Nonparametric Covariance Estimation for Longitudinal Data via Penalized Tensor Product Splines

Tayler A. Blake* Yoonkyung Lee†

March 6, 2018

1 Performance assessment via simulation study

1.1 Performance benchmarking with complete data

In this section we compare bivariate spline estimators of the Cholesky factor to other methods of covariance estimation. Our primary comparisons are that with the parametric polynomial estimator proposed by citetpourahmadi1999joint, Pan and Mackenzie [2003], and Pourahmadi and Daniels [2002], which is also based on the modified Cholesky decomposition, and with the oracle estimator, which effectively gives a lower bound on the risk for given covariance structure. As a benchmark, we also include the sample covariance matrix, and two regularized variants of it: the tapered sample covariance matrix (Cai et al. [2010]) and the soft thresholding estimator (Rothman et al. [2009]), which does not rely on a natural ordering among the variables.

Simulations were carried out for five covariance structures: the diagonal covariance with homogenous variances, a heterogenous autoregressive process with linear varying coefficient function, the same heterogeneous process but truncated to zero to band the inverse covariance matrix, the rational quadratic covariance model, and the compound symmetric model. The two-dimensional surfaces corresponding to each of these are shown left to right in Figure **??**. The first row of image plots display the surface which coincides with the appropriate discrete covariance matrix, and in the second row are the surface maps of the corresponding Cholesky factors. Precise models used for simulations are defined in Table 1.1.

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

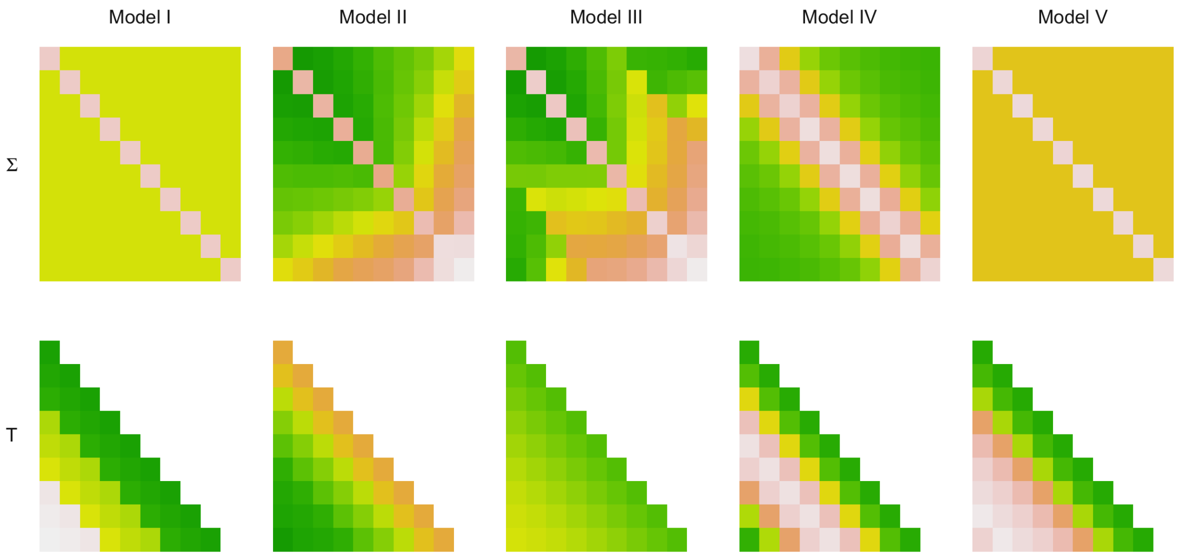


Figure 1: True covariance surfaces (row 1) under simulation Model I - Model V and their corresponding Cholesky factor T (row 2).

Connecting the covariance matrices in first row of Figure **??** with their Cholesky factor in the second row, covariance structures exhibiting sparsity or parsimony do not necessarily exhibit the same simplicity in the components of the Cholesky decomposition. The Cholesky factor for Model III, the truncated linear varying coefficient AR model, is sparse, with elements on the outer half of the subdiagonals equal to zero. While this corresponds to a banded inverse covariance structure, Σ itself is not sparse. The compound symmetric model has simple structure and is parsimonious; its dependence parameters can be expressed as the evaluation of a function which is constant in time t. However, the elements of the Cholesky factor and diagonal matrix $D = T\Sigma T'$ do not exhibit such elementary structure, the elements of which are nonlinear in t.

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: $\Sigma = \mathrm{I}$, 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. $k_{1/2}$-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. Rational quadratic covariance: $\Sigma = (\sigma_{ij})$ where

$$Cov\left(y\left(t_i\right), y\left(t_j\right)\right) = \left(1 + \frac{\left(t_i - t_j\right)^2}{2\alpha k^2}\right)^{-\alpha}, \tag{1}$$

with $k = 0.6$ and $\alpha = 1$.

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

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

For each of the covariance models, we generated a set of observations of sample size $N = 50, 100$ from a multivariate normal distribution, and considered three different values of within-subject sample size $M = 10, 20, 30$. The estimators were computed with tuning parameters selected using both leave-one-subject-out cross validation losoCV (λ) and unbiased risk estimate $\mathrm{U}(\lambda)$. Given the selected values of the tuning parameters, we computed the estimated covariance matrix and compared it to the true covariance matrix via entropy loss and quadratic loss.

1.1.1 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:

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

$$\Delta_2\left(\Sigma, \hat{\Sigma}\right) = tr\left(\Sigma^{-1}\hat{\Sigma}\right) - log|\Sigma^{-1}\hat{\Sigma}| - M. \tag{3}$$

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.

1.1.2 Alternative estimators

The following estimators serve as benchmarks for performance under the five simulation settings outlined above: the MCD polynomial estimator $\hat{\Sigma}_{poly}$, the sample covariance matrix S, the soft thresholding estimator S^λ, and the tapering estimator S^ω. We will review the general definitions of these, but for detailed discussion of the construction and properties of these estimators, see Chapter 1, Section **??**, **??** and **??**.

In the spirit of the GLM, the MCD polynomial 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:

$$\begin{aligned} \phi_{jk} &= z_{jk}'\gamma \\ \log \sigma_{jk}^2 &= z_i'\lambda, \end{aligned}$$

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

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

where polynomial orders p, q are chosen by BIC. Rothman et al. [2009] presented a class of generalized thresholding estimators, including the soft-thresholding estimator 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. The tapering estimator proposed by Cai et al. [2010] is given by

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

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 controlled by a tuning parameter, k, which can take integer values between 0 and M. Without loss of generality, we assume that $k_h = k/2$ is even. The weights may be rewritten as

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

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.

Figure **??** provides a visual representation of the qualitative differences between the estimates resulting from each of the eight methods of estimation for the five covariance structures used for simulation. The first row in the grid shows the surface plot of each of the true covariance structures, and each row thereafter corresponds to the five covariance estimates for the given estimation method. The surface plots of the oracle estimate in the second row serve as a point of reference for the 'gold standard' in each scenario, since the oracle estimates were constructed assuming that the functional form of the covariance is known (either the full covariance structure or the components of the Cholesky decomposition.) The corresponding estimates of the Cholesky factor T for the estimators based on the modified Cholesky decomposition are shown in Figure **??**, and the decomposition of the $\hat{T}$ corresponding to the smoothing spline ANOVA estimator $\hat{\Sigma}_{SS}$ into functional components is displayed in Figure **??**

Figure 2: 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.

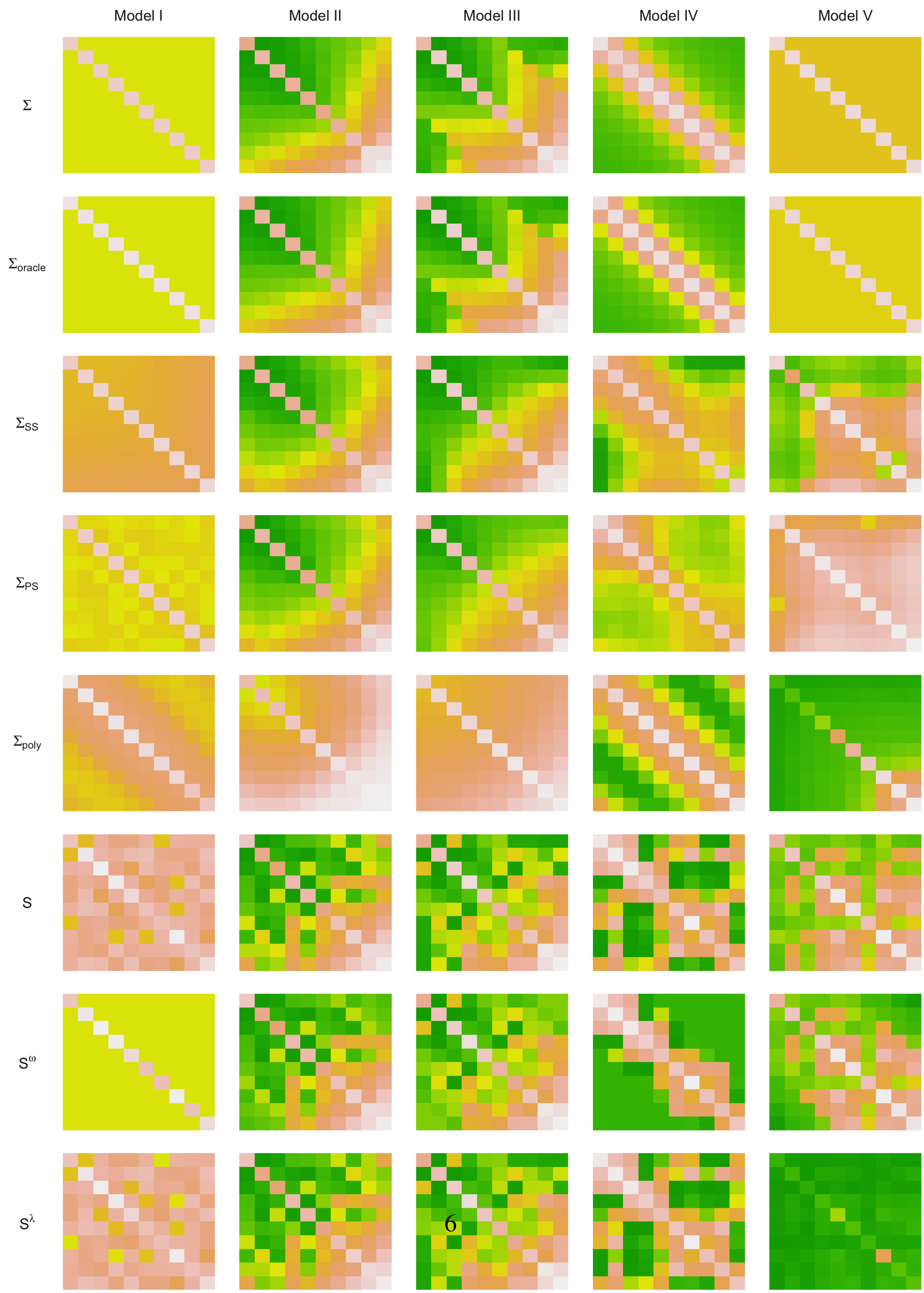


Figure 3: The true lower triangle of Cholesky factor T corresponding to Model I - Model V and estimates of the same surface for estimators based on the modified Cholesky decomposition. The true covariance structure is displayed across the top row.

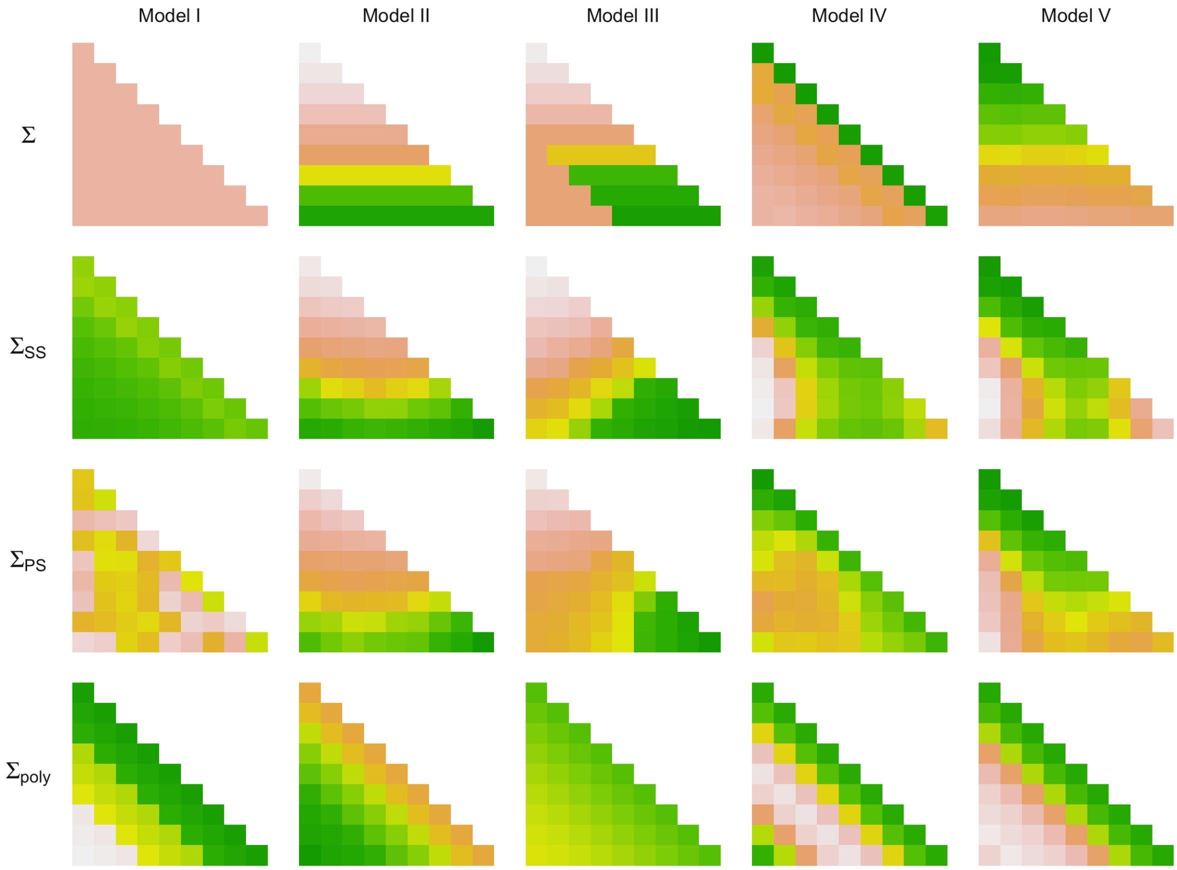
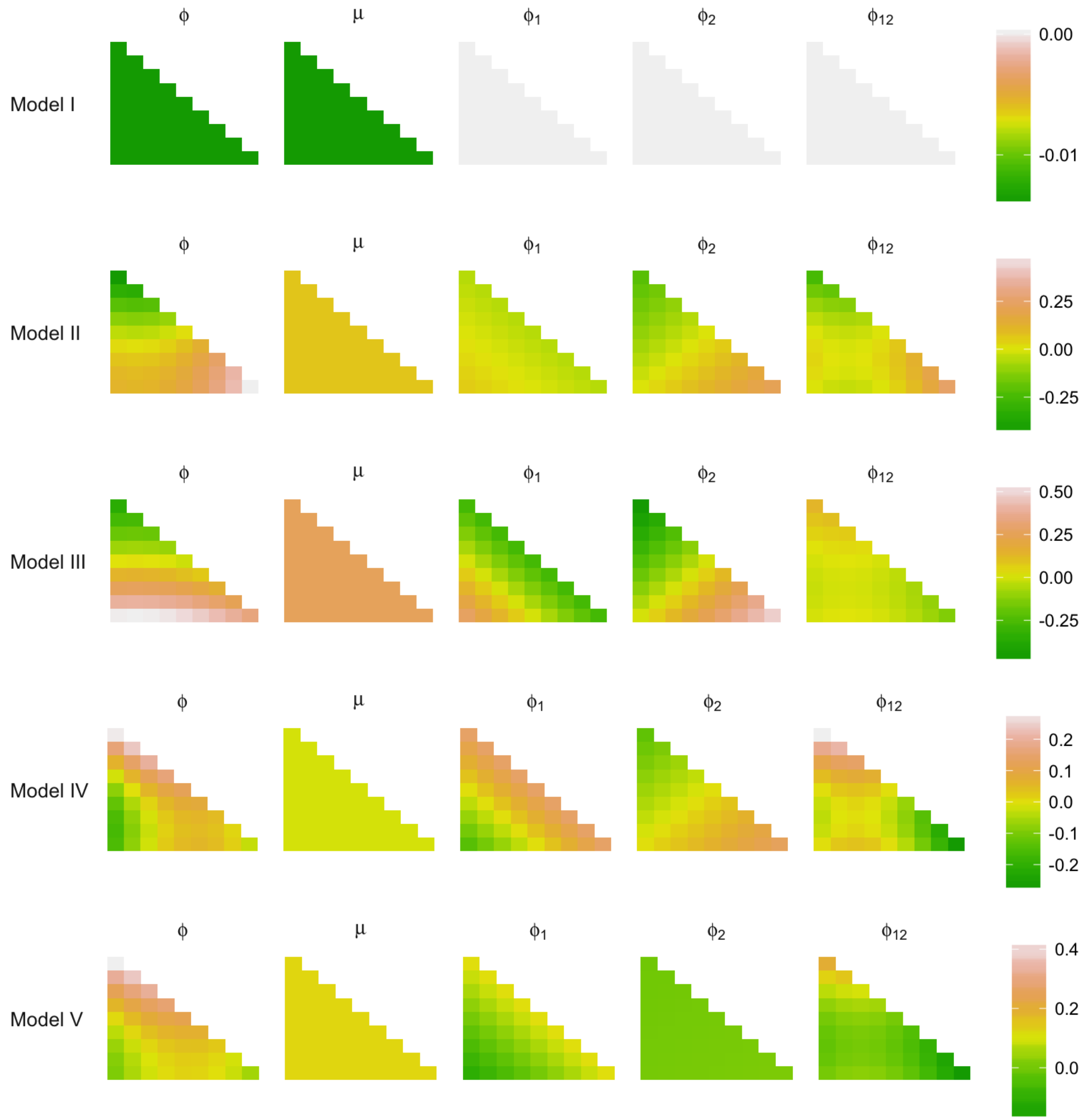


Figure 4: Estimated functional components of the smoothing spline ANOVA decomposition $\phi = \phi_1 + \phi_2 + \phi_{12}$ for $\hat{\Sigma}_{SS}$ under each simulation model I - V.



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 **??**, tables **??** - **??**. Risk estimates under quadratic loss lead to similar conclusions as those made using entropy loss and are left to the Appendix, Table **??**-**??**.

Since the regularized sample covariance estimators assume that Σ is either sparse or has entries which decrease in absolute value in distance from the diagonal, both estimators perform poorly under Model II, which exhibits neither of these characteristics. 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 MCD polynomial estimator fails for the same reason, since it assumes that ϕ can be modeled as a function of l only. The non-zero elements of the covariance matrix under Model III have similar structure and presenting similar difficulties for the three alternative estimators. The sample covariance matrix far outperforms both of its regularized variants almost uniformly across subject sample sizes N for moderate within-subject sampling rates ($M = 20, 30$.)

Table 1: Multivariate normal simulations for Model I. Estimated entropy risk is reported for our smoothing spline ANOVA estimator and P-spline estimator, the oracle estimator for each covariance structure, the parametric polynomial estimator of Pan and MacKenzie (2003), the sample covariance matrix, the tapered sample covariance matrix, and the soft thresholding estimator.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.0749	0.1261	0.0135	0.1102	1.2047	0.5369	1.1742
	20	0.0872	0.1713	0.0229	0.1096	4.9850	1.3957	4.7796
	30	0.1102	0.1969	0.0196	0.1127	12.5517	2.8019	11.3175
$N = 100$	10	0.0451	0.0671	0.0105	0.0531	0.5685	0.2045	0.5236
	20	0.0425	0.0965	0.0105	0.0512	2.2831	0.5724	2.1358
	30	0.0431	0.1148	0.0139	0.0472	5.2770	1.2430	4.9126

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.0899	0.3423	0.0581	4.7673	1.2832	1.4644	1.1770
	20	0.0949	1.3640	0.0439	97.2334	5.1665	21.6407	39.3522
	30	0.0811	2.6485	0.0627	1539.6646	12.3582	55.3674	133.9980
$N = 100$	10	0.0457	0.2945	0.0386	4.7911	0.5812	0.8335	0.5628
	20	0.0416	1.2875	0.0269	98.1989	2.3364	10.1841	10.0864
	30	0.0367	2.4365	0.0288	1582.4795	5.2389	33.5207	62.5030

Table 2: Multivariate normal simulations for model III.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.3416	0.1065	0.0619	3.0108	1.2030	1.1460	1.1467
	20	1.1140	0.2555	0.0695	62.7522	4.9824	17.2244	14.9189
	30	2.3215	0.6242	0.0576	1091.1933	12.4792	49.9135	121.7795
$N = 100$	10	0.2904	0.0579	0.0268	3.0383	0.5699	0.5545	0.5371
	20	1.1963	0.2011	0.0275	62.8960	2.2700	11.8274	9.5217
	30	2.2811	0.3845	0.0221	1105.0449	5.2234	29.1693	60.3529

Table 3: Multivariate normal simulations for model IV.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.3422	0.1966	0.0217	0.7144	1.2218	0.7397	1.1921
	20	0.9208	0.3499	0.0286	1.4588	4.9091	1.9786	4.9206
	30	1.5992	0.5100	0.0283	2.2173	12.6114	3.7440	12.1489
$N = 100$	10	0.3047	0.2237	0.0125	0.6958	0.5570	0.3168	0.5515
	20	0.8911	0.3704	0.0105	1.4813	2.2659	0.9365	2.2474
	30	1.5213	0.5282	0.0134	2.2228	5.2106	1.9312	5.2111

Table 4: Multivariate normal simulations for model V.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.2743	0.2464	0.0986	1.2420	1.2023	18.5222	2.9824
	20	0.7526	0.8772	0.2512	2.8557	5.0195	34.6618	13.8690
	30	1.1776	0.9791	0.2641	4.5791	12.3460	46.5437	26.1364
$N = 100$	10	0.2416	0.1722	0.0520	1.1491	0.5821	16.4081	1.7397
	20	0.7286	0.2965	0.0827	2.9080	2.2918	32.5295	5.4649
	30	1.1813	0.4291	0.1799	4.4402	5.2197	39.2914	15.4295

Tuning parameter selection for the regularized versions of the sample covariance matrix was performed using cross validation. Under certain conditions pertaining to the ratio 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

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

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.

As discussed in Chapter 1, in the limit, soft thresholding produces a positive definite estimator with probability tending to 1 (Rothman et al. [2009]), however element-wise shrinkage estimators of the covariance matrix, including the soft thresholding estimator, are not guaranteed to be positive definite. 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 3 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].

1.2 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

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

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 1.1. To induce sparsity, we subsample from the complete data $\{y_i\left(t_j\right)\}$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, randomly omitting an observation $y_i\left(t_j\right)$ 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 $\mathcal{H}_0$.

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?

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 5: 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 6: 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 7: 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 8: 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 9: 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.

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?

1.3 Appendix

1.3.1 Quadratic risk estimates for simulation study 1.

Table 10: Multivariate normal simulations for model I. Estimated quadratic risk is reported for our smoothing spline ANOVA estimator and P-spline estimator, the oracle estimator for each covariance structure, the parametric polynomial estimator of Pan and MacKenzie (2003), the sample covariance matrix, the tapered sample covariance matrix, and the soft thresholding estimator.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.0015	0.0052	0.0267	0.0912	0.3901	0.3864	0.3874
	20	0.0010	0.0043	0.0459	0.0757	0.8371	0.7710	0.7716
	30	0.0026	0.0036	0.0386	0.1109	1.2857	1.1937	1.2074
$N = 100$	10	0.0005	0.0010	0.0209	0.0426	0.2116	0.1676	0.1720
	20	0.0003	0.0011	0.0212	0.0376	0.4255	0.3902	0.3970
	30	0.0002	0.0011	0.0276	0.0313	0.5984	0.5790	0.5842

Table 11: Multivariate normal simulation-estimated quadratic risk for model II.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.0483	0.0623	0.0792	7.0137	0.6269	0.8108	0.5770
	20	0.7972	1.2456	0.4317	852.2787	2.7659	30.8197	36.1492
	30	6.7921	12.8700	7.2129	96997.8508	21.0228	365.0301	1804.9695
$N = 100$	10	0.0254	0.0525	0.0580	7.0482	0.2683	0.4351	0.2665
	20	0.2877	0.8153	0.2625	861.3937	1.3347	5.5170	7.3283
	30	2.7399	6.9793	3.6619	101509.5641	8.4769	66.9461	420.2973

Table 12: Multivariate normal simulation-estimated quadratic risk for model III.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.0656	0.0665	0.0697	3.4849	0.4977	0.6678	0.5858
	20	1.0095	0.9146	0.4706	426.0848	2.0716	4.8213	8.4099
	30	10.8782	8.1124	5.3699	50613.5638	16.5536	779.2829	1181.3770
$N = 100$	10	0.0486	0.0363	0.0328	3.5437	0.2437	0.2929	0.2791
	20	0.6260	0.3783	0.1958	416.1285	1.0193	1.5353	5.1553
	30	5.9367	3.4576	2.2121	50821.3671	7.9582	14.2394	253.4296

Table 13: Multivariate normal simulation-estimated quadratic risk for model IV.

	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.0153	0.0196	0.0053	0.2575	0.4420	0.4628	0.4620
	20	0.0450	0.0154	0.0073	0.4384	0.7951	0.9184	0.9177
	30	0.0893	0.0189	0.0072	0.6539	1.3363	1.3014	1.3013
$N = 100$	10	0.0112	0.0186	0.0031	0.2098	0.2136	0.2299	0.2295
	20	0.0420	0.0143	0.0027	0.4877	0.4509	0.4311	0.4307
	30	0.0792	0.0181	0.0035	0.6616	0.6263	0.6598	0.6589

Table 14: Multivariate normal simulation-estimated quadratic risk for model V.

N	M	$\hat{\Sigma}_{SS}$	$\hat{\Sigma}_{PS}$	$\hat{\Sigma}_{oracle}$	$\hat{\Sigma}_{poly}$	S	S^{ω}	S^{λ}
$N = 50$	10	0.3659	0.2456	0.1610	1.3738	0.8484	1.6174	0.8963
	20	1.0146	0.8206	0.5236	2.8419	1.7324	3.0233	1.6375
	30	1.5352	1.1507	0.4632	4.1877	2.5484	5.1546	2.6727
$N = 100$	10	0.3091	0.2678	0.0813	1.2439	0.4175	1.0431	0.4922
	20	0.9734	0.4111	0.1522	2.7280	0.7896	2.1932	0.8461
	30	1.6032	0.7701	0.3656	3.8905	1.2577	3.5722	1.3270

Σ	N	M	$\hat{\Sigma}^{ure}_{SS}$		$\hat{\Sigma}^{ure}_{PS}$		$\hat{\Sigma}_{oracle}$		$\hat{\Sigma}_{poly}$		S		S^{ω}		S^{λ}	
I	50	10	0.0015	(3e-040)	0.0052	(0.0010)	0.0267	(0.0045)	0.0912	(0.0103)	0.3901	(0.0247)	0.3864	(0.0221)	0.3874	(0.
	50	20	0.0010	(2e-040)	0.0043	(6e-040)	0.0459	(0.0083)	0.0757	(0.0098)	0.8371	(0.0325)	0.7710	(0.0392)	0.7716	(0.
	50	30	0.0026	(0.0018)	0.0036	(6e-040)	0.0386	(0.0065)	0.1109	(0.0152)	1.2857	(0.0498)	1.1937	(0.0472)	1.2074	(0.
	100	10	0.0005	(1e-040)	0.0010	(1e-040)	0.0209	(0.0031)	0.0426	(0.0051)	0.2116	(0.0124)	0.1676	(0.0090)	0.1720	(0.
	100	20	0.0003	(1e-040)	0.0011	(1e-040)	0.0212	(0.0042)	0.0376	(0.0042)	0.4255	(0.0161)	0.3902	(0.0164)	0.3970	(0.
	100	30	0.0002	(1e-040)	0.0011	(1e-040)	0.0276	(0.0041)	0.0313	(0.0033)	0.5984	(0.0262)	0.5790	(0.0211)	0.5842	(0.
II	50	10	0.0483	(0.0070)	0.0623	(0.0043)	0.0792	(0.0083)	7.0137	(0.3452)	0.6269	(0.0363)	0.8108	(0.0690)	0.5770	(0.
	50	20	0.7972	(0.1388)	1.2456	(0.1778)	0.4317	(0.0809)	852.279	(38.431)	2.7659	(0.2037)	30.820	(15.7299)	36.1492	(9.
	50	30	6.7921	(1.5850)	12.8700	(1.4200)	7.2129	(1.2710)	1997.851	(55.87)	21.0228	(2.2821)	365.030	(18.7437)	1804.970	(4
	100	10	0.0254	(0.0044)	0.0525	(0.0033)	0.0580	(0.0071)	7.0482	(0.2405)	0.2683	(0.0164)	0.4351	(0.0279)	0.2665	(0.
	100	20	0.2877	(0.0477)	0.8153	(0.1501)	0.2625	(0.0377)	861.394	(34.1825)	1.3347	(0.1086)	5.5170	(0.6241)	7.3283	(1.
	100	30	2.7399	(0.4745)	6.9793	(0.9114)	3.6619	(0.7715)	1509.564	(53.587)	8.4769	(0.7058)	66.9461	(6.0353)	420.297	(1
III	50	10	0.0656	(0.0053)	0.0665	(0.0033)	0.0697	(0.0102)	3.4849	(0.2297)	0.4977	(0.0265)	0.6678	(0.0645)	0.5858	(0.
	50	20	1.0095	(0.1420)	0.9146	(0.1113)	0.4706	(0.0731)	426.085	(26.445)	2.0716	(0.1360)	4.8213	(1.1130)	8.4099	(1.
	50	30	10.8782	(1.1771)	8.1124	(1.2342)	5.3699	(0.8475)	5613.564	(112.439)	16.5536	(1.8098)	779.283	(14.9847)	1181.377	(3
	100	10	0.0486	(0.0040)	0.0363	(0.0047)	0.0328	(0.0040)	3.5437	(0.1839)	0.2437	(0.0130)	0.2929	(0.0196)	0.2791	(0.
	100	20	0.6260	(0.0200)	0.3783	(0.0823)	0.1958	(0.0308)	416.129	(12.8666)	1.0193	(0.0701)	1.5353	(0.1560)	5.1553	(1.
	100	30	5.9367	(0.7791)	3.4576	(0.7345)	2.2121	(0.3658)	4821.367	(85.815)	7.9582	(0.8381)	14.239	(1.7202)	253.430	(7
IV	50	10	0.0153	(0.0010)	0.0196	(0.0039)	0.0053	(0.0012)	0.2575	(0.0340)	0.4420	(0.0293)	0.4628	(0.0365)	0.4620	(0.
	50	20	0.0450	(6e-040)	0.0154	(0.0024)	0.0073	(0.0012)	0.4384	(0.0416)	0.7951	(0.0447)	0.9184	(0.0397)	0.9177	(0.
	50	30	0.0893	(0.0022)	0.0189	(0.0030)	0.0072	(0.0011)	0.6539	(0.0557)	1.3363	(0.0485)	1.3014	(0.0462)	1.3013	(0.
	100	10	0.0112	(5e-040)	0.0186	(0.0029)	0.0031	(6e-040)	0.2098	(0.0185)	0.2136	(0.0109)	0.2299	(0.0134)	0.2295	(0.
	100	20	0.0420	(4e-040)	0.0143	(0.0014)	0.0027	(4e-040)	0.4877	(0.0325)	0.4509	(0.0167)	0.4311	(0.0159)	0.4307	(0.
	100	30	0.0792	(4e-040)	0.0181	(0.0020)	0.0035	(6e-040)	0.6616	(0.0327)	0.6263	(0.0215)	0.6598	(0.0207)	0.6589	(0.
V	50	10	0.3659	(0.0123)	0.2456	(0.0206)	0.1610	(0.0332)	1.3738	(0.0999)	0.8484	(0.0549)	1.6174	(0.1133)	0.8963	(0.
	50	20	1.0146	(0.0102)	0.8206	(0.0213)	0.5236	(0.1373)	2.8419	(0.1751)	1.7324	(0.0802)	3.0233	(0.1872)	1.6375	(0.
	50	30	1.5352	(0.0088)	1.1507	(0.0176)	0.4632	(0.0755)	4.1877	(0.2390)	2.5484	(0.0975)	5.1546	(0.3173)	2.6727	(0.
	100	10	0.3091	(0.0047)	0.2678	(0.0112)	0.0813	(0.0133)	1.2439	(0.0664)	0.4175	(0.0258)	1.0431	(0.0556)	0.4922	(0.
	100	20	0.9734	(0.0075)	0.4111	(0.0084)	0.1522	(0.0331)	2.7280	(0.1010)	0.7896	(0.0306)	2.1932	(0.0929)	0.8461	(0.
	100	30	1.6032	(0.0088)	0.7701	(0.0098)	0.3656	(0.0968)	3.8905	(0.1447)	1.2577	(0.0466)	3.5722	(0.1457)	1.3270	(0.

Table 15: Risk estimates under quadratic loss and corresponding standard errors based on 100 Monte Carlo simulations.

Σ	N	M	$\hat{\Sigma}^{ure}_{SS}$		$\hat{\Sigma}^{ure}_{PS}$		$\hat{\Sigma}_{oracle}$		$\hat{\Sigma}_{poly}$		S		S^{ω}		S^{λ}	
I		10	0.0749	(0.0072)	0.1261	(0.0107)	0.0135	(0.0023)	0.1102	(0.0083)	1.2047	(0.0286)	0.5369	(0.0563)	1.1742	(0.036
	50	20	0.0872	(0.0081)	0.1713	(0.0095)	0.0229	(0.0041)	0.1096	(0.0087)	4.9850	(0.0644)	1.3957	(0.1859)	4.7796	(0.120
	50	30	0.1102	(0.0229)	0.1969	(0.0118)	0.0196	(0.0034)	0.1127	(0.0108)	12.5517	(0.1322)	2.8019	(0.4332)	11.3175	(0.355
	100	10	0.0451	(0.0035)	0.0671	(0.0042)	0.0105	(0.0015)	0.0531	(0.0038)	0.5685	(0.0151)	0.2045	(0.0235)	0.5236	(0.017
	100	20	0.0425	(0.0062)	0.0965	(0.0048)	0.0105	(0.0020)	0.0512	(0.0031)	2.2831	(0.0285)	0.5724	(0.0744)	2.1358	(0.060
	100	30	0.0431	(0.0044)	0.1148	(0.0062)	0.0139	(0.0021)	0.0472	(0.0033)	5.2770	(0.0472)	1.2430	(0.1569)	4.9126	(0.120
II	50	10	0.0899	(0.0069)	0.3423	(0.0082)	0.0581	(0.0055)	4.7673	(0.0919)	1.2832	(0.0334)	1.4644	(0.0475)	1.1770	(0.034
	50	20	0.0949	(0.0080)	1.3640	(0.0158)	0.0439	(0.0051)	97.2334	(2.4537)	5.1665	(0.0610)	21.6407	(1.2914)	39.3522	(8.160
	50	30	0.0811	(0.0075)	2.6485	(0.0472)	0.0627	(0.0063)	1539.665	(39.7267)	12.3582	(0.1070)	55.3674	(3.8362)	133.9980	(19.20
	100	10	0.0457	(0.0050)	0.2945	(0.0059)	0.0386	(0.0034)	4.7911	(0.0638)	0.5812	(0.0134)	0.8335	(0.0293)	0.5628	(0.015
	100	20	0.0416	(0.0038)	1.2875	(0.0100)	0.0269	(0.0027)	98.1989	(2.0835)	2.3364	(0.0316)	10.1841	(0.8276)	10.0864	(1.118
	100	30	0.0367	(0.0033)	2.4365	(0.0293)	0.0288	(0.0031)	1582.479	(36.0484)	5.2389	(0.0475)	33.5207	(0.9390)	62.5030	(14.77
III	50	10	0.3416	(0.0091)	0.1065	(0.0090)	0.0619	(0.0079)	3.0108	(0.0709)	1.2030	(0.0312)	1.1460	(0.0472)	1.1467	(0.034
	50	20	1.1140	(0.0100)	0.2555	(0.0109)	0.0695	(0.0075)	62.7522	(2.1710)	4.9824	(0.0689)	17.2244	(0.6234)	14.9189	(2.704
	50	30	2.3215	(0.0132)	0.6242	(0.0390)	0.0576	(0.0071)	1091.193	(31.2219)	12.4792	(0.1182)	49.9135	(7.7026)	121.7795	(18.39
	100	10	0.2904	(0.0045)	0.0579	(0.0050)	0.0268	(0.0027)	3.0383	(0.0559)	0.5699	(0.0142)	0.5545	(0.0162)	0.5371	(0.013
	100	20	1.1963	(0.1239)	0.2011	(0.0057)	0.0275	(0.0036)	62.8960	(1.1460)	2.2700	(0.0306)	11.8274	(0.7008)	9.5217	(1.016
	100	30	2.2811	(0.0079)	0.3845	(0.0169)	0.0221	(0.0024)	1105.045	(21.8998)	5.2234	(0.0462)	29.1693	(0.6585)	60.3529	(14.24
IV	50	10	0.3422	(0.0085)	0.1966	(0.0118)	0.0217	(0.0049)	0.7144	(0.0141)	1.2218	(0.0319)	0.7397	(0.0436)	1.1921	(0.031
	50	20	0.9208	(0.0054)	0.3499	(0.0174)	0.0286	(0.0046)	1.4588	(0.0179)	4.9091	(0.0676)	1.9786	(0.1650)	4.9206	(0.061
	50	30	1.5992	(0.0154)	0.5100	(0.0152)	0.0283	(0.0044)	2.2173	(0.0238)	12.6114	(0.1179)	3.7440	(0.3991)	12.1489	(0.190
	100	10	0.3047	(0.0047)	0.2237	(0.0125)	0.0125	(0.0025)	0.6958	(0.0080)	0.5570	(0.0130)	0.3168	(0.0142)	0.5515	(0.014
	100	20	0.8911	(0.0036)	0.3704	(0.0185)	0.0105	(0.0017)	1.4813	(0.0140)	2.2659	(0.0305)	0.9365	(0.0686)	2.2474	(0.033
	100	30	1.5213	(0.0029)	0.5282	(0.0163)	0.0134	(0.0022)	2.2228	(0.0141)	5.2106	(0.0473)	1.9312	(0.1746)	5.2111	(0.058
V	50	10	0.2743	(0.0068)	0.2464	(0.0108)	0.0986	(0.0200)	1.2420	(0.0294)	1.2023	(0.0318)	18.5222	(0.6731)	2.9824	(0.382
	50	20	0.7526	(0.0042)	0.8772	(0.0128)	0.2512	(0.0580)	2.8557	(0.0646)	5.0195	(0.0695)	34.6618	(0.6202)	13.8690	(0.891
	50	30	1.1776	(0.0051)	0.9791	(0.0125)	0.2641	(0.0474)	4.5791	(0.0914)	12.3460	(0.1112)	46.5437	(0.7836)	26.1364	(0.324
	100	10	0.2416	(0.0039)	0.1722	(0.0049)	0.0520	(0.0090)	1.1491	(0.0202)	0.5821	(0.0111)	16.4081	(0.4280)	1.7397	(0.036
	100	20	0.7286	(0.0028)	0.2965	(0.0046)	0.0827	(0.0170)	2.9080	(0.0383)	2.2918	(0.0244)	32.5295	(0.5786)	5.4649	(0.549
	100	30	1.1813	(0.0051)	0.4291	(0.0065)	0.1799	(0.0420)	4.4402	(0.0655)	5.2197	(0.0465)	39.2914	(0.2195)	15.4295	(0.846

Table 16: Risk estimates under entropy loss and corresponding standard errors based on 100 Monte Carlo simulations.

References

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 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.

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.