Nonparametric Covariance Estimation for Longitudinal Data via Penalized Tensor Product Splines

Tayler A. Blake* Yoonkyung Lee†

February 13, 2018

1 Simulation Studies

In this section, we evaluate the performance of the spline estimator under varying data generation distributions using two model selection criteria: the unbiased risk estimate and leave-one-subject-out cross validation. To gauge the utility of our method, we compare the performance of the estimator based on complete data to that of three alternative estimators. A complete dataset is one in which all subjects share a common set of evenly-spaced observation times $t_1, \ldots, t_M$, and there are no observations missing for any patient. To examine the robustness of our method to sparsity in the data, we also compare our performance in the ideal sampling case to the performance of the estimator based on incomplete data by subsampling observations and treating the remaining unused observations as missing data.

1.1 Alternative estimators for benchmarking

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 Rothman et al. [2009], and the tapering estimator of Cai et al. [2010]. The soft-thresholding estimator proposed in Rothman et al. [2009] is given by

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

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. Cai et al. [2010] derived optimal rates of convergence under the operator norm for the tapering estimator:

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

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

where the ω_{ij}^k are given by

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

The weights ω_{ij}^k are indexed with superscript to indicate that they are controlled by a tuning parameter, k, which can take integer values between 0 and M, the dimension of the covariance matrix. It controls the amount of shrinkage applied to the elements of the sample covariance matrix by defining a window from the main diagonal of the estimator, which determines three sets of off-diagonals. Different shrinkage is applied to elements of the subdiagonals belonging to each the three sets. The set of subdiagonals closest to the main diagonal receives no shrinkage penalty, and shrinkage increases as distance from the main diagonal increases. Without loss of generality, we assume that $k_h = k/2$ is even. The weights may be rewritten as

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

[discuss the implementation and R package here]

1.2 Loss functions for evaluating estimators

To assess performance of estimator $\hat{\Sigma}$, we consider two commonly used loss functions:

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

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

where Σ is the true covariance matrix and $\hat{\Sigma}$ is an $M \times M$ positive definite matrix. Each of these loss functions is 0 when $\hat{\Sigma} = \Sigma$ and is positive when $\hat{\Sigma} \neq \Sigma$. Both measures of loss are scale invariant. If we let random vector Y have covariance matrix Σ, and define the transformation Z as

$$Z = CY.$$

for some $M \times M$ matrix C, then Z has covariance matrix $\Sigma_z = C\Sigma C'$. Given an estimator $\hat{\Sigma}$ of Σ, one immediately obtains an estimator for Σ_z, $\hat{\Sigma}_z = C\hat{\Sigma}C'$. If C is invertible, then the loss functions Δ_1 and Δ_2 satisfy

$$\Delta_i\left(\Sigma, \hat{\Sigma}\right) = \Delta_i\left(C\Sigma C', C\hat{\Sigma}C'\right).$$

The first loss Δ_1 is commonly referred to as the entropy loss; it gives the Kullback-Leibler divergence of two multivariate Normal densities with the same mean corresponding to the two covariance matrices. The second loss Δ_2, or the quadratic loss, measures the difference between

$\left(\Sigma^{-1}\hat{\Sigma}\right)$ and the identity matrix with the squared Frobenius norm. The Frobenius norm of a symmetric matrix A is given by

$$||A||^2 = \text{tr}\left(AA'\right).$$

The quadratic loss consequently penalizes overestimates more than underestimates, so "smaller" estimates are favored more under Δ_2 than Δ_1. Among the class of estimators comprised of scalar multiples cS of the sample covariance matrix, it has been established Haff [1980] that S is optimal under Δ_2, while the smaller estimator $\frac{nS}{n+p+1}$ is optimal under Δ_1.

Given Σ, the corresponding values of the risk functions are obtained by taking expectations:

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

We prefer one estimator $\hat{\Sigma}_1$ to another $\hat{\Sigma}_2$ if it has smaller risk. Given Σ, we estimate the risk of an estimator via Monte Carlo approximation.

1.3 Simulation study design

To understand the strengths and weaknesses of our method and how its performance compares to that of the aforementioned alternative estimators, in the first portion of our simulation study, we examine performance for five underlying covariance structures across varying numbers of subjects, N, and within-subject sample sizes, M. Subjects share a common set of M regularly-spaced observation times so as to permit comparison with the estimators based on the sample covariance matrix, which cannot accommodate irregularly spaced observations. In the second portion of the study, our primary concern is studying the stability of our estimator as the irregularity in the observed time points across subjects increases. For fixed N, we observe performance when the data are generated from the same underlying covariance structures for varying within-subject sample sizes M and varying levels of data sparsity.

We study estimator performance for five covariance structures, each exhibiting varying degrees of structural complexity. We first consider the covariance structure corresponding to mutual independence, which is both the simplest and sparsest structure, corresponding to $\phi\left(t, s\right) = 0$ for all t, s and constant $\log\sigma^2\left(t\right)$. We then consider a class of covariance structures which we define explicitly in terms of the GARPs and IVs, having varying coefficient function which is linear in t: $\phi\left(t, s\right) \propto bt$ and constant innovation variance function. We then define three different covariance structures by banding the corresponding Cholesky factor at varying distances $k_l \in [0, 1]$ from the diagonal, which results in a set of inverse covariance matrices which are also banded at the same distances; see Bickel and Levina [2008]. Equivalently, we take $\phi\left(t, s\right) \propto bt$ for t, s such that $t - s \leq k_l$ and $\phi\left(t, s\right) = 0$ for t, s such that $t - s > k_l$. Lastly, we consider the compound symmetry model, a commonly utilized parametric model for longitudinal data; while the structure is of the overall covariance matrix is parsimonious, the varying coefficient function and innovation variance function of the corresponding Cholesky decomposition are nonlinear in t. Given covariance matrix Σ, risk estimates are obtained from$N_{sim} = 100$ samples from an M-dimensional multivariate

Normal distribution with mean zero and the same covariance. The results of the simulations for data on a regular grid are given in tables **??** - **??**; results for simulations with sparsely sampled data are given in tables **??** - **??**.

1.4 Results

1.4.1 Simulation study 1: performance comparison with complete data

Table 1: Risk estimates and corresponding standard errors for our proposed estimator under entropy loss, Δ_2 when the data are generated according to model **??**.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.0684	0.0678	1.2339	0.4451	1.1760
	20	0.0799	0.0720	5.0827	1.6504	4.7847
	30	0.0668	0.0740	12.5162	1.9975	11.0434
$N = 100$	10	0.0405	0.0379	0.5854	0.1783	0.5201
	20	0.0356	0.0378	2.3038	0.4394	1.9637
	30	0.0396	0.0322	5.2641	0.6717	4.5410

Table 2: Risk estimates and corresponding standard errors for our proposed estimator under entropy loss, Δ_2 when the data are generated according to model II.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.0647	0.0696	1.2431	1.4242	1.1195
	20	0.0884	0.0969	5.0437	17.0220	13.5290
	30	0.0702	0.0894	12.4559	39.9769	159.0521
$N = 100$	10	0.0307	0.0302	0.5403	0.7659	0.5609
	20	0.0357	0.0350	2.3195	10.0140	12.1431
	30	0.0372	0.0334	5.2817	35.0353	108.1015

Table 3: Risk estimates and corresponding standard errors for our proposed estimator under entropy loss, Δ_2 when the data are generated according to model III.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.3354	0.3174	1.1947	1.1073	1.1649
	20	1.1144	1.1143	5.0966	17.0220	12.6171
	30	2.3247	2.3168	12.4905	50.3684	101.8245
$N = 100$	10	0.2826	0.2955	0.5446	0.5410	0.5531
	20	1.0690	1.0627	2.3514	12.8490	11.4934
	30	2.2737	2.2767	5.4204	27.2736	30.5818

Table 4: Risk estimates and corresponding standard errors for our proposed estimator under entropy loss, Δ_2 when the data are generated according to model IV.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.2605	.2743	1.1692	0.5899	1.1126
	20	0.8836	.8764	5.0899	1.8834	4.6363
	30	1.6087	1.6195	12.5844	3.1902	11.4818
$N = 100$	10	0.2193	0.2183	0.5642	0.2902	0.5456
	20	0.8468	0.8491	2.2607	0.7869	2.2028
	30	1.5743	1.5802	5.2437	1.1974	4.8555

Table 5: Risk estimates and corresponding standard errors for our proposed estimator under entropy loss, Δ_2 when the data are generated according to model V.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.2837	0.2766	1.1943	17.3871	1.2122
	20	0.7551	0.7657	5.0283	35.4067	5.1671
	30	1.1936	1.1927	12.5871	46.5337	12.4110
$N = 100$	10	0.2449	0.2390	0.5734	16.2705	0.5796
	20	0.7231	0.7299	2.2678	31.3226	2.2988
	30	1.1780	1.1813	5.2562	39.2108	5.2592

1.5 Simulation study 2: irregularly sampled data

M	% subsampling	$\hat{\Delta}_1$	$se\left(\hat{\Delta}_1\right)$	$\hat{\Delta}_2$	$se\left(\hat{\Delta}_2\right)$
10	0.05	0.0016	0.0002	0.0760	0.0059
10	0.07	0.0017	0.0002	0.0824	0.0055
10	0.09	0.0015	0.0002	0.0776	0.0058
15	0.05	0.0020	0.0003	0.1027	0.0085
15	0.07	0.0024	0.0004	0.1135	0.0100
15	0.09	0.0021	0.0004	0.1013	0.0087
20	0.05	0.0011	0.0001	0.0878	0.0069
20	0.07	0.0011	0.0001	0.0971	0.0071
20	0.09	0.0013	0.0002	0.0998	0.0073

Table 6: Risk estimates and corresponding standard errors for our proposed estimator when the data are generated according to model **??** for varying data dimension and subsampling rates.

M	% subsampling	$\hat{\Delta}_1$	$se\left(\hat{\Delta}_1\right)$	$\hat{\Delta}_2$	$se\left(\hat{\Delta}_2\right)$
10	0.05	0.0520	0.0063	0.0940	0.0076
10	0.07	0.0462	0.0061	0.0949	0.0085
10	0.09	0.0676	0.0088	0.1124	0.0101
15	0.05	0.4004	0.0548	0.1434	0.0111
15	0.07	0.7398	0.1168	0.1895	0.0161
15	0.09	1.3971	0.1984	0.3201	0.0332
20	0.05	5.1618	0.6220	0.2705	0.0218
20	0.07	9.9945	1.0978	0.3894	0.0306
20	0.09	19.6154	2.0697	0.7071	0.0520

Table 7: Risk estimates and corresponding standard errors for our proposed estimator when the data are generated according to model II for varying data dimension and subsampling rates.

M	% subsampling	$\hat{\Delta}_1$	$se\left(\hat{\Delta}_1\right)$	$\hat{\Delta}_2$	$se\left(\hat{\Delta}_2\right)$
10	0.05	0.0617	0.0041	0.3451	0.0078
10	0.07	0.0681	0.0043	0.3498	0.0074
10	0.09	0.0574	0.0041	0.3427	0.0085
15	0.05	0.2226	0.0193	0.6905	0.0257
15	0.07	0.4622	0.0680	0.6909	0.0253
15	0.09	0.6438	0.0708	0.8038	0.0463
20	0.05	3.6000	0.4421	1.2193	0.0208
20	0.07	8.6383	1.1900	1.3306	0.0316
20	0.09	10.0914	1.4934	1.3546	0.0369

Table 8: Risk estimates and corresponding standard errors for our proposed estimator when the data are generated according to model III for varying data dimension and subsampling rates.

M	% subsampling	$\hat{\Delta}_1$	$se\left(\hat{\Delta}_1\right)$	$\hat{\Delta}_2$	$se\left(\hat{\Delta}_2\right)$
10	0.05	0.0116	0.0006	0.2573	0.0051
10	0.07	0.0126	0.0007	0.2665	0.0064
10	0.09	0.0113	0.0006	0.2537	0.0056
15	0.05	0.0325	0.0012	0.5596	0.0077
15	0.07	0.0421	0.0027	0.6065	0.0131
15	0.09	0.0365	0.0014	0.5835	0.0082
20	0.05	0.0659	0.0019	0.9159	0.0105
20	0.07	0.0603	0.0009	0.8904	0.0066
20	0.09	0.0615	0.0012	0.8935	0.0078

Table 9: Risk estimates and corresponding standard errors for our proposed estimator when the data are generated according to model IV for varying data dimension and subsampling rates.

M	% subsampling	$\hat{\Delta}_1$	$se\left(\hat{\Delta}_1\right)$	$\hat{\Delta}_2$	$se\left(\hat{\Delta}_2\right)$
10	0.05	0.4202	0.0165	0.3159	0.0099
10	0.07	0.4674	0.0187	0.3349	0.0100
10	0.09	0.6244	0.0363	0.3887	0.0149
15	0.05	0.7857	0.0262	0.6157	0.0137
15	0.07	0.8649	0.0260	0.6548	0.0145
15	0.09	1.0203	0.0425	0.7163	0.0195
20	0.05	1.0288	0.0203	0.8323	0.0156
20	0.07	1.1388	0.0343	0.9065	0.0247
20	0.09	1.3248	0.0593	1.0355	0.0351

Table 10: Risk estimates and corresponding standard errors for our proposed estimator when the data are generated according to model V for varying data dimension and subsampling rates.

1.6 Discussion

The following gives the precise covariance structures for the data generating distribution in the simulation settings for which the results are presented above.

I. Mutual independence: $\Sigma = \mathrm{I}$, where

$$\begin{aligned} \phi(t,s) &= 0, \quad 0 \le s < t \le 1, \\ \sigma^2(t) &= 1, \quad 0 \le t \le 1. \end{aligned}$$

II. Linear varying coefficient model with constant innovation variance: $\Sigma^{-1} = T'D^{-1}T$, where

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

III. AR (k) model with linear varying coefficient: $\Sigma^{-1} = T'D^{-1}T$, where $k = \lfloor M/2 \rfloor + 1$ and

$$\begin{aligned} \phi(t,s) &= \begin{cases} t - \frac{1}{2}, & t - s \le 0.5 \\ 0, & t - s > 0.5 \end{cases}, \\ \sigma^2(t) &= 0.1^2, \quad 0 \le t \le 1. \end{aligned}$$

IV. AR (1) model with linear varying coefficient: $\Sigma^{-1} = T'D^{-1}T$ where

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

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

$$\begin{aligned} \phi_{ts} &= -\frac{\rho}{1 + (t-1)\rho}, \quad t = 2, \ldots, M, \quad s = 1, \ldots, t-1 \\ \sigma_t^2 &= \begin{cases} 1, & t = 1 \\ 1 - \frac{(t-1)\rho^2}{1+(t-1)\rho}, & t = 2, \ldots, M \end{cases} \end{aligned}$$

1.6.1 Adjustments made for non-positive definite shrinkage estimators

Like other element-wise shrinkage estimators of the covariance matrix, the soft thresholding estimator is not guaranteed to be positive definite, though Rothman et al. [2009] established that in the limit, soft thresholding produces a positive definite estimator with probability tending to 1. We observed simulations runs which yielded a soft thresholding estimator that was indeed not positive definite. Evaluation of the entropy loss 2 is undefined at an estimator having at least one eigenvalue that is not greater than zero. To enable the evaluation of the entropy loss, we coerced these estimates to the "nearest" positive definite estimate via application of the technique presented in Cheng and Higham [1998]. For a symmetric matrix A, which is not positive definite, a modified Cholesky algorithm produces a symmetric perturbation matrix E such that $A + E$ is positive definite.

1.6.2 Tuning parameter selection for elementwise shrinkage estimators

As discussed in **??**, the soft thresholding estimator can be written as the solution to the optimization problem

$$s_\lambda(z) = \arg\min_\sigma \left[\frac{1}{2}(\sigma - z)^2 + J_\lambda(\sigma)\right], \tag{3}$$

so that estimation of the covariance matrix can be accomplished by solving multiple univariate Lasso-penalized least squares problems. The Frobenius is a natural measure of the accuracy of an estimator; it quantifies the sum over the unique elements of Σ of the the first term in 3,

$$||\hat{\Sigma}^\lambda - \Sigma||^2 = \left(\sum_{i,j}\left(\hat{\sigma}_{ij}^\lambda - \sigma_{ij}\right)^2\right)^{1/2} \tag{4}$$

If Σ were available, one would choose the value of the tuning parameter λ which minimizes **??**. In practice, one tries to first approximate the risk, or

$$E_\Sigma\left[||\hat{\Sigma}^\lambda - \Sigma||^2\right],$$

and then choose the optimal value of λ. As in regression methods, cross validation and a number of its variants have become popular choices for tuning parameter selection in covariance estimation. K-fold cross validation requires first splitting the data into folds $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K$. The value of the tuning parameter is selected to minimize

$$\text{CV}_F(\lambda) = \arg\min_\lambda K^{-1}\sum_{k=1}^K ||\hat{\Sigma}^{(-k)} - \tilde{\Sigma}^{(k)}||_F^2, \tag{5}$$

where $\tilde{\Sigma}^{(k)}$ is the unregularized estimator based on based on $\mathcal{D}_k$, and $\hat{\Sigma}^{(-k)}$ is the regularized estimator under consideration based on the data after holding $\mathcal{D}_k$ out. Using this approach, the size of the training data set is approximately $(K-1)N/K$, and the size of the validation set is approximately N/K (though these quantities are only relevant when subjects have equal numbers of observations). For linear models, it has been shown that cross validation is asymptotically consistent is the ratio of the validation data set size over the training set size goes to 1. See Shao [1993]. This result motivates the reverse cross validation criterion, which is defined as follows:

$$\text{rCV}_F(\lambda) = \arg\min_\lambda K^{-1}\sum_{k=1}^K ||\hat{\Sigma}^{(k)} - \tilde{\Sigma}^{(-k)}||_F^2, \tag{6}$$

where $\tilde{\Sigma}^{(-k)}$ is the unregularized estimator based on based on the data after holding out $\mathcal{D}_k$, and $\hat{\Sigma}^{(k)}$ is the regularized estimator under consideration based on $\mathcal{D}_k$. Per the suggested approach of Fang et al. [2016] based on an extensive simulation study, we use $K = 10$-fold cross validation to select the tuning parameters for both the tapering estimator S^ω and the soft thresholding estimator S^λ. They implement cross validation for a number of element-wise shrinkage estimators for covariance matrices in the Wang [2014] R package, which was used to produce the risk estimates for S^ω and S^λ.

2 Appendix

2.1 Quadratic risk estimates for simulation study 1

Table 11: Risk estimates and corresponding standard errors under quadratic loss, Δ_1 when data are generated according to model **??** for varying combinations of sample size, data dimension and subsampling rates.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.0010	0.0013	0.4702	0.3926	0.3871
	20	0.0007	0.0006	0.8495	0.8301	0.8287
	30	0.0003	0.0004	1.1449	1.1926	1.1924
$N = 100$	10	0.0004	0.0004	0.2072	0.1802	0.1777
	20	0.0002	0.0002	0.3920	0.3858	0.3817
	30	0.0001	0.0001	0.5712	0.6191	0.6109

Table 12: 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 $K = 10$-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.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.0314	0.0411	0.5726	0.5810	0.7758
	20	0.3266	0.7265	2.3130	5.5964	2.7545
	30	5.0696	4.9073	15.1096	765.7206	28.6820
$N = 100$	10	0.0156	0.0147	0.2479	0.2501	0.3544
	20	0.1894	0.2017	1.3177	5.1945	4.7634
	30	2.3876	1.6465	9.8175	488.6801	85.9508

Table 13: 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 $K = 10$-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.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.0562	0.0547	0.5237	0.5810	0.5313
	20	0.7832	0.8934	2.1419	9.5721	9.1421
	30	8.2650	10.6855	15.2842	407.3659	129.7459
$N = 100$	10	0.0376	0.0449	0.2546	0.2556	0.2661
	20	0.6260	0.5967	1.3751	3.3281	1.2759
	30	5.7635	6.2824	7.4750	203.6710	10.0634

Table 14: 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 $K = 10$-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.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.0134	0.0145	0.4169	0.3987	0.3985
	20	0.0590	0.0574	0.8810	0.9078	0.9073
	30	0.1351	0.1362	1.2571	1.2570	1.2575
$N = 100$	10	0.0077	0.0078	0.2263	0.2111	0.2104
	20	0.0549	0.0534	0.4309	0.4127	0.4120
	30	0.1331	0.1320	0.6819	0.6579	0.6565

Table 15: Simulation results for the compound symmetry model under quadratic loss, Δ_1. 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.

		$\hat{\Sigma}_{ssanova}$		S	S^{λ}	S^{ω}
	M	LosoCV	URE			
$N = 50$	10	0.3688	0.3599	0.7872	0.8058	1.4774
	20	0.9770	0.9954	1.6167	1.7840	3.4516
	30	1.6067	1.6151	2.5548	2.4837	4.9027
$N = 100$	10	0.3210	0.3168	0.3913	0.3819	0.8958
	20	0.9793	0.9774	0.8714	0.8479	2.2110
	30	1.6177	1.6032	1.2967	1.2293	3.4968

References

Peter J Bickel and Elizaveta Levina. Regularized estimation of large covariance matrices. *The Annals of Statistics*, pages 199–227, 2008.

T Tony Cai, Cun-Hui Zhang, Harrison H Zhou, et al. Optimal rates of convergence for covariance matrix estimation. *The Annals of Statistics*, 38(4):2118–2144, 2010.

Sheung Hun Cheng and Nicholas J Higham. A modified cholesky algorithm based on a symmetric indefinite factorization. *SIAM Journal on Matrix Analysis and Applications*, 19(4):1097–1110, 1998.

Yixin Fang, Binhuan Wang, and Yang Feng. Tuning-parameter selection in regularized estimations of large covariance matrices. *Journal of Statistical Computation and Simulation*, 86(3):494–509, 2016.

LR Haff. Empirical bayes estimation of the multivariate normal covariance matrix. *The Annals of Statistics*, pages 586–597, 1980.

Adam J Rothman, Elizaveta Levina, and Ji Zhu. Generalized thresholding of large covariance matrices. *Journal of the American Statistical Association*, 104(485):177–186, 2009.

Jun Shao. Linear model selection by cross-validation. *Journal of the American statistical Association*, 88(422):486–494, 1993.

Binhuan Wang. *CVTuningCov: Regularized Estimators of Covariance Matrices with CV Tuning*, 2014. URL `https://CRAN.R-project.org/package=CVTuningCov`. R package version 1.0.