Nonparametric Covariance Estimation for Longitudinal Data

Tayler A. Blake

The Ohio State University, Department of Statistics

April 30, 2018

Background

Let

$$Y = (y_1, \ldots, y_p)', \quad (t_1, \ldots, t_p)'$$

denote the random vector of observations and their associated measurement times, where

$$Cov\,(Y) = \Sigma = [\sigma_{ij}]$$

- Dimensionality: the number of parameters is quadratic in p.

Background

Let

$$Y = (y_1, \ldots, y_p)', \quad (t_1, \ldots, t_p)'$$

denote the random vector of observations and their associated measurement times, where

$$Cov\,(Y) = \Sigma = [\sigma_{ij}]$$

- Dimensionality: the number of parameters is quadratic in p.
- Covariance estimates should satisfy the positive definite constraint:

$$c'\Sigma c = \sum_{i,j=1}^{p} c_i c_j \sigma_{ij} \geq 0$$

Background

Let

$$Y = (y_1, \ldots, y_p)', \quad (t_1, \ldots, t_p)'$$

denote the random vector of observations and their associated measurement times, where

$$Cov\,(Y) = \Sigma = [\sigma_{ij}]$$

- Dimensionality: the number of parameters is quadratic in p.
- Covariance estimates should satisfy the positive definite constraint:

$$c'\Sigma c = \sum_{i,j=1}^{p} c_i c_j \sigma_{ij} \geq 0$$

- Observation times may be irregular or subject-specific.

Outline

- Background: A Review of Covariance Estimation for Longitudinal Data
    - The Cholesky Decomposition
- A Reproducing Kernel Hilbert Space Framework for Covariance Estimation
- Tensor Product P-splines for Covariance Estimation
- Simulation Studies
- Data Analysis: Cattle Weights

Common approaches to covariance estimation

low bias	high bias
high variance	low variance

S

- Improvements on S
  - Stein's estimator, Ledoit et. al. (2000)
- (non)stationary banded models
- linear, log linear models

Parametric models

The Modified Cholesky decomposition

For any positive definite Σ, there exists a unique lower-triangular matrix $C = [c_{ij}]$, $c_{ii} > 0$:

$$\Sigma = CC',$$

Let $D^{1/2} = diag\,(c_{11}, \ldots, c_{pp})$, $L = D^{-1/2}C$, then

$$\Sigma = LDL'.$$

The **modified Cholesky decomposition** (MCD) of Σ is given by

$$D = T\Sigma T', \tag{1}$$

where $T = L^{-1}$. The lower triangular entries of T are *unconstrained*.

Statistical Interpretation of (T, D)

Let $\hat{y}_t$ be the linear least-squares predictor of y_t based on previous measurements $y_{t-1}, \dots, y_1$ and $\epsilon_t = y_t - \hat{y}_t$ denote the corresponding mean zero prediction error with variance $Var\left(\epsilon_t\right) = \sigma_t^2$. We can find unique scalars ϕ_{tj}:

$$y_t = \begin{cases} \epsilon_t, & t = 1 \\ \sum_{j=1}^{t-1} \phi_{tj} y_j + \epsilon_t, & t = 2, \dots, p, \end{cases}$$

where $D = Cov\left(\epsilon\right) = diag\left(\sigma_1^2, \dots, \sigma_p^2,\right)$. Then

$$\underbrace{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_{p-1} \\ \epsilon_p \end{bmatrix}}_{\epsilon} = \underbrace{\begin{bmatrix} 1 & & & & \\ -\phi_{21} & 1 & & & \\ -\phi_{31} & -\phi_{32} & 1 & & \\ \vdots & & & \ddots & \\ -\phi_{p1} & -\phi_{p2} & \dots & -\phi_{p,p-1} & 1 \end{bmatrix}}_{T} \underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{p-1} \\ y_p \end{bmatrix}}_{Y}$$

Taking the covariance on both sides gives the MCD (1).

The coefficients and prediction error variances of successive regressions are unconstrained.

The **generalized autoregressive parameters** ϕ_{tj} and **log innovation variances** $\log \sigma_j^2$ are unconstrained but

$$\hat{\Sigma}^{-1} = \hat{T}'\hat{D}^{-1}T$$

is guaranteed to be positive definite.

$$\begin{array}{cccccc}
y_1 & y_2 & y_3 & \cdots & y_{p-1} & y_p \\
\hline
1 & & & & & \\
\phi_{21} & 1 & & & & \\
\phi_{31} & \phi_{32} & 1 & & & \\
\vdots & \vdots & & \ddots & & \\
\vdots & \vdots & & & \ddots & \\
\phi_{p1} & \phi_{p2} & \cdots & \cdots & \phi_{p,p-1} & 1 \\
\hline
\sigma_1^2 & \sigma_2^2 & \cdots & \cdots & \sigma_{p-1}^2 & \sigma_p^2
\end{array}$$

The ideal form of longitudinal data:

		Time					
		1	2	$\dots$	t	$\dots$	p
Unit	1	y_{11}	y_{12}	$\cdots$	y_{1t}	$\cdots$	y_{1p}
	2	y_{21}	y_{22}	$\cdots$	y_{2t}	$\cdots$	y_{2p}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	i	y_{i1}	y_{i2}	$\cdots$	y_{it}	$\cdots$	y_{ip}
	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$
	N	y_{N1}	y_{N2}	$\cdots$	y_{Nt}	$\cdots$	y_{Np}

Maximum Normal Likelihood Estimation for the Cholesky Decomposition

The MLE for (T, D) has closed form under the Gaussian likelihood.

For $Y_1, \ldots, Y_N \sim N(0_p, \Sigma)$ and $S = N^{-1} \sum_{i=1}^{N} Y_i Y_i'$,

$$\begin{aligned} -2\ell(\Sigma | Y_1, \ldots, Y_N) &= \sum_{i=1}^{N} \left(\log|\Sigma| + Y_i' \Sigma^{-1} Y_i \right) \\ &= N \log|D| + N \mathrm{tr}\left(D^{-1} T S T' \right) \end{aligned} \tag{2}$$

is quadratic in T for fixed D, so the MLE for the ϕ_{tj} has closed form. Similarly, the MLE for D for fixed T has closed form.

Parametric Models for the Cholesky Decomposition

Pourahmadi (2000), Pan andMackenzie (2003) suggest modeling ϕ_{tj}, σ_t^2 with covariates, letting

$$\phi_{jk} = x'_{jk}\gamma$$
$$\log \sigma_j^2 = z'_j\lambda.$$

Common choices for the covariates x_{jk} and z_j are

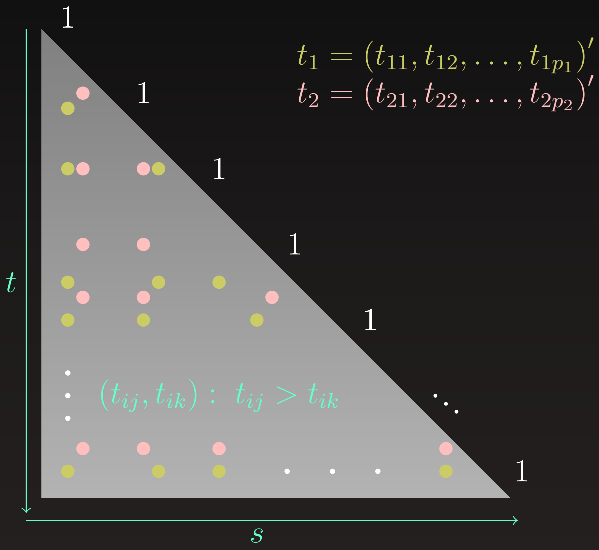
$$x'_{jk} = \left(1, t_j - t_k, (t_j - t_k)^2, \ldots, (t_j - t_k)^{d-1}\right)',$$
$$z'_j = \left(1, t_j, \ldots, t_j^{q-1}\right)'.$$

Polynomial orders d and q are tuning parameters chosen by a model selection criterion (BIC, AIC).

Simulations

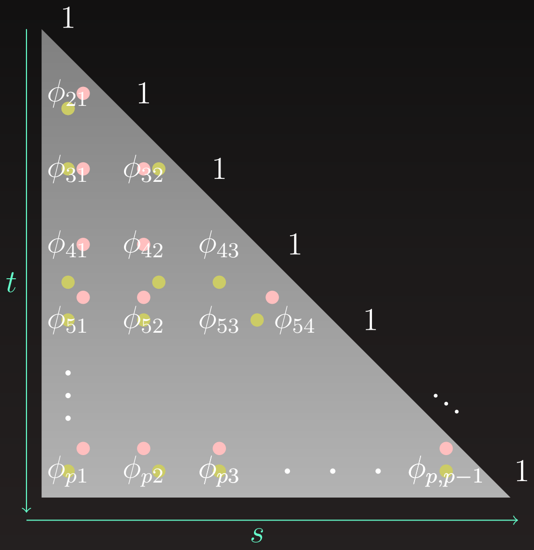
Accommodating Unbalanced Data

Subjects 1 and 2 have different measurement times.



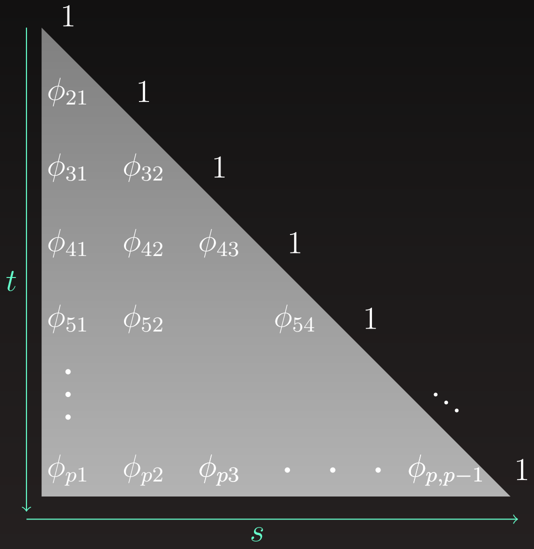
Accommodating Unbalanced Data

Subjects 1 and 2 have different measurement times.



Accommodating Unbalanced Data

by treating ϕ as a continuous bivariate function.



Accommodating Unbalanced Data

by treating ϕ as a continuous bivariate function.

$$
\begin{array}{ccccccc}
1 & & & & & & \\
\phi_{21} & 1 & & & & & \\
\phi_{31} & \phi_{32} & 1 & & & & \\
\phi_{41} & \phi_{42} & \phi_{43} & 1 & & & \\
\phi_{51} & \phi_{52} & \bullet & \phi_{54} & 1 & & \\
\vdots & & & & & \ddots & \\
\phi_{p1} & \phi_{p2} & \phi_{p3} & \cdots & & \phi_{p,p-1} & 1
\end{array}
$$

t (vertical axis), s (horizontal axis)

$$\phi_{53} = \tilde{\phi}\left(t_5, t_3\right)$$

In general, $\phi_{ts} = \tilde{\phi}\left(t, s\right), \quad 0 \leq s < t \leq 1$

$$\sigma_t^2 = \sigma^2\left(t\right), \quad 0 \leq t \leq 1$$

A functional varying coefficient model for ϕ

Assume measurements $Y_i = (y_{i1}, \dots, y_{i,p_i})'$ arise from $Y(t)$ observed at

$$t_i = \{t_{i1} < \cdots < t_{i,p_i}\} \subset \mathcal{T} = [0,1]$$

$$\boxed{y(t_{ij}) = \sum_{k<j} \tilde{\phi}(t_{ij}, t_{ik})\, y(t_{ik}) + \epsilon(t_{ij}), \qquad \begin{array}{l} i = 1, \dots, N \\ j = 2, \dots, p_i, \end{array}}$$

where $\epsilon(t) \sim N\left(0, \sigma^2(t)\right)$. Transform $l = t - s, \quad m = \frac{t+s}{2}$, let

$$\phi(l, m) = \phi\left(t - s, \frac{1}{2}(s+t)\right) = \tilde{\phi}(t, s),$$

so that

$$-2\ell\left(\phi, \sigma^2 | Y_1, \dots, Y_N\right) = \sum_{i=1}^{N} \sum_{j=2}^{p_i} \left[\log \sigma_{ij}^2 + \frac{1}{\sigma_{ij}^2}\left(y_{ij} - \sum_{k<j} \tilde{\phi}(t_{ij}, t_{ik})\, y_{ik}\right)^2\right]$$

The Smoothing Spline Model Space

$J(f)$ induces an orthogonal decomposition of $\mathcal{H}$:

$$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}_1$$

$$\begin{aligned} \mathcal{H}_0 &= \{f : \ J(f) = 0\} \text{ is the null space of } J, \\ \mathcal{H}_1 &= \left\{f \in \mathcal{H} : ||f||^2 = J(f)\right\} \end{aligned}$$

For the cubic smoothing spline defined on $\chi = [0, 1]$,

$$J(f) = \int_0^1 \left(f''(x)\right)^2 \ dx,$$

$$\mathcal{H} = C^{(2)}[0,1] = \left\{ f : \int_0^1 \left(f''(x)\right)^2 \ dx < \infty \right\}$$

with inner product

$$\langle f, g \rangle_{\mathcal{H}} = (M_0 f)(M_0 g) + (M_1 f)(M_1 g) + \int_0^1 f''(x)\, g''(x) \ dx$$

where $M_i f = \int_0^1 f^{(i)}(x) \ dx$.

ANOVA Decomposition of $\mathcal{H}$

$$\begin{aligned}\mathcal{H} &= \mathcal{H}_0 \oplus \mathcal{H}_1 \\ &= \underbrace{\mathcal{H}_{00}}_{\text{mean}} \oplus \underbrace{\mathcal{H}_{01}}_{\text{parametric contrast}} \oplus \underbrace{\mathcal{H}_1}_{\text{nonparametric constrast}}\end{aligned}$$

$$\mathcal{H}_0 = \mathcal{H}_{00} \oplus \mathcal{H}_{01} = \{f:\ f \propto 1\} \oplus \{f:\ f \propto k_1\}$$

$$\mathcal{H}_1 = \{f : M_0 f = M_1 f = 0,\ \ \int_0^1 \left(f''(x)\right)^2\ dx < \infty\}.$$

With reproducing kernel $K = K_{00} + K_{01} + K_1$:

$$\begin{aligned}K_{00}(x,y) &= 1, \\ K_{01}(x,y) &= k_1(x)\, k_1(y)\,, \text{ and} \\ K_1(x,y) &= k_2(x)\, k_2(y) - k_4(x-y)\,.\end{aligned}$$

where k_1, k_2, k_4 are the first, second, and fourth scaled Bernoulli polynomials.

The Tensor Product Smoothing Spline Space

Tensor Product Cubic Spline

	$\mathcal{H}_{00[2]}$	$\mathcal{H}_{01[2]}$	$\mathcal{H}_{1[2]}$
$\mathcal{H}_{00[1]}$	$\mathcal{H}_{00[1]} \otimes \mathcal{H}_{00[2]}$	$\mathcal{H}_{00[1]} \otimes \mathcal{H}_{01[2]}$	$\mathcal{H}_{00[1]} \otimes \mathcal{H}_{1[2]}$
$\mathcal{H}_{01[1]}$	$\mathcal{H}_{01[1]} \otimes \mathcal{H}_{00[2]}$	$\mathcal{H}_{01[1]} \otimes \mathcal{H}_{01[2]}$	$\mathcal{H}_{01[1]} \otimes \mathcal{H}_{1[2]}$
$\mathcal{H}_{1[1]}$	$\mathcal{H}_{1[1]} \otimes \mathcal{H}_{00[2]}$	$\mathcal{H}_{1[1]} \otimes \mathcal{H}_{01[2]}$	$\mathcal{H}_{1[1]} \otimes \mathcal{H}_{1[2]}$

	$\{1\}$	$\{k_1\}$	$\mathcal{H}_{1[2]}$
$\{1\}$	mean	p-main effect	np-main effect
$\{k_1\}$	p-main effect	$p \times p$-interaction	$p \times np$-interaction
$\mathcal{H}_{1[1]}$	np-main effect	$np \times p$-interaction	$np \times np$-interaction

A Reproducing Kernel Hilbert Space Framework for ϕ

Let

$$\begin{aligned}\phi \in \mathcal{H} &= \mathcal{H}_{[l]} \otimes \mathcal{H}_{[m]} \\ &= \mathcal{H}_0 \oplus \mathcal{H}_1.\end{aligned}$$

with RK $K = K_{[l]} K_{[m]}$. For fixed $\sigma_{ij}^2 = \sigma^2(t_{ij})$, find ϕ minimizing

$$-2\ell\left(\phi | Y_1, \ldots, Y_N, \sigma^2\right) + \lambda J(\phi) = \sum_{i=1}^{N} \sum_{j=2}^{p_i} \frac{1}{\sigma_{ij}^2} \left(y_{ij} - \sum_{k<j} \phi\left(\boldsymbol{v}_{ijk}\right) y_{ik} \right)^2 + \lambda J(\phi) \tag{3}$$

where $J(\phi) = ||P_1 \phi||^2$ and $\boldsymbol{v}_{ijk} \in \mathcal{V} = [0,1]^2$,

$$\begin{aligned}\boldsymbol{v}_{ijk} &= \left(t_{ij} - t_{ik}, \frac{1}{2}\left(t_{ij} + t_{ik}\right)\right) \\ &= \left(l_{ijk}, m_{ijk}\right)\end{aligned}$$

Define the set of unique within-subject pairs of observation times:

$$V \equiv \bigcup_{i,j,k} \left\{\boldsymbol{v}_{ijk}\right\} = \left\{\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{|V|}\right\}$$

A Representer Theorem

Theorem

Let $\{\nu_1, \ldots, \nu_{\mathcal{N}_0}\}$ *span* $\mathcal{H}_0$*, the null space of* $J(\phi) = ||P_1\phi||^2$*. Let* B *denote the* $|V| \times \mathcal{N}_0$ *matrix having* i^{th} *column equal to* ν_i *evaluated at the observed* $\boldsymbol{v} \in V$*, and assume that* B *has full column rank. Then the minimizer* ϕ_λ *of* (3) *is given by*

$$\phi_\lambda(\boldsymbol{v}) = \sum_{i=1}^{\mathcal{N}_0} d_i \nu_i(\boldsymbol{v}) + \sum_{j=1}^{|V|} c_j K_1(\boldsymbol{v}_j, \boldsymbol{v}), \tag{4}$$

where $K_1(\boldsymbol{v}_j, \boldsymbol{v})$ *denotes the reproducing kernel for* $\mathcal{H}_1$ *evaluated at* $\boldsymbol{v}_j$*, the* j^{th} *element of* V.

Obtaining the solution ϕ_λ

By the Representer Theorem, (3) becomes

$$-2\ell\left(c,d|\tilde{Y},\tilde{B},\tilde{K}_V\right)+\lambda J(\phi)=\left[\tilde{Y}-\tilde{B}d-\tilde{K}_V c\right]'\left[\tilde{Y}-\tilde{B}d-\tilde{K}_V c\right]+\lambda c' K_V c,$$

where

$$\begin{aligned}
Y &= (Y_1', Y_2', \ldots, Y_N')' = (y_{12}, y_{13}, \ldots, y_{1p_1}, \ldots, y_{N2}, \ldots, y_{Np_N})' \\
D &= diag\left(\sigma_{12}^2, \sigma_{13}^2, \ldots, \sigma_{1p_1}^2, \ldots, \sigma_{N2}^2, \ldots, \sigma_{Np_N}^2\right) \\
X_i &= (p_i - 1) \times |V| \text{ matrix of AR covariates for Subject } i \\
K_V &= |V| \times |V| \text{ matrix with } (i,j) \text{ element } K_1(\boldsymbol{v}_i, \boldsymbol{v}_j) \\
B &= |V| \times \mathcal{N}_0 \text{ matrix with } (i,j) \text{ element } \nu_j(\boldsymbol{v}_i)
\end{aligned}$$

and $\tilde{Y} = D^{-1/2}Y$, $\tilde{B} = D^{-1/2}XB$, $\tilde{K}_V = D^{-1/2}XK_V$,
$X = \begin{bmatrix} X_1' & X_2' & \ldots & X_N' \end{bmatrix}'$

Obtaining the solution ϕ_λ

Setting derivatives equal to zero, for fixed λ, c and d satisfy

$$\begin{bmatrix} \tilde{B}'\tilde{Y} \\ \tilde{K}_V'\tilde{Y} \end{bmatrix} = \underbrace{\begin{bmatrix} \tilde{B}'\tilde{B} & \tilde{B}'\tilde{K}_V \\ \tilde{K}_V'\tilde{B} & \tilde{K}_V'\tilde{K}_V + \lambda K_V \end{bmatrix}}_{C'C} \begin{bmatrix} d \\ c \end{bmatrix}$$

$$\Longrightarrow \begin{bmatrix} \hat{d} \\ \hat{c} \end{bmatrix} = C^{-1}(C')^{-1} \begin{bmatrix} \tilde{B}' \\ \tilde{K}_V' \end{bmatrix} \tilde{Y}.$$

The fitted values are given by $\widehat{\tilde{Y}} = \tilde{A}_{\lambda,\boldsymbol{\theta}}\tilde{Y}$, where the smoothing matrix is

$$\tilde{A}_{\lambda,\boldsymbol{\theta}} = \begin{bmatrix} \tilde{B} & \tilde{K}_V \end{bmatrix} C^{-1}(C')^{-1} \begin{bmatrix} \tilde{B}' \\ \tilde{K}_V' \end{bmatrix}$$

Smoothing Parameter Selection

- Let $\mu_{ij} = E\left[y_{ij}|y_{i1}, \ldots, y_{i,j-1}\right] = \sum_{k<j} \phi\left(\boldsymbol{v}_{ijk}\right) y_{ik}$.

$$U\left(\lambda\right) = \tilde{Y}'\left(I - \tilde{A}_{\lambda,\boldsymbol{\theta}}\right)^2 \tilde{Y} + 2\,\mathrm{tr}\left(\tilde{A}_{\lambda,\boldsymbol{\theta}}\right)$$

  is an unbiased estimator of the risk $E\left[L\left(\lambda\right)\right] = E\left[\sum_{i,j} \frac{1}{\sigma_{ij}^2}\left(\hat{y}_{ij}^{\lambda} - \mu_{ij}\right)^2\right]$.

- Let $\widehat{\tilde{\mu}}_i^{[-i]}$ denote the estimate of $E\left[\tilde{Y}_i|X_i\right]$ based on the data when $\tilde{Y}_i$ is omitted. The leave-one-subject-out cross validation score

$$V_{loso}\left(\lambda\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Y}_i - \widehat{\tilde{\mu}}_i^{[-i]}\right)'\left(\tilde{Y}_i - \widehat{\tilde{\mu}}_i^{[-i]}\right)$$

  approximates $MSPE = \frac{1}{N}\sum_{i=1}^{N} E\left[||\tilde{Y}_i^* - \widehat{\tilde{\mu}}_i||^2\right]$, where $\tilde{Y}_i^*$ denotes a vector of new observations $\tilde{y}_{i1}^*, \tilde{y}_{i1}^*, \ldots, \tilde{y}_{i,p_i}^*$. The V_{loso} satisfies

$$V_{loso}\left(\lambda\right) = \frac{1}{N}\sum_{i=1}^{N}\left(\tilde{Y}_i - \widehat{\tilde{Y}}_i\right)'\left(I_{p_i} - \tilde{A}_{ii}\right)^{-2}\left(\tilde{Y}_i - \widehat{\tilde{Y}}_i\right),$$

A RKHS Framework for $\log\sigma^2$

For fixed $\phi(\cdot)$, let $Z_i = (z_{i1}, \ldots, z_{ip_i})'$, $z_{ij} = \epsilon_{ij}^2$, where

$$\epsilon_{ij} = y_{ij} - \sum_{k<j} \phi(\boldsymbol{v}_{ijk})\, y_{ik}.$$

The log likelihood of the squared working innovations $Z_1, \ldots, Z_N$ coincides with a Gamma distribution with scale parameter $\alpha = 2$:

$$-2\ell\left(\sigma^2 | Z_1, \ldots, Z_N\right) = \sum_{i=1}^{N}\sum_{j=1}^{p_i} \eta_{ij} + \sum_{i=1}^{N}\sum_{j=1}^{p_i} z_{ij} e^{-\eta_{ij}},$$

where $\eta_{ij} = \eta(t_{ij}) = \log\sigma^2(t_{ij})$.

A RKHS Framework for $\log \sigma^2$

Take the estimator of $\eta(t) = \log \sigma^2(t) \in \mathcal{H}$ to minimize

$$-2\ell(\eta|Z_1, \ldots, Z_N) + \lambda J(\eta) = \sum_{i=1}^{N} \sum_{j=1}^{p_i} \eta(t_{ij}) + \sum_{i=1}^{N} \sum_{j=1}^{p_i} z_{ij} e^{-\eta(t_{ij})} + \lambda J(\eta),$$

for $\eta \in \mathcal{H}$, where $J(\eta) = ||P_1 \eta||^2$. Let

$$\mathcal{T} = \bigcup_{i,j} \{t_{ij}\},$$

Theorem 1 gives that the minimizer has the form

$$\eta_\lambda(t) = \sum_{i=1}^{\mathcal{N}_0} d_i \nu_i(t) + \sum_{j=1}^{|\mathcal{T}|} c_j K_1(t_j, t),$$

where $\{\nu_i\}$ span $\mathcal{H}_0$, $K_1(t_j, t)$ is the RK for $\mathcal{H}_1$ evalutated at t_j, the j^{th} element of $\mathcal{T}$.

Bivariate smoothing with tensor product P-splines

Penalized B-splines are a flexible alternative to smoothing splines.

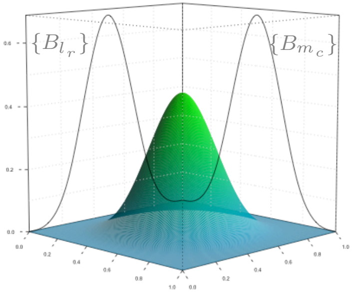
Given B-splines bases for l and m, let

$$\phi(l,m) = \sum_{r=1}^{k_l}\sum_{c=1}^{k_m} \theta_{rc} B_{l_r}(l) B_{m_c}(m).$$

For fixed σ_{ij}^2, take ϕ_λ to minimize

$$\begin{aligned} -2\ell(\phi|Y_1,\ldots,Y_N) + J_\lambda(\phi) \\ = (Y - XB\theta)' D^{-1} (Y - XB\theta) \\ + \lambda_l ||P_l\theta||^2 + \lambda_m ||P_m\theta||^2, \end{aligned}$$

where $||P_l\theta||^2$, $||P_m\theta||^2$ are finite difference penalties on the θ_{rc} along the l and m axes.



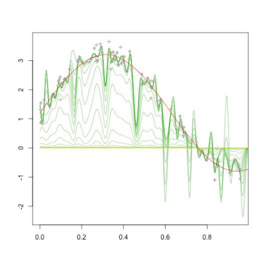
$T_{rc}(l,m) = B_{l_r}(l) B_{m_c}(m)$

Bivariate smoothing with tensor product P-splines

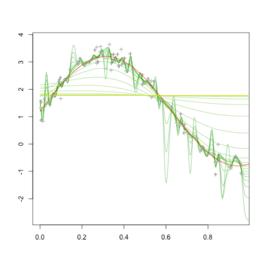
The penalties are trivial to construct for any difference order.

For $f''(x) = \sum_{j=1}^{k} \theta_j B_j''(x)$,

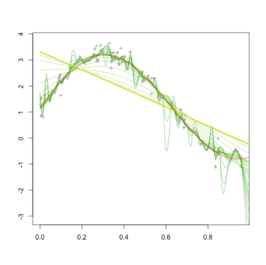
$$\int_0^1 \left(f''(x)\right)^2 \, dx = \int_0^1 \sum_{j=1}^{k} \theta_i B_j''(x) \, dx \approx c_1 \sum_j \left(\Delta^2 \theta_j\right)^2$$



(a) difference penalty of order $d = 0$



(b) difference penalty of order $d = 1$



(c) difference penalty of order $d = 2$

Simulation Studies

Simulation Conditions

I. Complete data

Σ	Model I, II, III, IV, V
N	$50, 100$
p	$10, 20, 30$

II. Unbalanced data, $N = 50$

Σ	Model I, II, III, IV, V
p	$10, 20$
$\%$ missing	$0, 0.1, 0.2, 0.3$

Performance is measured with loss functions

$$\Delta_1\left(\Sigma, \hat{\Sigma}\right) = tr\left(\left(\Sigma^{-1}\hat{\Sigma} - \mathrm{I}\right)^2\right), \quad \Delta_2\left(\Sigma, \hat{\Sigma}\right) = tr\left(\Sigma^{-1}\hat{\Sigma}\right) - log|\Sigma^{-1}\hat{\Sigma}| - p$$

Use Monte Carlo simulation to estimate risk

$$R_i\left(\Sigma, \hat{\Sigma}\right) = E_{\Sigma}\left[\Delta_i\left(\Sigma, \hat{\Sigma}\right)\right], \quad i = 1, 2$$

In Study I, we compare performance to that of the oracle estimator, Polynomial MCD GLM $\hat{\Sigma}_{poly}$, the sample covariance matrix $S = [s_{ij}]$, Shrinkage estimators S^{ω}, S^{λ}

Applying Elementwise Shrinkage to S

Tapering Estimators

- **The Banded Sample Covariance Matrix**

$$B_k\left(S\right) = \left[s_{ij}1\left(\left|i-j\right| \leq k\right)\right] = R_B * S, \quad 0 < k < p.$$

- **The Tapered Sample Covariance Matrix**

$$S^{\omega} = \left[\omega_{ij}^{k} s_{ij}\right],$$

where $0 < k < p$, and if $k_h = k/2$

$$\omega_{ij}^{k} = k_h^{-1}\left[\left(k - \left|i-j\right|\right)_+ - \left(k_h - \left|i-j\right|\right)_+\right].$$

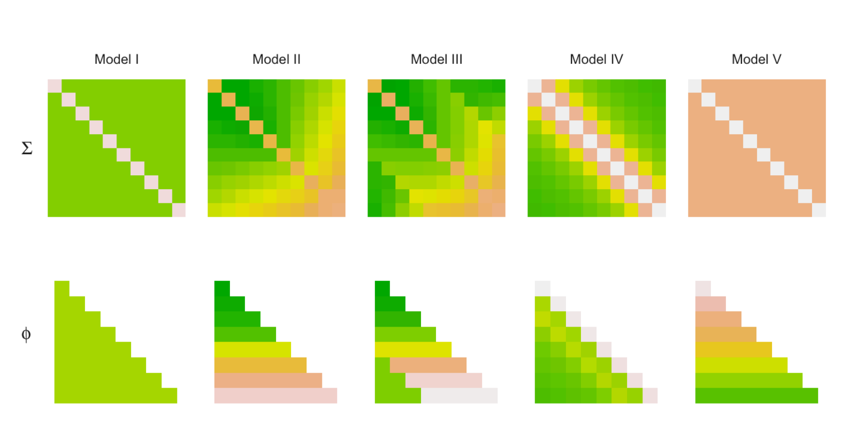
- **Soft Thresholding Estimator:**

$$S^{\lambda} = \left[\text{sign}\left(s_{ij}\right)\left(s_{ij} - \lambda\right)_+\right]$$

Simulations

Simulation Studies

Data Generation Models

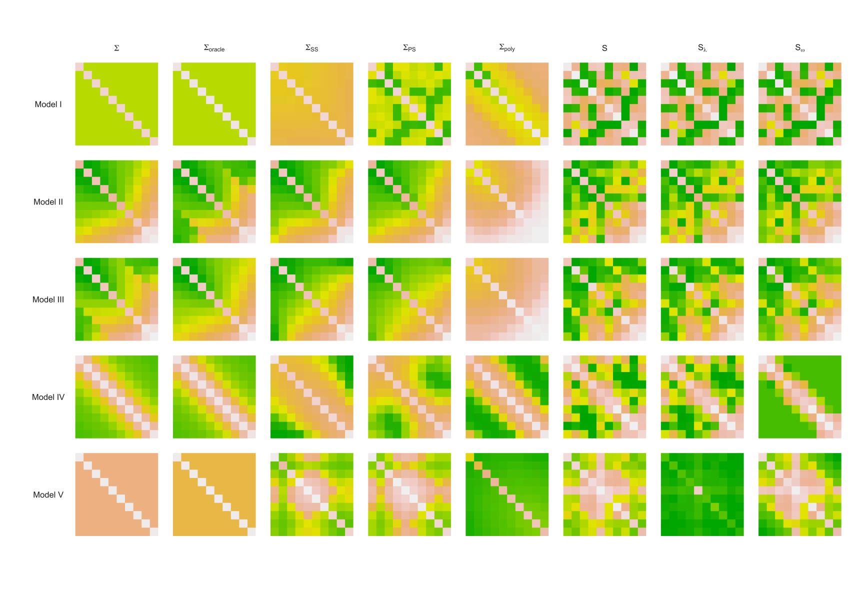


Simulation Studies

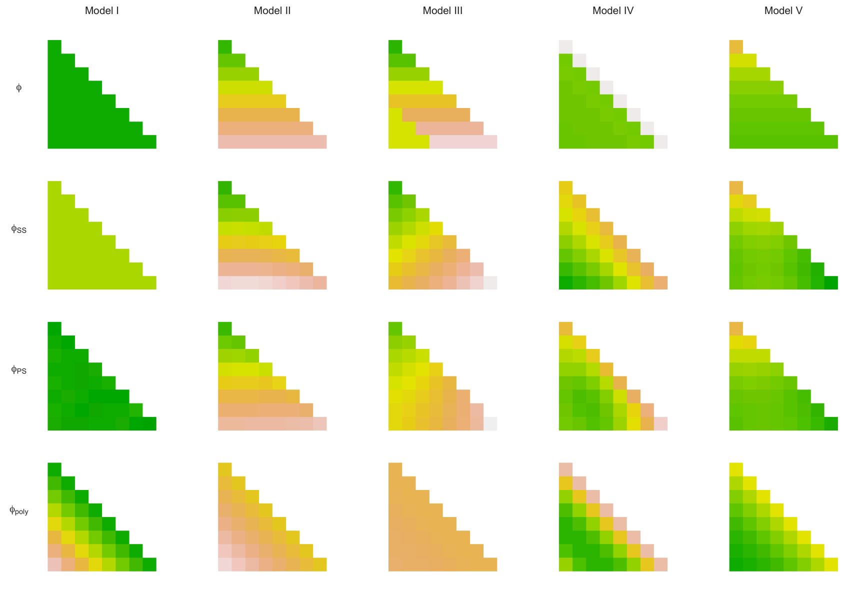
Data Generation Settings

I. $\Sigma = \mathrm{I}$	$\phi(t,s) = 0, 0 \le s < t \le 1$	$\sigma^2(t) = 1, 0 \le t \le 1$
II. $\Sigma = T^{-1} D T'^{-1}$	$\phi(t,s) = t - \frac{1}{2}, \quad 0 \le t \le 1$	$\sigma^2(t) = 0.1^2, 0 \le t \le 1$
III. $\Sigma = T^{-1} D T'^{-1}$	$\phi(t,s) = \begin{cases} t - \frac{1}{2}, & t - s \le 0.5 \\ 0, & t - s > 0.5 \end{cases}$	$\sigma^2(t) = 0.1^2, 0 \le t \le 1$
IV. $\Sigma = [\sigma_{ij}]$	$\sigma_{ij} = \left(1 + \frac{(t_i - t_j)^2}{2k^2}\right)^{-1}$	$k = 0.6, 0 < t_i, t_j < 1$
V. $\Sigma = \rho \mathrm{J} + (1-\rho)\mathrm{I}$, $\rho = 0.7$	$\phi(t,s) = \frac{\rho}{1+(t-2)\rho}, t = 2, \ldots, p$	$\sigma^2(t) = 1 - \frac{(\max(t,2)-2)\rho^2}{1+(\max(t,2)-2)\rho}$

Simulation Studies

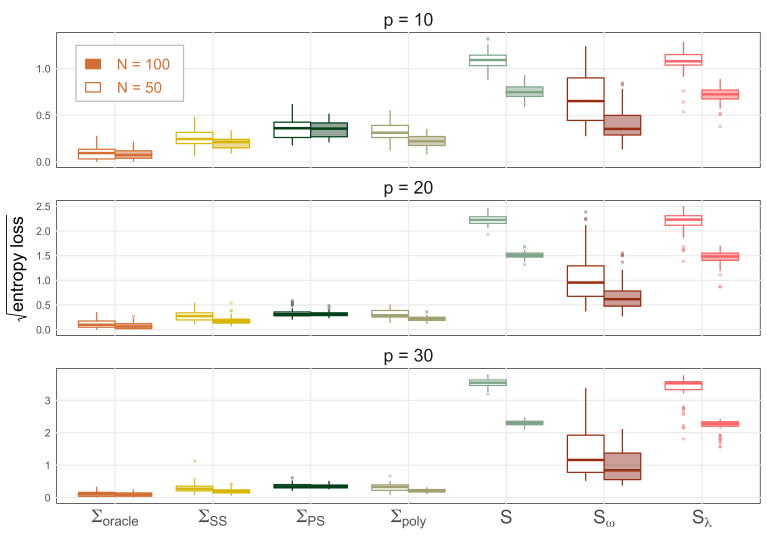


Simulation Studies



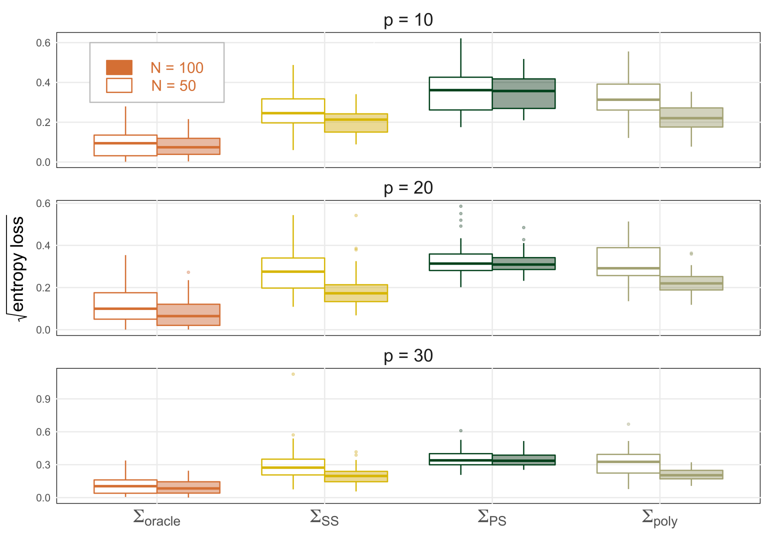
Simulation Studies

Results with complete data: Model I



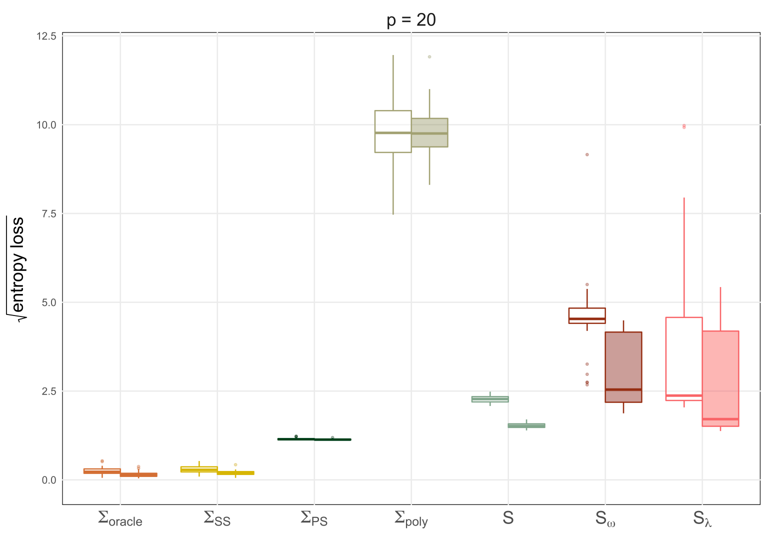
Simulation Studies

Results with complete data: Model I



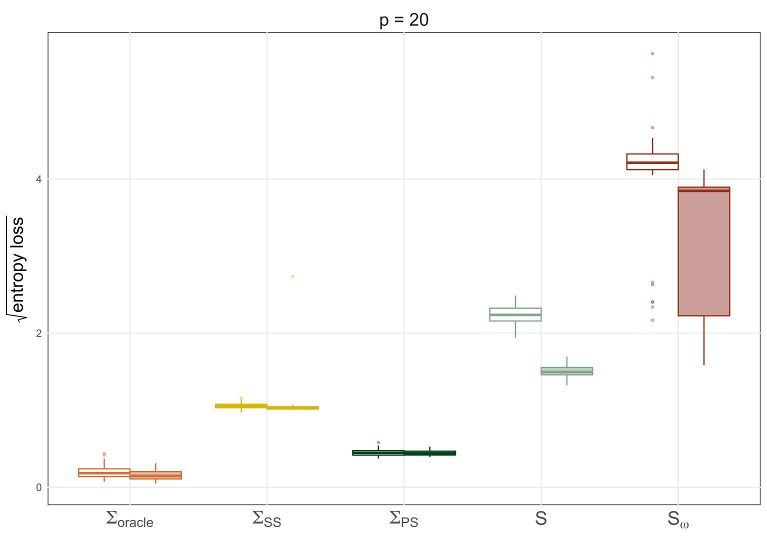
Simulation Studies

Results with complete data: Model II



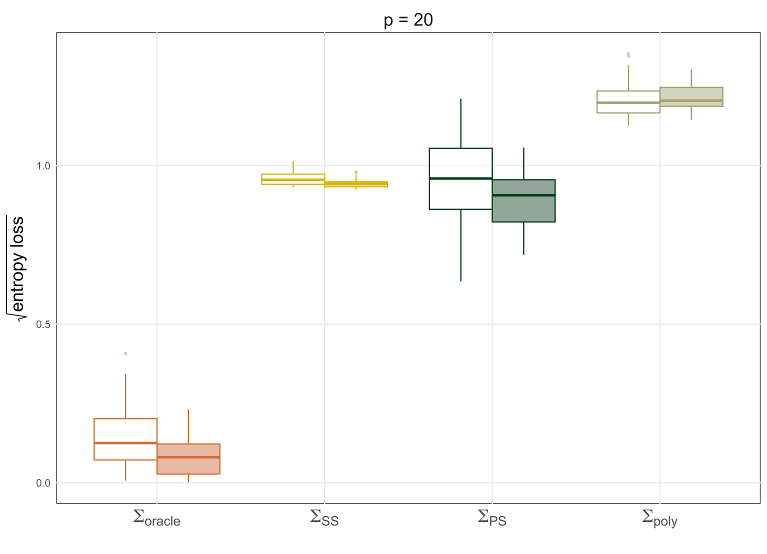
Simulation Studies

Results with complete data: Model III



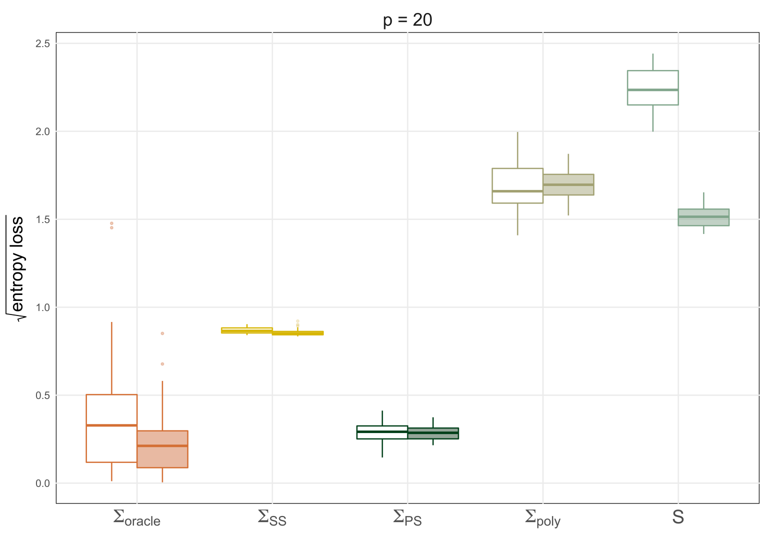
Simulation Studies

Results with complete data: Model IV



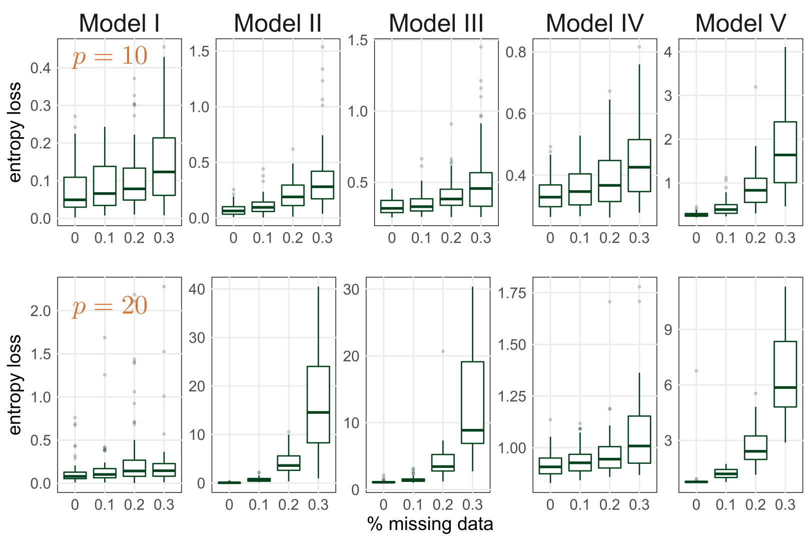
Simulation Studies

Results with complete data: Model V



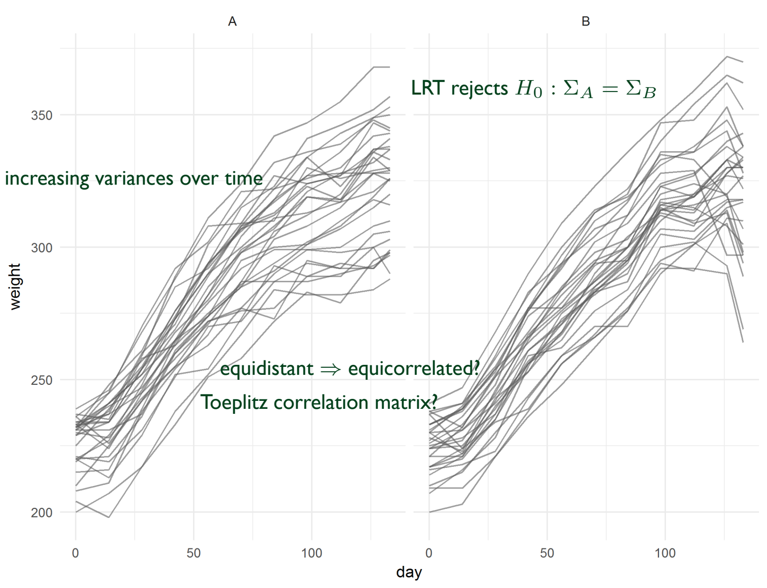
Simulation Studies

Results with incomplete data, $N = 50$



Kenward Cattle Data

Suppressed immune resistance in animals infected with roundworm can significantly impact growth.



LRT rejects $H_0 : \Sigma_A = \Sigma_B$

increasing variances over time

equidistant $\Rightarrow$ equicorrelated?

Toeplitz correlation matrix?

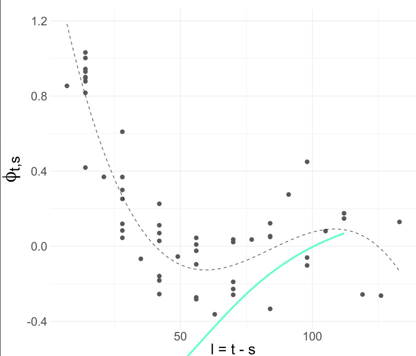
Kenward Cattle Data

Sample correlations

	day										
	0	14	28	42	56	70	84	98	112	126	133
0	1.00										
14	0.82	1.00									
28	0.76	0.91	1.00								
42	0.65	0.86	0.93	1.00							
56	0.63	0.83	0.89	0.93	1.00						
70	0.58	0.75	0.85	0.90	0.94	1.00					
84	0.51	0.64	0.75	0.80	0.85	0.92	1.00				
98	0.52	0.68	0.77	0.82	0.88	0.93	0.92	1.00			
112	0.51	0.61	0.71	0.74	0.81	0.89	0.92	0.96	1.00		
120	0.46	0.59	0.69	0.70	0.77	0.85	0.86	0.94	0.96	1.00	
133	0.46	0.56	0.67	0.67	0.74	0.81	0.84	0.91	0.95	0.98	1.00

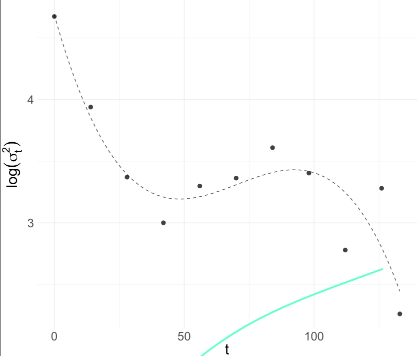
The regressogram and innovation variogram

Sample generalized autoregressive parameters



$$\phi_{ts} = x'_{ts}\gamma = \gamma_0 + \gamma_1 (t-s) + \gamma_2 (t-s)^2 + \gamma_3 (t-s)^3$$

Sample innovation variances



$$\log \sigma_t^2 = z'_t\xi = \left[\xi_0 + \xi_1 t + \xi_2 t^2 + \xi_3 t^3\right]$$

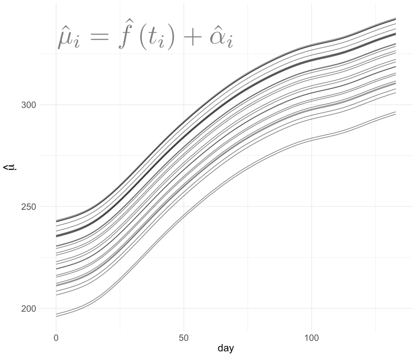
A joint mean-covariance model for the cattle weights

$$y_{ij} = f(t_i) + \alpha_i + \epsilon^*_{ij},$$
$$i = 1, \ldots, N = 30$$
$$j = 1, \ldots, 11,$$

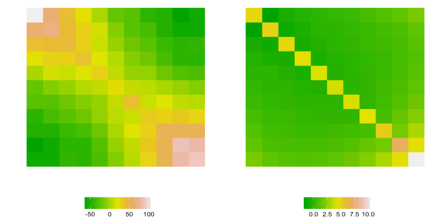
where the $\alpha_i \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2_\alpha)$ mutually independent of $\epsilon^*_{ij} = \left(\epsilon^*_{i1}, \ldots, \epsilon^*_{ip_i}\right)'$,

$$\epsilon^*_i \sim N(0, \Sigma),$$
$$f \in \mathcal{H} = \mathcal{C}^2[0,1],$$

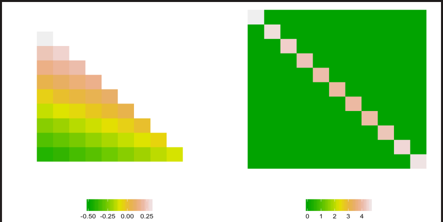
and $J(f) = \int_0^1 \left(f''(x)\right)^2 \, dx$.



Modeling the Cholesky decomposition



S $\qquad$ $\hat{\Sigma} = \hat{T}'^{-1} \hat{D} \hat{T}^{-1}$



$\hat{\phi}(t, s)$ $\qquad$ $\hat{\sigma}^2(t)$

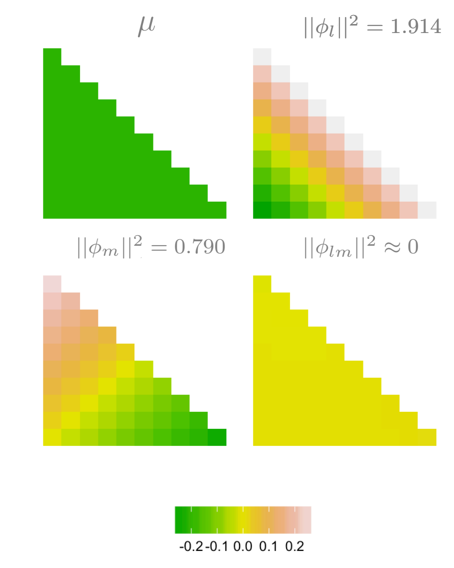
Model

$$\epsilon^*(t_{ij}) = \sum_{k<j} \phi(t_{ij}, t_{ik}) \, \epsilon^*(t_{ik}) + \epsilon(t_{ij})$$

where $\epsilon(t) \sim N\left(0, \sigma^2(t)\right)$ and

$$\begin{aligned}
\phi \in \mathcal{H} &= \mathcal{H}_{[l]} \otimes \mathcal{H}_{[m]} \\
\mathcal{H}_{[l]} &= \left\{ \phi : \phi'' = 0 \right\} \\
&\quad \oplus \left\{ \phi : \phi(0), \phi'(0) = 0; \; \phi'' \in \mathcal{L}_2[0,1] \right\} \\
\mathcal{H}_{[m]} &= \left\{ \phi : \phi \propto 1 \right\} \\
&\quad \oplus \left\{ \phi : \int_0^1 \phi \, dx = 0, \;\; \phi' \in \mathcal{L}_2[0,1] \right\}
\end{aligned}$$

The functional components of ϕ



Concluding Remarks and Future Work

- Joint mean-covariance estimation for longitudinal data
- Use additional covariates to simultaneously model covariance matrices for k groups $\Sigma_1, \ldots, \Sigma_k$
- Employ shrinkage alongside smoothing with P-splines
- P-spline mixed model framework with REML estimation

Thank you!