Supporting Information for "A Compound Decision Approach to Covariance Matrix Estimation" by Huiqin Xin and Sihai Dave Zhao

1 Introduction

In this supporting information, we present more details about the simulations shown in the main manuscript as well as additional experiments. We used the six models described in Section 4.1 of the main text. We report the medians and interquartile ranges of the Frobenius errors across our replications. We used 200 replications unless otherwise specified. We also present proofs of the propositions in the main text.

2 Clustering-based exemplar algorithm

The following table gives exact numerical results corresponding to Figure 1 of the main text.

Table 1: Numerical results corresponding to Figure 1

Model	Method	p=30	p=100	p=200
Sparse	Exemplar	3.02 (1.21)	7.35 (1.52)	11.4 (1.73)
	$K = 2p$	3.26 (1.25)	7.51 (1.47)	11.52 (1.78)
	$K = p$	3.38 (1.33)	7.86 (1.53)	11.93 (1.8)
	$K = p/2$	3.59 (1.34)	7.98 (1.69)	11.94 (1.86)
	$K = p/4$	3.95 (1.45)	8.4 (1.77)	11.97 (1.76)
Hypercorrelated	Exemplar	5.19 (6.86)	15.75 (19.95)	32.23 (34.18)
	$K = 2p$	5.42 (7.2)	15.51 (20.15)	32.61 (33.87)
	$K = p$	5.34 (7.03)	15.25 (19.32)	32.06 (33.43)
	$K = p/2$	5.4 (6.75)	15.48 (18.94)	33.58 (34.96)
	$K = p/4$	5.22 (6.59)	16.31 (18.37)	33.03 (34.54)
Dense-0.7	Exemplar	3.76 (3.81)	12.38 (13.37)	19.93 (21.49)
	$K = 2p$	3.59 (3.79)	12.45 (13.44)	20.07 (21.51)
	$K = p$	3.72 (3.6)	12.27 (12.87)	20.54 (21.65)
	$K = p/2$	3.68 (3.72)	12.57 (13.39)	20.08 (22.77)
	$K = p/4$	3.83 (3.81)	12.9 (12.75)	21.71 (22.82)
Dense-0.9	Exemplar	4.46 (5.23)	14.63 (21.61)	24.35 (37.1)
	$K = 2p$	4.48 (5.28)	14.83 (22.3)	24.58 (37.05)
	$K = p$	4.65 (5.21)	14.85 (21.99)	24.39 (36.78)
	$K = p/2$	4.48 (5.29)	14.43 (21.76)	24.24 (38.92)
	$K = p/4$	4.49 (5.2)	14.44 (21.0)	25.01 (39.0)
Orthogonal	Exemplar	5.72 (0.27)	13.02 (0.16)	20.23 (0.25)
	$K = 2p$	5.71 (0.28)	13.01 (0.16)	20.22 (0.26)
	$K = p$	5.71 (0.25)	13.01 (0.18)	20.24 (0.26)
	$K = p/2$	5.78 (0.29)	13.01 (0.16)	20.25 (0.24)
	$K = p/4$	6.05 (0.39)	13.04 (0.16)	20.27 (0.26)
Spiked	Exemplar	2.62 (0.21)	3.71 (0.12)	4.76 (0.15)
	$K = 2p$	2.62 (0.2)	3.7 (0.12)	4.74 (0.16)
	$K = p$	2.63 (0.21)	3.71 (0.11)	4.75 (0.16)
	$K = p/2$	2.66 (0.2)	3.71 (0.11)	4.77 (0.16)
	$K = p/4$	2.77 (0.19)	3.77 (0.13)	4.77 (0.18)

3 Estimation accuracies

The following table gives exact numerical results corresponding to Figure 2 of the main text.

Table 2: Numerical results corresponding to Figure 2

Model	Method	p=30	p=100	p=200
Sparse	MSG	3.46 (1.39)	7.97 (1.67)	12.18 (1.87)
	MSGCor	3.42 (1.41)	7.75 (1.72)	11.79 (1.9)
	Adap	5.27 (1.53)	14.24 (1.61)	23.04 (1.67)
	Linear	4.91 (1.21)	15.28 (1.19)	28.06 (1.24)
	QIS	4.81 (1.34)	14.49 (1.15)	26.77 (1.11)
	NERCOME	4.94 (1.32)	14.54 (1.34)	26.61 (1.21)
	CorShrink	3.97 (1.49)	8.99 (1.53)	13.53 (1.74)
	Sample	4.98 (1.03)	16.56 (1.52)	33.21 (1.8)
	OracNonlin	3.99 (0.74)	13.5 (0.95)	25.65 (1.01)
	OracMSG	2.03 (0.76)	6.12 (1.35)	10.09 (1.49)
Hypercorrelated	MSG	5.26 (5.04)	17.71 (19.52)	29.12 (42.35)
	MSGCor	5.26 (5.04)	17.71 (19.52)	29.12 (42.35)
	Adap	13.59 (2.5)	55.12 (6.47)	114.28 (13.71)
	Linear	7.3 (3.24)	25.88 (13.03)	44.84 (28.3)
	QIS	7.33 (4.14)	24.04 (12.94)	47.96 (26.64)
	NERCOME	6.97 (4.13)	23.95 (15.33)	46.8 (31.58)
	CorShrink	6.73 (3.96)	23.0 (13.92)	42.03 (30.29)
	Sample	7.36 (3.66)	25.7 (13.44)	46.21 (28.97)
	OracNonlin	4.9 (1.15)	16.35 (2.03)	32.98 (4.15)
	OracMSG	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Dense-0.7	MSG	3.6 (3.96)	12.37 (13.53)	23.42 (28.93)
	MSGCor	3.6 (3.96)	12.37 (13.53)	23.42 (28.93)
	Adap	4.97 (2.99)	18.51 (10.55)	46.22 (32.62)
	Linear	5.09 (3.1)	16.57 (9.56)	33.6 (19.65)
	QIS	4.74 (2.73)	15.67 (8.83)	31.34 (18.12)
	NERCOME	5.01 (3.09)	16.84 (10.99)	30.67 (23.03)
	CorShrink	4.25 (3.16)	14.94 (11.29)	28.68 (24.66)
	Sample	5.03 (2.83)	17.18 (10.01)	32.92 (20.91)
	OracNonlin	3.27 (0.56)	10.83 (1.41)	21.3 (2.44)
	OracMSG	0.01 (0.08)	0.23 (0.61)	0.9 (1.52)
Dense-0.9	MSG	4.64 (5.06)	14.23 (16.91)	30.05 (34.95)
	MSGCor	4.64 (5.06)	14.22 (16.91)	30.05 (34.95)
	Adap	4.77 (4.71)	16.17 (15.34)	33.03 (29.22)
	Linear	4.78 (4.77)	15.18 (15.51)	31.13 (29.03)
	QIS	4.68 (4.27)	15.75 (13.87)	31.53 (27.94)
	NERCOME	4.95 (4.85)	16.42 (16.58)	34.18 (32.12)
	CorShrink	4.72 (4.73)	15.83 (15.48)	33.28 (31.55)
	Sample	4.8 (4.74)	16.14 (14.96)	33.86 (30.36)
	OracNonlin	2.01 (0.43)	6.8 (0.81)	13.81 (1.6)
	OracMSG	0.0 (0.01)	0.0 (0.03)	0.01 (0.11)
Orthogonal	MSG	4.24 (0.19)	8.89 (0.19)	13.85 (0.39)
	MSGCor	4.24 (0.19)	8.89 (0.19)	13.85 (0.39)
	Adap	5.0 (0.19)	9.54 (0.11)	12.93 (0.16)
	Linear	4.12 (0.12)	8.48 (0.09)	11.71 (0.1)
	QIS	4.25 (0.27)	10.01 (0.31)	12.42 (0.25)
	NERCOME	4.09 (0.15)	8.52 (0.06)	11.69 (0.05)
	CorShrink	4.52 (0.22)	9.07 (0.13)	12.71 (0.15)
	Sample	8.15 (0.42)	24.43 (0.68)	51.86 (0.81)
	OracNonlin	3.79 (0.12)	8.34 (0.06)	11.53 (0.03)
	OracMSG	4.13 (0.13)	8.47 (0.07)	11.64 (0.03)
Spiked	MSG	2.61 (0.2)	3.71 (0.09)	4.77 (0.2)
	MSGCor	2.61 (0.2)	3.71 (0.09)	4.77 (0.2)
	Adap	3.59 (0.07)	3.95 (0.06)	4.24 (0.08)
	Linear	2.58 (0.19)	3.52 (0.06)	3.71 (0.05)
	QIS	2.28 (0.31)	3.94 (0.22)	3.84 (0.19)
	NERCOME	2.24 (0.27)	3.25 (0.3)	3.66 (0.07)

4 Correlation matrix estimation

To fairly compare our methods to the CorShrink procedure of Dey and Stephens [2018], which was designed for estimating correlation matrices, we also studied the performance of these methods for estimating correlation matrices under Frobenius norm loss. For each method in Section 4.2, we calculated a correlation matrix estimate by scaling the corresponding covariance matrix estimate such that the estimated variances were equal to one. The results in Table 3 show that CorShink performs slightly better than our methods except in Model 2, likely because the additional flexibility that our method trades off lower bias for higher variance.

Table 3: Estimation errors for correlation matrices.

Model	Method	p=30	p=100	p=200
Sparse	MSG	1.5 (0.44)	3.59 (0.55)	5.8 (0.66)
	MSGCor	1.49 (0.46)	3.54 (0.52)	5.68 (0.69)
	Adap	1.98 (0.58)	5.81 (0.61)	9.63 (0.78)
	Linear	2.58 (0.28)	8.21 (0.45)	14.85 (0.45)
	QIS	2.30 (0.34)	7.68 (0.37)	14.40 (0.36)
	NERCOME	2.35 (0.37)	7.76 (0.4)	14.12 (0.4)
	CorShrink	1.38 (0.4)	3.35 (0.55)	5.08 (0.62)
	Sample	2.78 (0.3)	9.75 (0.48)	19.83 (0.46)
	OracNonlin	2.19 (0.36)	7.58 (0.44)	14.29 (0.39)
	OracMSG	0.89 (0.39)	2.78 (0.64)	4.56 (0.68)
Hypercorrelated	MSG	1.0 (0.61)	2.28 (1.47)	4.34 (2.9)
	MSGCor	1.0 (0.61)	2.25 (1.43)	4.36 (2.92)
	Adap	6.51 (1.15)	26.75 (2.07)	56.71 (2.84)
	Linear	2.21 (0.43)	7.38 (1.09)	14.86 (2.2)
	QIS	1.75 (0.61)	5.68 (1.61)	11.14 (2.94)
	NERCOME	1.66 (0.45)	5.54 (1.09)	11.01 (2.02)
	CorShrink	2.01 (0.31)	6.53 (0.59)	12.91 (1.07)
	Sample	2.31 (0.35)	7.65 (0.75)	15.35 (1.25)
	OracNonlin	1.52 (0.37)	5.17 (0.77)	10.3 (1.44)
	OracMSG	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Dense-0.7	MSG	0.76 (0.74)	1.97 (2.2)	4.74 (4.55)
	MSGCor	0.76 (0.7)	2.01 (2.22)	4.64 (4.56)
	Adap	1.4 (0.48)	5.16 (2.51)	13.69 (10.33)
	Linear	1.5 (0.82)	4.94 (2.3)	9.88 (4.98)
	QIS	1.88 (0.97)	6.87 (2.98)	15.20 (6.06)
	NERCOME	1.74 (0.71)	7.04 (2.67)	14.37 (5.81)
	CorShrink	0.64 (0.83)	1.82 (2.37)	4.53 (4.44)
	Sample	1.39 (0.42)	4.55 (1.12)	9.41 (2.44)
	OracNonlin	1.5 (0.25)	6.41 (0.42)	13.66 (0.58)
	OracMSG	0.01 (0.08)	0.21 (0.57)	0.73 (1.89)
Dense-0.9	MSG	0.36 (0.27)	0.96 (0.86)	1.98 (2.25)
	MSGCor	0.32 (0.27)	0.94 (0.88)	1.88 (2.19)
	Adap	0.49 (0.17)	1.63 (0.56)	3.38 (1.56)
	Linear	0.76 (0.47)	2.59 (1.63)	5.49 (4.09)
	QIS	0.79 (0.49)	3.00 (1.59)	6.82 (3.31)
	NERCOME	0.72 (0.34)	2.95 (1.32)	6.51 (3.49)
	CorShrink	0.29 (0.32)	0.84 (1.04)	1.91 (2.17)
	Sample	0.49 (0.17)	1.63 (0.55)	3.36 (1.37)
	OracNonlin	0.6 (0.05)	2.78 (0.1)	6.01 (0.15)
	OracMSG	0.0 (0.02)	0.01 (0.11)	0.03 (0.35)
Orthogonal	MSG	1.54 (0.07)	3.64 (0.09)	5.34 (0.14)
	MSGCor	1.54 (0.07)	3.64 (0.09)	5.34 (0.14)
	Adap	1.72 (0.03)	3.68 (0.0)	4.6 (0.0)
	Linear	1.5 (0.04)	3.47 (0.04)	4.52 (0.04)
	QIS	1.55 (0.10)	4.09 (0.14)	4.79 (0.10)
	NERCOME	1.48 (0.06)	3.48 (0.03)	4.51 (0.01)
	CorShrink	1.51 (0.05)	3.47 (0.03)	4.5 (0.02)
	Sample	2.95 (0.14)	9.98 (0.15)	20.03 (0.15)
	OracNonlin	1.39 (0.06)	3.41 (0.02)	4.46 (0.01)
	OracMSG	1.5 (0.05)	3.46 (0.03)	4.5 (0.01)
	MSG	1.99 (0.15)	3.35 (0.08)	4.53 (0.17)
	MSGCor	1.99 (0.15)	3.35 (0.08)	4.53 (0.17)
	Adap	2.69 (0.03)	3.31 (0.0)	3.49 (0.0)
	Linear	1.98 (0.14)	3.16 (0.05)	3.48 (0.05)
	QIS	1.71 (0.21)	3.61 (0.20)	3.63 (0.18)

5 Model misspecification

Our estimator assumes that data are generated from multivariate Gaussian distribution. To investigate the performance of our methods when the data are non-normal, we generated $\mathbf{Y}$ from either $U(0,1)$ or a negative binomial distribution with size 10 and mean 4. The $\mathbf{Y}$ were then normalized by their theoretical standard deviations to have unit variance. Finally, we generated the observed data following $\mathbf{X} = \mathbf{LY}$, where $\mathbf{L}$ is the Cholesky decomposition of the desired covariance matrix of $\mathbf{Y}$. We simulated our data using the covariance matrices defined in Section 4.1. The results in Tables 4 and 5 suggest that MSG and MSGCor still have excellent performance under model misspecification for both continuous and discrete data, though they can be outperformed in Models 5 and 6.

Table 4: Estimation errors for covariance matrices from uniformly distributed data.

Model	Method	p=30	p=100	p=200
Sparse	MSG	3.39 (1.58)	7.64 (1.64)	11.92 (1.81)
	MSGCor	3.36 (1.58)	7.48 (1.79)	11.58 (1.81)
	Adap	5.17 (1.4)	13.71 (1.43)	22.21 (1.54)
	Linear	4.84 (1.22)	15.15 (1.23)	28.15 (1.04)
	QIS	4.67 (1.37)	14.45 (1.25)	26.72 (1.15)
	NERCOME	4.82 (1.44)	14.49 (1.19)	26.71 (1.15)
	CorShrink	3.89 (1.52)	8.91 (1.42)	13.51 (1.78)
	Sample	4.97 (1.25)	16.46 (1.56)	33.05 (1.79)
	OracNonlin	4.01 (0.81)	13.42 (1.01)	25.66 (0.93)
	OracMSG	1.9 (0.91)	6.01 (1.16)	10.07 (1.64)
Hypercorrelated	MSG	4.18 (3.87)	14.2 (16.03)	28.91 (31.48)
	MSGCor	4.18 (3.87)	14.2 (16.03)	28.91 (31.48)
	Adap	12.14 (2.75)	50.89 (5.48)	110.43 (9.1)
	Linear	6.54 (2.74)	22.22 (9.85)	47.03 (20.36)
	QIS	6.73 (2.90)	20.21 (10.45)	41.69 (21.84)
	NERCOME	6.3 (2.66)	21.98 (10.21)	43.9 (21.92)
	CorShrink	5.83 (2.71)	20.68 (10.95)	42.38 (21.83)
	Sample	6.64 (2.4)	23.08 (10.13)	47.14 (20.15)
	OracNonlin	4.8 (0.85)	16.2 (2.04)	33.0 (3.52)
	OracMSG	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Dense-0.7	MSG	3.4 (3.29)	10.07 (11.07)	21.37 (25.63)
	MSGCor	3.4 (3.29)	10.05 (11.07)	21.37 (25.63)
	Adap	4.75 (2.23)	15.17 (6.74)	31.85 (16.09)
	Linear	4.68 (2.49)	15.11 (7.61)	30.87 (15.31)
	QIS	4.31 (2.08)	13.59 (7.92)	28.44 (13.86)
	NERCOME	4.54 (2.39)	14.44 (6.97)	30.14 (18.25)
	CorShrink	3.96 (2.54)	12.19 (8.2)	26.07 (19.72)
	Sample	4.73 (2.09)	15.08 (6.66)	30.69 (16.01)
	OracNonlin	3.17 (0.66)	10.79 (1.4)	21.13 (2.43)
	OracMSG	0.0 (0.02)	0.07 (0.31)	0.34 (0.83)
Dense-0.9	MSG	3.23 (3.58)	11.32 (11.95)	20.44 (24.74)
	MSGCor	3.23 (3.59)	11.32 (11.95)	20.44 (24.74)
	Adap	3.72 (3.08)	13.4 (10.75)	24.37 (19.79)
	Linear	3.7 (3.15)	13.61 (11.23)	24.28 (20.68)
	QIS	3.48 (2.52)	12.68 (8.65)	22.28 (20.26)
	NERCOME	4.0 (2.99)	13.49 (10.66)	24.32 (19.74)
	CorShrink	3.75 (2.9)	13.0 (10.73)	24.31 (20.56)
	Sample	3.86 (2.79)	13.4 (10.41)	25.16 (19.93)
	OracNonlin	2.03 (0.39)	6.89 (0.94)	13.97 (1.4)
	OracMSG	0.0 (0.0)	0.0 (0.01)	0.0 (0.01)
Orthogonal	MSG	4.14 (0.14)	8.68 (0.13)	13.11 (0.23)
	MSGCor	4.14 (0.14)	8.68 (0.13)	13.11 (0.23)
	Adap	4.79 (0.13)	9.23 (0.09)	12.37 (0.1)
	Linear	4.07 (0.12)	8.46 (0.08)	11.7 (0.09)
	QIS	3.48 (0.22)	9.89 (0.32)	12.34 (0.27)
	NERCOME	4.04 (0.17)	8.51 (0.07)	11.68 (0.05)
	CorShrink	4.23 (0.15)	8.72 (0.11)	12.13 (0.1)
	Sample	7.92 (0.43)	24.31 (0.49)	51.82 (0.7)
	OracNonlin	3.74 (0.15)	8.34 (0.05)	11.53 (0.02)
	OracMSG	4.09 (0.13)	8.49 (0.07)	11.64 (0.03)
Spiked	MSG	2.6 (0.21)	3.61 (0.08)	4.44 (0.1)
	MSGCor	2.6 (0.21)	3.61 (0.08)	4.44 (0.1)
	Adap	3.55 (0.06)	3.8 (0.05)	3.94 (0.04)
	Linear	2.59 (0.22)	3.51 (0.06)	3.71 (0.05)
	QIS	2.22 (0.26)	3.89 (0.18)	3.79 (0.18)

Table 5: Estimation errors for covariance matrices from negative binomial data.

Model	Method	p=30	p=100	p=200
Sparse	MSG	3.56 (1.39)	8.37 (2.14)	12.75 (1.86)
	MSGCor	3.5 (1.41)	8.2 (2.25)	12.45 (1.98)
	Adap	5.49 (1.62)	14.8 (1.69)	24.13 (1.98)
	Linear	5.03 (1.28)	15.44 (1.36)	28.33 (1.12)
	QIS	4.70 (1.17)	14.60 (1.17)	26.99 (1.05)
	NERCOME	4.98 (1.37)	14.85 (1.36)	26.82 (1.18)
	CorShrink	3.93 (1.47)	9.37 (1.65)	14.08 (1.73)
	Sample	5.04 (1.36)	16.65 (1.59)	33.02 (2.07)
	OracNonlin	3.96 (0.78)	13.68 (0.89)	25.83 (1.12)
	OracMSG	2.07 (0.89)	6.41 (1.46)	10.29 (1.55)
Hypercorrelated	MSG	5.13 (5.59)	15.17 (21.1)	38.68 (45.87)
	MSGCor	5.13 (5.59)	15.16 (21.13)	38.68 (45.87)
	Adap	14.02 (3.18)	56.02 (9.97)	121.63 (26.82)
	Linear	7.44 (3.8)	23.47 (13.98)	52.9 (33.53)
	QIS	7.71 (4.95)	24.20 (17.19)	49.38 (30.91)
	NERCOME	6.94 (4.38)	22.88 (14.15)	50.61 (36.64)
	CorShrink	6.83 (4.17)	21.38 (15.8)	49.5 (35.57)
	Sample	7.36 (4.08)	23.61 (14.1)	52.89 (34.0)
	OracNonlin	4.83 (0.99)	16.34 (2.41)	32.87 (4.19)
	OracMSG	0.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Dense-0.7	MSG	3.68 (4.3)	12.52 (11.77)	22.57 (27.91)
	MSGCor	3.68 (4.3)	12.52 (11.8)	22.56 (27.91)
	Adap	5.42 (3.46)	21.91 (15.03)	58.59 (41.83)
	Linear	4.83 (2.82)	17.7 (9.35)	32.7 (18.85)
	QIS	4.91 (2.93)	16.12 (11.55)	30.89 (20.99)
	NERCOME	4.95 (3.53)	15.86 (9.61)	31.35 (21.77)
	CorShrink	4.53 (3.3)	14.91 (10.05)	28.64 (22.39)
	Sample	5.07 (2.94)	17.19 (8.85)	32.61 (19.27)
	OracNonlin	3.23 (0.64)	10.66 (1.6)	20.95 (2.89)
	OracMSG	0.02 (0.13)	0.52 (1.08)	1.54 (2.48)
Dense-0.9	MSG	4.66 (6.0)	17.77 (20.8)	34.8 (37.69)
	MSGCor	4.65 (6.0)	17.77 (20.81)	34.79 (37.69)
	Adap	5.09 (5.39)	21.68 (21.41)	41.3 (36.24)
	Linear	5.01 (5.01)	19.26 (18.01)	36.06 (35.59)
	QIS	5.89 (5.16)	16.88 (19.65)	35.57 (32.09)
	NERCOME	5.46 (5.44)	19.21 (19.57)	35.68 (37.16)
	CorShrink	4.87 (5.24)	18.97 (19.15)	36.71 (35.06)
	Sample	4.93 (5.19)	19.25 (19.16)	37.16 (34.44)
	OracNonlin	2.02 (0.49)	6.93 (1.21)	14.01 (1.99)
	OracMSG	0.0 (0.02)	0.01 (0.2)	0.05 (0.5)
Orthogonal	MSG	4.34 (0.3)	9.13 (0.28)	14.66 (0.48)
	MSGCor	4.34 (0.3)	9.13 (0.28)	14.66 (0.48)
	Adap	5.18 (0.32)	10.25 (0.51)	14.93 (0.74)
	Linear	4.13 (0.13)	8.49 (0.08)	11.7 (0.09)
	QIS	4.31 (0.25)	10.08 (0.34)	12.45 (0.30)
	NERCOME	4.09 (0.18)	8.53 (0.07)	11.7 (0.05)
	CorShrink	4.66 (0.27)	9.29 (0.19)	13.04 (0.19)
	Sample	8.25 (0.49)	24.45 (0.75)	52.11 (0.99)
	OracNonlin	3.79 (0.13)	8.35 (0.06)	11.54 (0.03)
	OracMSG	4.13 (0.13)	8.49 (0.07)	11.64 (0.03)
Spiked	MSG	2.66 (0.23)	3.82 (0.11)	5.13 (0.23)
	MSGCor	2.66 (0.23)	3.82 (0.11)	5.13 (0.23)
	Adap	3.64 (0.1)	4.33 (0.24)	5.22 (0.38)
	Linear	2.6 (0.23)	3.52 (0.06)	3.71 (0.06)
	QIS	2.30 (0.24)	4.01 (0.23)	3.87 (0.19)
	NERCOME	2.28 (0.27)	3.28 (0.24)	3.66 (0.07)
	CorShrink	2.75 (0.24)	3.91 (0.1)	4.39 (0.11)

6 Large dimension

In previous simulations, we took $p = 30, 100$ and 200. We also studied the performance of the various methods when $p = 1000$. Because of the computational burden, here we only performed 50 replications. The results in Table 6 show that our estimator remains competitive in Models 1 through 4, but can be substantially outperformed in Models 5 and 6.

Table 6: Simulations investigating behavior when $p = 1000$.

Method	Sparse	Hypercorrelated	Dense-0.7	Dense-0.9	Orthogonal	Spiked
MSG	39.11(2.51)	210.24(205.80)	111.39(141.12)	176.32(151.84)	49.36(1.07)	16.71(0.45)
MSGCor	37.81(2.48)	210.22(205.80)	111.34(141.08)	176.27(151.84)	49.36(1.07)	16.71(0.45)
Adap	66.52(1.89)	645.60(161.94)	472.83 (321.31)	202.18(154.34)	30.22(0.28)	6.28(0.17)
Linear	97.60(0.61)	275.21(141.56)	152.10(107.99)	172.98(135.58)	28.35(0.21)	4.87(0.18)
QIS	98.54(0.69)	270.95(181.24)	173.91(120.91)	162.85(133.79)	27.86(0.09)	4.19(0.16)
NERCOME	97.46(0.83)	271.85(183.10)	157.60 (84.65)	173.08(160.65)	27.46(0.01)	3.83(0.01)
CorShrink	33.40(1.50)	256.70(151.64)	138.11(101.14)	190.83(152.22)	29.63(0.12)	5.86(0.11)
Sample	163.29(3.08)	278.28(153.45)	161.04 (88.05)	194.30(149.77)	250.26(1.89)	101.12(0.88)
OracNonlin	96.31(0.73)	167.39(141.56)	105.08(10.96)	68.40(7.33)	27.33(0.01)	3.73(0.01)
OracMSG	31.33(1.75)	0.0(0.0)	7.15(6.41)	0.52(1.16)	27.37(0.01)	3.74(0.00)

7 Proofs

7.1 Proof of Proposition 1

Using the fact that when $\mathbb{E}\mathbf{X} = \mathbf{0}$, $\mathbb{E}s_{jk} = \sigma_{jk}$ for all $j, k = 1, \ldots, p$. Therefore for the class of linear decision rules (3), the scaled Frobenius risk (1) equals

$$
\begin{aligned}
R(\boldsymbol{\Sigma}, \delta) &= \frac{1}{p^2} \sum_{j,k=1}^{p} \mathbb{E}\{(\beta_S s_{jk} + \beta_I u_{jk} - \sigma_{jk})^2\} \\
&= \frac{1}{p^2} \sum_{j,k=1}^{p} \mathbb{E}[s_{jk} - \sigma_{jk} - \{(1-\beta_S)s_{jk} - \beta_I u_{jk}\}]^2 \\
&= \frac{1}{p^2} \sum_{j,k=1}^{p} [\mathbb{E}\{(s_{jk} - \sigma_{jk})^2\} + \mathbb{E}\{(1-\beta_S)s_{jk} - \beta_I u_{jk}\}^2 - 2(1-\beta_S)\mathbb{E}\{s_{jk}(s_{jk} - \sigma_{jk})\}] \\
&= \frac{1}{p^2} \sum_{j,k=1}^{p} [(2\beta_S - 1)\mathrm{Var}(s_{jk}) + \mathbb{E}\{(1-\beta_S)s_{jk} - \beta_I u_{jk}\}^2] \\
&= \frac{1}{p^2} \sum_{j,k=1}^{p} \mathbb{E}[(2\beta_S - 1)\frac{n}{n-1}\hat{\Delta}_{jk}^2 + \{(1-\beta_S)s_{jk} - \beta_I u_{jk}\}^2],
\end{aligned}
$$

with $\hat{\Delta}_{jk}$ defined in Proposition 2. Therefore

$$
\begin{aligned}
\mathbb{E}\hat{R}_L(\beta_S, \beta_I) - R(\boldsymbol{\Sigma}, \delta) &= \frac{-1}{n-1}(2\beta_S - 1)\frac{1}{p^2} \sum_{j,k=1}^{p} \mathbb{E}\hat{\Delta}_{jk}^2 = \frac{-1}{n}(2\beta_S - 1)\frac{1}{p^2} \sum_{j,k=1}^{p} \mathrm{Var}(s_{jk}) \\
&= \frac{-1}{n}(2\beta_S - 1)\frac{1}{p^2}\mathbb{E}\|\mathbf{S} - \boldsymbol{\Sigma}\|_F^2 \to 0,
\end{aligned}
$$

where the last result follows because by assumption, $p^{-2}\mathbb{E}\|\mathbf{S} - \boldsymbol{\Sigma}\|_F^2$ is bounded as $n \to \infty$.

7.2 Proof of Proposition 2

We first rewrite the risk estimate $\hat{R}_L(\beta_S, \beta_I)$. Define $\mathbf{M} = (\sum_{j,k=1}^p \hat{\Delta}_{jk}^2, 0)^\top$, $\beta = (\beta_S, \beta_I)^\top$, and the vectorized covariance matrices $\mathbf{v_S} = (s_{11}, \dots, s_{pp})^\top$, $\mathbf{v_I} = (u_{11}, \dots, u_{pp})^\top$, and $\mathbf{v_\Sigma} = (\sigma_{11}, \dots, \sigma_{pp})^\top$. Then the risk estimate can be re-written as

$$p^2 \hat{R}_L(\beta_S, \beta_I) = \beta^\top (\mathbf{Z}^\top \mathbf{Z}) \beta - 2(\mathbf{Z}^\top \mathbf{v_S} - \mathbf{M})^\top \beta - \mathbf{1}^\top \mathbf{M} + \mathbf{v}_S^\top \mathbf{v}_S,$$

where $\mathbf{Z} = (\mathbf{v_S}, \mathbf{v_I})$. Therefore

$$\begin{aligned} \hat{\beta} &= \arg\min_\beta \hat{R}_L(\beta_S, \beta_I) = (\mathbf{Z}^\top \mathbf{Z})^{-1} (\mathbf{Z}^\top \mathbf{v_S} - \mathbf{M}), \\ \hat{\mathbf{v}}_{\mathbf{\Sigma}} &= \mathbf{Z} \hat{\beta} = \mathbf{v_S} - \mathbf{Z} (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{M}. \end{aligned}$$

Since

$$\mathbf{Z}^\top \mathbf{Z} = \begin{pmatrix} s_{11} & \dots & s_{pp} \\ u_{11} & \dots & u_{pp} \end{pmatrix} \begin{pmatrix} s_{11} & u_{11} \\ \dots & \dots \\ s_{pp} & u_{pp} \end{pmatrix} = \begin{pmatrix} \sum_{j,k=1}^p s_{jk}^2 & \sum_{j=1}^p s_{jj} \\ \sum_{j=1}^p s_{jj} & p \end{pmatrix}$$

and $\det(\mathbf{Z}^\top \mathbf{Z}) = p \sum_{j,k=1}^p s_{jk}^2 - (\sum_{j=1}^p s_{jj})^2 = p^2 d_n^2$, it follows that

$$(\mathbf{Z}^\top \mathbf{Z})^{-1} = \frac{1}{p^2 d_n^2} \begin{pmatrix} p & -\sum_{j=1}^p s_{jj} \\ -\sum_{j=1}^p s_{jj} & \sum_{j,k=1}^p s_{jk}^2 \end{pmatrix},$$

and in addition

$$\mathbf{Z}^\top \mathbf{v_S} = \begin{pmatrix} \sum_{j,k=1}^p s_{jk}^2 \\ \sum_{j=1}^p s_{jj} \end{pmatrix}, \quad \mathbf{Z}^\top \mathbf{v_S} - \mathbf{M} = \begin{pmatrix} \sum_{j,k=1}^p s_{jk}^2 - \hat{\Delta}_{jk}^2 \\ \sum_{j=1}^p s_{jj} \end{pmatrix}.$$

Therefore

$$\begin{aligned} \hat{\beta} &= (\mathbf{Z}^\top \mathbf{Z})^{-1}(\mathbf{Z}^\top \mathbf{v_S} - \mathbf{M}) \\ &= \frac{1}{p^2 d_n^2} \begin{pmatrix} p & -\sum_{j=1}^p s_{jj} \\ -\sum_{j=1}^p s_{jj} & \sum_{j,k=1}^p s_{jk}^2 \end{pmatrix} \begin{pmatrix} \sum_{j,k=1}^p (s_{jk}^2 - \hat{\Delta}_{jk}^2) \\ \sum_{j=1}^p s_{jj} \end{pmatrix} \\ &= \frac{1}{p^2 d_n^2} \begin{pmatrix} p \sum_{j,k=1}^p s_{jk}^2 - p \sum_{j,k=1}^p \hat{\Delta}_{jk}^2 - (\sum_{j=1}^p s_{jj})^2 \\ (\sum_{j=1}^p s_{jj})(\sum_{j,k=1}^p \hat{\Delta}_{jk}^2) \end{pmatrix}. \end{aligned}$$

The second component of $\hat{\beta}$ equals $\hat{\beta}_I$, and

$$\begin{aligned} \hat{\beta}_I &= \frac{1}{p^2 d_n^2} (\sum_{j=1}^p s_{jj})(\sum_{j,k=1}^p \hat{\Delta}_{jk}^2) \\ &= \{(\sum_{j=1}^p s_{jj})/p\}\{(\sum_{j,k=1}^p \hat{\Delta}_{jk}^2)/p\}/d_n^2 = \hat{\mu}\frac{b_n^2}{d_n^2}, \end{aligned}$$

so $\min(\hat{\mu}, \hat{\beta}_I) = \hat{\mu} b_n^2/d_n^2$, which is the coefficient of the second term in the Ledoit and Wolf estimator (5). Furthermore,

$$\begin{aligned} \hat{\beta}_I/\hat{\mu} + \hat{\beta}_S &= \frac{1}{p^2 d_n^2} \sum_{j,k=1}^p \{p s_{jk}^2 - p \sum_{j,k=1}^p \hat{\Delta}_{jk}^2 - (\sum_{j=1}^p s_{jj})^2 + p \sum_{j,k=1}^p \hat{\Delta}_{jk}^2\} \\ &= \frac{1}{p^2 d_n^2} \{p \sum_{j,k=1}^p s_{jk}^2 - (\sum_{j=1}^p s_{jj})^2\} = 1, \end{aligned}$$

so $\min(\hat{\beta}_S, 0) = 1 - b_n^2/d_n^2$, which is the coefficient of the first term in (5).

7.3 Proof of Proposition 3

For any $\delta \in \mathbf{S}$ (7), the Frobenius risk (1) can be written as

$$
\begin{aligned}
R(\boldsymbol{\Sigma}, \boldsymbol{\delta}) = & \frac{2}{p^2} \sum_{1 \leq k < j \leq p}^{p} \int \{t_{od}(\mathbf{X}, \mathbf{X}') - \sigma_{jk}\}^2 f_{2n}(\mathbf{X}, \mathbf{X}' \mid \sigma_j, \sigma_k, r_{jk}) d\mathbf{X} d\mathbf{X}' + \\
& \frac{1}{p^2} \sum_{j=1}^{p} \int \{t_d(\mathbf{X}) - \sigma_{jj}\}^2 f_{1n}(\mathbf{X} \mid s_{jj}) d\mathbf{X} \\
= & \frac{p-1}{p} \int \int \{t_{od}(\mathbf{X}, \mathbf{X}') - ab\gamma\}^2 f_{2n}(\mathbf{X}, \mathbf{X}' \mid a, b, \gamma) dG_{od}(a, b, \gamma) d\mathbf{X} d\mathbf{X}' + \\
& \frac{1}{p} \int \int \{t_d(\mathbf{X}) - a^2\}^2 f_{1n}(\mathbf{X} \mid a) dG_d(a) d\mathbf{X}.
\end{aligned}
$$

Therefore for any t_{od},

$$
\begin{aligned}
& \int \{t_{od}(\mathbf{X}, \mathbf{X}') - ab\gamma\}^2 f_{2n}(\mathbf{X}, \mathbf{X}' \mid a, b, \gamma) dG_{od}(a, b, \gamma) \\
= & \int \{t_{od}(\mathbf{X}, \mathbf{X}') - t_{od}^{\star}(\mathbf{X}, \mathbf{X}')\}^2 f_{2n}(\mathbf{X}, \mathbf{X}' \mid a, b, \gamma) dG_{od}(a, b, \gamma) + \\
& 2\{t_{od}(\mathbf{X}, \mathbf{X}') - t_{od}^{\star}(\mathbf{X}, \mathbf{X}')\} \int \{t_{od}^{\star}(\mathbf{X}, \mathbf{X}') - ab\gamma\} f_{2n}(\mathbf{X}, \mathbf{X}' \mid a, b, \gamma) dG_{od}(a, b, \gamma) + \\
& \int \{t_{od}^{\star}(\mathbf{X}, \mathbf{X}') - ab\gamma\}^2 f_{2n}(\mathbf{X}, \mathbf{X}' \mid a, b, \gamma) dG_{od}(a, b, \gamma) \\
\geq & \int \{t_{od}^{\star}(\mathbf{X}, \mathbf{X}') - ab\gamma\}^2 f_{2n}(\mathbf{X}, \mathbf{X}' \mid a, b, \gamma) dG_{od}(a, b, \gamma),
\end{aligned}
$$

because the middle cross term is equal to zero by the definition of $t_{od}^{\star}$. Similarly, it can be shown that for any t_d,

$$
\int \{t_d(\mathbf{X}) - a^2\}^2 f_{1n}(\mathbf{X} \mid a) dG_d(a) \geq \int \{t_d^{\star}(\mathbf{X}) - a^2\}^2 f_{1n}(\mathbf{X} \mid a) dG_d(a).
$$

This implies that $R(\boldsymbol{\Sigma}, \delta) \geq R(\boldsymbol{\Sigma}, \delta^{\star})$ for any $\delta \in \mathcal{S}$.

References

K. K. Dey and M. Stephens. Corshrink: Empirical bayes shrinkage estimation of correlations, with applications. *bioRxiv*, page 368316, 2018.