Nonparametric Covariance Estimation for Longitudinal Data via Penalized Tensor Product Splines

Tayler A. Blake[*] Yoonkyung Lee[†]

February 3, 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

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

[*]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. 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. The first loss Δ_1 is commonly referred to as the entropy loss; it gives the Kullback-Leibler divergence of two multivariate Normal densities corresponding to the two covariance matrices. The second loss Δ_2, or the quadratic loss, measures the Euclidean or Frobenius norm of its matrix argument, and consequently penalizes overestimates more than underestimates, so "smaller" estimates are favored more under Δ_2 than Δ_1. 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 four Cholesky covariance structures for the underlying generating distribution:

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

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

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

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

III. AR (k) model with linear varying coefficient: $\Sigma_3 = T^{-T}D^2T^{-1}$ where $k = \lfloor M/2 \rfloor + 1$ and

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

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

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

V. 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 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
	10	0.0010		0.4702
$N = 50$	20	0.0007		0.8495
	30	0.0003		1.1449
	10	0.0004		0.2072
$N = 100$	20	0.0002		0.3920
	30	0.0001		0.5712

2 Discussion

See **?** section 3.1 for further discussion of loss functions

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
$N = 50$	10	0.0684		1.2339
	20	0.0799		5.0827
	30	0.0668		12.5162
$N = 100$	10	0.0405		0.5854
	20	0.0356		2.3038
	30	0.0396		5.2641

Table 2: 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
$N = 50$	10	0.0314		0.5726
	20	0.3266		2.3130
	30	5.0696		15.1096
$N = 100$	10	0.0156		0.2479
	20	0.1894		1.3177
	30	2.3876		8.3983

Table 3: 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
$N = 50$	10	0.0647		1.2431
	20	0.0884		5.0437
	30	0.0702		12.4559
$N = 100$	10	0.0307		0.5403
	20	0.0357		2.3195
	30	0.0372		5.2817

Table 4: Simulation results for Σ_3, the k-banded linear varying coefficient AR model with $k = \lfloor M/2 \rfloor + 1$, 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
	10	0.0562		0.5237
$N = 50$	20	0.7832		2.1419
	30	8.2650		15.2842
	10	0.0376		0.2546
$N = 100$	20	0.6260		1.3751
	30	5.7635		7.4750

Table 5: Simulation results for Σ_3, the k-banded linear varying coefficient AR model with $k = \lfloor M/2 \rfloor + 1$, 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
	10	0.3354		1.1947
$N = 50$	20	1.1144		5.0966
	30	2.3247		12.4905
	10	0.2826		0.5446
$N = 100$	20	1.0690		2.3514
	30	2.2737		5.4204

Table 6: Simulation results for Σ_4, the 2-banded 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	
	10	0.0134		0.4169
$N = 50$	20	0.0590		0.8810
	30	0.1351		1.2571
	10	0.0077		0.2263
$N = 100$	20	0.0549		0.4309
	30	0.1331		0.6819

Table 7: Simulation results for Σ_4, the 2-banded 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	S
	10	0.2605		1.1692
$N = 50$	20	0.8836		5.0899
	30	1.6087		12.5844
	10	0.2193		0.5642
$N = 100$	20	0.8468		2.2607
	30	1.5743		5.2437

Table 8: Simulation results for Σ_5, 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	S
	10	0.3688		0.7872
$N = 50$	20	0.9770		1.6167
	30	1.6067		2.5548
	10	0.3210		0.3913
$N = 100$	20	0.9793		0.8385
	30	1.6177		1.2383

Table 9: Simulation results for Σ_5, 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. 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.

	M	$\hat{\Sigma}_{ssanova}$		S
		LosoCV	URE	S
	10	0.2837		1.1943
$N = 50$	20	0.7551		5.0283
	30	1.1936		12.5871
	10	0.2449		0.5734
$N = 100$	20	0.7231		2.2678
	30	1.1780		5.2562