{1}{2}\{-\boldsymbol{\xi}_n^T \mathbf{P}(\boldsymbol{\theta}_0)\}. \tag{3}$$

For any $K > 0$, if $\|\hat{\boldsymbol{\theta}}_n\| \geq K$ we have

$$\frac{1}{n}\sum_{i=1}^{n} g(Z_i;\hat{\boldsymbol{\theta}}_n) \leq \frac{1}{n}\sum_{i=1}^{n} \psi(Z_i) \tag{4}$$

with

$$\psi(z) = \sup_{\|\boldsymbol{\theta}\| \geq K} g(z;\boldsymbol{\theta}).$$

By Law of Large Numbers $n^{-1}\sum_{i=1}^{n}\psi(Z_i) \xrightarrow{P} E_{\boldsymbol{\theta}_0}\{\psi(Z)\}$, and by Bounded Convergence Theorem we can switch supremum and expectation:

$$E_{\boldsymbol{\theta}_0}\{\psi(Z)\} = \sup_{\|\boldsymbol{\theta}\| \geq K} E_{\boldsymbol{\theta}_0}\{g(Z;\boldsymbol{\theta})\}.$$

Now, as in the proof of Lemma 1, by Jensen's Inequality we have

$$E_{\boldsymbol{\theta}_0}\{g(Z;\boldsymbol{\theta})\} \leq \log E_{\boldsymbol{\theta}_0}\left\{\frac{f(Z;\boldsymbol{\theta})+f(Z;\boldsymbol{\theta}_0)}{2f(Z;\boldsymbol{\theta}_0)}\right\} = 0 = E_{\boldsymbol{\theta}_0}\{g(Z;\boldsymbol{\theta}_0)\}$$

with strict inequality for any $\boldsymbol{\theta} \neq \boldsymbol{\theta}_0$. So $\max E_{\boldsymbol{\theta}_0}\{g(Z;\boldsymbol{\theta})\} = 0$ and the maximum is attained only at the $\boldsymbol{\theta}_0$s. We can rule out the possibility of $E_{\boldsymbol{\theta}_0}\{g(Z;\boldsymbol{\theta})\}$ approaching zero at infinity because $\lim_{\|\boldsymbol{\theta}\|\to\infty} f(z;\boldsymbol{\theta}) = 0$ and then

$$\lim_{\|\boldsymbol{\theta}\|\to\infty} E_{\boldsymbol{\theta}_0}\{g(Z;\boldsymbol{\theta})\} = E_{\boldsymbol{\theta}_0}\{\lim_{\|\boldsymbol{\theta}\|\to\infty} g(Z;\boldsymbol{\theta})\} = \log(\frac{1}{2}) < 0.$$

Therefore, there exists an $\varepsilon > 0$ and a $K > 0$ such that $E_{\boldsymbol{\theta}_0}\{\psi(Z)\} < -\varepsilon$. This fact together with (3) and (4) imply that $P(\|\hat{\boldsymbol{\theta}}_n\| \geq K)$ goes to zero as $n \to \infty$. □

Lemma 3. *Under assumptions A1–A3,* $\hat{\boldsymbol{\theta}}_n \xrightarrow{P} \boldsymbol{\theta}_0$ *as* $n \to \infty$.

Proof. By Lemma 2, for any $\varepsilon > 0$ we can choose $K > 0$ such that $P\{\|\hat{\boldsymbol{\theta}}_n\| > K\} < \varepsilon/2$ for all n, and we can choose it so that $K \geq \|\boldsymbol{\theta}_0\|$. On the other hand, for $\|\hat{\boldsymbol{\theta}}_n\| \leq K$ we have

$$\hat{\boldsymbol{\theta}}_n = \underset{\Theta\cap\{\|\boldsymbol{\theta}\|\leq K\}}{\operatorname{argmax}} \ell_n(\boldsymbol{\theta}).$$

The penalty function $\mathbf{P}(\boldsymbol{\theta})$ is continuous and therefore uniformly continuous on compact sets, and the process $M_n(\boldsymbol{\theta})$ is stochastically equicontinuous (Pollard, 1984, ch. 7), so $\ell_n(\boldsymbol{\theta})$ converges in probability to $M(\boldsymbol{\theta})$ uniformly

over bounded sets. Then by the Argmax Theorem (Van der Vaart, 2000, ch. 5.9),

$$\underset{\Theta\cap\{\|\boldsymbol{\theta}\|\leq K\}}{\operatorname{argmax}}\ \ell_n(\boldsymbol{\theta}) \xrightarrow{P} \underset{\Theta\cap\{\|\boldsymbol{\theta}\|\leq K\}}{\operatorname{argmax}}\ M(\boldsymbol{\theta}) = \boldsymbol{\theta}_0,$$

so for any $\delta > 0$ we can choose N such that $P\{\|\hat{\boldsymbol{\theta}}_n\| \leq K, \|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| > \delta\} < \varepsilon/2$ for every $n \geq N$. This completes the proof. □

0.2.3 Asymptotic normality

To prove the asymptotic normality of $\hat{\boldsymbol{\theta}}_n$ we will follow the approach of Geyer (1994), which makes use of the tangent cone of the parameter space. The definition and properties of tangent cones can be found in Rockafellar and Wets (1998, ch. 6). Using Theorem 6.31 of Rockafellar and Wets (1998), the tangent cone of Θ at $\boldsymbol{\theta}_0$ is

$$\begin{aligned}\mathscr{T}_0 \;=\; &\{\boldsymbol{\delta} \in \mathbb{R}^s : \nabla h_{kl}^C(\boldsymbol{\theta}_0)^T\boldsymbol{\delta} = 0, \quad k = 1,\ldots,l, \quad l = 1,\ldots,p_1,\\ &\nabla h_{kl}^D(\boldsymbol{\theta}_0)^T\boldsymbol{\delta} = 0, \quad k = 1,\ldots,l, \quad l = 1,\ldots,p_2\}.\end{aligned}$$

Note that $c_{k1,0}$ and $d_{k1,0}$ are strictly positive, so they do not contribute restrictions to the tangent cone. The explicit forms of $\nabla h_{kl}^C(\boldsymbol{\theta})$ and $\nabla h_{kl}^D(\boldsymbol{\theta})$ are derived in Section 0.2.4. Fisher's information matrix $\mathbf{F}_0$, which appears in the results below, was derived in Section 0.2.1.

Lemma 4. *$\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| = O_P(n^{-1/2})$ if $\sqrt{n}\|\boldsymbol{\xi}_n\| = O_P(1)$.*

Proof. The estimator $\hat{\boldsymbol{\theta}}_n$ maximizes $\ell_n(\boldsymbol{\theta})$, or equivalently

$$\tilde{\ell}_n(\boldsymbol{\theta}) = n\{\ell_n(\boldsymbol{\theta}) - \ell_n(\boldsymbol{\theta}_0)\},$$

over $\boldsymbol{\theta} \in \Theta$. Let $r(z, \boldsymbol{\theta}, \boldsymbol{\theta}_0)$ be such that

$$\begin{aligned} \log f(z, \boldsymbol{\theta}) &= \log f(z, \boldsymbol{\theta}_0) + \nabla \log f(z, \boldsymbol{\theta}_0)^T (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &\quad + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \, r(z, \boldsymbol{\theta}, \boldsymbol{\theta}_0), \end{aligned}$$

and $M(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}_0}\{\log f(Z; \boldsymbol{\theta})\}$ as above. Then

$$\begin{aligned} \tilde{M}_n(\boldsymbol{\theta}) &= \sum_{i=1}^{n} \nabla \log f(Z_i, \boldsymbol{\theta}_0)^T (\boldsymbol{\theta} - \boldsymbol{\theta}_0) \\ &\quad + \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \sum_{i=1}^{n} [r(Z_i, \boldsymbol{\theta}, \boldsymbol{\theta}_0) - E_{\boldsymbol{\theta}_0}\{r(Z, \boldsymbol{\theta}, \boldsymbol{\theta}_0)\}] \\ &\quad + n\{M(\boldsymbol{\theta}) - M(\boldsymbol{\theta}_0)\} - n\boldsymbol{\xi}_n^T \{\mathbf{P}(\boldsymbol{\theta}) - \mathbf{P}(\boldsymbol{\theta}_0)\}. \end{aligned}$$

Note that $E_{\theta_0}\{\nabla \log f(Z, \boldsymbol{\theta}_0)\} = \nabla M(\boldsymbol{\theta}_0) = \mathbf{0}$ because $f(z, \boldsymbol{\theta})$ is a density function; the fact that $\boldsymbol{\theta}_0$ maximizes $M(\boldsymbol{\theta})$ does not necessarily imply $\nabla M(\boldsymbol{\theta}_0) = \mathbf{0}$ because $\boldsymbol{\theta}_0$ may be on the border of Θ. Let

$$R_n(\boldsymbol{\theta}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [r(Z_i, \boldsymbol{\theta}, \boldsymbol{\theta}_0) - E_{\boldsymbol{\theta}_0}\{r(Z, \boldsymbol{\theta}, \boldsymbol{\theta}_0)\}]$$

and

$$\mathbf{Z}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \nabla \log f(Z_i, \boldsymbol{\theta}_0).$$

Since $\tilde{\ell}_n(\hat{\boldsymbol{\theta}}_n) \geq \tilde{\ell}_n(\boldsymbol{\theta}_0) = 0$,

$$\begin{aligned} &\sqrt{n}\mathbf{Z}_n^T(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + \sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| R_n(\hat{\boldsymbol{\theta}}_n) - n\boldsymbol{\xi}_n^T\{\mathbf{P}(\hat{\boldsymbol{\theta}}_n) - \mathbf{P}(\boldsymbol{\theta}_0)\} \\ \geq \; & -n\{M(\hat{\boldsymbol{\theta}}_n) - M(\boldsymbol{\theta}_0)\}. \end{aligned} \tag{5}$$

Clearly $\|\mathbf{Z}_n\| = O_P(1)$ because $\mathbf{Z}_n \xrightarrow{D} \mathrm{N}(0, \mathbf{F}_0)$. The mean value theorem applied to $\mathbf{P}(\boldsymbol{\theta})$ implies

$$\begin{aligned} n\boldsymbol{\xi}_n^T\{\mathbf{P}(\hat{\boldsymbol{\theta}}_n) - \mathbf{P}(\boldsymbol{\theta}_0)\} &= n\|\boldsymbol{\xi}_n\| O_P(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|) \\ &= \sqrt{n}\|\boldsymbol{\xi}_n\| O_P(1)\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|. \end{aligned}$$

The process $R_n(\boldsymbol{\theta})$ is equicontinuous in $\boldsymbol{\theta}$ (Pollard, 1984, ch. 7) and $R_n(\boldsymbol{\theta}) \xrightarrow{D} \mathrm{N}(0, v(\boldsymbol{\theta}, \boldsymbol{\theta}_0))$ with $v(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = 0$, so $R_n(\hat{\boldsymbol{\theta}}_n) \xrightarrow{P} 0$. Then it follows from (5) that

$$\begin{aligned} & \{O_P(1) + o_P(1) - \sqrt{n}\|\boldsymbol{\xi}_n\| O_P(1)\}\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \\ \geq\ & -n\{M(\hat{\boldsymbol{\theta}}_n) - M(\boldsymbol{\theta}_0)\}. \end{aligned}$$

Now,

$$M(\hat{\boldsymbol{\theta}}_n) - M(\boldsymbol{\theta}_0) = \frac{1}{2}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)^T \nabla^2 M(\boldsymbol{\theta}_0)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) + o_P(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|^2)$$

and $\nabla^2 M(\boldsymbol{\theta}_0) = -\mathbf{F}_0$, so if $\lambda_1 > 0$ is the smallest eigenvalue of $\mathbf{F}_0$,

$$-n\{M(\hat{\boldsymbol{\theta}}_n) - M(\boldsymbol{\theta}_0)\} \geq n\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|^2\lambda_1 - no_P(\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|^2).$$

Then from the last two inequalities we have

$$\{O_P(1) + o_P(1) - \sqrt{n}\|\boldsymbol{\xi}_n\| O_P(1)\}\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| \geq n\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|^2\{\lambda_1 - o_P(1)\},$$

which implies $\sqrt{n}\|\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| = O_P(1)$. □

Theorem 5. *Under assumption A4,* $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \xrightarrow{D} \boldsymbol{\delta}(\mathbf{Z})$*, with* $\boldsymbol{\delta}(\mathbf{Z})$ *the maximizer of*

$$W(\boldsymbol{\delta}) = \{\mathbf{Z}^T - \boldsymbol{\kappa}^T \mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)\}\boldsymbol{\delta} - \frac{1}{2}\boldsymbol{\delta}^T \mathbf{F}_0 \boldsymbol{\delta}$$

over $\boldsymbol{\delta} \in \mathscr{T}_0$*, where* $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{F}_0)$.

Proof. Let $W_n(\boldsymbol{\delta}) = \tilde{\ell}_n(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n})$ with $\tilde{\ell}_n(\boldsymbol{\theta})$ as in the previous lemma. Then

$$\begin{aligned} W_n(\boldsymbol{\delta}) &= \mathbf{Z}_n^T \boldsymbol{\delta} \\ &\quad + \|\boldsymbol{\delta}\| \, R_n(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}) \\ &\quad + n\{M(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}) - M(\boldsymbol{\theta}_0)\} \\ &\quad - n\boldsymbol{\xi}_n^T \{\mathbf{P}(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}) - \mathbf{P}(\boldsymbol{\theta}_0)\}, \end{aligned}$$

and $\hat{\boldsymbol{\delta}}_n = \sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0)$ maximizes $W_n(\boldsymbol{\delta})$ over $\mathscr{T}_n = \sqrt{n}(\Theta - \{\boldsymbol{\theta}_0\})$. Having already proved that $\|\hat{\boldsymbol{\delta}}_n\| = O_P(1)$, given $\varepsilon > 0$ we can take K such that $P(\|\hat{\boldsymbol{\delta}}_n\| \leq K) \geq 1 - \varepsilon$ for every n, and focus on the set $\mathscr{T}_n \cap \{\|\boldsymbol{\delta}\| \leq K\}$. In the limit, as $n \to \infty$, we have:

$$\mathscr{T}_n \to \mathscr{T}_0, \text{ the tangent cone of } \Theta \text{ at } \boldsymbol{\theta}_0$$

(Geyer, 1994);

$$\mathbf{Z}_n \xrightarrow{D} \mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{F}_0);$$

$$R_n(\boldsymbol{\theta}_0 + \boldsymbol{\delta}_n/\sqrt{n}) \xrightarrow{P} 0 \text{ for any bounded sequence } \{\boldsymbol{\delta}_n\}$$

by stochastic equicontinuity of $R_n(\boldsymbol{\theta})$;

$$n\{M(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}) - M(\boldsymbol{\theta}_0)\} = \frac{1}{2}\boldsymbol{\delta}^T\{-\mathbf{F}_0 + o_P(1)\}\boldsymbol{\delta};$$

and

$$n\boldsymbol{\xi}_n^T\{\mathbf{P}(\boldsymbol{\theta}_0 + \boldsymbol{\delta}/\sqrt{n}) - \mathbf{P}(\boldsymbol{\theta}_0)\} = \sqrt{n}\boldsymbol{\xi}_n^T\{\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0) + o_P(1)\}\boldsymbol{\delta}.$$

All this implies that $W_n(\boldsymbol{\delta}) \xrightarrow{D} W(\boldsymbol{\delta})$ with

$$W(\boldsymbol{\delta}) = \{\mathbf{Z}^T - \boldsymbol{\kappa}^T\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)\}\boldsymbol{\delta} - \frac{1}{2}\boldsymbol{\delta}^T\mathbf{F}_0\boldsymbol{\delta},$$

and the convergence is uniform in $\boldsymbol{\delta}$, i.e. $\sup_{\mathscr{T}_n \cap \{\|\boldsymbol{\delta}\| \leq K\}} |W_n(\boldsymbol{\delta}) - W(\boldsymbol{\delta})| \xrightarrow{P} 0$. Then

$$\underset{\mathscr{T}_n \cap \{\|\boldsymbol{\delta}\| \leq K\}}{\operatorname{argmax}} W_n(\boldsymbol{\delta}) \xrightarrow{D} \underset{\mathscr{T}_0}{\operatorname{argmax}} W(\boldsymbol{\delta}),$$

which implies that $\hat{\boldsymbol{\delta}}_n \xrightarrow{D} \boldsymbol{\delta}(\mathbf{Z})$ as stated. □

We have $\mathscr{T}_0 = \{\boldsymbol{\delta} \in \mathbb{R}^s : \mathbf{A}\boldsymbol{\delta} = \mathbf{0}\}$, with $\mathbf{A}$ the $\{p_1(p_1+1)/2 + p_2(p_2+1)/2\} \times s$ matrix with rows $\nabla h_{kl}^C(\boldsymbol{\theta}_0)^T$ and $\nabla h_{kl}^D(\boldsymbol{\theta}_0)^T$. Let $s_1 = \{p_1(p_1+1)/2 + p_2(p_2+1)/2\}$. Then a $\boldsymbol{\delta} \in \mathscr{T}_0$ is of the form $\boldsymbol{\delta} = \mathbf{B}^T\tilde{\boldsymbol{\delta}}$ with $\mathbf{B}$ an orthogonal $(s - s_1) \times s$ matrix with rows orthogonal to those of $\mathbf{A}$ and $\tilde{\boldsymbol{\delta}} \in \mathbb{R}^{s-s_1}$ free. So we can reparameterize the process $W(\boldsymbol{\delta})$ above in terms of $\tilde{\boldsymbol{\delta}}$:

$$W(\boldsymbol{\delta}) = W(\mathbf{B}^T\tilde{\boldsymbol{\delta}}) = \{\mathbf{Z}^T - \boldsymbol{\kappa}^T\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)\}\mathbf{B}^T\tilde{\boldsymbol{\delta}} - \frac{1}{2}\tilde{\boldsymbol{\delta}}^T\mathbf{B}\mathbf{F}_0\mathbf{B}^T\tilde{\boldsymbol{\delta}},$$

which is maximized by $\tilde{\boldsymbol{\delta}}(\mathbf{Z}) = (\mathbf{B}\mathbf{F}_0\mathbf{B}^T)^{-1}\mathbf{B}\{\mathbf{Z} - \mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)^T\boldsymbol{\kappa}\}$, and then $\boldsymbol{\delta}(\mathbf{Z}) = \mathbf{B}^T\tilde{\boldsymbol{\delta}}(\mathbf{Z})$. Since $\mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{F}_0)$, we have $\tilde{\boldsymbol{\delta}}(\mathbf{Z}) \sim \mathrm{N}(-(\mathbf{B}\mathbf{F}_0\mathbf{B}^T)^{-1}\mathbf{B}\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)^T\boldsymbol{\kappa}, (\mathbf{B}\mathbf{F}_0\mathbf{B}^T)^{-1})$ and then

$$\boldsymbol{\delta}(\mathbf{Z}) \sim \mathrm{N}(-\mathbf{V}\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)^T\boldsymbol{\kappa}, \mathbf{V})$$

with $\mathbf{V} = \mathbf{B}^T(\mathbf{B}\mathbf{F}_0\mathbf{B}^T)^{-1}\mathbf{B}$. The explicit form of $\mathsf{D}\mathbf{P}(\boldsymbol{\theta}_0)$ is derived in Section 0.2.4.

0.2.4 Derivatives of constraints and smoothness penalties

The explicit forms of $\nabla h_{kl}^C(\boldsymbol{\theta})$ and $\nabla h_{kl}^D(\boldsymbol{\theta})$ can be derived as follows. Let $\mathbf{K}_{\mathbf{c}_k}$ be the $q \times s$ matrix that "extracts" $\mathbf{c}_k$ from $\boldsymbol{\theta}$, that is, $\mathbf{c}_k = \mathbf{K}_{\mathbf{c}_k}\boldsymbol{\theta}$. Then we can write $h_{kl}^C(\boldsymbol{\theta}) = \boldsymbol{\theta}^T\mathbf{K}_{\mathbf{c}_k}^T\mathbf{J}\mathbf{K}_{\mathbf{c}_l}\boldsymbol{\theta} - \delta_{kl}$ and it follows that

$$\nabla h_{kl}^C(\boldsymbol{\theta}) = (\mathbf{K}_{\mathbf{c}_k}^T\mathbf{J}\mathbf{K}_{\mathbf{c}_l} + \mathbf{K}_{\mathbf{c}_l}^T\mathbf{J}\mathbf{K}_{\mathbf{c}_k})\boldsymbol{\theta}.$$

Similarly, if $\mathbf{K}_{\mathbf{d}_k}$ is the $q \times s$ matrix such that $\mathbf{d}_k = \mathbf{K}_{\mathbf{d}_k}\boldsymbol{\theta}$, we have $h_{kl}^D(\boldsymbol{\theta}) = \boldsymbol{\theta}^T\mathbf{K}_{\mathbf{d}_k}^T\mathbf{J}\mathbf{K}_{\mathbf{d}_k}\boldsymbol{\theta} - \delta_{kl}$ and then

$$\nabla h_{kl}^D(\boldsymbol{\theta}) = (\mathbf{K}_{\mathbf{d}_k}^T\mathbf{J}\mathbf{K}_{\mathbf{d}_l} + \mathbf{K}_{\mathbf{d}_l}^T\mathbf{J}\mathbf{K}_{\mathbf{d}_k})\boldsymbol{\theta}.$$

The explicit form of $\mathsf{D}\mathbf{P}(\boldsymbol{\theta})$ is derived in a similar way. Using extraction matrices $\mathbf{K}$ as above and the smoothing matrix $\boldsymbol{\Omega}$ derived in Section 0.1.3,

we have $\mathbf{P}(\boldsymbol{\theta}) = (P(\mu), \sum_{k=1}^{p_1} P(\phi_k), P(\nu), \sum_{k=1}^{p_2} P(\psi_k))^T$ with

$$\begin{aligned} P(\mu) &= \mathbf{c}_0^T \boldsymbol{\Omega} \mathbf{c}_0 \\ &= \boldsymbol{\theta}^T \mathbf{K}_{\mathbf{c}_0}^T \boldsymbol{\Omega} \mathbf{K}_{\mathbf{c}_0} \boldsymbol{\theta}, \end{aligned}$$

$$\begin{aligned} \sum_{k=1}^{p_1} P(\phi_k) &= \operatorname{tr}(\mathbf{C}^T \boldsymbol{\Omega} \mathbf{C}) \\ &= \operatorname{vec} \mathbf{C}^T (\mathbf{I}_{p_1} \otimes \boldsymbol{\Omega}) \operatorname{vec} \mathbf{C} \\ &= \boldsymbol{\theta}^T \mathbf{K}_{\operatorname{vec} \mathbf{C}}^T (\mathbf{I}_{p_1} \otimes \boldsymbol{\Omega}) \mathbf{K}_{\operatorname{vec} \mathbf{C}} \boldsymbol{\theta}, \end{aligned}$$

$$\begin{aligned} P(\nu) &= \mathbf{d}_0^T \boldsymbol{\Omega} \mathbf{d}_0 \\ &= \boldsymbol{\theta}^T \mathbf{K}_{\mathbf{d}_0}^T \boldsymbol{\Omega} \mathbf{K}_{\mathbf{d}_0} \boldsymbol{\theta}, \end{aligned}$$

and

$$\begin{aligned} \sum_{k=1}^{p_2} P(\psi_k) &= \operatorname{tr}(\mathbf{D}^T \boldsymbol{\Omega} \mathbf{D}) \\ &= \boldsymbol{\theta}^T \mathbf{K}_{\operatorname{vec} \mathbf{D}}^T (\mathbf{I}_{p_2} \otimes \boldsymbol{\Omega}) \mathbf{K}_{\operatorname{vec} \mathbf{D}} \boldsymbol{\theta}. \end{aligned}$$

Then

$$\mathsf{D}\mathbf{P}(\boldsymbol{\theta}) = \begin{bmatrix} 2\boldsymbol{\theta}^T \mathbf{K}_{\mathbf{c}_0}^T \boldsymbol{\Omega} \mathbf{K}_{\mathbf{c}_0} \\ 2\boldsymbol{\theta}^T \mathbf{K}_{\operatorname{vec} \mathbf{C}}^T (\mathbf{I}_{p_1} \otimes \boldsymbol{\Omega}) \mathbf{K}_{\operatorname{vec} \mathbf{C}} \\ 2\boldsymbol{\theta}^T \mathbf{K}_{\mathbf{d}_0}^T \boldsymbol{\Omega} \mathbf{K}_{\mathbf{d}_0} \\ 2\boldsymbol{\theta}^T \mathbf{K}_{\operatorname{vec} \mathbf{D}}^T (\mathbf{I}_{p_2} \otimes \boldsymbol{\Omega}) \mathbf{K}_{\operatorname{vec} \mathbf{D}} \end{bmatrix}.$$

0.3 Simulations

For the simulations reported in the paper, we used the following choices of smoothing parameters $\boldsymbol{\xi}_n = (\xi_{n1}, \xi_{n2}, \xi_{n3}, \xi_{n4})$. For five-knot spline bases we took $\boldsymbol{\xi}_{50} = (10^{-4}, 10^{-5}, 10^{-4}, 10^{-4})$, $\boldsymbol{\xi}_{100} = (10^{-5}, 10^{-5}, 10^{-5}, 10^{-5})$, $\boldsymbol{\xi}_{200} = (10^{-7}, 10^{-6}, 10^{-7}, 10^{-7})$ and $\boldsymbol{\xi}_{400} = (10^{-7}, 10^{-6}, 10^{-8}, 10^{-8})$. For ten-knot spline bases we took $\boldsymbol{\xi}_{50} = (10^{-4}, 10^{-5}, 10^{-4}, 10^{-4})$, $\boldsymbol{\xi}_{100} = (10^{-4}, 10^{-5}, 10^{-4}, 10^{-4})$, $\boldsymbol{\xi}_{200} = (10^{-4}, 10^{-5}, 10^{-5}, 10^{-5})$ and $\boldsymbol{\xi}_{400} = (10^{-5}, 10^{-6}, 10^{-6}, 10^{-6})$. The same $\boldsymbol{\xi}$s were used for both rates r.

Estimation errors for ten-knot spline estimators and variance proportion $\alpha = .75$ are reported in Table 1, for five-knot spline estimators and $\alpha = .60$ in Table 2, and for ten-knot spline estimators and $\alpha = .60$ in Table 3.

True finite-sample standard deviations of the elements of $\hat{\boldsymbol{\Sigma}}_{uv}$ along with median and median absolute errors of their asymptotic estimators are given in Table 4 for estimators based on ten-knot splines and models with variance proportion $\alpha = .75$; for five-knot splines and $\alpha = .60$ they are given in Table 5, and for ten-knot splines and $\alpha = .60$ in Table 6.

Figures 1–6 are plots of functional estimators based on five-knot splines, for variance proportion $\alpha = .75$, rates 10 and 30, and sample sizes between 50 to 200. In each figure, the six rows correspond to $\hat{\mu}$, $\hat{\phi}_1$, $\hat{\phi}_2$, $\hat{\nu}$, $\hat{\psi}_1$ and $\hat{\psi}_2$, respectively; the first column shows the 300 simulated estimators, the

second column shows the pointwise mean of the estimators (solid line) and the true functions (dashed line), and the third column shows the pointwise standard deviation of the estimators.

0.4 Application: online auction data

The plots of functional estimators obtained for different smoothing parameters are shown in Figures 7 and 8. They suggest the choices $\xi_1 = \xi_2 = \xi_4 = 10^{-4}$ and $\xi_3 = 10^{-6}$ as reasonable values for the parameters, but other choices are clearly possible since the estimators do not change much for nearby ξs.

Normal probability plots of the estimated component scores $\hat{u}_{ik}$s and $\hat{v}_{ik}$s are shown in Figure 9 and of the residuals $\hat{\eta}_{ij}$ in Figure 10. The latter shows tails somewhat heavier than Gaussian. Figure 11 shows root mean squared errors of the $\hat{\eta}_{ij}$s for each i; no gross outliers are evident.

References

Dempster, A.P., Laird, N.M., and Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society Series B* **39** 1–38.

	$r = 10$			$r = 30$		
Parameter	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\Sigma_{uv,11}$	.054	.031	.023	.037	.026	.018
$\Sigma_{uv,21}$	.056	.038	.023	.027	.017	.012
$\Sigma_{uv,12}$	.036	.023	.014	.021	.014	.010
$\Sigma_{uv,22}$	.022	.018	.013	.014	.009	.006
μ	.122	.087	.068	.097	.077	.063
ν	.126	.098	.086	.163	.144	.136
ϕ_1	.735	.516	.359	.430	.263	.175
ϕ_2	.883	.723	.566	.585	.391	.279
ψ_1	.219	.259	.233	.139	.105	.062
ψ_2	.212	.238	.206	.146	.108	.068
σ_{u1}	.068	.057	.051	.039	.028	.020
σ_{u2}	.067	.070	.060	.034	.024	.018
σ_{v1}	.069	.056	.091	.062	.047	.036
σ_{v2}	.062	.093	.093	.037	.033	.018
σ_η	.057	.084	.076	.013	.021	.010
u_{i1}	.218	.184	.170	.154	.139	.134
u_{i2}	.163	.140	.122	.118	.104	.097
v_{i1}	.167	.157	.149	.168	.151	.143
v_{i2}	.150	.152	.129	.105	.085	.072

Table 1: Simulation Results. Root mean squared errors of estimators based on ten-knot B-splines under different baseline rates r and sampling sizes n, for model with variance proportion $\alpha = .75$.

	$r = 10$			$r = 30$		
Parameter	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\Sigma_{uv,11}$	.063	.041	.037	.047	.032	.020
$\Sigma_{uv,21}$	.058	.050	.046	.041	.035	.028
$\Sigma_{uv,12}$	.057	.046	.039	.039	.032	.025
$\Sigma_{uv,22}$	.044	.030	.024	.030	.021	.012
μ	.123	.100	.089	.093	.081	.070
ν	.113	.095	.081	.145	.130	.126
ϕ_1	.900	.773	.688	.672	.539	.414
ϕ_2	.957	.849	.719	.719	.565	.433
ψ_1	.510	.351	.295	.418	.311	.163
ψ_2	.507	.325	.241	.419	.312	.164
σ_{u1}	.089	.047	.032	.035	.023	.019
σ_{u2}	.042	.031	.026	.030	.022	.016
σ_{v1}	.058	.041	.062	.054	.041	.032
σ_{v2}	.090	.078	.080	.047	.033	.023
σ_η	.068	.063	.069	.012	.011	.011
u_{i1}	.237	.204	.189	.174	.156	.143
u_{i2}	.194	.178	.166	.158	.144	.129
v_{i1}	.259	.193	.177	.229	.190	.149
v_{i2}	.276	.205	.175	.225	.174	.109

Table 2: Simulation Results. Root mean squared errors of estimators based on five-knot B-splines under different baseline rates r and sampling sizes n, for model with variance proportion $\alpha = .60$.

	$r = 10$			$r = 30$		
Parameter	$n = 50$	$n = 100$	$n = 200$	$n = 50$	$n = 100$	$n = 200$
$\Sigma_{uv,11}$	.061	.041	.036	.047	.033	.020
$\Sigma_{uv,21}$	.059	.051	.046	.042	.035	.028
$\Sigma_{uv,12}$	.055	.046	.040	.040	.032	.025
$\Sigma_{uv,22}$	.042	.030	.025	.030	.021	.013
μ	.120	.087	.065	.095	.076	.062
ν	.114	.092	.077	.145	.129	.126
ϕ_1	.902	.775	.707	.678	.543	.416
ϕ_2	.958	.848	.746	.724	.568	.438
ψ_1	.507	.349	.296	.418	.312	.163
ψ_2	.494	.326	.250	.420	.313	.164
σ_{u1}	.072	.048	.030	.036	.024	.019
σ_{u2}	.043	.032	.027	.029	.021	.015
σ_{v1}	.058	.042	.034	.055	.041	.032
σ_{v2}	.090	.079	.084	.047	.033	.023
σ_η	.069	.068	.072	.013	.011	.011
u_{i1}	.230	.204	.191	.175	.156	.143
u_{i2}	.194	.179	.169	.158	.144	.130
v_{i1}	.258	.189	.167	.230	.190	.149
v_{i2}	.275	.205	.176	.225	.174	.109

Table 3: Simulation Results. Root mean squared errors of estimators based on ten-knot B-splines under different baseline rates r and sampling sizes n, for model with variance proportion $\alpha = .60$.

	$n = 100$			$n = 200$			$n = 400$		
Parameter	True	Med	MAE	True	Med	MAE	True	Med	MAE
				$r = 10$					
$\Sigma_{uv,11}$	.31	1.35	1.04	.23	.33	.10	.17	.18	.01
$\Sigma_{uv,21}$	.38	1.49	1.11	.23	.41	.17	.15	.22	.07
$\Sigma_{uv,12}$	.23	.87	.63	.17	.25	.08	.10	.13	.03
$\Sigma_{uv,22}$	.18	.64	.46	.12	.17	.05	.13	.09	.04
				$r = 30$					
$\Sigma_{uv,11}$	.25	.87	.62	.18	.26	.07	.12	.14	.02
$\Sigma_{uv,21}$	.17	.65	.47	.12	.18	.07	.08	.10	.03
$\Sigma_{uv,12}$	.14	.49	.35	.10	.14	.04	.06	.08	.02
$\Sigma_{uv,22}$	.09	.36	.27	.06	.11	.04	.04	.06	.01

Table 4: Simulation Results. True standard deviations and median and median absolute errors of estimated asymptotic standard deviations ($\times 10$) of estimators under different baseline rates r and sampling sizes n, for estimators based on ten-knot B-splines and variance proportion $\alpha = .75$

	$r = 10$								
	$n = 100$			$n = 200$			$n = 400$		
Parameter	True	Med	MAE	True	Med	MAE	True	Med	MAE
$\Sigma_{uv,11}$	.39	.77	.38	.36	.35	.11	.23	.19	.07
$\Sigma_{uv,21}$	.50	.77	.27	.46	.48	.14	.36	.32	.11
$\Sigma_{uv,12}$	.46	.65	.20	.39	.39	.12	.32	.26	.10
$\Sigma_{uv,22}$	.23	.46	.23	.20	.23	.05	.18	.13	.06
	$r = 30$								
$\Sigma_{uv,11}$	.32	.45	.13	.20	.21	.02	.12	.12	.01
$\Sigma_{uv,21}$	.35	.54	.19	.28	.29	.07	.18	.18	.03
$\Sigma_{uv,12}$	.32	.49	.17	.25	.26	.06	.17	.17	.03
$\Sigma_{uv,22}$	.18	.26	.08	.11	.13	.02	.07	.08	.01

Table 5: Simulation Results. True standard deviations and median and median absolute errors of estimated asymptotic standard deviations ($\times 10$) of estimators under different baseline rates r and sampling sizes n, for estimators based on five-knot B-splines and variance proportion $\alpha = .60$

	$r = 10$								
	$n = 100$			$n = 200$			$n = 400$		
Parameter	True	Med	MAE	True	Med	MAE	True	Med	MAE
$\Sigma_{uv,11}$	.39	1.64	1.24	.33	.44	.11	.23	.21	.06
$\Sigma_{uv,21}$	.51	1.56	1.05	.46	.53	.15	.36	.34	.10
$\Sigma_{uv,12}$	.46	1.35	.89	.40	.45	.13	.32	.28	.10
$\Sigma_{uv,22}$	.23	.99	.76	.20	.27	.08	.18	.14	.06
	$r = 30$								
$\Sigma_{uv,11}$	.32	.94	.62	.20	.25	.05	.12	.13	.01
$\Sigma_{uv,21}$	.35	1.10	.74	.28	.34	.07	.18	.20	.04
$\Sigma_{uv,12}$	.32	.98	.66	.25	.31	.07	.17	.18	.03
$\Sigma_{uv,22}$	.18	.53	.35	.11	.15	.04	.07	.08	.01

Table 6: Simulation Results. True standard deviations and median and median absolute errors of estimated asymptotic standard deviations ($\times 10$) of estimators under different baseline rates r and sampling sizes n, for estimators based on ten-knot B-splines and variance proportion $\alpha = .60$

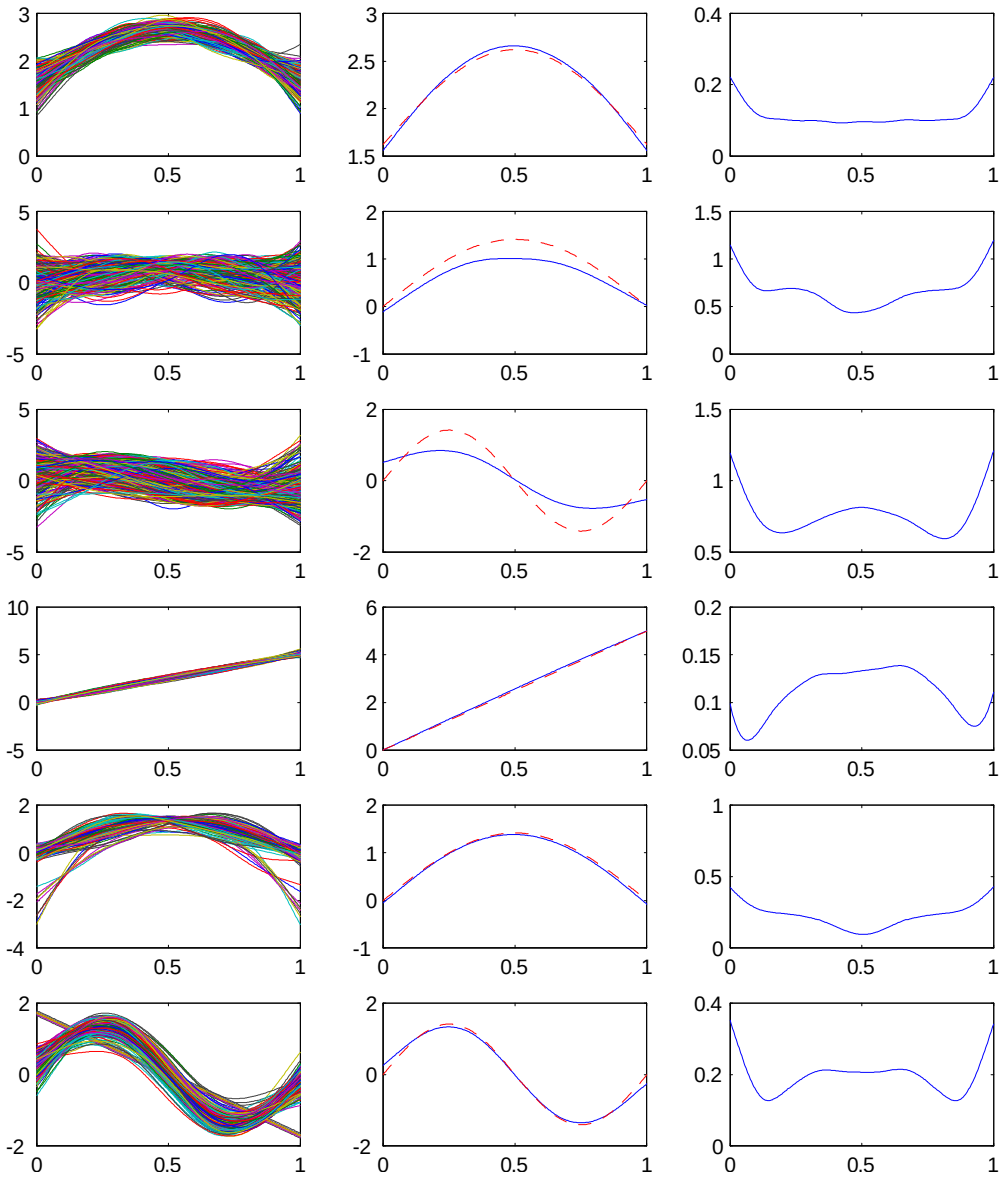

Figure 1: Simulation Results. Plots for $r = 10$ and $n = 50$.

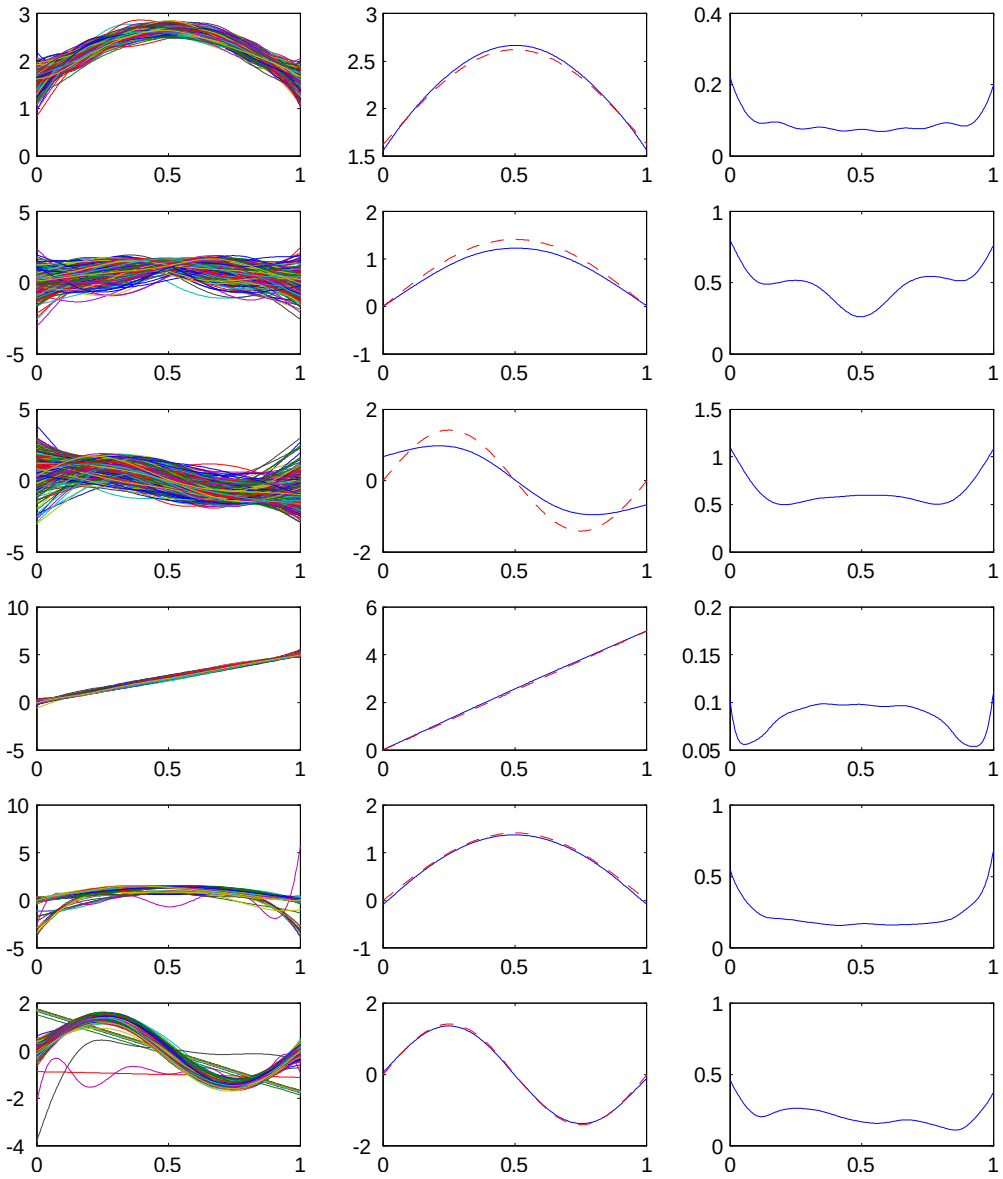

Figure 2: Simulation Results. Plots for $r = 10$ and $n = 100$.

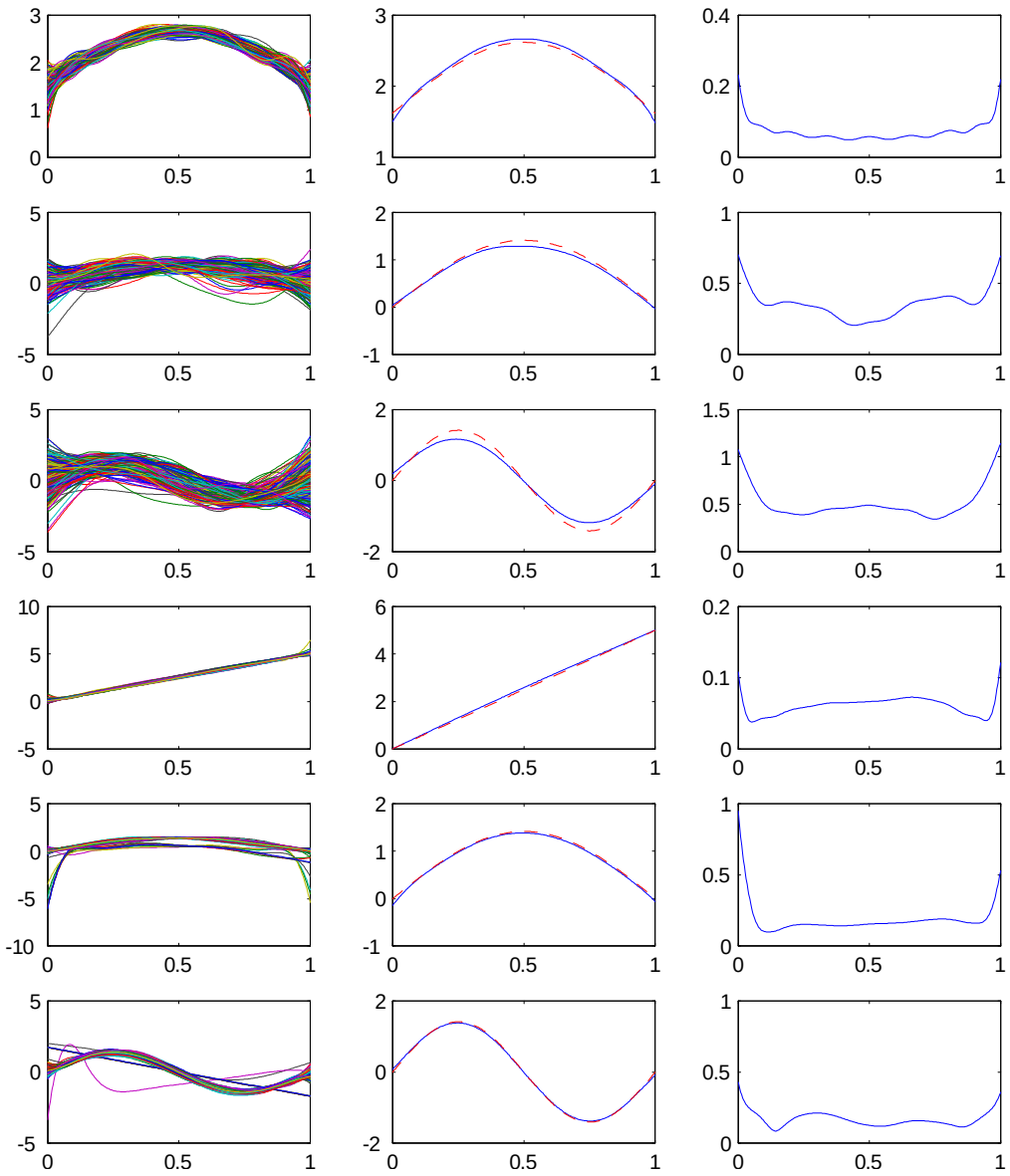

Figure 3: Simulation Results. Plots for $r = 10$ and $n = 200$.

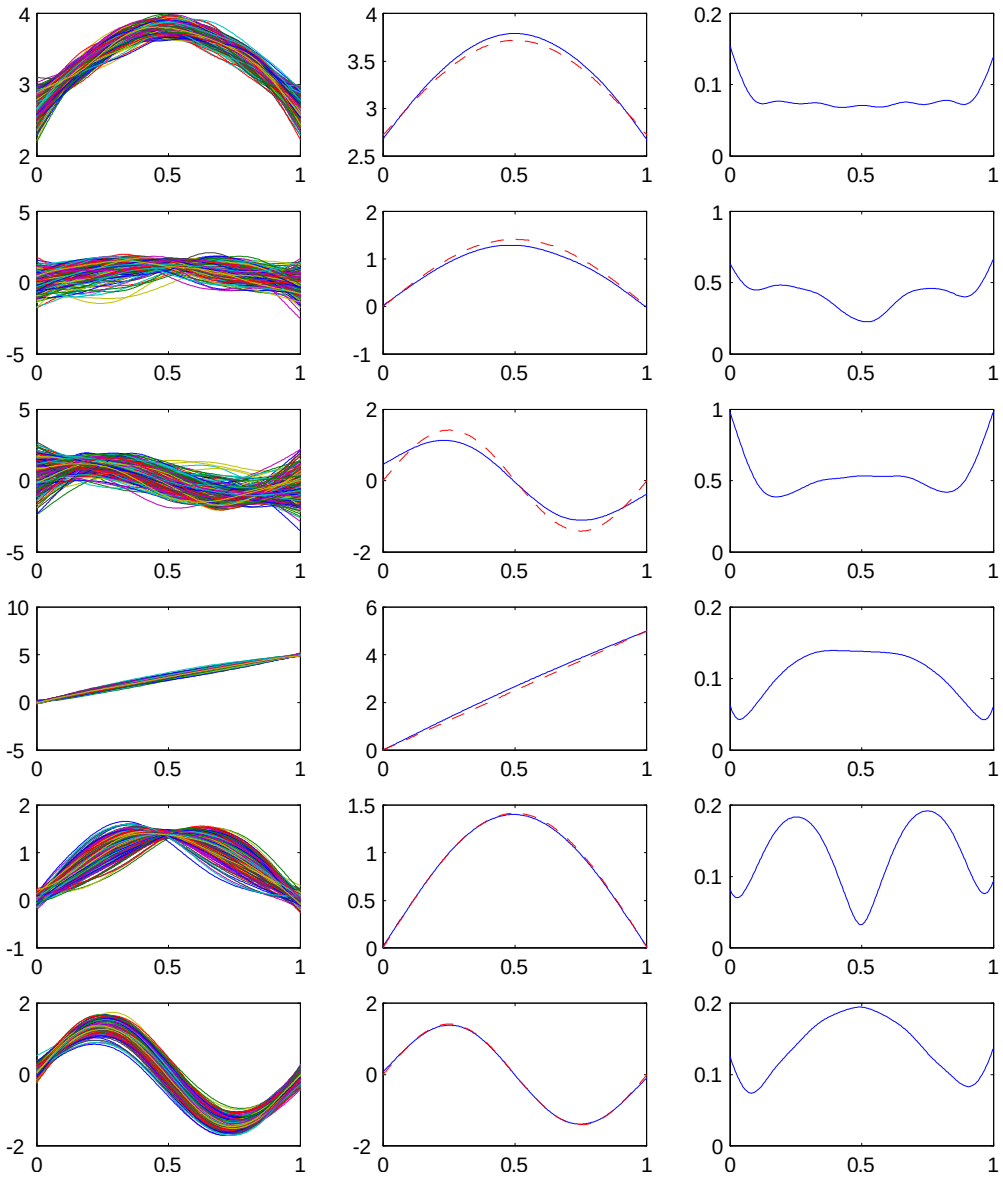

Figure 4: Simulation Results. Plots for $r = 30$ and $n = 50$.

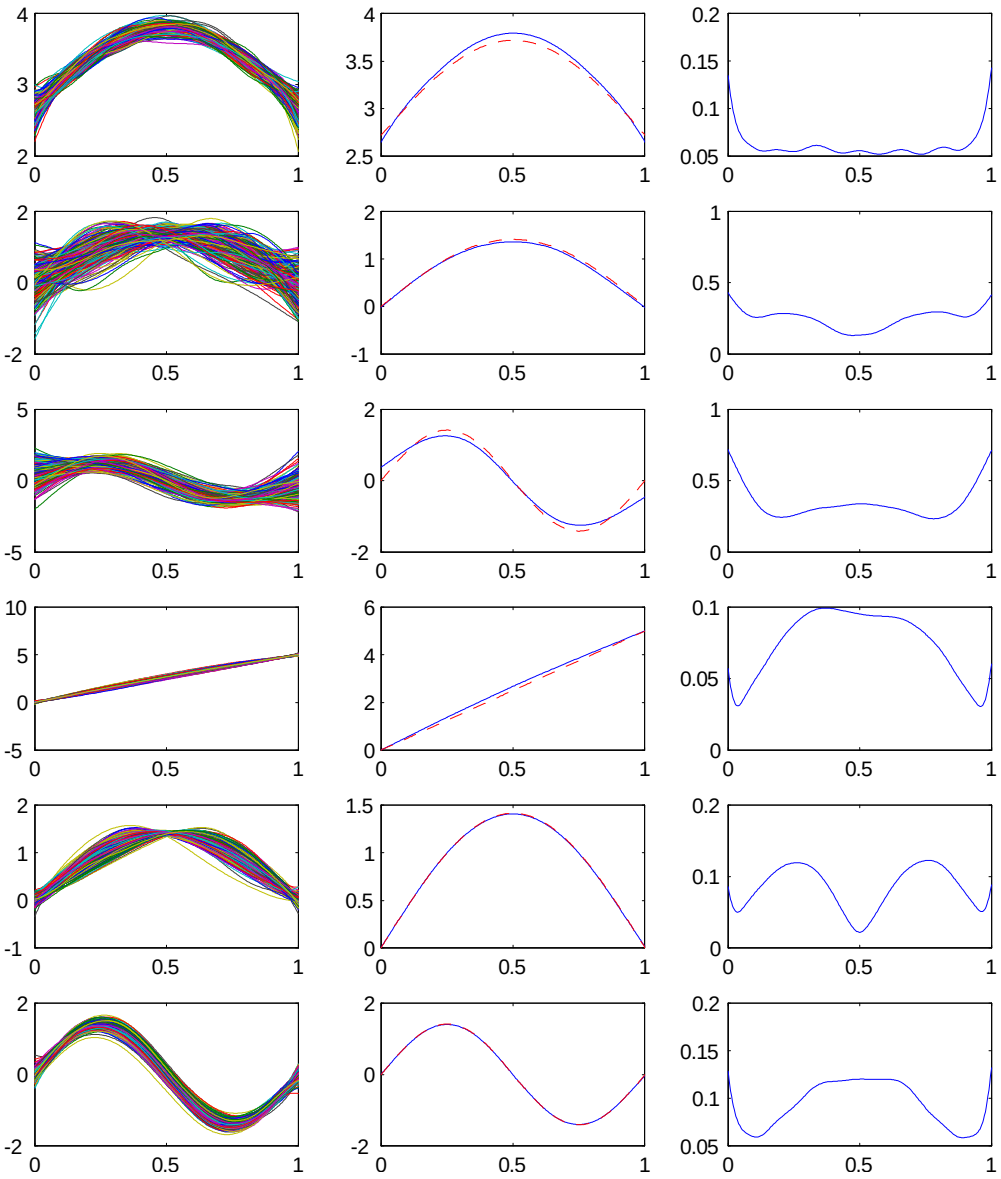

Figure 5: Simulation Results. Plots for $r = 30$ and $n = 100$.

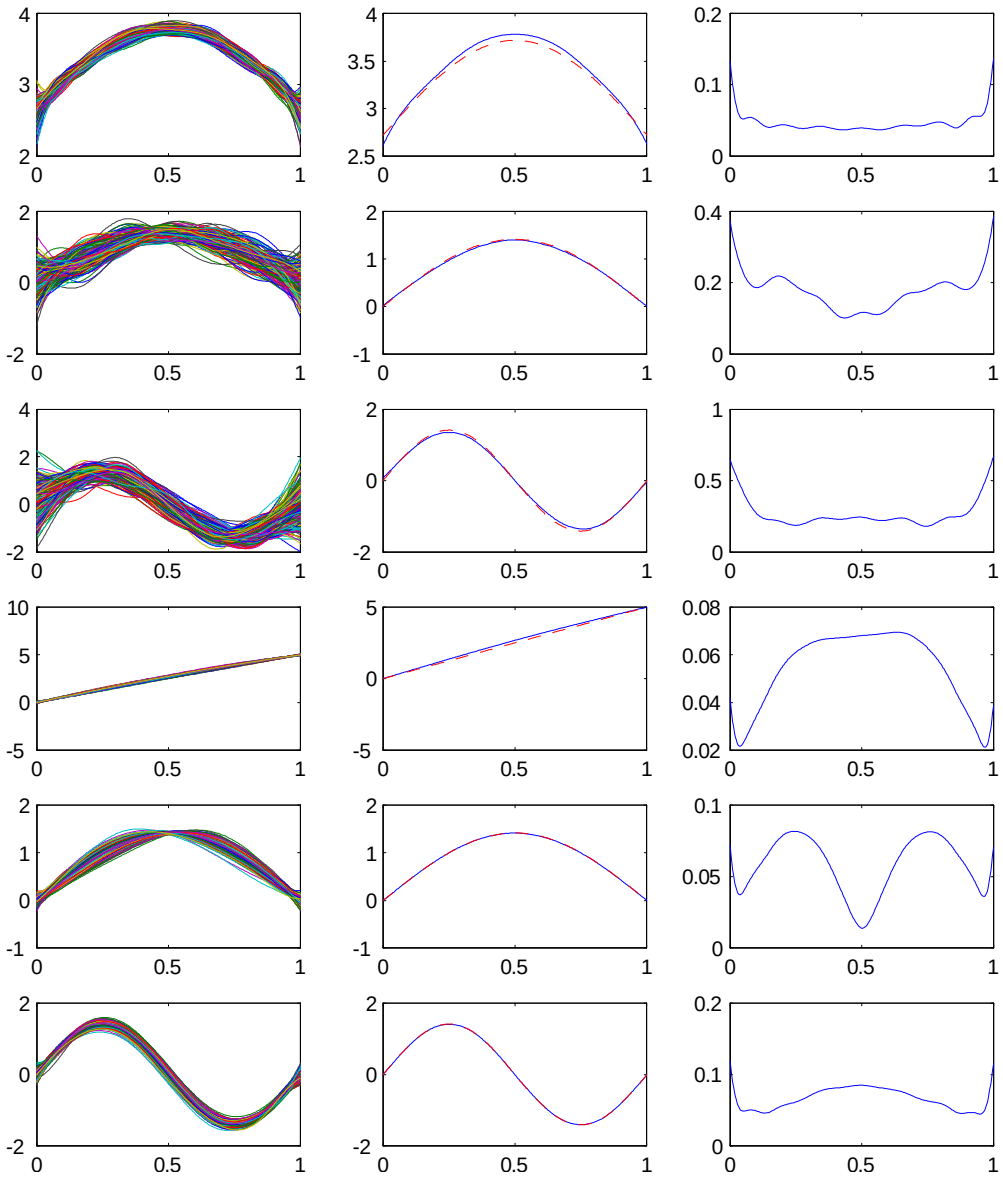

Figure 6: Simulation Results. Plots for $r = 30$ and $n = 200$.

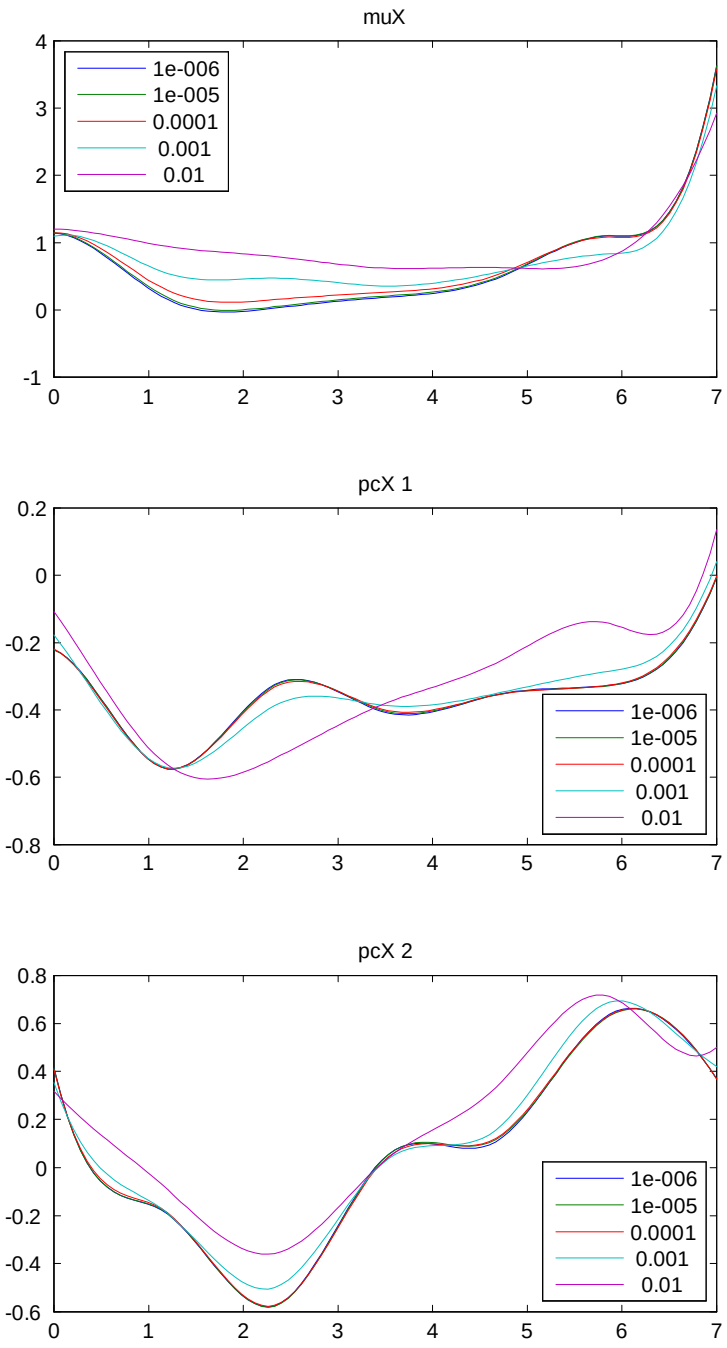

Figure 7: Online Auction Data. Estimators of mean and components of X process for various smoothing parameters ξ.

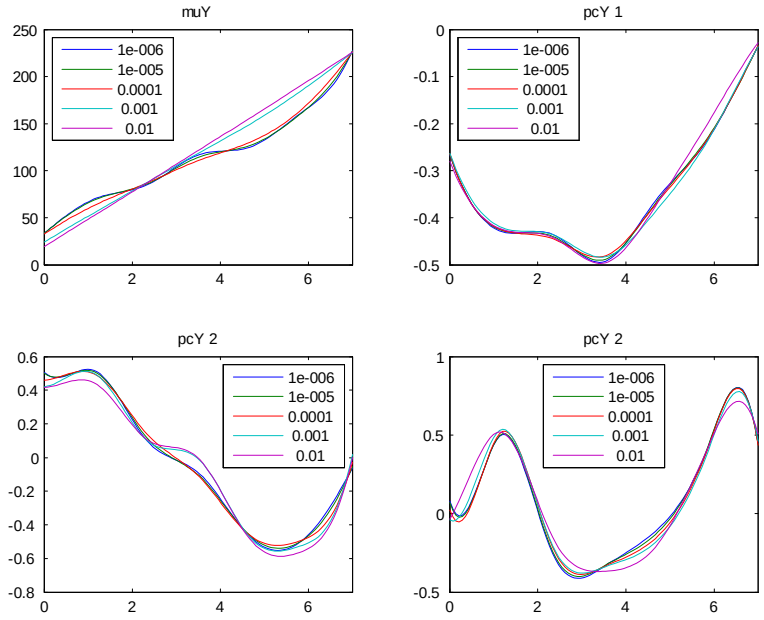

Figure 8: Online Auction Data. Estimators of mean and components of Y process for various smoothing parameters ξ.

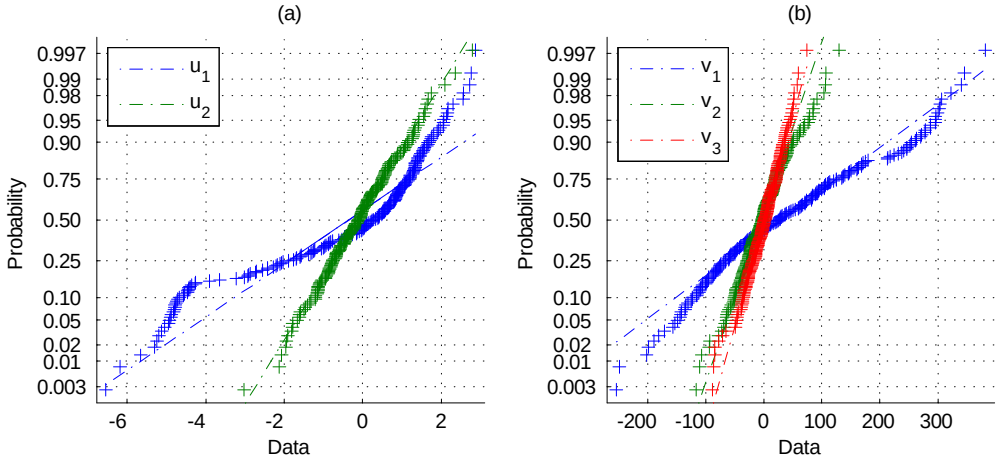

Figure 9: Online Auction Data. Normal probability plots of the estimated component scores of (a) the bid-time process and (b) the bid-price process.

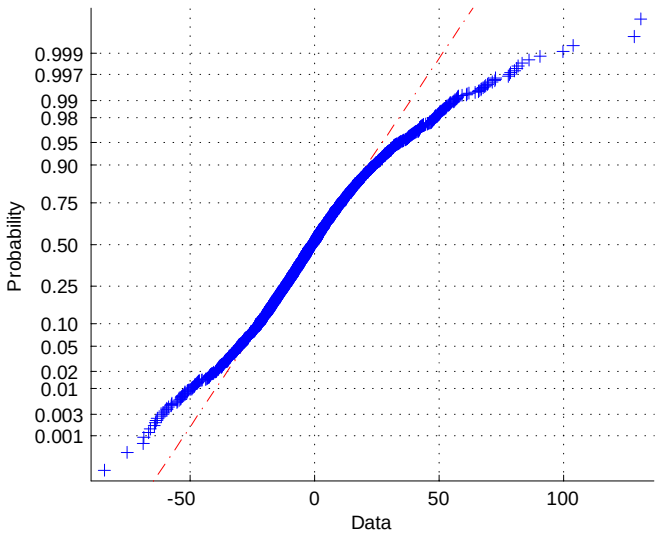

Figure 10: Online Auction Data. Normal probability plot of the residuals of fitted bid prices.

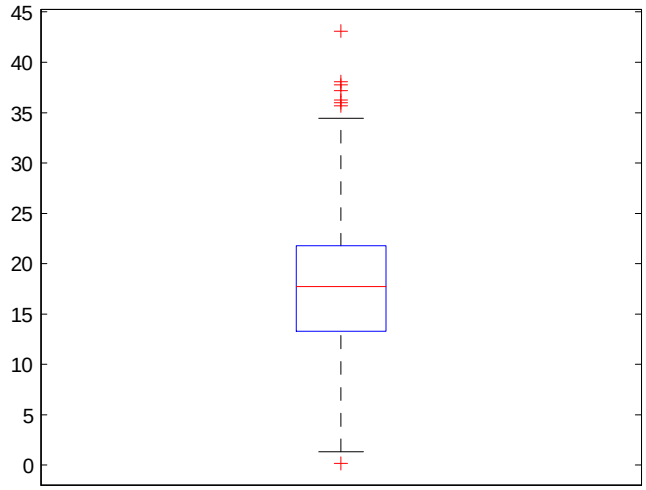

Figure 11: Online Auction Data. Boxplot of root mean squared errors of individual bid price trajectories.

Gervini, D. (2017). Multiplicative component models for replicated point processes. *ArXiv* **1705.09693**.

Geyer, C.J. (1994). On the asymptotics of constrained M-estimation. *The Annals of Statistics* **22** 1993–2010.

James, G., Hastie, T., and Sugar, C. (2000). Principal component models for sparse functional data. *Biometrika* **87** 587–602.

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

Rockafellar, R.T., and Wets, R.J. (1998). *Variational Analysis.* Springer, New York.

Van der Vaart, A. (2000). *Asymptotic Statistics.* Cambridge University Press, Cambridge, UK.