Supplementary Material to

Empirical Bayes data integration for multi-response regression

S.1 Convergence of the loss

We conduct a thorough simulation experiment to study the asymptotic approximation given by Theorem 5. Specifically, we consider three scenarios:

1. **Independence:** Suppose $X_{ij} \overset{iid}{\sim} \pi$ such that $\mathbb{E}(X_{ij}) = 0$ and $\mathbb{E}(X_{ij}^2) = 1$ and $\mathbb{E}(X_{ij}^{12}) < \infty$ for $i = 1, \ldots, n$, $j = 1, \ldots, p$. Here, the population covariance matrix is $\Sigma_n = \mathrm{I}_p$. This is satisfied by a large class of distributions, e.g. the exponential family. We set π to be $\mathrm{N}(0, 1)$. Let $S_n = n^{-1}X^{\mathrm{T}}X$. For this setting, F_n converges to the Marcenko-Pastur distribution when $p/n \to c \in (0, 1)$. Moreover, it can be shown that $\mathbf{L}_{1,n} = \mathrm{tr}[(\widetilde{\Sigma}_n^{-1} - \Sigma_n^{-1})^2 S_n] \to 0$ as $n \to \infty$. Thus $\mathbf{L}_1 = \lim_{n\to\infty} \mathbf{L}_{1,n} = 0$. We set $\widetilde{\Sigma}_n$ to be the proposed Stein shrinkage estimator from Theorem 6.

2. **Weak dependence:** Suppose $X_i^{p\times 1} \overset{iid}{\sim} \mathrm{N}(0, \Sigma_n)$ for $i = 1, \ldots, n$ where $\Sigma_{n,jj} = 1$ and $\Sigma_{n,jk} = \rho^{|j-k|}$ when $j \neq k$ for $|\rho| < 1$. This is an AR(1) structure. Unlike the previous case, here an analytical form of $\mathbf{L}_1$ is unknown although the limiting spectral distribution is known (Gray et al., 2006). We call it the weak dependence model since $\mathrm{Cov}(X_{ij}, X_{ik}) \to 0$ as $|j - k| \to \infty$. For this case, we look at $\mathbf{L}_{1,n}$ for different choices of n. The estimator $\widetilde{\Sigma}_n$ is the same as the previous case.

3. **Strong dependence:** Suppose $X_i^{p\times 1} \overset{iid}{\sim} \mathrm{N}(0, \Sigma_n)$ for $i = 1, \ldots, n$ where $\Sigma_n = \mathrm{I}_p + \rho \mathbf{1}_p \mathbf{1}_p'$. We set $\rho = 0.5$. This is a spike-covariance model. Here also, $\mathbf{L}_1$ is not

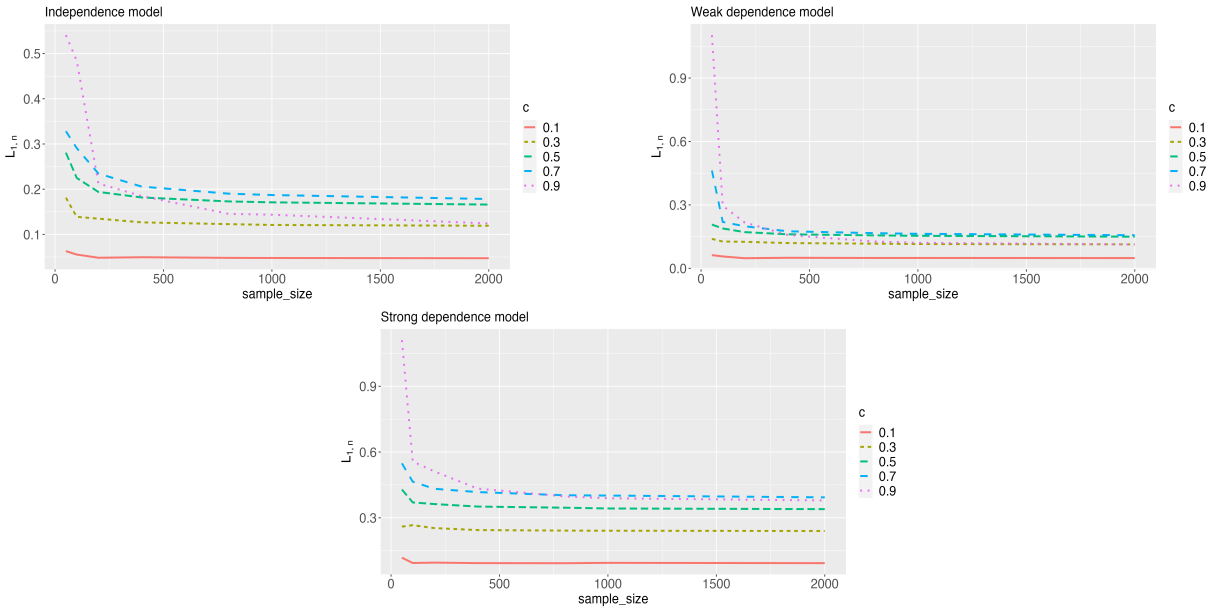

Figure S.1: Behavior of $\mathbf{L}_{1,n}$ with n for concentration ratios $c = 0.1, 0.3, 0.5, 0.7, 0.9$. The sample size $n = 50, 100, 200, 400, 800, 1000, 2000$.

available analytically. We note however that F_n still converges to the Marcenko-Pastur law (Baik and Silverstein, 2006). This is a strong dependence model since the $\text{Cov}(X_{ij}, X_{ik})$ is constant with respect to $|j - k|$. Similar to the previous setting, we look at the behaviour of $\mathbf{L}_{1,n}$ as n increases. The same estimator $\widetilde{\Sigma}_n$ is also used here.

For all these cases, we report the behavior of $\mathbf{L}_{1,n}$ with n in Figure S.1 where we vary the concentration ratio c from 0.1 to 0.9 with increments of 0.2 and the sample size is varied within $\{50, 100, 200, 400, 800, 1000, 2000\}$. The results have been averaged over 100 replications. As can be seen from the results, the difference between the empirical and the limiting loss is getting smaller for the **Independence model** as n grows, however, the convergence is very slow. This is perhaps not surprising given that almost sure convergence of even F_n to F can be as slow as $n^{-1/6}$ (Bai and Silverstein, 2010, Theorem 8.23), especially when c is close to 1. For the other two models, the empirical loss also seems to be converging.

S.2 Discussion on the risk of the proposed estimators

In this section, we provide a discussion on the risk of the proposed shrinkage rules. The posterior expectation $\mathbb{E}(\beta^{(t)} \mid \hat{\beta}^{(t)}) = (\mathrm{I} - C)\hat{\beta}^{(t)}$ where $C = Q^{1/2}(\mathrm{I}+\Omega)^{-1}Q^{-1/2}$, which is the basis of the linear shrinkage rule. Equivalently, in matrix notation we obtain the decision rule $\widetilde{B}(\widehat{B}) = \widehat{B} - \widehat{B}[Q^{1/2}(\mathrm{I} + \Omega)Q^{1/2}]^{-1}Q$. To put things into a general framework, in the following discussion we use the matrix-variate normal distribution Gupta and Nagar (2018). A matrix $X^{n\times p}$ is said to have a matrix-variate normal distribution with mean matrix $M^{n\times p}$ and covariance matrices $U^{n\times n}$, $V^{p\times p}$ (denoted as $X \sim \mathrm{MN}(M, U, V)$) if its density function is of the form:

$$p(X \mid M, U, V) \propto \exp\left[-\frac{1}{2}\mathrm{tr}\left\{V^{-1}(X - M)^{\mathrm{T}}U^{-1}(X - M)\right\}\right].$$

Our current model and prior can be equivalently characterized as $\widehat{B} \mid B \sim \mathrm{MN}(B, \mathrm{I}_n, Q)$ and $B \sim \mathrm{MN}(0, \mathrm{I}_n, Q^{1/2}\Omega Q^{1/2})$, where $\widehat{B}$ and B are $n \times p$ matrices with each row containing the vector $\hat{\beta}^{(t)}$ and $\beta^{(t)}$, respectively. The marginal distribution of $\widehat{B}$ is also a matrix-variate normal distribution: $\widehat{B} \sim \mathrm{MN}(0, \mathrm{I}_n, Q^{1/2}(\mathrm{I} + \Omega)Q^{1/2})$. We write $m(\widehat{B})$ as the density of the marginal distribution. The linear shrinkage rule $\widetilde{B}(\widehat{B}) = \widehat{B} + [\nabla \log m(\widehat{B})]Q$. This is a generalization of Tweedie's formula (Efron, 2011) to the matrix-variate case. For the sequel, let us assume $Q = \mathrm{I}$. As long as Q is known, this assumption does not lose generality. An adaptation of Stein's risk result (Stein, 1981) for estimators $\widetilde{B}(\widehat{B}) = \widehat{B} + \nabla \log m(\widehat{B})$ in this context gives

$$\mathbf{R}(\widetilde{B}, B) = \mathbb{E}_B[\mathrm{tr}(\widetilde{B} - B)(\widetilde{B} - B)^{\mathrm{T}}] = np - \mathbb{E}_B[D\widetilde{B}(\widehat{B})],$$

where

$$D\widetilde{B}(\widehat{B}) = \|\nabla \log m(\widehat{B})\|_F^2 - 2\nabla^2 m(\widehat{B})/m(\widehat{B}), \quad \nabla^2 m(\widehat{B}) = \sum_{j,k} \nabla^2_{jk} m(\widehat{B}).$$

In the above expression, the expectation is taken with respect to the distribution of $\widehat{B} \mid B$. The equality for $\mathbf{R}(\widetilde{B}, B)$ quantifies the potential risk reduction for the estimator $\widetilde{B}$ compared to the maximum likelihood estimator, which has np. For example, if $\nabla^2 m(\widehat{B}) < 0$ for all $\widehat{B}$, then $\widetilde{B}$ strictly improves over the maximum likelihood estimator.

In the mixture setup, the marginal distribution of B is $m_\star(\widehat{B}) = \sum_{k=1}^K \pi_k m_k(\widehat{B})$, where each $m_k(B)$ is the density of a matrix-variate normal distribution with mean 0, and covariance matrices I_n, $Q^{1/2}\Omega_k Q^{1/2}$. The posterior mean also has a similar decomposition $\widetilde{B}_{mix}(\widehat{B}) = \widehat{B} + \sum_{k=1}^K \pi_k^\star(\widehat{B}) \nabla \log m_k(\widehat{B})$ with $\pi_k^\star(\widehat{B}) = \pi_k m_k(\widehat{B})/m_\star(\widehat{B})$, i.e. $\widetilde{B}_{mix}(\widehat{B}) = \sum_{k=1}^K \pi_k^\star(\widehat{B}) \widetilde{B}_k(\widehat{B})$. The risk of $\widetilde{B}_{mix}(\widehat{B})$ is then $\mathbf{R}(\widetilde{B}_{mix}, B) = np - \mathbb{E}_B[D\widetilde{B}_{mix}(\widehat{B})]$. This can be characterized following (George, 1986, Corollary 3), which gives

$$D\widetilde{B}_{mix}(\widehat{B}) = \sum_{k=1}^K \pi_k^\star(\widehat{B}) \left[D\widetilde{B}_k(\widehat{B}) - \frac{1}{2} \sum_{l=1}^K \pi_l^\star(\widehat{B}) \|\widetilde{B}_k(\widehat{B}) - \widetilde{B}_l(\widehat{B})\|_F^2 \right]$$

Thus, the risk gains of $\widetilde{B}_{mix}$ is a (posterior) weighted combination of risk gains of $D\widetilde{B}_k$ and a term that gives the shrinkage conflict between rules $\widetilde{B}_k$ and $\widetilde{B}_l$.

S.3 Covariance estimation simulation

In this section, our main focus is to compare the risk improvement in estimating a covariance matrix Σ using the proposed estimator $\widetilde{\Sigma}(h) = U\Delta(h)U^{\mathrm{T}}$ where the observed sample covariance S has the spectral decomposition $S = U\Lambda U^{\mathrm{T}}$. Here, $\Delta(h) = \mathrm{diag}(\delta(h))$ and $\delta(h)$ is constructed following Theorem 6. We choose h according to Ledoit and Wolf (2022) denoted as h_0, and the proposed data-dependent method developed in Section 3.1, denoted as $\hat{h}$.

Additionally, we consider an oracle choice of h which is computed as follows. Given a range of values of h, we compute $\widetilde{\Sigma}(h)$, and compute the corresponding risk $\mathbb{E}[\mathbf{L}(\Sigma^{-1}, \widetilde{\Sigma}(h)^{-1})]$ approximated by taking Monte Carlo averages over 100 replications. We then select the oracle h for which the risk is minimum. We write the oracle choice of h as $\underline{h}$ and the resulting estimate of Σ as $\underline{\Sigma} = \widetilde{\Sigma}(\underline{h})$. Clearly, this selection of h requires knowledge of the true Σ. This forms the baseline of risk improvement that we would consider while computing the percentage relative improvement in average loss (PRIAL) defined as

$$\text{PRIAL}[\Sigma(h)] = \frac{\mathbb{E}[\mathbf{L}(\Sigma^{-1}, S^{-1})] - \mathbb{E}[\mathbf{L}(\Sigma^{-1}, \widetilde{\Sigma}(h)^{-1}]}{\mathbb{E}[\mathbf{L}(\Sigma^{-1}, S^{-1})] - \mathbb{E}[\mathbf{L}(\Sigma^{-1}, \underline{\Sigma}^{-1}]} \times 100\% \tag{S.1}$$

We report the PRIAL as it varies over the concentration ratio p/n over the interval $\{0.1, 0.2, \ldots, 0.9\}$. Here, the product np is fixed at 20000, and then (n, p) are chosen such that the concentration ratio is closest to elements of the set above. The true covariance matrix in all these cases were assumed to have a factor structure, i.e. $\Sigma = \Xi\Xi^{\mathrm{T}} + \mathrm{I}_p$ where $\Xi \in \mathbb{R}^{p \times k}$. We set $k = 5$ for all cases and generate elements of Ξ independently from $\mathrm{N}(0, 1)$. All expectations are approximated by 100 independent Monte Carlo replications.

The result is summarized in Figure S.2, which shows that for smaller concentration ratios, the default choice suggested by Ledoit-Wolf gives better results compared to the data-dependent method developed here. However, the benefit of the data-dependent method becomes apparent as we move into cases with a larger p/n, where the risk improvements are significant.

S.4 Simulations for high-dimensional settings

In this section, we extend our experiments to settings where the number of tissues/sources is $n = 200$ and $p = 100, 120, 150, 500$. The number of data points within each tissue/source

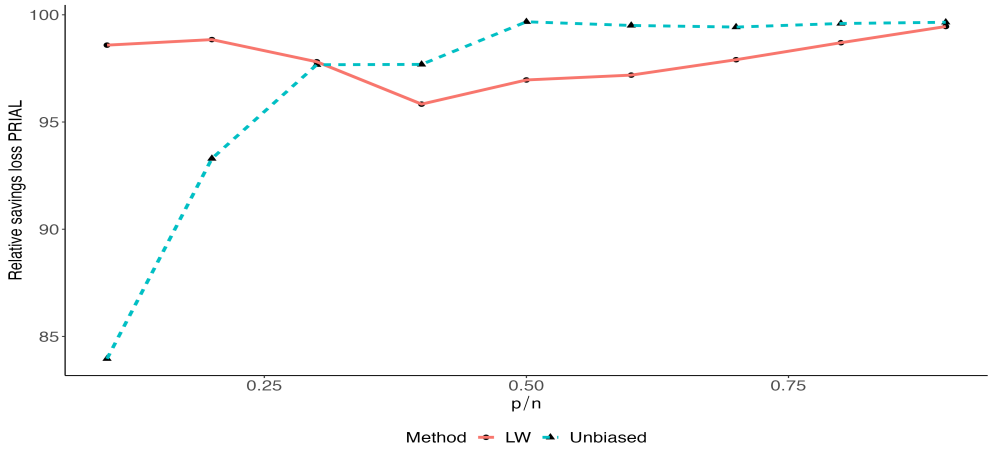

Figure S.2: Relative savings loss PRIAL with h chosen following Ledoit and Wolf (2022) (LW) and the unbiased risk estimate procedure (unbiased) when the concentration ratio p/n varies over the interval $\{0.1, 0.2, \ldots, 0.9\}$.

is $N = 1000$ when $(n, p) = (200, 500)$. For all other cases, we set $N = 200$. This is to ensure that the OLS estimator exists for all of these settings. For these high-dimensional examples, we compared our linear shrinkage estimator with default choice of h (LS), the linear shrinkage estimator with the data-dependent choice of h (ULS) versus other methods - UTMOST, ISA. We do not report the proposed local linear shrinkage estimator as the results were almost identical with LS/ULS. Also, the existing implementation of MASH frequently ran into convergence issues in these high-dimensional settings. Hence, we do not report results for this method as well. Finally, we note that for the case $(n, p) = (200, 500)$, the ULS estimator does not work since it assumes $p < n$. These results are reported in Tables S.1, S.2, S.3, and S.4. Our conclusions from these new experiments remain consistent from our previous experiments - the proposed method performs significantly better across all different settings.

		Low-rank				Approximately sparse			
		LS	ULS	UTMOST	ISA	LS	ULS	UTMOST	ISA
	$\rho = 0$	0.0009	0.0008	2.89	0.0003	0.002	0.002	0.013	0.039
	$\rho = 0.5$	0.001	0.001	2.83	0.0006	0.004	0.004	0.015	0.04
$p = 100$	$\rho = 0.8$	0.004	0.004	4.79	0.001	0.01	0.01	0.03	0.04
	$\rho = 0$	0.001	0.0009	2.99	0.0003	0.002	0.002	0.015	0.039
	$\rho = 0.5$	0.002	0.001	3.16	0.0007	0.004	0.004	0.015	0.041
$p = 120$	$\rho = 0.8$	0.005	0.004	5.08	0.001	0.01	0.01	0.03	0.04
	$\rho = 0$	0.001	0.001	2.63	0.0006	0.002	0.002	0.015	0.04
	$\rho = 0.5$	0.002	0.002	3.55	0.001	0.005	0.005	0.016	0.045
$p = 150$	$\rho = 0.8$	0.005	0.005	5.79	0.002	0.012	0.012	0.031	0.049
	$\rho = 0$	0.001	-	2.78	0.0002	0.002	-	0.016	0.07
	$\rho = 0.5$	0.002	-	4.90	0.0003	0.004	-	0.027	0.09
$p = 500$	$\rho = 0.8$	0.005	-	6.85	0.001	0.009	-	0.036	0.09

Table S.1: MSE of different estimators over 20 replications in high-dimensional settings.

		Horseshoe				Mixture			
		LS	ULS	UTMOST	ISA	LS	ULS	UTMOST	ISA
	$\rho = 0$	0.002	0.002	13.01	14.57	0.002	0.002	23.93	73.27
	$\rho = 0.5$	0.004	0.004	14. 21	11.54	0.005	0.005	22.86	71.60
$p = 100$	$\rho = 0.8$	0.01	0.01	16.88	13.78	0.01	0.01	28.37	71.05
	$\rho = 0$	0.002	0.002	17.42	15.83	0.002	0.002	23.03	68.91
	$\rho = 0.5$	0.005	0.005	18.21	15.21	0.005	0.005	24.01	69.68
$p = 120$	$\rho = 0.8$	0.013	22.81	16.37	0.012	0.012	25.77	72.31	
	$\rho = 0$	0.003	0.003	15.21	13.87	0.01	0.01	22.71	53.65
	$\rho = 0.5$	0.005	0.005	15.52	14.50	0.005	0.005	24.93	57.90
$p = 150$	$\rho = 0.8$	0.01	0.01	16.90	14.66	0.01	0.01	28.72	65.22
	$\rho = 0$	0.003	-	150.01	45.04	0.002	-	25.03	67.32
	$\rho = 0.5$	0.004	-	153.77	44.39	0.004	-	48.57	74.01
$p = 500$	$\rho = 0.8$	0.01	-	152.87	59.31	0.005	-	53.97	76.29

Table S.2: MSE of different estimators over 20 replications in high-dimensional settings.

		Low-rank				Approximately sparse			
		LS	ULS	UTMOST	ISA	LS	ULS	UTMOST	ISA
	$\rho = 0$	1.09	1.08	146.92	1.01	1.22	1.20	2.43	5.01
	$\rho = 0.5$	1.07	1.05	149.99	1.01	1.27	1.25	1.77	6.35
$p = 100$	$\rho = 0.8$	1.12	1.08	86.27	1.24	1.20	1.14	1.54	5.54
	$\rho = 0$	1.07	1.06	436.4	1.01	1.33	1.32	2.72	5.88
	$\rho = 0.5$	1.12	1.06	189.70	1.06	1.20	1.17	1.73	4.85
$p = 120$	$\rho = 0.8$	1.16	1.12	133.76	1.22	1.28	1.26	1.85	7.41
	$\rho = 0$	1.15	1.13	167.41	1.06	1.43	1.35	3.33	7.10
	$\rho = 0.5$	1.21	1.05	183.3	1.11	1.48	1.43	2.33	7.14
$p = 150$	$\rho = 0.8$	1.15	1.12	182.83	1.10	1.47	1.40	2.05	5.42
	$\rho = 0$	1.47	-	713.4	1.08	1.93	-	8.60	20.17
	$\rho = 0.5$	1.57	-	881.4	1.12	1.96	-	8.03	16.94
$p = 500$	$\rho = 0.8$	1.47	-	525.2	1.20	1.91	-	4.65	36.13

Table S.3: PE of different estimators over 20 replications in high-dimensional settings.

		Horseshoe				Mixture			
		LS	ULS	UTMOST	ISA	LS	ULS	UTMOST	ISA
	$\rho = 0$	1.23	1.20	406.1	1115.2	1.28	1.21	1072.9	4938.9
	$\rho = 0.5$	1.26	1.19	401.7	1025.5	1.24	1.07	1158.5	5877.7
$p = 100$	$\rho = 0.8$	1.26	1.24	352.2	1067.2	1.23	1.17	953.2	6744.6
	$\rho = 0$	1.29	1.25	652.8	1724.1	1.39	1.33	2921.2	8493.4
	$\rho = 0.5$	1.35	1.27	768.1	1672.5	1.29	1.12	1402.7	9736.2
$p = 120$	$\rho = 0.8$	1.31	1.29	807.3	1227.7	1.30	1.26	1276.3	8506.1
	$\rho = 0$	1.47	1.42	1382.8	2207.8	1.39	1.33	1610.5	8432.9
	$\rho = 0.5$	1.53	1.50	1872.2	1483.6	1.41	1.32	1849.1	7992.8
$p = 150$	$\rho = 0.8$	1.55	1.52	1923.5	1973.1	1.46	1.44	1689.6	8112.6
	$\rho = 0$	2.07	-	5×10^4	23328.9	1.93	-	11913.1	31440.7
	$\rho = 0.5$	2.43	-	5×10^4	20461.4	1.96	-	12135.4	34729.4
$p = 500$	$\rho = 0.8$	2.63	-	5×10^4	22646.5	1.99	-	12282.4	33947.3

Table S.4: PE of different estimators over 20 replications in high-dimensional settings.

S.5 Real data application on the Yeast Cell Cycle dataset

In this section, we apply the proposed LS method to the Yeast Cell Cycle dataset Chun and Keleş (2010), and compare it with the existing ordinary least squares (OLS), the Unified Test for MOlecular SignaTures (UTMOST) (Hu et al., 2019), the Iterated stable autoencoder (ISA) (Josse and Wager, 2016), and the Multivariate Adaptive Shrinkage (MASH) (Kim et al., 2024) methods. The Yeast Cell Cycle dataset contains 18 responses and 106 covariates for 542 genes (i.e., sample size $N = 542$). Each of the responses represents mRNA levels measured at every 7 minutes during 119 minutes. The covariates consist of the binding information for 106 transcription factors.

We conduct a 10-fold cross-validation analysis for each response to evaluate the prediction performance of each method. Specifically, for each response, we randomly split all the observed samples into 10 equally sized folds, and name them Folds $1 - 10$. For each $i = 1, \ldots, 10$, we treat Fold i in all the tissues as a testing set, and the remaining folds in all the tissues together as a training set. We then use the average prediction mean squared error (PMSE) across all the folds and responses for the evaluation of prediction accuracy.

The average PMSEs of all the methods are provided in Table S.5. The results show that the proposed method outperforms all the existing methods in terms of the average PMSE. We also provide the PMSE of each response in the left plot of Figure S.3, which shows that the proposed method produces the smallest PMSE among all the methods for most of responses. In addition, we calculate the Pearson correlation between the predicted values and true values for each response and each method, and present the results in the right plot of Figure S.3. We can observe that the correlation corresponding to the proposed method is higher than those of other methods for most responses.

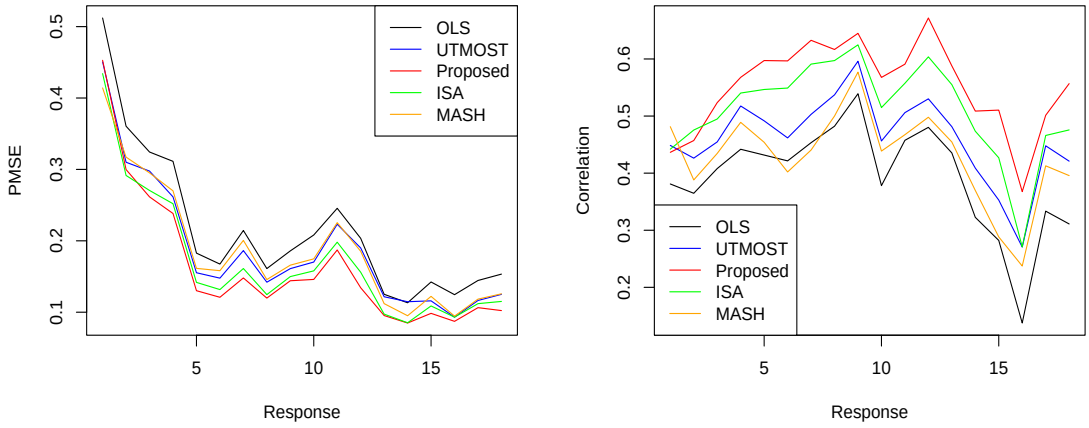

Figure S.3: Prediction mean squared error (PMSE) and correlation for each response and each the method.

In summary, the proposed method works well in the multi-response regression for the Yeast Cell Cycle dataset which is not related to TWAS.

Methods	Proposed (LS)	OLS	UTMOST	ISA	MASH
PMSE	0.164 (0.095)	0.216 (0.103)	0.188 (0.091)	0.171 (0.089)	0.188 (0.087)

Table S.5: Average prediction mean squared errors (PMSEs) for different methods, with standard deviation (SD) in the parentheses. "PMSE" represents the average PMSE across all the 18 responses.

S.6 Proofs of Sections 2 and 3

S.6.1 Proof of Proposition 1

Without loss of generality, set $Q = \mathrm{I}$. Under the assumed model $\mathbb{E}(B \mid \widehat{B}) = \widehat{B}(\mathrm{I} - C)$. Thus,

$$
\begin{aligned}
&\mathbb{E}[\mathbf{L}(B, \widetilde{B})] = \mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}[\mathbf{L}(B, \widetilde{B})] \\
=&\mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}\,\mathrm{tr}[\{\widehat{B}(\mathrm{I} - \widetilde{C}) - B\}\{\widehat{B}(\mathrm{I} - \widetilde{C}) - B\}^{\mathrm{T}}] \\
=&\mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}[\mathrm{tr}(BB^{\mathrm{T}})] - \mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}\left[\mathrm{tr}\{B(I - \widetilde{C}^{\mathrm{T}})\widehat{B}^{\mathrm{T}})\}\right] - \mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}\left[\mathrm{tr}\{\widehat{B}(\mathrm{I} - \widetilde{C})B^{\mathrm{T}}\}\right] + \\
&+ \mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}[\mathrm{tr}\{\widehat{B}(\mathrm{I} - \widetilde{C})(\mathrm{I} - \widetilde{C}^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}] \\
=&\mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}[\mathrm{tr}(BB^{\mathrm{T}})] - \mathbb{E}_{\widehat{B}}\left[\mathrm{tr}\{\widehat{B}(\mathrm{I} - C)(\mathrm{I} - \widetilde{C}^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}\right] - \mathbb{E}_{\widehat{B}}\left[\mathrm{tr}\{\widehat{B}(\mathrm{I} - \widetilde{C})(\mathrm{I} - C^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}\right] \\
&+\mathbb{E}_{\widehat{B}}[\mathrm{tr}\{\widehat{B}(\mathrm{I} - \widetilde{C})(\mathrm{I} - \widetilde{C}^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}].
\end{aligned}
$$

Since $\mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}[\mathrm{tr}(BB^{\mathrm{T}})] = \mathbb{E}_{\widehat{B}}\left\{\mathbb{E}_{B|\widehat{B}}\left[\sum_{t=1}^{T}\beta^{(t)^{\mathrm{T}}}\beta^{(t)}\right]\right\}$, and

$$
\mathbb{E}_{B|\widehat{B}}[\mathrm{tr}(B\Psi B^{\mathrm{T}})] = T\,\mathrm{tr}[(\mathrm{I} - C^{\mathrm{T}})] + \mathbb{E}_{\widehat{B}}[\mathrm{tr}(\widehat{B}(\mathrm{I} - C)(\mathrm{I} - C^{\mathrm{T}})\widehat{B}^{\mathrm{T}})],
$$

we have

$$
\begin{aligned}
\mathbb{E}_{\widehat{B}}\mathbb{E}_{B|\widehat{B}}[\mathbf{L}(B, \widetilde{B})] &= \mathbb{E}_{\widehat{B}}[\mathrm{tr}(\widehat{B}(\mathrm{I} - C)(\mathrm{I} - C^{\mathrm{T}})\widehat{B}^{\mathrm{T}})] - \mathbb{E}_{\widehat{B}}\left[\mathrm{tr}\{\widehat{B}(\mathrm{I} - C)(\mathrm{I} - \widetilde{C}^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}\right] \\
&\quad -\mathbb{E}_{\widehat{B}}\left[\mathrm{tr}\{\widehat{B}(\mathrm{I} - \widetilde{C})(\mathrm{I} - C^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}\right] + \mathbb{E}_{\widehat{B}}[\mathrm{tr}\{\widehat{B}(\mathrm{I} - \widetilde{C})(\mathrm{I} - \widetilde{C}^{\mathrm{T}})\widehat{B}^{\mathrm{T}}\}] + \text{constant} \\
&= \mathbb{E}_{\widehat{B}}[\mathrm{tr}\{(\mathrm{I} - C) - (\mathrm{I} - \widetilde{C})\}\{(\mathrm{I} - C) - (\mathrm{I} - \widetilde{C})\}^{\mathrm{T}}\widehat{B}^{\mathrm{T}}\widehat{B}] + \text{constant} \\
&= \mathbb{E}_{\widehat{B}}[\mathrm{tr}\{(\widetilde{C} - C)(\widetilde{C} - C)^{\mathrm{T}}\}\widehat{B}^{\mathrm{T}}\widehat{B}] + \text{constant} \\
&= \mathbb{E}_{\widehat{B}}[\mathrm{tr}\{(\widetilde{\Sigma}^{-1} - \Sigma^{-1})^2\}\widehat{B}^{\mathrm{T}}\widehat{B}] + \text{constant},
\end{aligned}
$$

which was to show.

S.6.2 Proof of Theorem 5

Recall that

$$\mathbf{L}_{m,n}(\Sigma_n^{-1}, \tilde{\Sigma}_n^{-1}; \mathrm{I}) = \int_{-\infty}^{\infty} x^m d\Phi_n^{(-2)}(x) - 2\int_{-\infty}^{\infty} \frac{x^m}{\delta_n(x)} d\Phi_n^{(-1)}(x) + \int_{-\infty}^{\infty} \frac{x^m}{\delta_n^2(x)} dF_n(x).$$

Since x^m is a continuous function and $\Phi_n^{(-2)}(x)$ converges weakly to $\Phi^{(-2)}(x)$ by Lemma 1,

$$\int_{-\infty}^{\infty} x^m d\Phi_n^{(-2)}(x) \overset{a.s.}{\to} \int_{-\infty}^{\infty} x^m d\Phi^{(-2)}(x).$$

Assumption 4 and the continuous mapping theorem imply that

$$\frac{x^m}{\delta_n(x)} \overset{a.s.}{\to} \frac{x^m}{\delta(x)} \quad \text{and} \quad \frac{x^m}{\delta_n^2(x)} \overset{a.s.}{\to} \frac{x^m}{\delta^2(x)}$$

for $x \in \mathrm{Supp}(F)$. In addition, the convergence is uniform for $x \in \cup_{k=1}^{K}[a_k + \eta, b_k - \eta]$ for any small $\eta > 0$. Furthermore, there exists a finite nonrandom constant $\widetilde{M}$ such that $|x^m/\delta_n(x)|$ and $|x^m/\delta_n^2(x)|$ are uniformly bounded from above by $\widetilde{M}$ almost surely for all $x \in \cup_{k=1}^{K}[a_k - \eta, b_k + \eta]$, large n, and small $\eta > 0$.

By (Ledoit and Wolf, 2018, Lemma 11.1), under our working assumptions, $\Phi_n^{(-1)}(x)$ converges weakly to $\Phi^{(-1)}(x)$, $\Phi^{(-1)}(x)$ is continuously differentiable on $\mathbb{R}$, and

$$\Phi^{(-1)}(x) = \int_{-\infty}^{x} \phi^{(-1)}(\xi) dF(\xi),$$

for $\forall x \in \mathbb{R}$. Note that, by Silverstein and Choi (1995), Silverstein and Bai (1995), and Silverstein (1995), we also have

$$F_n(x) \overset{a.s.}{\to} F(x) \quad \forall x \in \mathbb{R},$$

and $F(x)$ is continuously differentiable. Thus, similar to the proofs of (Ledoit and Wolf, 2018, Lemma 11.2), we can have

$$\int_{-\infty}^{\infty} \frac{x^m}{\delta_n(x)} d\Phi_n^{(-1)}(x) \overset{a.s.}{\to} \sum_{k=1}^{K} \int_{a_k}^{b_k} \frac{x^m}{\delta(x)} \phi^{(-1)}(x) dF(x),$$

and

$$\int_{-\infty}^{\infty} \frac{x^m}{\delta_n^2(x)} dF_n(x) \xrightarrow{a.s.} \sum_{k=1}^{K} \int_{a_k}^{b_k} \frac{x^m}{\delta^2(x)} dF(x).$$

S.6.3 Proof of Corollary 1

To find a function $\delta(x)$ that minimizes the limit in Equation (3.4), for each fixed x, we take derivative of

$$-2\frac{x^m}{\delta(x)}\phi^{(-1)}(x) + \frac{x^m}{\delta^2(x)}$$

with respect to δ, and let it equal zero. Here we do not consider the first term in Equation (3.4) since it does not involve $\delta(x)$. Then, the minimizer is $1/\phi^{(-1)}(x)$.

S.6.4 Proof of Theorem 6

By the proof of (Ledoit and Wolf, 2022, Theorem 3.1), we obtain $\delta_n^*(x) \xrightarrow{p} \delta^*(x)$ for any $x \in \text{Supp}(F)$.

S.7 Proof of Theorem 7

We record the following results from Haddouche et al. (2021); Boukehil et al. (2021) which will be useful in proving Theorem 7.

Lemma 2. *Suppose $Q \sim \text{W}_p(\Sigma, n)$, $n > p$ and $G(Q)$ is a $p \times p$ weakly differentiable matrix function. If $\mathbb{E}_\Sigma\left[\,|\,\text{tr}\{\Sigma Q G(Q)\}\,|\,\right] < \infty$, then*

$$\mathbb{E}_\Sigma\left[\text{tr}\{\Sigma^{-1}QG(Q)\}\right] = \mathbb{E}_\Sigma\left[\text{tr}\left\{(n-p-1)G(Q) + 2D_Q(G(Q)^{\text{T}}Q)\right\}\right],$$

where D_Q is a differential operator defined as $D_Q = \left\{\frac{1}{2}(1+d_{ij})\frac{\partial}{\partial Q_{ij}}\right\}$ for $1 \le i, j \le p$ with $d_{ij} = 1$ if $i = j$ and 0 otherwise.

Lemma 3. *Suppose $S = PLP^{\mathrm{T}}$ and $G(S) = P\Psi(L)P^{\mathrm{T}}$ are the spectral decomposition of S and $G(S)$, respectively for symmetric positive definite S. Then,*

$$D_S\{G(S)\} = P\Psi^{(1)}P^{\mathrm{T}} + \frac{1}{2}\mathrm{tr}\{L^{-1}\Psi(L)\}(\mathrm{I}_p - PP^{\mathrm{T}}),$$

where the j-th element of the diagonal matrix $\Psi^{(1)}$ is $\psi_j^{(1)} = \frac{\partial \psi_i}{\partial l_i} + \frac{1}{2}\sum_{i\neq j}^{p} \frac{\psi_j - \psi_i}{l_j - l_i}$.

We are now ready to prove Theorem 7. Setting $G(\underline{S}) = \widetilde{\Sigma}(h)^{-1}$ and applying Lemma 2 we get

$$\begin{aligned}
\mathbb{E}_\Sigma\left[\mathrm{tr}\left(\Sigma^{-1}\underline{S}\widetilde{\Sigma}(h)^{-1}\right)\right] &= \mathbb{E}_\Sigma\left[\mathrm{tr}\left\{(n-p-1)G(\underline{S}) + 2D_{\underline{S}}(G(\underline{S})^{\mathrm{T}}\underline{S})\right\}\right] \\
&= \mathbb{E}_\Sigma\left[\mathrm{tr}\left\{(n-p-1)G(\underline{S}) + 2D_{\underline{S}}(\widetilde{\Sigma}_h^{-1}\underline{S})\right\}\right] \\
&= \mathbb{E}_\Sigma\left[\mathrm{tr}\left\{(n-p-1)G(\underline{S}) + 2D_{\underline{S}}(U\Delta^{-1}\Lambda^* U^{\mathrm{T}})\right\}\right] \\
&= \mathbb{E}_\Sigma\left[\mathrm{tr}\left\{(n-p-1)G(\underline{S}) + 2D_{\underline{S}}(U\zeta(\Lambda^*)U^{\mathrm{T}})\right\}\right] \\
&= \mathbb{E}_\Sigma\left[\mathrm{tr}\left\{(n-p-1)G(\underline{S}) + 2U\zeta^{(1)}(\Lambda^*)U^{\mathrm{T}} + \mathrm{tr}\{\Lambda^{*-1}\zeta(\Lambda^*)\}(\mathrm{I_p} - \mathrm{UU^{\mathrm{T}}})\right\}\right]
\end{aligned}$$

where $\zeta(\Lambda^*)$ is a diagonal matrix with j-th element $\zeta_j = \lambda_j^*/\delta_j$, $j = 1, \ldots, p$, and the last equality follows by applying Lemma 3. Now, recall the definition of δ:

$$\delta_j^{-1} = c_1\lambda_j^{-1} + c_2\lambda_j^{-1}g_n^*\left(\lambda_j^{-1}\right),\ g_n^*(x) = \frac{1}{p}\sum_{i=1}^{p}\lambda_i^{-1}\frac{\lambda_i^{-1} - x}{(\lambda_i^{-1} - x)^2 + h^2\lambda_i^{-2}},$$

where $c_1 = (1 - p/n)$ and $c_2 = 2(p/n)$. Hence,

$$\frac{\partial\delta_j^{-1}}{\partial\lambda_j^{-1}} = c_1 + c_2\hat{\theta}\left(\lambda_j^{-1}\right) + c_2\lambda_j^{-1}\frac{dg_n^*(\lambda_j^{-1})}{d\lambda_j^{-1}}, \quad \frac{dg_n^*(\lambda_j^{-1})}{d\lambda_j^{-1}} = \frac{1}{p}\sum_{i=1}^{p}\lambda_i^{-1}\frac{(\lambda_i^{-1} - \lambda_j^{-1})^2 - h^2\lambda_i^{-2}}{\{(\lambda_i^{-1} - \lambda_j^{-1})^2 + h^2\lambda_i^{-2}\}^2}.$$

Therefore,

$$\frac{\partial\zeta_j}{\partial\lambda_j^*} = \frac{1}{\delta_j} - \frac{\lambda_j^*}{\delta_j^2}\frac{\partial\delta_j}{\partial\lambda_j^*} = \frac{1}{\delta_j} - \frac{\lambda_j^*}{n\delta_j^2}\frac{\partial\delta_j}{\partial\lambda_j} = \frac{1}{\delta_j} - \frac{1}{\lambda_j}\frac{\partial\delta_j^{-1}}{\partial\lambda_j^{-1}}$$

S.8 Proofs of Section 3.2

S.8.1 Proof of Theorem 8

The difference between the cases $m = 0$ and $m > 1$ comes from the presence of the term x^m for $m > 1$. Specifically, for $m > 1$, using Lemma from (Ledoit and Wolf, 2018, Lemma 14.1) we get that $\mathbf{L}_{m,n}$ has the almost sure limit

$$\begin{aligned}\mathbf{L}_m &= \int_{-\infty}^{\infty} x^m d\Phi^{(-2)}(x) - \frac{2}{c}\sum_{k=1}^{K}\int_{a_k}^{b_k}\frac{x^m}{\delta(x)}\phi^{(-1)}(x)d\underline{F}(x) + \frac{1}{c}\sum_{k=1}^{K}\int_{a_k}^{b_k}\frac{x^m}{\delta^2(x)}d\underline{F}(x) \\ &\quad + \frac{c-1}{c}\left[\frac{0^m}{\delta^2(0)}\phi^{(-1)}(0) - 2\frac{\phi^{(-1)}(0)0^m}{\delta(0)}\right] \\ &= \int_{-\infty}^{\infty} x^m d\Phi^{(-2)}(x) - 2\sum_{k=1}^{K}\int_{a_k}^{b_k}\frac{x^m}{\delta(x)}\phi^{(-1)}(x)dF(x) + \sum_{k=1}^{K}\int_{a_k}^{b_k}\frac{x^m}{\delta^2(x)}dF(x),\end{aligned}$$

since $dF(x) = (1/c)d\underline{F}(x)$. For $m = 0$, $\mathbf{L}_{m,n} = \int_{-\infty}^{\infty} d\Phi_n^{(-2)}(x) - 2\int_{-\infty}^{\infty}\frac{1}{\delta_n(x)}d\Phi_n^{(-1)}(x) + \int_{-\infty}^{\infty}\frac{1}{\delta_n^2(x)}dF_n(x)$. The limit can then be calculated using similar arguments as for the case $m > 1$.

S.8.2 Proof of Corollary 2

The proof is similar to Corollary 1.

S.8.3 Proof of Theorem 9

Recall that $\breve{m}_{\underline{F}}(x) = c\breve{m}_F(x) + (c-1)/x$ when $p > n$. Define $\underline{\Phi}(x) = 1 - \underline{F}(1/x)$ if $x > 0$ and 0 otherwise. Let $\underline{\Psi}(x) = \int_{-\infty}^{x} t d\underline{\Phi}(t)$, and for any real-valued function g,

$$\mathcal{H}_g(x) = \frac{1}{\pi}PV\int_{-\infty}^{\infty} g(t)\frac{1}{t-x}dt$$

denote the Hilbert transform of g. The following relations are true (Ledoit and Wolf, 2022, Appendix C, D)

$$\mathrm{Re}[\breve{m}_{\underline{\Psi}}(1/x)] = -x\mathrm{Re}[\breve{m}_{\underline{F}}(x)]\,\forall x \in \mathrm{Supp}(\underline{F}),\ \mathrm{Re}[\breve{m}_{\underline{\Psi}}(x)] = \pi\mathcal{H}_{\underline{\psi}}(x),$$

where $\psi = d\Psi$. Then,

$$\begin{aligned}
\delta^*(x) &= \frac{x}{1-c-2cx\mathrm{Re}[\breve{m}_F(x)]} \\
&= \frac{x}{1-c-2cx\mathrm{Re}[(1/c)\breve{m}_{\underline{F}}(x) - \{(c-1)/cx\}]} \\
&= \frac{x}{c-1-2x\mathrm{Re}[\breve{m}_{\underline{F}}(x)]} \\
&= \frac{x}{c-1+2\mathrm{Re}[\breve{m}_{\underline{\Psi}}(1/x)]} \\
&= \frac{x}{c-1+2\pi\mathcal{H}_{\underline{\psi}}(1/x)} = \frac{1}{(c-1)x^{-1}+2\pi\mathcal{H}_{\underline{\psi}}(1/x)x^{-1}}.
\end{aligned}$$

Next consider the shrinkage function

$$\delta_n^*(x) = \left[\left(\frac{p}{n}-1\right)x^{-1} + 2x^{-1}g_n^*(x^{-1})\right]^{-1} = \left[\left(\frac{p}{n}-1\right)x^{-1} + 2x^{-1}\pi\mathcal{H}_{\underline{\psi}_n}(x^{-1}))\right]^{-1},$$

where $\underline{\Phi}_n(x) = 1 - \underline{F}_n(1/x)$, $\underline{\Psi}_n(x) = \int_{-\infty}^{x} t d\underline{\Phi}_n(t)$, and $\underline{\psi}_n = d\underline{\Psi}_n$. The result follows from (Ledoit and Wolf, 2022, Theorem D.1).

S.9 Algorithm to estimate (π_k, Σ_k) from Section 4

Input: A matrix $\widehat{B}_\star$ with $\widehat{\beta}_\star^{(t)}$ as the t-th row for $t = 1, \ldots, n$.

Output: Mean of $\mathbb{E}(\beta^{(t)} \mid \widehat{\beta}^{(t)})$ across T samples.

1. Initialize $z_1, \ldots, z_n \in \{0,1\}^{K\times 1}$, where z_i is the latent indicator vector corresponding to the i-th row in $\widehat{B}_\star$.

2. Set $\widehat{B}_k$ as a matrix by extracting the i-th row in $\widehat{B}_\star$ if and only if $z_{ik} = 1$, and let $n_k =$ number of rows of $\widehat{B}_k$, where z_{ik} is the k-th element in z_i.

3. If $n_k > p$, estimate Σ_k using the estimator from Theorem 6 with data $\widehat{B}_k$.

4. If $n_k < p$, estimate Σ_k using the estimator from Theorem 9 with data $\widehat{B}_k$.

5. Estimate $\pi_k = \sum_{t=1}^{n} z_k/n$.

6. Sample $z^{(t)} \sim \text{Multinomial}(p_1^{(t)}, \ldots, p_K^{(t)})$ where $p_k^{(t)} \propto \pi_k f(\widehat{\beta}_\star^{(t)}; 0, \Sigma_k)$

7. Compute $\mathbb{E}(\beta^{(t)} \mid \widehat{\beta}^{(t)}) = \sum_{k=1}^{K} p_k^{(t)}(\mathrm{I} - C_k)\beta^{(t)}$ where $C_k = Q^{-1/2}\Sigma_k Q^{-1/2}$.

8. Repeat steps 2-7 T times.

			Low-rank								Approximately sparse							
			LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH	LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH
$n = 40$	$p = 10$	$\rho = 0$	0.005	0.005	0.001	0.001	0.001	3.363	3.631	7.955	0.005	0.005	0.001	0.002	0.003	0.018	0.04	0.005
		$\rho = 0.5$	0.006	0.006	0.005	0.005	0.005	2.42	0.193	5.067	0.008	0.009	0.003	0.003	0.003	0.02	0.043	0.04
		$\rho = 0.8$	0.015	0.015	0.016	0.012	0.022	1.992	0.333	3.579	0.017	0.017	0.012	0.013	0.012	0.029	0.037	0.036
	$p = 20$	$\rho = 0$	0.004	0.004	0.003	0.004	0.003	3.333	0.003	8.156	0.006	0.005	0.002	0.002	0.001	0.018	0.043	0.005
		$\rho = 0.5$	0.007	0.007	0.009	0.044	0.005	3.598	0.006	8.229	0.01	0.01	0.008	0.002	0.005	0.018	0.04	0.02
		$\rho = 0.8$	0.017	0.017	0.023	0.011	0.017	4.099	0.014	7.684	0.02	0.02	0.016	0.019	0.021	0.027	0.04	0.039
	$p = 30$	$\rho = 0$	0.005	0.004	0.003	0.003	0.003	3.141	0.003	7.796	0.006	0.006	0.001	0.002	0.002	0.018	0.04	0.006
		$\rho = 0.5$	0.009	0.009	0.006	0.006	0.006	3.153	0.005	7.52	0.011	0.011	0.009	0.004	0.005	0.018	0.04	0.037
		$\rho = 0.8$	0.02	0.02	0.018	0.011	0.011	3.549	0.012	8.07	0.023	0.023	0.059	0.016	0.081	0.026	0.04	0.04
$n = 50$	$p = 20$	$\rho = 0$	0.003	0.003	0.008	0.003	0.004	3.384	0.003	8.635	0.005	0.005	0.002	0.002	0.008	0.017	0.039	0.007
		$\rho = 0.5$	0.006	0.006	0.027	0.01	0.005	3.409	0.005	8.039	0.009	0.009	0.005	0.013	0.006	0.018	0.041	0.039
		$\rho = 0.8$	0.015	0.015	0.013	0.013	0.012	4.211	0.013	8.235	0.02	0.019	0.02	0.029	0.017	0.026	0.041	0.041
	$p = 30$	$\rho = 0$	0.004	0.004	0.009	0.008	0.003	3.004	0.002	7.678	0.005	0.005	0.001	0.001	0.001	0.017	0.04	0.007
		$\rho = 0.5$	0.007	0.007	0.005	0.009	0.005	3.638	0.004	8.861	0.01	0.01	0.003	0.009	0.003	0.018	0.039	0.038
		$\rho = 0.8$	0.017	0.016	0.012	0.012	0.012	3.97	0.011	8.115	0.022	0.022	0.044	0.019	0.235	0.025	0.04	0.04
	$p = 40$	$\rho = 0$	0.006	0.005	0.004	0.004	0.004	3.088	0.002	7.727	0.006	0.006	0.001	0.002	0.001	0.017	0.039	0.009
		$\rho = 0.5$	0.008	0.008	0.006	0.006	0.006	3.116	0.004	7.681	0.011	0.011	0.027	0.005	0.007	0.018	0.04	0.039
		$\rho = 0.8$	0.025	0.028	0.017	0.017	0.017	3.684	0.01	7.607	0.026	0.026	0.033	0.014	0.012	0.026	0.04	0.04

Table S.6: Mean of MSE for low-rank and approximately sparse settings when the number of tissues $n = 40, 50$. For $n = 40$, the number of covariates considered are $p = 10, 20, 30$, and for $n = 50$ they are $p = 20, 30, 40$. The ρ denotes correlation among the covariates.

			Horseshoe								Mixture							
			LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH	LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH
$n = 40$	$p = 10$	$\rho = 0$	0.005	0.005	0.001	0.001	0.003	0.389	0.222	41.78	0.005	0.005	0.00009	0.0004	0.0004	18.734	69.816	69.815
		$\rho = 0.5$	0.009	0.009	0.002	0.001	0.002	0.099	0.467	11.757	0.009	0.009	0.0004	0.001	0.001	22.493	67.585	67.584
		$\rho = 0.8$	0.019	0.019	0.008	0.009	0.012	0.646	0.225	43.664	0.025	0.025	0.001	0.002	0.004	36.026	68.666	68.665
	$p = 20$	$\rho = 0$	0.005	0.005	0.001	0.003	0.001	5.455	0.291	29.18	0.006	0.006	0.001	0.001	0.0004	15.983	45.682	54.527
		$\rho = 0.5$	0.01	0.01	0.003	0.004	0.002	2.262	0.205	21.588	0.01	0.01	0.001	0.004	0.005	24.738	71.846	72.975
		$\rho = 0.8$	0.024	0.024	0.011	0.009	0.01	2.616	0.224	69.25	0.027	0.027	0.003	0.017	0.012	41.648	72.82	73.879
	$p = 30$	$\rho = 0$	0.006	0.006	0.002	0.001	0.0003	0.34	0.15	21.305	0.006	0.006	0.0003	0.0002	0.0001	22.668	54.901	65.854
		$\rho = 0.5$	0.011	0.011	0.002	0.001	0.001	0.45	0.149	5.335	0.011	0.011	0.002	0.0004	0.0003	24.979	58.82	71.139
		$\rho = 0.8$	0.027	0.027	0.014	0.004	0.015	0.765	0.155	9.985	0.028	0.028	0.003	0.001	0.001	37.602	57.827	72.047
$n = 50$	$p = 20$	$\rho = 0$	0.005	0.005	0.0003	0.001	0.001	0.18	0.316	29.052	0.006	0.006	0.0001	0.014	0.0004	20.947	69.986	69.986
		$\rho = 0.5$	0.01	0.01	0.002	0.002	0.002	0.288	0.263	13.965	0.011	0.011	0.001	0.001	0.002	23.517	70.569	70.569
		$\rho = 0.8$	0.023	0.023	0.005	0.008	0.01	14.937	0.201	33.898	0.026	0.026	0.001	0.004	0.003	41.638	72.403	72.403
	$p = 30$	$\rho = 0$	0.006	0.006	0.004	0.001	0.0003	2.107	0.119	10.879	0.006	0.006	0.003	0.001	0.0001	24.988	68.049	74.969
		$\rho = 0.5$	0.007	0.007	0.005	0.009	0.005	3.638	0.004	8.861	0.011	0.011	0.002	0.001	0.001	23.57	62.323	67.754
		$\rho = 0.8$	0.028	0.028	0.491	0.0003	0.002	83.803	4.601	227.233	0.029	0.029	0.009	0.002	0.002	25.804	62.456	67.89
	$p = 40$	$\rho = 0$	0.006	0.006	0.004	0.0005	0.0005	20.473	3.221	3841.965	0.006	0.006	0.0002	0.0001	0.0001	30.739	59.232	72.502
		$\rho = 0.5$	0.0121	0.0121	0.0002	0.0001	0.0004	10.9984	4.515	1253.4127	0.012	0.012	0.001	0.0004	0.0003	25.182	51.436	68.82
		$\rho = 0.8$	0.031	0.031	0.003	0.0001	0.00009	7.493	5.558	238.306	0.03	0.03	0.002	0.001	0.001	36.298	51.744	69.121

Table S.7: Mean of MSE for Horseshoe and mixture settings with the same settings considered above.

			Low-rank								Approximately sparse							
			LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH	LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH
$n = 40$	$p = 10$	$\rho = 0$	1.052	1.052	1.056	1.06	1.052	32.243	35.378	75.498	1.061	1.061	1.071	1.074	1.084	1.195	1.429	1.06
		$\rho = 0.5$	1.042	1.042	1.046	1.046	1.044	14.891	2.652	45.399	1.038	1.038	1.044	1.051	1.049	1.099	1.408	1.168
		$\rho = 0.8$	1.025	1.025	1.037	1.183	1.03	16.589	2.394	59.654	1.039	1.039	1.055	1.049	1.045	1.1	1.405	1.297
	$p = 20$	$\rho = 0$	1.065	1.063	1.082	1.109	1.064	59.323	1.048	146.886	1.077	1.076	1.121	1.23	1.106	1.361	1.781	1.073
		$\rho = 0.5$	1.052	1.051	1.116	1.232	1.063	38.53	1.058	159.744	1.094	1.093	1.135	1.15	1.14	1.189	1.857	1.419
		$\rho = 0.8$	1.064	1.064	1.101	1.075	1.1	27.509	1.129	157.787	1.037	1.036	1.068	1.092	1.088	1.124	1.728	1.616
	$p = 30$	$\rho = 0$	1.146	1.138	1.146	1.167	1.147	99.17	1.081	249.926	1.185	1.181	1.188	1.198	1.207	1.552	2.202	1.186
		$\rho = 0.5$	1.135	1.128	1.135	167.986	1.134	51.3	1.087	205.935	1.157	1.153	1.269	1.176	1.19	1.265	2.253	1.828
		$\rho = 0.8$	1.137	1.136	1.185	1.137	1.137	22.861	1.203	207.818	1.134	1.131	1.357	1.161	1.467	1.176	2.105	2.07
$n = 50$	$p = 20$	$\rho = 0$	1.078	1.078	1.229	1.092	1.112	67.854	1.062	169.632	1.079	1.078	1.114	1.105	1.181	1.326	1.768	1.114
		$\rho = 0.5$	1.068	1.069	1.468	1.136	1.068	36.253	1.083	149.564	1.104	1.104	1.125	1.233	1.148	1.201	1.787	1.703
		$\rho = 0.8$	1.076	1.075	1.084	1.092	1.08	26.82	1.163	176.953	1.09	1.089	1.137	1.209	1.128	1.157	1.911	1.896
	$p = 30$	$\rho = 0$	1.072	1.069	1.469	1.166	1.072	85.227	1.04	217.24	1.207	1.207	1.215	1.238	1.212	1.622	2.358	1.277
		$\rho = 0.5$	1.078	1.078	1.078	1.16	1.078	61.998	1.05	282.444	1.168	1.166	1.18	1.255	1.169	1.285	2.235	2.039
		$\rho = 0.8$	1.073	1.07	1.073	1.073	1.073	28.372	1.13	240.155	1.137	1.136	1.437	1.22	2.042	1.146	2.073	2.057
	$p = 40$	$\rho = 0$	1.196	1.183	1.196	1.196	1.196	126.587	1.087	308.133	1.291	1.288	1.287	1.295	1.267	1.726	2.621	1.429
		$\rho = 0.5$	1.239	1.235	1.239	1.239	1.239	62.742	1.189	325.571	1.28	1.271	1.973	1.313	1.336	1.408	2.714	2.561
		$\rho = 0.8$	1.236	1.248	1.236	1.236	1.236	29.764	1.239	284.038	1.239	1.234	1.397	1.266	1.247	1.288	2.496	2.493

Table S.8: Mean of PE for low-rank and approximately sparse settings when the number of tissues $n = 40, 50$. For $n = 40$, the number of covariates considered are $p = 10, 20, 30$, and for $n = 50$ they are $p = 20, 30, 40$. The ρ denotes correlation among the covariates.

			Horseshoe								Mixture							
			LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH	LS	ULS	LLS-2	LLS-3	LLS-4	UTMOST	ISA	MASH
$n = 40$	$p = 10$	$\rho = 0$	1.03	1.03	1.033	1.031	1.06	7.905	3.253	62.921	1.072	1.072	1.072	1.077	1.071	174.86	663.571	663.568
		$\rho = 0.5$	1.087	1.087	1.088	1.102	1.089	2.074	3.477	106.468	1.044	1.044	1.043	1.044	1.045	123.324	704.839	704.835
		$\rho = 0.8$	1.052	1.052	1.051	1.057	1.063	3.455	3.216	349.183	1.042	1.042	1.041	1.038	1.037	155.127	744.975	744.971
	$p = 20$	$\rho = 0$	1.112	1.113	1.128	1.168	1.134	85.393	6.364	320.396	1.107	1.107	1.103	1.155	1.15	444.867	1376.433	1387.943
		$\rho = 0.5$	1.118	1.118	1.139	1.146	1.129	57.086	5.168	507.273	1.131	1.131	1.133	1.19	1.153	247.869	1341.508	1355.575
		$\rho = 0.8$	1.114	1.114	1.148	1.136	1.143	14.528	5.073	1240.998	1.098	1.098	1.099	1.138	1.112	251.947	1518.572	1539.977
	$p = 30$	$\rho = 0$	1.167	1.167	1.197	1.188	1.163	15.187	5.683	631.574	1.167	1.166	1.16	1.161	1.16	680.277	1632.223	1934.98
		$\rho = 0.5$	1.154	1.153	1.173	1.162	1.17	14.385	5.784	158.461	1.163	1.163	1.174	1.156	1.153	379.048	1743.015	1979.399
		$\rho = 0.8$	1.169	1.167	1.239	1.181	1.266	8.398	5.099	270.007	1.19	1.19	1.198	1.18	1.18	332.021	1780.478	2203.79
$n = 50$	$p = 20$	$\rho = 0$	1.109	1.109	1.11	1.118	1.12	4.448	7.113	504.101	1.111	1.111	1.111	1.295	1.111	428.534	1447.038	1447.033
		$\rho = 0.5$	1.13	1.13	1.139	1.145	1.143	4.844	6.933	304.907	1.106	1.107	1.105	1.11	1.117	242.846	1477.281	1477.276
		$\rho = 0.8$	1.099	1.099	1.096	1.113	1.118	143.691	5.566	711.598	1.068	1.068	1.064	1.069	1.065	253.064	1492.048	1492.043
	$p = 30$	$\rho = 0$	1.149	1.15	1.214	1.158	1.153	50.395	4.671	358.337	1.142	1.142	1.217	1.164	1.137	742.768	2078.155	2299.509
		$\rho = 0.5$	1.182	1.181	1.2	1.203	1.198	15.029	6.539	121.082	1.152	1.152	1.164	1.138	1.135	389.566	2042.295	2123.863
		$\rho = 0.8$	1.144	1.144	2.894	1.149	1.162	960.661	122.802	3322.108	1.149	1.149	1.162	1.129	1.13	160.469	1464.253	1678.723
	$p = 40$	$\rho = 0$	1.247	1.247	1.344	1.251	1.252	340.666	130.216	7640.402	1.25	1.249	1.24	1.229	1.234	1212.75	2345.017	2818.027
		$\rho = 0.5$	1.194	1.194	1.2	1.193	1.193	317.62	185.923	55318.149	1.237	1.237	1.234	1.231	1.227	535.298	2061.745	2699.554
		$\rho = 0.8$	1.229	1.228	1.27	1.228	1.227	79.532	286.479	11073.277	1.204	1.202	1.199	1.192	1.193	356.081	1434.132	2051.96

Table S.9: Mean of PE for Horseshoe and mixture settings with the same settings considered above.

References

Bai, Z. and J. W. Silverstein (2010). Spectral analysis of large dimensional random matrices. Springer.

Baik, J. and J. W. Silverstein (2006). Eigenvalues of large sample covariance matrices of spiked population models. Journal of multivariate analysis 97(6), 1382–1408.

Boukehil, D., D. Fourdrinier, F. Mezoued, and W. E. Strawderman (2021). Estimation of the inverse scatter matrix for a scale mixture of wishart matrices under Efron–Morris type losses. Journal of Statistical Planning and Inference 215, 368–387.

Chun, H. and S. Keleş (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 72(1), 3–25.

Efron, B. (2011). Tweedie's formula and selection bias. Journal of the American Statistical Association 106(496), 1602–1614.

George, E. I. (1986). Minimax multiple shrinkage estimation. The Annals of Statistics 14(1), 188–205.

Gray, R. M. et al. (2006). Toeplitz and circulant matrices: A review. Foundations and Trends® in Communications and Information Theory 2(3), 155–239.

Gupta, A. K. and D. K. Nagar (2018). Matrix variate distributions. Chapman and Hall/CRC.

Haddouche, A. M., D. Fourdrinier, and F. Mezoued (2021). Scale matrix estimation of an elliptically symmetric distribution in high and low dimensions. Journal of Multivariate Analysis 181, 104680.

Hu, Y., M. Li, Q. Lu, H. Weng, J. Wang, S. M. Zekavat, Z. Yu, B. Li, J. Gu, S. Muchnik, et al. (2019). A statistical framework for cross-tissue transcriptome-wide association analysis. Nature Genetics 51(3), 568–576.

Josse, J. and S. Wager (2016). Bootstrap-based regularization for low-rank matrix estimation. Journal of Machine Learning Research 17(124), 1–29.

Kim, Y., W. Wang, P. Carbonetto, and M. Stephens (2024). A flexible empirical Bayes approach to multiple linear

regression and connections with penalized regression. Journal of Machine Learning Research 25(185), 1–59.

Ledoit, O. and M. Wolf (2018). Optimal estimation of a large-dimensional covariance matrix under Stein's loss. Bernoulli 24(4B), 3791 – 3832.

Ledoit, O. and M. Wolf (2022). Quadratic shrinkage for large covariance matrices. Bernoulli 28(3), 1519–1547.

Silverstein, J. W. (1995). Strong convergence of the empirical distribution of eigenvalues of large dimensional random matrices. Journal of Multivariate Analysis 55(2), 331–339.

Silverstein, J. W. and Z. Bai (1995). On the empirical distribution of eigenvalues of a class of large dimensional random matrices. Journal of Multivariate Analysis 54(2), 175–192.

Silverstein, J. W. and S.-I. Choi (1995). Analysis of the limiting spectral distribution of large dimensional random matrices. Journal of Multivariate Analysis 54(2), 295–309.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. The Annals of Statistics, 1135–1151.