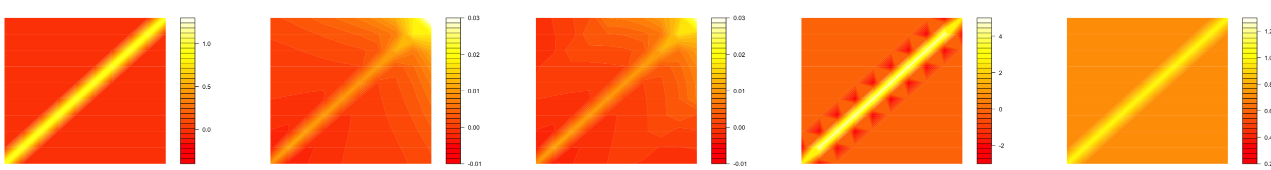
Nonparametric Covariance Estimation for Longitudinal Data via Penalized Tensor Product Splines

Tayler A. Blake* Yoonkyung Lee†

March 2, 2018

1 Performance assessment via simulation study

To understand the strengths and weaknesses of our method, 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. We study estimator performance for five covariance structures, which were chosen to exhibit varying degrees of structural complexity. The two-dimensional surfaces corresponding to each of these are show in Figure **??**. At one end of the spectrum, we consider covariance corresponding to mutual independence. It is both the simplest and sparsest structure, having constant zero-valued varying coefficient function, and constant innovation variance function. The second covariance structure is that of a heterogeneous AR model, where the autoregressive coefficient are a linear function of t, and the IVs are constant over the time domain. The third covariance structure corresponds to the same linear autoregressive model, but we truncate the autoregressive coefficients for large $l = t - s \in [0, 1]$. Lastly, we consider the compound symmetric model, shown rightmost in Figure **??**.

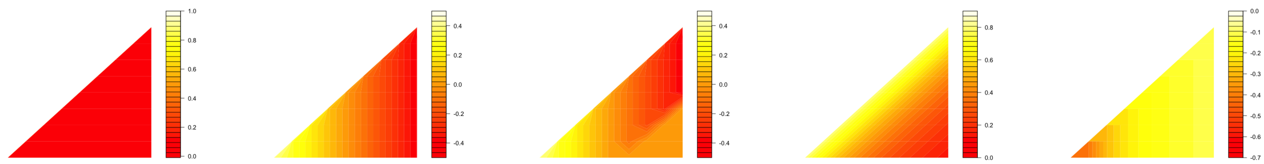


(a) I (b) II (c) III (d) IV (e) V

Figure 1: True covariance surfaces under simulation Model I - Model V

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

The two-dimensional surface corresponding to the Cholesky factor T associated with the inverse of each of the underlying covariances used for simulation are shown in Figure **??**. Comparison with Figure **??** connects the Cholesky surface (and the innovation variance function) with its corresponding covariance surface.



(a) I (b) II (c) III (d) IV (e) V

Figure 2: True cholesky surfaces corresponding to Model I - Model V

Covariance structures exhibiting sparsity or parsimony do not necessarily exhibit the same simplicity in the components of the Cholesky decomposition. For example, Model V in Figure **??** and Figure **??** show the covariance surface and the Cholesky factor for the compound symmetric model. It can be represented with exactly two parameters. However, the ϕ and $\log \sigma^2$ corresponding to the Cholesky decomposition $D = T\Sigma T'$ are nonlinear in t.

Similarly, sparsity in the covariance matrix itself does not imply sparsity in the inverse or in the components of T. For example, the covariance matrix corresponding to the moving average model of order $q = 1$ is banded, which elements $\sigma_{ij} = 0$ when $|i - j| > 1$. Its surface is shown in column IV of Figure **??**. The elements on the first off diagonal are non-zero, and zero on every other off-diagonal. However, the inverse covariance matrix as well as the corresponding Cholesky factor are not sparse. The elements of both T and Σ^{-1} decay geometrically in distance from the main diagonal, $|i - j|$. The Cholesky factor for the MA(1) model is shown in column IV of Figure **??**. A similar relationship holds for the relationship between the Cholesky factor and covariance surface for Model III, where the Cholesky structure exhibits sparsity, but the corresponding covariance structure does not.

For each of the general covariance structures outlined in the previous simulation study description, data were simulated according to multivariate normal distributions with the following covariance matrices:

I. Mutual independence: $\Sigma =$ I, where

$$\phi(t, s) = 0, \quad 0 \le s < t \le 1,$$
$$\sigma^2(t) = 1, \quad 0 \le t \le 1.$$

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

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

TODO: How do we describe the structures for models II-V in terms of continuous $t \in [0,1]$?

III. $k_{1/2}$-banded linear varying coefficient model with constant innovation variance: $\Sigma^{-1} = T'D^{-1}T$, where

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

IV. geometrically decaying GARPs with constant innovation variance: $\Sigma^{-1} = T'D^{-1}T$ where

$$\phi(t,s) = 0.6^{t-s}, \quad 0 \leq s < t \leq 1,$$
$$\sigma^2(t) = \frac{4}{3}, \quad 0 \leq t \leq 1.$$

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

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

2 Loss functions and corresponding risk measures

Regularized estimators are typically obtained by minimizing appropriate norms or risk functions. To assess performance of an estimator $\hat{\Sigma}$, we consider two loss functions commonly used when the total number of observations n_Y is greater than the dimension M:

$$\Delta_1\left(\Sigma, \hat{\Sigma}\right) = tr\left(\left(\Sigma^{-1}\hat{\Sigma} - \mathrm{I}\right)^2\right), \tag{1}$$

$$\Delta_2\left(\Sigma, \hat{\Sigma}\right) = tr\left(\Sigma^{-1}\hat{\Sigma}\right) - log|\Sigma^{-1}\hat{\Sigma}| - M. \tag{2}$$

Σ denotes 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 discrepancy 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

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

The quadratic loss consequently penalizes overestimates more than underestimates, so "smaller" estimates are favored more under Δ_2 than Δ_1. For example, among the class of estimators comprised of scalar multiples cS of the sample covariance matrix, Haff [1980] established 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.

3 Performance benchmarking with complete data

The first of our two primary goals in this simulation study is to assess the utility of our proposed methods under the covariance structures discussed in the previous section. We gauge the performance of our estimator by comparing it to that of four additional covariance estimators serving as points of reference in benchmarking performance: the sample covariance matrix S and two of its regularized variants: the soft thresholding estimator of Rothman et al. [2009], S^λ, and the tapering estimator of Cai et al. [2010], S^ω, as well as the MCD polynomial estimator proposed by Pourahmadi [1999], Pan and [2006], Pourahmadi and Daniels [2002].

Rothman et al. [2009] presented a class of generalized thresholding estimators, including the soft-thresholding estimator given by

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

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:

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

where the ω^k_{ij} are given by

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

The weights ω^k_{ij} 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. Without loss of generality, we assume that $k_h = k/2$ is even. The weights may be rewritten as

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

This expression of the weights makes it clear how the selection of k controls the amount of shrinkage applied to different elements of the sample covariance matrix. The estimator applies no shrinkage to elements of S belonging to the subdiagonals closest to the main diagonal. As one moves away from the main diagonal, shrinkage increases. A shrinkage factor of $2 - \frac{i-j}{k_h}$ is applied to elements belonging to subdiagonals $k_h, \ldots, k-1, k$, and elements further than k subdiagonals from the main diagonal are shrunk to zero.

In the spirit of the GLM, the MCD polynimal estimator is a particular case of estimators which model the components of the Cholesky decomposition using covariates. The polynomial estimator takes the GARPs and IVs to be polynomials of lag and time, respectively:

$$\begin{aligned} \phi_{jk} &= z'_{jk}\gamma \\ \log \sigma^2_{jk} &= z'_i \lambda, \end{aligned}$$

for $j = 1, \ldots, M$, $k = 1, \ldots, j-1$. The vectors z_j and z_{jk} are of dimension $q \times 1$ and $p \times 1$ which hold covariates

$$\begin{aligned} z'_{jk} &= \left(1, t_j - t_k, (t_j - t_k)^2, \ldots, (t_j - t_k)^{p-1}\right)', \\ z'_i &= \left(1, t, \ldots, t^{q-1}\right)'. \end{aligned}$$

where polynomial orders p, q are chosen by BIC. For detailed discussion of these estimators, see See Chapter 1, Section **??** and Section **??**. Since construction of the sample covariance matrix S, S^{ω}, and S^{λ} rely on having an equal number of regularly-spaced observations on each subject, these simulations were conducted using complete data with common measurement times across all N subjects. Performance is evaluated for varying sample size N and within-subject sample size M.

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 complete data under entropy loss are presented in Section Section 4.1.1, tables **??** - **??**. We also obtained risk estimates under quadratic loss, which echo conclusions made based on entropy loss. There are left to the Appendix, Table **??**-**??**.

Figure **??** provides a visual summary of how well our estimator as well as each of the alternative estimators correctly identify the underlying covariance structure. The first row in the grid shows the surface plot of each of the true covariance structures. The suface plots of the oracle estimate for each of the corresponding models shown in the second row serve as a point of reference.

Our estimator is stable across all of the underlying covariance structures for the differing number of sampled trajectories $N = 50, 100$, while the performance of the alternative estimators markedly improves when the subject sample size is doubled for each of the generating structures, particularly for the case of $M = 30$. Irrespective of tuning parameter selection method, our estimator is preferable to all three of the alternative estimators, except under Model IV when N is large and within-subject sampling rates are moderate. Under this model, both the inverse covariance as well as the covariance matrix itself are sparse. Specifically, the inverse is banded so that $\sigma^{ij} = 0$ for $|i - j| > 1$, but the non-zero elements are quite large. Inversion results in a covariance matrix which decays quickly as distance from the main diagonal increases, which is in concordance with the assumed structure of the softthresholding estimator.

The covariance matrix corresponding to Model II is highly nonstationary. It is neither sparse, nor has entries which decrease in absolute value as the time between observations increases, which is in discordance with the assumed structure of both element-wise shrinkage estimators. However, on every subdiagonal are entries which are very small. The tapering estimator performs abysmally for this structure, since for almost any choice of k, it will incorrectly be shrinking many entries which are large in absolute values to zero. The soft thresholding estimator assumes no implicit structure of the M measurements which make up the random vector (it does not assume that $y_1, \ldots, y_M$ are time-ordered.) While the covariance is nonstationary, the elements of Σ are highly structured, but the soft-thresholding esitmator fails to exploit this structure which results in S^{λ} having 0s spuriously placed. The covariance matrix under Model III has similar structure, presenting similar difficulties for both estimators. The sample covariance matrix far outperforms both of its regularized renditions almost uniformly across subject sample sizes N for moderate within-subject sampling rates ($M = 20, 30$.)

Review of generalized thresholding estimators, including the soft thresholding estimator is presented in in **??**. Recall that S^{λ} 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.

Under certain conditions pertaining to the ration of sample sizes of the training and validation datasets, the K-fold cross validation criterion is a consistent estimator of the Frobenius norm risk. It is defined

$$\text{CV}_{F}(\lambda) = \arg\min_{\lambda} K^{-1} \sum_{k=1}^{K} ||\hat{\Sigma}^{(-k)} - \tilde{\Sigma}^{(k)}||_{F}^{2}, \tag{4}$$

There is little established about the optimal method for tuning parameter selection in for the class of estimators based on element-wise shrinkage of the sample covariance matrix. However, based on the results of an extensive simulation study presented in Fang et al. [2016], 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 authors implement cross validation for a number of element-wise shrinkage estimators for covariance matrices in the Wang [2014] R package, which was used to calculate the risk estimates for S^{ω} and S^{λ}.

Element-wise shrinkage estimators of the covariance matrix, including the soft thresholding estimator, are 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. In this case, the estimate has at least one eigenvalue less than or equal to zero, and the evaluation of the entropy loss 2 is undefined. 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.

Pan and Mackenzie [2003] present an iterative procedure for estimating coefficient vectors λ, γ of the polynomial model **??**. Their algorithm uses a quasi-Newton step for computing the MLE under the multivariate normal likelihood. Their work is is implemented in the JMCM package for R, which we used to compute the polynomial MCD estimates. For implementation details, see Pan and Pan [2017].

We need to decide which tables will be included in the non-appendix numerical results. In the actual dissertation, section 6 will not immediate follow section 5.

Figure 3: Covariance Model I - Model V used for simulation and corresponding estimates. The columns in the grid correspond to each simulation model. The first row of shows the true covariance structure, and each row beneath corresponds to each of the estimators.

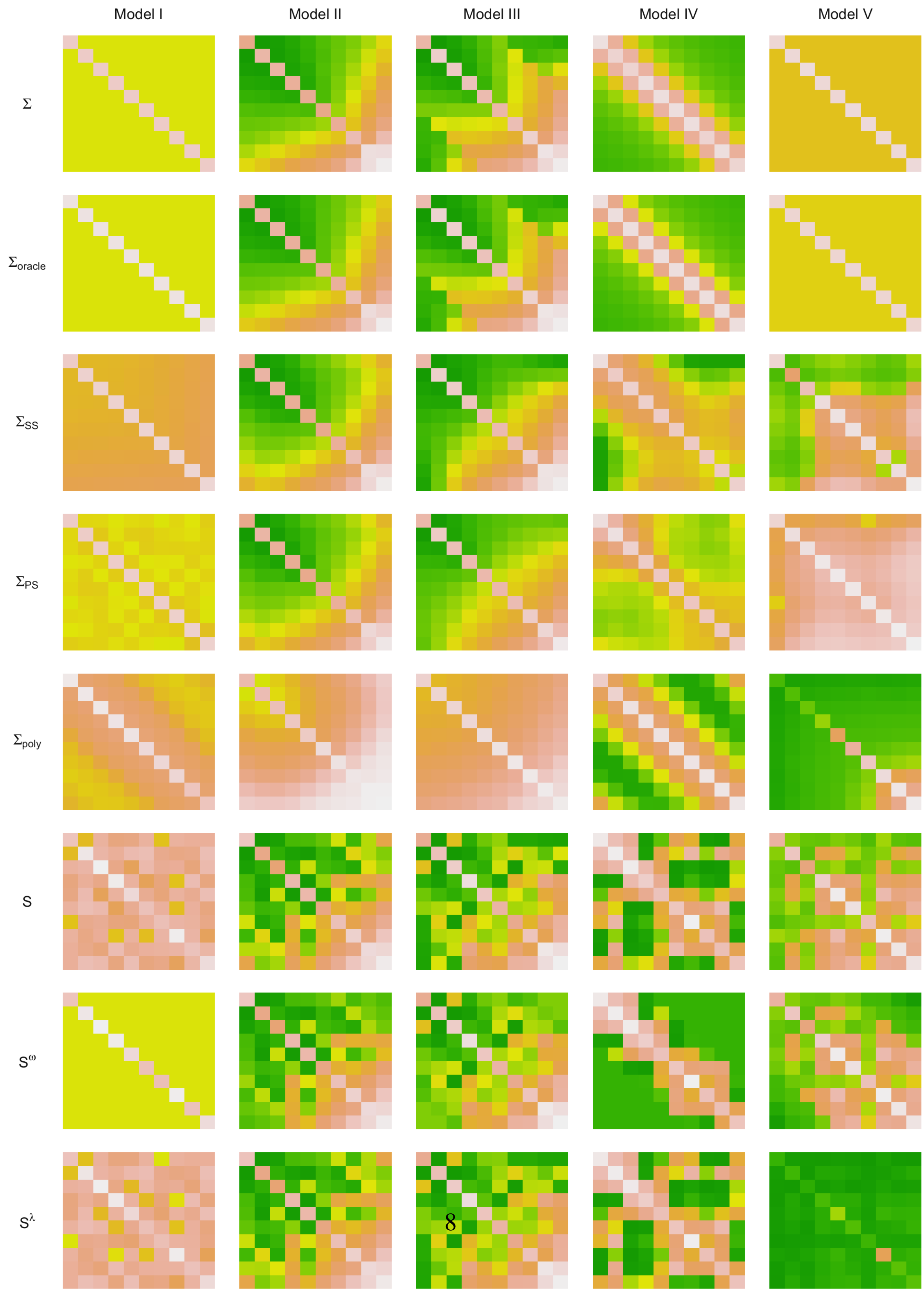


Figure 4: The true lower triangle of Cholesky factor T corresponding to Model I - Model V and estimates of the same surface for estimators based on the modified Cholesky decomposition. The true covariance structure is displayed across the top row.

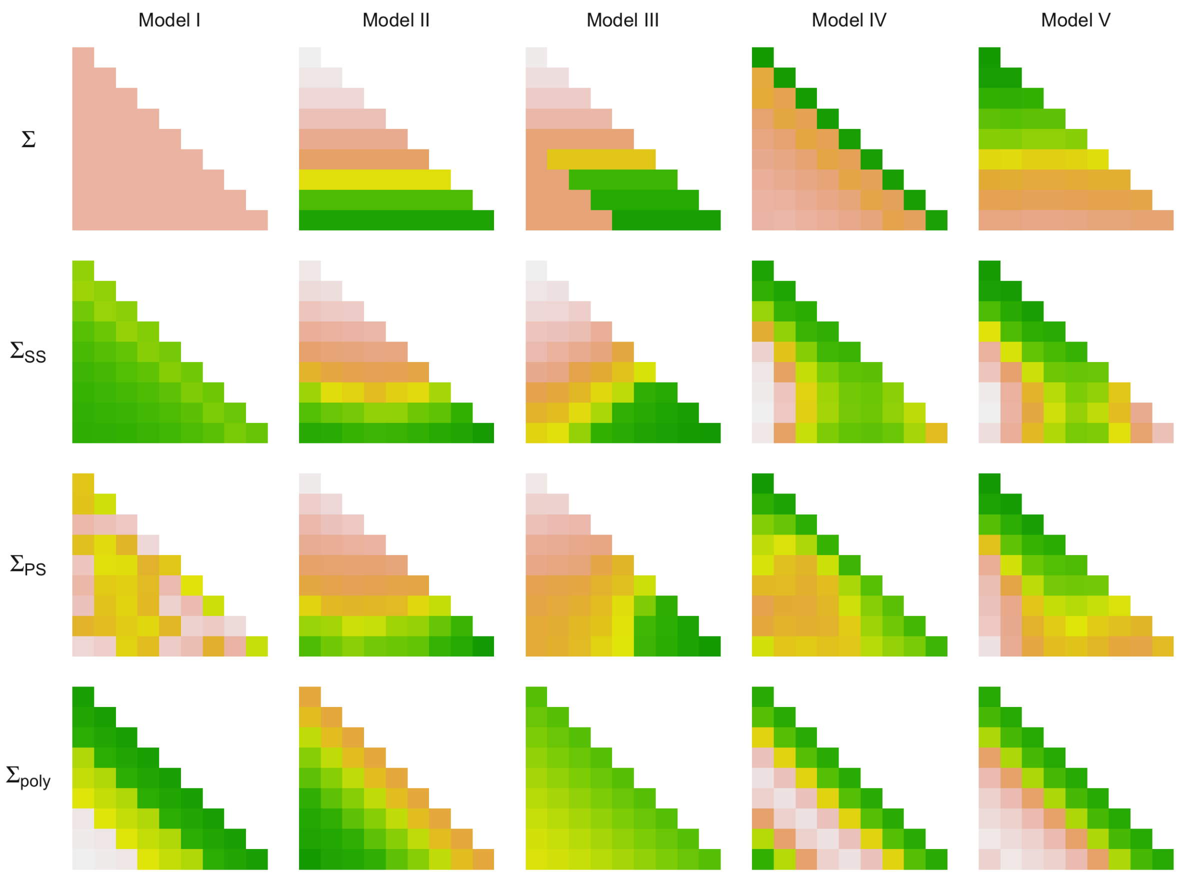
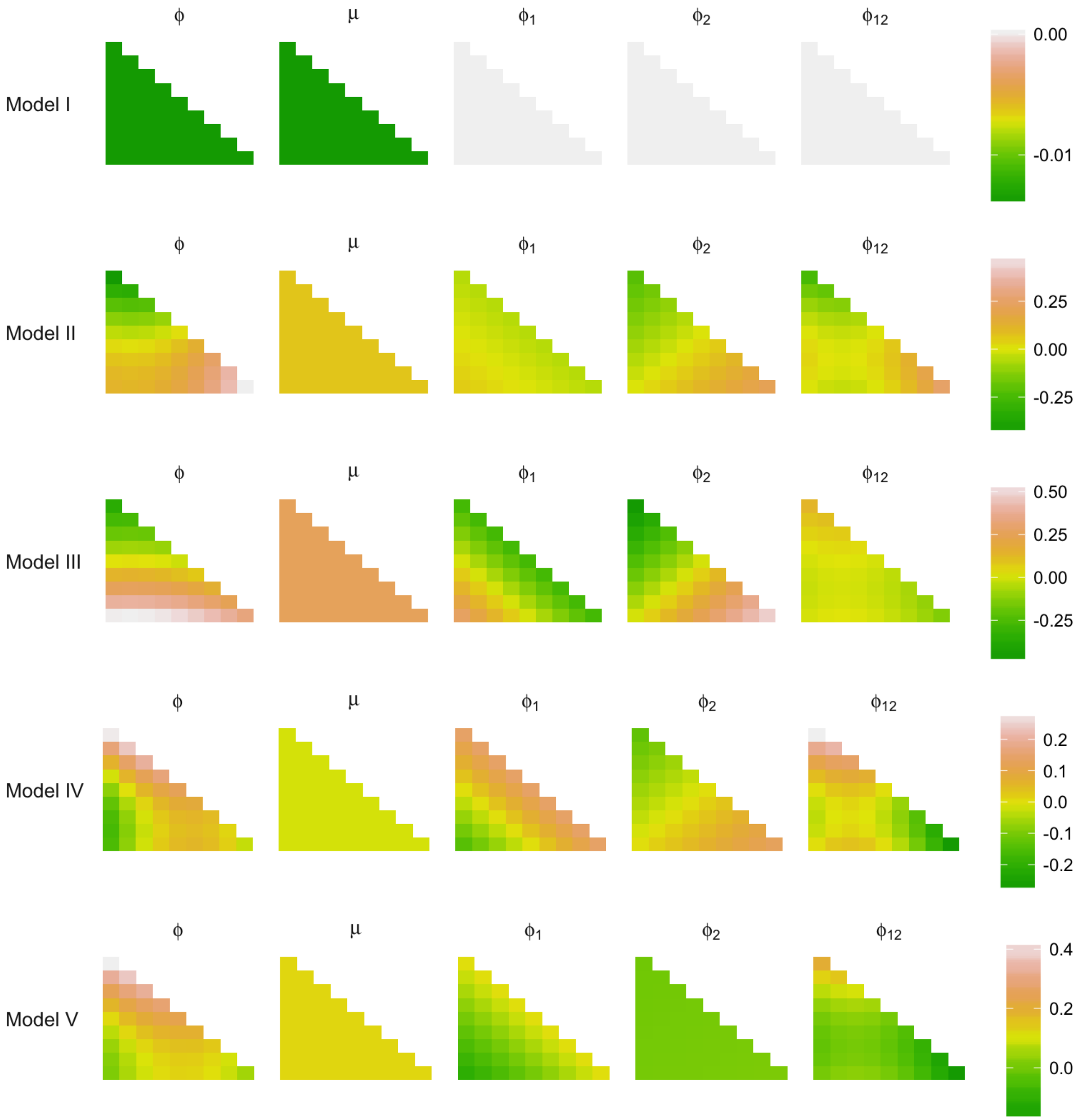


Figure 5: Estimated functional components of the smoothing spline ANOVA decomposition $\phi = \phi_1 + \phi_2 + \phi_{12}$ of the MCD smoothing spline estimator for each of simulation model.



4 Performance with irregularly sampled data

Our second concern in evaluation of our methods is how performance changes when the data exhibit varying degrees of sparsity. We fix the number of sampled trajectories N and vary M, the size of the set of possible measurement times

$$t_1, \ldots, t_M.$$

We generate irregular data by first generating a complete dataset

$$\begin{aligned}
Y_1 &= (y_1(t_1), y_1(t_2), \ldots, y_1(t_M))' \\
Y_2 &= (y_2(t_1), y_2(t_2), \ldots, y_2(t_M))' \\
&\vdots \\
Y_N &= (y_N(t_1), y_N(t_2), \ldots, y_N(t_M))',
\end{aligned}$$

where $Y_1, \ldots, Y_N$ are independently and identically distributed according to an M-dimensional multivariate Normal distribution with mean zero and having covariance structure identical to one of Models I - V in 1. To induce sparsity, we subsample from the complete data $\{y_i(t_j)\}$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, randomly omitting an observation $y_i(t_j)$ with probability 0.05, 0.07, and 0.09.

Results under quadratic loss and entropy loss are given in Section **??**, tables **??** - **??**. Standard errors of the risk estimates are left to the appendix; see Table **??** and Table **??**. Performance degradation of the estimator in the presence of missing data is highly dependent on the underlying structure of the Cholesky factor of the inverse covariance matrix. For the identity matrix and for the non-truncated linear varying coefficient GARP model, we observe little change in estimated entropy risk for within subject sample sizes $M = 10$ and $M = 20$ with downsampling as compared to the estimated risk for both sample sizes in the complete data case.

Making the same comparison for the banded Cholesky factor having linear varying coefficient function truncated at $t = 0.5$, we see only slight decreases in performance for $M = 10$: an estimated entropy risk of 0.3174 with no missing data versus 0.3451 (0.3498, 0.3437) with 5% (7%, 9%) missing data. The degredation is more pointed for the moderate sample size of $M = 20$. The rate of missing observations has the greatest impact for the simulation conducted using the compound symmetric model. This is not surprising, since it corresponds to the Cholesky factor having the most complex structure. While the functions defining the Cholesky factors of Models III and IV do not belong to the null space defined by the cubic smoothing spline penalty, they are both piecewise functions with each piece itself belonging to $\mathcal{H}_0$.

Should the discussion that immediately follows be moved to after the tables containing non-appendix numerical results?

Should the discussion of study # 1 be with the tablse for study 1, separate from the discussion + tables for study 2?

TODO: remember to cite the nlme package for fitting the MA(1) and CS oracle models.

Performance degradation of the estimator in the presence of missing data is highly dependent on the underlying structure of the Cholesky factor of the inverse covariance matrix. For the identity matrix and for the non-truncated linear varying coefficient GARP model, we observe little change in estimated entropy risk for within subject sample sizes $M = 10$ and $M = 20$ with downsampling as compared to the estimated risk for both sample sizes in the complete data case. Making the same comparison for the banded Cholesky factor having linear varying coefficient function truncated at $t = 0.5$, we see only slight decreases in performance for $M = 10$: an estimated entropy risk of 0.3174 with no missing data versus 0.3451 (0.3498, 0.3437) with 5% (7%, 9%) missing data. The degredation is more pointed for the moderate sample size of $M = 20$. The rate of missing observations has the greatest impact for the simulation conducted using the compound symmetric model. This is not surprising, since it corresponds to the Cholesky factor having the most complex structure. While the functions defining the Cholesky factors of Models III and IV do not belong to the null space defined by the cubic smoothing spline penalty, they are both piecewise functions with each piece itself belonging to $\mathcal{H}_0$.

Should the discussion that immediately follows be moved to after the tables containing non-appendix numerical results?

Should the discussion of study # 1 be with the tablse for study 1, separate from the discussion + tables for study 2?

Review of generalized thresholding estimators, including the soft thresholding estimator is presented in in **??**. Recall that S^{λ} can be written as the solution to the optimization problem

$$s_{\lambda}\left(z\right) = \arg\min_{\sigma} \left[\frac{1}{2}\left(\sigma - z\right)^2 + J_{\lambda}\left(\sigma\right)\right], \tag{5}$$

so that estimation of the covariance matrix can be accomplished by solving multiple univariate Lasso-penalized least squares problems.

Under certain conditions pertaining to the ration of sample sizes of the training and validation datasets, the K-fold cross validation criterion is a consistent estimator of the Frobenius norm risk. It is defined

$$\text{CV}_{F}\left(\lambda\right) = \arg\min_{\lambda} K^{-1} \sum_{k=1}^{K} ||\hat{\Sigma}^{(-k)} - \tilde{\Sigma}^{(k)}||_{F}^2, \tag{6}$$

There is little established about the optimal method for tuning parameter selection in for the class of estimators based on element-wise shrinkage of the sample covariance matrix. However, based on the results of an extensive simulation study presented in Fang et al. [2016], 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 authors implement cross validation for a number of element-wise shrinkage estimators for covariance matrices in the Wang [2014] R package, which was used to calculate the risk estimates for S^{ω} and S^{λ}.

Element-wise shrinkage estimators of the covariance matrix, including the soft thresholding estimator, are 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. In this case, the estimate has at least one eigenvalue less than or equal to zero, and the evaluation of the entropy loss 2 is undefined. 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.

Pan and Mackenzie [2003] present an iterative procedure for estimating coefficient vectors λ, γ of the polynomial model **??**. Their algorithm uses a quasi-Newton step for computing the MLE under the multivariate normal likelihood. Their work is is implemented in the JMCM package for R, which we used to compute the polynomial MCD estimates. For implementation details, see Pan and Pan [2017].

We need to decide which tables will be included in the non-appendix numerical results. In the actual dissertation, section 6 will not immediate follow section 5.

4.1 Numerical results

4.1.1 Simulation study 1: 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 I.

		$\hat{\Sigma}_{SS}$		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.

		$\hat{\Sigma}_{SS}$		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	8.5141	11.3740
	30	0.0372	0.0334	5.2817	16.5003	89.3414

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.

		$\hat{\Sigma}_{SS}$		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.

		$\hat{\Sigma}_{SS}$		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.

	M	$\hat{\Sigma}_{SS}$ LosoCV	URE	S	S^{λ}	S^{ω}
$N = 50$	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

4.2 Simulation study 2: irregularly sampled data

M	% subsampling	$\hat{\Delta}_1$		$\hat{\Delta}_2$	
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 I and smoothing parameters are selected using the unbiased risk estimate.

M	% subsampling	$\hat{\Delta}_1$		$\hat{\Delta}_2$	
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 and smoothing parameters are selected using the unbiased risk estimate.

M	% subsampling	$\hat{\Delta}_1$		$\hat{\Delta}_2$	
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 and smoothing parameters are selected using the unbiased risk estimate.

M	% subsampling	$\hat{\Delta}_1$		$\hat{\Delta}_2$	
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 and smoothing parameters are selected using the unbiased risk estimate.

M	% subsampling	$\hat{\Delta}_1$		$\hat{\Delta}_2$	
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 and smoothing parameters are selected using the unbiased risk estimate.

5 Appendix

5.1 Quadratic risk estimates for simulation study 1

??

Table 11: Risk estimates and corresponding standard errors for our proposed estimator under quadratic loss, Δ_1 when the data are generated according to model I.

		$\hat{\Sigma}_{SS}$		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: Risk estimates and corresponding standard errors for our proposed estimator under quadratic loss, Δ_1 when the data are generated according to model II.

		$\hat{\Sigma}_{SS}$		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: Risk estimates and corresponding standard errors for our proposed estimator under quadratic loss, Δ_1 when the data are generated according to model III.

		$\hat{\Sigma}_{SS}$		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: Risk estimates and corresponding standard errors for our proposed estimator under quadratic loss, Δ_1 when the data are generated according to model IV.

		$\hat{\Sigma}_{SS}$		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: Risk estimates and corresponding standard errors for our proposed estimator under quadratic loss, Δ_1 when the data are generated according to model V.

		$\hat{\Sigma}_{SS}$		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

5.2 Simulation study 1: risk estimates (standard errors)

Should we decide to keep the following tables in the chapter, the analogous tables for S^{λ} and S^{ω} will also be included here as well.

Model	N	M	$\hat{\Delta}_1(\hat{\Sigma}_{SS})$		$\hat{\Delta}_2(\hat{\Sigma}_{SS})$		$\hat{\Delta}_1(S)$		$\hat{\Delta}_2(\hat{\Sigma}_{SS})$	
1	50	10	0.0023	(0.0006)	0.0843	(0.0099)	0.3915	(0.0262)	1.1924	(0.0348)
1	50	20	0.0009	(0.0002)	0.0804	(0.0062)	0.7953	(0.0365)	5.0792	(0.0676)
1	50	30	0.0006	(0.0001)	0.0822	(0.0064)	1.2408	(0.0460)	12.2989	(0.1174)
1	100	10	0.0007	(0.0001)	0.0453	(0.0042)	0.1901	(0.0107)	0.5800	(0.0133)
1	100	20	0.0003	(0.0001)	0.0421	(0.0034)	0.4025	(0.0199)	2.3150	(0.0273)
1	100	30	0.0002	(0.0001)	0.0407	(0.0040)	0.5914	(0.0224)	5.2940	(0.0476)
2	50	10	0.0709	(0.0063)	0.3298	(0.0082)	0.5168	(0.0359)	1.2156	(0.0318)
2	50	20	1.0948	(0.1234)	1.1179	(0.0108)	2.3802	(0.1604)	5.0130	(0.0664)
2	50	30	13.9982	(1.9602)	2.3284	(0.0161)	22.5542	(2.8650)	12.3822	(0.1101)
2	100	10	0.0449	(0.0022)	0.2955	(0.0047)	0.2515	(0.0145)	0.5566	(0.0127)
2	100	20	0.6322	(0.0421)	1.0638	(0.0064)	1.1628	(0.0925)	2.3893	(0.0252)
2	100	30	7.1979	(0.6928)	2.2805	(0.0091)	10.7818	(1.4529)	5.2753	(0.0418)
3	50	10	0.0573	(0.0086)	0.0753	(0.0061)	0.5234	(0.0369)	1.2228	(0.0308)
3	50	20	0.8747	(0.1197)	0.1025	(0.0085)	2.8719	(0.2644)	5.0775	(0.0733)
3	50	30	8.1496	(1.5069)	0.0958	(0.0077)	24.8586	(3.9217)	12.5350	(0.1127)
3	100	10	0.0200	(0.0028)	0.0329	(0.0028)	0.2642	(0.0217)	0.5750	(0.0165)
3	100	20	0.3360	(0.0624)	0.0387	(0.0038)	1.4008	(0.1128)	2.3517	(0.0290)
3	100	30	3.6555	(1.0573)	0.0382	(0.0043)	9.6946	(1.1953)	5.2919	(0.0413)
4	50	10	0.0170	(0.0015)	0.2812	(0.0067)	0.4254	(0.0273)	1.2228	(0.0338)
4	50	20	0.0600	(0.0012)	0.8899	(0.0082)	0.9665	(0.0423)	5.1032	(0.0639)
4	50	30	0.1378	(0.0015)	1.6220	(0.0074)	1.1690	(0.0417)	12.3825	(0.1139)
4	100	10	0.0088	(0.0005)	0.2252	(0.0046)	0.1941	(0.0102)	0.5676	(0.0187)
4	100	20	0.0543	(0.0007)	0.8514	(0.0043)	0.4281	(0.0221)	2.2750	(0.0250)
4	100	30	0.1333	(0.0009)	1.5826	(0.0037)	0.6650	(0.0219)	5.2777	(0.0408)
5	50	10	0.3956	(0.0207)	0.2900	(0.0078)	0.8750	(0.0619)	1.2395	(0.0400)
5	50	20	0.9995	(0.0110)	0.7660	(0.0063)	1.8312	(0.0745)	5.0307	(0.0719)
5	50	30	1.6198	(0.0134)	1.1976	(0.0090)	2.5880	(0.1102)	12.4199	(0.0979)
5	100	10	0.3194	(0.0065)	0.2407	(0.0037)	0.4209	(0.0284)	0.5530	(0.0115)
5	100	20	0.9774	(0.0060)	0.7299	(0.0041)	0.8714	(0.0339)	2.2297	(0.0283)
5	100	30	1.6032	(0.0088)	1.1813	(0.0051)	1.2967	(0.0474)	5.3014	(0.0526)

Table 16: Risk estimates and corresponding standard errors for our proposed estimator and the sample covariance matrix under models I - V where smoothing parameters are selected using the unbiased risk estimate.

The following table will be formatted just as the one immediately before it if we decide to include it in the chapter.

Model	N	M	$\hat{\Delta}_1(\hat{\Sigma}_{SS})$		$\hat{\Delta}_2(\hat{\Sigma}_{SS})$		$\hat{\Delta}_1(S)$		$\hat{\Delta}_2(S)$	
1	50	10	0.0031	0.0009	0.0941	0.0115	0.4624	0.0249	1.2584	0.0316
1	50	20	0.0015	0.0003	0.1065	0.0105	0.8609	0.0543	5.0851	0.0660
1	50	30	0.0005	0.0001	0.0778	0.0075	1.1356	0.0396	12.5491	0.1248
1	100	10	0.0006	0.0001	0.0477	0.0045	0.2092	0.0124	0.5911	0.0143
1	100	20	0.0003	0.0000	0.0410	0.0037	0.3891	0.0161	2.3162	0.0278
1	100	30	0.0002	0.0000	0.0457	0.0042	0.5724	0.0211	5.2742	0.0498
2	50	10	0.0715	0.0073	0.3476	0.0091	0.5318	0.0415	1.2190	0.0301
2	50	20	0.8640	0.1042	1.1204	0.0117	2.2754	0.1813	5.0921	0.0713
2	50	30	14.6715	2.0991	2.3440	0.0144	16.2024	1.6967	12.4749	0.1194
2	100	10	0.0415	0.0026	0.2854	0.0034	0.2661	0.0170	0.5510	0.0138
2	100	20	0.6597	0.0495	1.0718	0.0071	1.3733	0.1096	2.3569	0.0304
2	100	30	6.8953	0.6368	2.2727	0.0082	7.3618	0.8281	5.4390	0.0428
3	50	10	0.0498	0.0083	0.0736	0.0064	0.5675	0.0425	1.2447	0.0353
3	50	20	0.4704	0.0763	0.0972	0.0088	2.5606	0.2129	5.0612	0.0705
3	50	30	8.2884	1.6398	0.0823	0.0089	18.9040	2.3564	12.4416	0.1086
3	100	10	0.0245	0.0046	0.0379	0.0042	0.2558	0.0163	0.5506	0.0134
3	100	20	0.3433	0.0729	0.0404	0.0041	1.4803	0.1345	2.3231	0.0304
3	100	30	3.3560	0.6491	0.0392	0.0036	9.3395	0.8295	5.2862	0.0421
4	50	10	0.0163	0.0019	0.2692	0.0081	0.4266	0.0286	1.1800	0.0319
4	50	20	0.0590	0.0010	0.8836	0.0074	0.8810	0.0399	5.0899	0.0724
4	50	30	0.1365	0.0015	1.6123	0.0084	1.2551	0.0433	12.5609	0.1101
4	100	10	0.0088	0.0008	0.2246	0.0046	0.2216	0.0131	0.5639	0.0168
4	100	20	0.0551	0.0007	0.8476	0.0037	0.4292	0.0198	2.2649	0.0295
4	100	30	0.1342	0.0010	1.5769	0.0035	0.6775	0.0222	5.2374	0.0480
5	50	10	0.4017	0.0216	0.2988	0.0117	0.7943	0.0523	1.2061	0.0395
5	50	20	0.9817	0.0105	0.7555	0.0051	1.6034	0.0775	5.0172	0.0601
5	50	30	1.6266	0.0135	1.2043	0.0083	2.5378	0.0861	12.5483	0.1092
5	100	10	0.3307	0.0101	0.2507	0.0058	0.3969	0.0235	0.5751	0.0133
5	100	20	6.0835	4.9816	1.3637	0.5844	0.8470	0.0327	2.2673	0.0300
5	100	30	3.2806	1.6330	1.5457	0.3612	1.2599	0.0439	5.2507	0.0499

Table 17: Risk estimates and corresponding standard errors for our proposed estimator and the sample covariance matrix under models I - V where smoothing parameters are selected using losoCV.

Model	N	M	$\hat{\Delta}_1(\hat{\Sigma}^{ure}_{SS})$		$\hat{\Delta}_1(\hat{\Sigma}^{loso}_{SS})$		$\hat{\Delta}_1(\hat{\Sigma}^{loso}_{PS})$		$\hat{\Delta}_1(S)$		$\hat{\Delta}_1(S^{\omega})$		$\hat{\Delta}_1(S^{\lambda})$	
1	50	10	0.0023	(6e-040)	0.0017	(3e-040)	0.0052	(0.0010)	0.3915	(0.0262)	0.3655	(0.0197)	0.3703	(0.0
1	50	20	0.0009	(2e-040)	0.0024	(7e-040)	0.0043	(6e-040)	0.7953	(0.0365)	0.7986	(0.0360)	0.7983	(0.0
1	50	30	0.0006	(1e-040)	0.0015	(9e-040)	0.0036	(6e-040)	1.2408	(0.0460)	1.2364	(0.0504)	1.2519	(0.0
1	100	10	0.0007	(1e-040)	0.0006	(1e-040)	0.0010	(1e-040)	0.1901	(0.0107)	0.1938	(0.0130)	0.1961	(0.0
1	100	20	0.0003	(000000)	0.0003	(000000)	0.0011	(1e-040)	0.4025	(0.0199)	0.3944	(0.0181)	0.4001	(0.0
1	100	30	0.0002	(000000)	0.0002	(000000)	0.0011	(1e-040)	0.5914	(0.0224)	0.5695	(0.0232)	0.5865	(0.0
2	50	10	0.0709	(0.0063)	0.0666	(0.0056)	0.0623	(0.0043)	0.5168	(0.0359)	0.5642	(0.0393)	0.5153	(0.0
2	50	20	1.0948	(0.1234)	0.8882	(0.0952)	3.3623	(0.2241)	2.3802	(0.1604)	5.1056	(1.3636)	15.5901	(3.0
2	50	30	13.9982	(1.9602)	12.6406	(1.3045)	110.7368	(12.1364)	22.5542	(2.8650)	97.0180	(35.8215)	1507.7082	(330
2	100	10	0.0449	(0.0022)	0.0415	(0.0026)	0.0525	(0.0033)	0.2515	(0.0145)	0.2516	(0.0144)	0.2467	(0.0
2	100	20	0.6322	(0.0421)	0.6597	(0.0495)	3.3059	(0.2199)	1.1628	(0.0925)	1.5125	(0.1383)	3.3054	(0.
2	100	30	7.1979	(0.6928)	6.8953	(0.6368)	71.0815	(5.4194)	10.7818	(1.4529)	12.5811	(1.6436)	474.4443	(116.
3	50	10	0.0573	(0.0086)	0.0457	(0.0056)	0.0665	(0.0081)	0.5234	(0.0369)	0.9081	(0.1014)	0.5890	(0.0
3	50	20	0.8747	(0.1197)	0.6152	(0.0951)	3.2706	(0.2149)	2.8719	(0.2644)	19.5830	(6.4061)	19.9996	(3.0
3	50	30	8.1496	(1.5069)	8.1968	(1.1830)	199.3705	(21.3899)	24.8586	(3.9217)	181.0963	(23.2385)	1323.0063	(351.
3	100	10	0.0200	(0.0028)	0.0245	(0.0046)	0.0363	(0.0047)	0.2642	(0.0217)	0.3915	(0.0285)	0.2841	(0.0
3	100	20	0.3360	(0.0624)	0.3433	(0.0729)	3.1740	(0.2060)	1.4008	(0.1128)	6.1001	(0.7095)	12.7102	(2.0
3	100	30	3.6555	(1.0573)	3.3560	(0.6491)	98.3929	(8.2390)	9.6946	(1.1953)	83.8858	(7.6284)	682.4404	(205.0
4	50	10	0.0170	(0.0015)	0.0157	(0.0012)	0.0088	(8e-040)	0.4254	(0.0273)	0.5219	(0.0364)	0.5227	(0.0
4	50	20	0.0600	(0.0012)	0.0601	(0.0013)	0.0393	(0.0013)	0.9665	(0.0423)	0.8922	(0.0426)	0.8920	(0.0
4	50	30	0.1378	(0.0015)	0.1445	(0.0039)	0.0944	(0.0014)	1.1690	(0.0417)	1.3120	(0.0451)	1.3116	(0.0
4	100	10	0.0088	(5e-040)	0.0088	(8e-040)	0.0061	(4e-040)	0.1941	(0.0102)	0.2511	(0.0120)	0.2514	(0.0
4	100	20	0.0543	(7e-040)	0.0551	(7e-040)	0.0331	(9e-040)	0.4281	(0.0221)	0.4297	(0.0197)	0.4296	(0.0
4	100	30	0.1333	(9e-040)	0.1342	(0.0010)	0.0896	(8e-040)	0.6650	(0.0219)	0.7035	(0.0248)	0.7036	(0.0
5	50	10	0.3956	(0.0207)	0.3588	(0.0096)	0.2056	(0.0206)	0.8750	(0.0619)	1.6562	(0.1361)	0.8844	(0.0
5	50	20	0.9995	(0.0110)	1.0380	(0.0138)	0.2298	(0.0213)	1.8312	(0.0745)	3.7009	(0.3637)	1.7680	(0.
5	50	30	1.6198	(0.0134)	1.5981	(0.0116)	0.2713	(0.0176)	2.5880	(0.1102)	5.0495	(0.2337)	2.5783	(0.0
5	100	10	0.3194	(0.0065)	0.3307	(0.0101)	0.1023	(0.0112)	0.4209	(0.0284)	1.0553	(0.0569)	0.4926	(0.0
5	100	20	0.9774	(0.0060)	6.0835	(4.9816)	0.1087	(0.0072)	0.8714	(0.0339)	2.3203	(0.0969)	0.8585	(0.0
5	100	30	1.6032	(0.0088)	3.2806	(1.6330)	0.1577	(0.0124)	1.2967	(0.0474)	3.2678	(0.1193)	1.3052	(0.0

Table 18: Quadratic risk estimates and corresponding standard errors based on N = 100 Monte Carlo simulations.

Model	N	M	$\hat{\Delta}_2(\hat{\Sigma}_{SS}^{ure})$		$\hat{\Delta}_2(\hat{\Sigma}_{SS}^{loso})$		$\hat{\Delta}_2(\hat{\Sigma}_{PS}^{loso})$		$\hat{\Delta}_2(S)$		$\hat{\Delta}_2(S^{\omega})$		$\hat{\Delta}_2(S^{\lambda})$	
1	50	10	0.0843	(0.0099)	0.0767	(0.0065)	0.1261	(0.0107)	1.1924	(0.0348)	0.4217	(0.0489)	1.1239	(0.0374)
1	50	20	0.0804	(0.0062)	0.1134	(0.0142)	0.1713	(0.0095)	5.0792	(0.0676)	1.3097	(0.1618)	4.6562	(0.1145)
1	50	30	0.0822	(0.0064)	0.1020	(0.0162)	0.1969	(0.0118)	12.2989	(0.1174)	3.1227	(0.4753)	11.1448	(0.3347)
1	100	10	0.0453	(0.0042)	0.0477	(0.0045)	0.0671	(0.0042)	0.5800	(0.0133)	0.2283	(0.0224)	0.5465	(0.0165)
1	100	20	0.0421	(0.0034)	0.0410	(0.0037)	0.0965	(0.0048)	2.3150	(0.0273)	0.4777	(0.0599)	2.0194	(0.0590)
1	100	30	0.0407	(0.0040)	0.0457	(0.0042)	0.1148	(0.0062)	5.2940	(0.0476)	0.8231	(0.1306)	4.3808	(0.1573)
2	50	10	0.3298	(0.0082)	0.3499	(0.0083)	0.3423	(0.0082)	1.2156	(0.0318)	1.1704	(0.0419)	1.1468	(0.0366)
2	50	20	1.1179	(0.0108)	1.1028	(0.0103)	1.3640	(0.0158)	5.0130	(0.0664)	17.7525	(0.7425)	27.9181	(5.6851)
2	50	30	2.3284	(0.0161)	2.3344	(0.0145)	2.6485	(0.0472)	12.3822	(0.1101)	44.2387	(2.2823)	154.2427	(18.4613)
2	100	10	0.2955	(0.0047)	0.2854	(0.0034)	0.2945	(0.0059)	0.5566	(0.0127)	0.5430	(0.0140)	0.5353	(0.0117)
2	100	20	1.0638	(0.0064)	1.0718	(0.0071)	1.2875	(0.0100)	2.3893	(0.0252)	11.3740	(0.7298)	8.5141	(0.8752)
2	100	30	2.2805	(0.0091)	2.2727	(0.0082)	2.4365	(0.0293)	5.2753	(0.0418)	28.7007	(0.5485)	89.3414	(16.5003)
3	50	10	0.0753	(0.0061)	0.0707	(0.0063)	0.1051	(0.0090)	1.2228	(0.0308)	1.4722	(0.0588)	1.1788	(0.0341)
3	50	20	0.1025	(0.0085)	0.0866	(0.0073)	0.2555	(0.0109)	5.0775	(0.0733)	20.9163	(0.9914)	25.0179	(4.7751)
3	50	30	0.0958	(0.0077)	0.0736	(0.0069)	0.6242	(0.0390)	12.5350	(0.1127)	51.1717	(1.6911)	129.4202	(18.3376)
3	100	10	0.0329	(0.0028)	0.0379	(0.0042)	0.0579	(0.0050)	0.5750	(0.0165)	0.7487	(0.0231)	0.5507	(0.0148)
3	100	20	0.0387	(0.0038)	0.0404	(0.0041)	0.2011	(0.0057)	2.3517	(0.0290)	10.8373	(0.8588)	13.7121	(2.1390)
3	100	30	0.0382	(0.0043)	0.0392	(0.0036)	0.3845	(0.0169)	5.2919	(0.0413)	34.7162	(1.1178)	82.9490	(16.6058)
4	50	10	0.2812	(0.0067)	0.2840	(0.0087)	0.1894	(0.0063)	1.2228	(0.0338)	0.7065	(0.0461)	1.2284	(0.0300)
4	50	20	0.8899	(0.0082)	0.8850	(0.0078)	0.7039	(0.0066)	5.1032	(0.0639)	2.1561	(0.1828)	5.0695	(0.0703)
4	50	30	1.6220	(0.0074)	1.6371	(0.0137)	1.3687	(0.0076)	12.3825	(0.1139)	4.1290	(0.4015)	12.3981	(0.1196)
4	100	10	0.2252	(0.0046)	0.2246	(0.0046)	0.1673	(0.0038)	0.5676	(0.0187)	0.3253	(0.0162)	0.5862	(0.0143)
4	100	20	0.8514	(0.0043)	0.8476	(0.0037)	0.6664	(0.0033)	2.2750	(0.0250)	0.9367	(0.0689)	2.3158	(0.0315)
4	100	30	1.5826	(0.0037)	1.5769	(0.0035)	1.3334	(0.0040)	5.2777	(0.0408)	2.0095	(0.1893)	5.2885	(0.0408)
5	50	10	0.2900	(0.0078)	0.2732	(0.0059)	0.1464	(0.0108)	1.2395	(0.0400)	18.4324	(0.9145)	3.2572	(0.4606)
5	50	20	0.7660	(0.0063)	0.7629	(0.0068)	0.1772	(0.0128)	5.0307	(0.0719)	35.2296	(0.6763)	14.9515	(0.8737)
5	50	30	1.1976	(0.0090)	1.2039	(0.0096)	0.2179	(0.0125)	12.4199	(0.0979)	48.0137	(0.8519)	26.7310	(0.1161)
5	100	10	0.2407	(0.0037)	0.2507	(0.0058)	0.0722	(0.0049)	0.5530	(0.0115)	16.7606	(0.2606)	1.7943	(0.0607)
5	100	20	0.7299	(0.0041)	1.3637	(0.5844)	0.0878	(0.0042)	2.2297	(0.0283)	32.2514	(0.6075)	4.8505	(0.4335)
5	100	30	1.1813	(0.0051)	1.5457	(0.3612)	0.1216	(0.0062)	5.3014	(0.0526)	39.0077	(0.1698)	15.7546	(0.8444)

Table 19: Entropy risk estimates and corresponding standard errors based on N = 100 Monte Carlo simulations.

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.

Jianxin Pan and Gilbert . Regression models for covariance structures in longitudinal studies. *Statistical Modelling*, 6(1):43–57, 2006.

Jianxin Pan and Gilbert Mackenzie. On modelling mean-covariance structures in longitudinal studies. *Biometrika*, 90(1):239–244, 2003.

Jianxin Pan and Yi Pan. jmcm: An r package for joint mean-covariance modeling of longitudinal data. *Journal of Statistical Software*, 82(1):1–29, 2017.

M Pourahmadi and MJ Daniels. Dynamic conditionally linear mixed models for longitudinal data. *Biometrics*, 58(1):225–231, 2002.

Mohsen Pourahmadi. Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterisation. *Biometrika*, 86(3):677–690, 1999.

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.

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.