Nonparametric Covariance Estimation for Longitudinal Data via Penalized Tensor Product Splines

Tayler A. Blake* Yoonkyung Lee†

January 19, 2018

1 Performance

In this section, we evaluate the performance of the spline estimator under different simulation settings when the tuning parameters are chosen by the unbiased risk estimate and leave-one-subject-out cross validation. We compare the performance of the maximum penalised likelihood estimator using the classical smoothness penalty under the smoothing spline representation to the performance of the tensor product P-spline estimator for varying orders of the penalty. We also compare performance under complete data to the performance under irregularly sampled data:

- All subjects share a common set of observation times $t_1, \ldots, t_M$.
- Observation times vary across subjects, with subject-specific deviation defined as follows:

For the case of common observation times across all subjects, we also consider three other methods of estimating a covariance matrix for comparison: the sample covariance matrix Σ^*, the soft thresholding estimator of **?**, and the tapering estimator of **?**. The soft-thresholding estimator proposed in **?** is given by

$$\hat{\Sigma}^{sthresh}(\lambda) = \left[\text{sign}\left(\sigma^*_{ij}\right)\left(\sigma^*_{ij} - \lambda\right)_+\right],$$

where σ^*_{ij} denotes the i-j^{th} entry of the sample covariance matrix, and λ is a penalty parameter controlling the amount of shrinkage applied to the empirical estimator. The tapering estimator presented in **?** is defined

$$\hat{\Sigma}^{taper}(\lambda) = \left[\omega^\lambda_{ij}\sigma^*_{ij}\right].$$

The weights ω^λ_{ij} are given by

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

*The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201
†The Ohio State University, 1958 Neil Avenue, Columbus, OH 43201

where $k_h = k/2$ is assumed to be even without loss of generality. These may be rewritten as

$$\omega_{ij} = \begin{cases} 1, & ||i-j|| \leq k_h \\ 2 - \frac{i-j}{k_h} & k_h < ||i-j|| \leq k, \\ 0 & \text{otherwise} \end{cases}$$

The subscript on the weights ω_{ij} serves to indicate that these are controlled by a tuning parameter which controls the amount of shrinkage applied to the elements of the sample covariance matrix.

[discuss the MCRE and CVTuningCov package]

[discuss the implementation and R package here]

To assess performance of estimator G, we consider two commonly used loss functions:

$$\Delta_1(\Sigma, G) = tr\left(\Sigma^{-1}G\right) - log|\Sigma^{-1}G| - M, \tag{1}$$

$$\Delta_2(\Sigma, G) = tr\left(\left(\Sigma^{-1}G - \mathrm{I}\right)^2\right) \tag{2}$$

where Σ is the true covariance matrix and G is an $M \times M$ positive definite matrix, which are commonly referred to as the entropy loss and the quadratic loss, respectively. Each of these loss functions are 0 when $G = \Sigma$ and is positive when $G! = \Sigma$. Both are invariant with respect to transformations

$$G^* = CGC', \quad \Sigma^* = C\Sigma C',$$

for a nonsingular matrix C. We obtain the corresponding risk functions by taking expectations,

$$R_i(\Sigma, G) = E_\Sigma\left[\Delta_i(\Sigma, G)\right], \quad i = 1, 2.$$

We prefer estimator $\hat{\Sigma}_1$ over another estimator $\hat{\Sigma}_2$ if $R_i\left(\Sigma, \hat{\Sigma}_2\right) < R_i\left(\Sigma, \hat{\Sigma}_2\right)$. We estimate the risk functions by Monte Carlo approximation, using $N_{sim} = 100$ simulation runs for each scenario outlined above. Estimation is performed on data generated according to an M-dimensional multivariate Normal distribution with mean zero; we consider three for the covariance structure defining the generating distribution:

I. Mutual independence: $\Sigma_1 = T^{-T}D^2T^{-1} = \mathrm{I}$ where

$$\begin{gathered} \phi(t, s) = 0, \quad 1 \leq t < s \leq M; \\ \sigma^2(t) = 1, \quad t = 1, \ldots, M. \end{gathered}$$

II. Linear varying coefficient model with constant innovation variance: $\Sigma_2 = T^{-T}D^2T^{-1}$ where

$$\begin{gathered} \phi(t, s) = t - \frac{1}{2M}, \quad 1 \leq s < t \leq M \\ \sigma^2(t) = 0.1, \quad t = 1, \ldots, M. \end{gathered}$$

Table 1: Simulation results for $\Sigma_1 = \mathrm{I}$ under quadratic loss, Δ_1. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
N	M				URE	losoCV	URE	losoCV
50	10	0.4043				0.0016		
	20	0.7761				0.0008		
	30	1.2350				0.0006		
100	10							
	20							
	30							

III. AR (1) model with linear varying coefficient: $\Sigma_3 = T^{-T} D^2 T^{-1}$ where

$$\phi(t,s) = \begin{cases} t - \frac{1}{2M}, & t - s = 1 \\ 0, & t - s > 1 \end{cases},$$
$$\sigma^2(t) = 0.1, \quad t = 1, \ldots, M.$$

IV. The compound symmetry model: $\Sigma_4 = \sigma^2(\rho \mathrm{J} + (1-\rho)\mathrm{I})$, $\rho = 0.7$, $\sigma^2 = 1$.

$$\phi_{ts} = -\frac{\rho}{1 + (t-1)\rho}, \quad t = 2, \ldots, M, \quad s = 1, \ldots, t-1$$
$$\sigma_t^2 = 1 - \frac{(t-1)\rho^2}{1 + (t-1)\rho}, \quad t = 2, \ldots, M.$$

Table 2: Simulation results for $\Sigma_1 = \mathrm{I}$ under entropy loss, Δ_2. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	1.2399				0.0783		
	20	5.0550				0.0800		
	30	12.3280				0.0735		
100	10							
	20							
	30							

Table 3: Simulation results for Σ_2, the linear varying coefficient AR model, under quadratic loss, Δ_1. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	0.4885				0.0567		
	20	2.6654				0.6851		
	30	23.0959				6.9789		
100	10							
	20							
	30							

Table 4: Simulation results for Σ_2, the linear varying coefficient AR model, under entropy loss, Δ_2. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	1.1861				0.0800		
	20	5.1155				0.0730		
	30	12.5243				0.0789		
100	10							
	20							
	30							

Table 5: Simulation results for Σ_3, the linear AR (1) model under quadratic loss, Δ_1. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	0.4086				0.0145		
	20	0.9926				0.0609		
	30	1.2884				0.1387		
100	10							
	20							
	30							

Table 6: Simulation results for Σ_3, the linear AR$\,(1)$ model, under entropy loss, Δ_2. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	1.2023				0.2750		
	20	5.0599				0.8759		
	30	12.3077				1.6266		
100	10							
	20							
	30							

Table 7: Simulation results for Σ_4, the compound symmetry model, under quadratic loss, Δ_1. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\Sigma}^{taper}$	$\hat{\Sigma}^{ST}$	$\hat{\Sigma}^{ssanova}$		$\hat{\Sigma}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	47.4073				4.8320		
	20	104.8177				5.5327		
	30	151.9395				5.6466		
100	10							
	20							
	30							

Table 8: Simulation results for Σ_4, the compound symmetry model, under entropy loss, Δ_2. The risk functions for the sample covariance matrix, the tapered estimator, the soft thresholding estimator, the SSANOVA Cholesky estimator, and the tensor product P-spline Cholesky estimator were estimated using Monte Carlo simulation, with $N_sim = 100$ simulation trials. The tuning parameters for the tapering estimator and the soft thresholding estimator were chosen using $K = 5$-fold cross validation. The performance of the spline estimators is evaluated when both the unbiased risk estimate and leave-one-subject-out cross validation are used to select the smoothing parameters.

		Σ^*	$\hat{\sum}^{taper}$	$\hat{\sum}^{ST}$	$\hat{\sum}^{ssanova}$		$\hat{\sum}^{tps}$	
N	M				URE	losoCV	URE	losoCV
50	10	14.6842				3.9489		
	20	36.5299				4.6406		
	30	59.5043				4.9214		
100	10							
	20							
	30							