Supplementary Note

Fisher Z-score

The population Fisher Z-score[21] is defined as

$$Z_{ij} = \frac{1}{2}\log\left[\frac{1+R_{ij}}{1-R_{ij}}\right] \tag{10}$$

where R is the population correlation matrix. The corresponding empirical Fisher Z-score is defined as follows

$$\hat{Z}_{ij} = \frac{1}{2}\log\left[\frac{1+\hat{R}_{ij}}{1-\hat{R}_{ij}}\right] \tag{11}$$

For bivariate normally distributed random variables X_i and X_j, the empirical Fisher Z-score $\hat{Z}_{ij}$ (based on n_{ij}-many samples) is normally distributed given the population counterpart Z_{ij} [22]:

$$\hat{Z}_{ij}|Z_{ij} \sim N\left(Z_{ij}, \frac{1}{n_{ij}-1} + \frac{2}{(n_{ij}-1)^2}\right); \tag{12}$$

and the Z-scores are conditionally independent. Dey and Stephens [5] assume an adaptive shrinkage prior on the population Fisher Z-scores for each pair of variables. Here we use property (12) in the context of directly estimating Σ or Ω with an ℓ_1-norm penalty.

Derivation of C

Here we show how we derive the analytical form of the upper bound C in (3) appearing in Problem (2).

Lemma 1. *Let $X^f_{N\times P}$ be the fully observed version of the data matrix X; and let every sample $X^f_{n,\star}$ follow a Multivariate Gaussian distribution with covariance matrix Σ and correlation matrix R. The samples are independent. Then, for any fixed $\epsilon > 0$ and for sufficiently large n_{ij}, there exists a $C'_{ij}(\epsilon)$ such that*

$$\Pr\left(|\hat{R}_{ij} - R_{ij}| \le C'_{ij}(\epsilon) \middle| R_{ij}\right) > (1-\epsilon) \tag{13}$$

where

$$C'_{ij}(\epsilon) := \min\left(2, \eta(n_{ij})M(\epsilon)\left\{(1-\hat{R}^2_{ij}) + \frac{2M(\epsilon)}{3\sqrt{3}}\eta(n_{ij})\right\}\right) \quad \forall i \ne j \tag{14}$$

and

$$\eta(n_{ij}) := \sqrt{\frac{1}{n_{ij}-1} + \frac{2}{(n_{ij}-1)^2}} \tag{15}$$

and $M(\epsilon)$ is a sufficiently large finite number.

Corollary 1. *For $\epsilon = 0.001$, $M(\epsilon)$ can be taken to be 3 in Lemma 1. Then*

$$\Pr\left(|\hat{R}_{ij} - R_{ij}| < C'_{ij} \middle| R_{ij}\right) \approx 1 \tag{16}$$

where

$$C'_{ij} := min\left(2, \eta(n_{ij})\left\{3(1-\hat{R}^2_{ij}) + 2\sqrt{3}\eta(n_{ij})\right\}\right) \quad \forall i \ne j \tag{17}$$

If n_i and n_j are sufficiently large, in which case $\hat{\sigma}_i \approx \sigma_i$ and $\hat{\sigma}_j \approx \sigma_j$, then Corollary 1 leads to the following probability inequality for the pairwise sample covariance:

$$\Pr\left(|\hat{\Sigma}_{ij} - \Sigma_{ij}| < C_{ij} \middle| \Sigma_{ij}\right) \approx 1 \tag{18}$$

where

$$C_{ij} := \hat{\sigma}_i\hat{\sigma}_j C'_{ij}. \tag{19}$$

Proof of Lemma 1 and Corollary 1

If a random variable $W \sim N(0,1)$, then for any small $\epsilon > 0$, we can get a number $M(\epsilon)$ such that

$$\Pr\big(|W| < M(\epsilon)\big) > (1-\epsilon) \tag{20}$$

Using (12) and (20), we have

$$\Pr\left(|\hat{Z}_{ij} - Z_{ij}| < M(\epsilon)\eta(n_{ij}) \middle| Z_{ij}\right) > (1-\epsilon). \tag{21}$$

The estimated and population correlations $\hat{R}_{ij}$ and R_{ij} (respectively) can be written in terms of the Z-scores using (10) as follows:

$$\hat{R}_{ij} = \frac{\exp(2\hat{Z}_{ij})-1}{\exp(2\hat{Z}_{ij})+1}, \qquad R_{ij} = \frac{\exp(2Z_{ij})-1}{\exp(2Z_{ij})+1}. \tag{22}$$

Applying a Taylor series expansion to R_{ij} as a function of Z_{ij} around $\hat{Z}_{ij}$, we get:

$$\begin{aligned}\frac{\exp(2Z_{ij})-1}{\exp(2Z_{ij})+1} &= \frac{\exp(2\hat{Z}_{ij})-1}{\exp(2\hat{Z}_{ij})+1} + 4\frac{\exp(2\hat{Z}_{ij})}{\exp(2\hat{Z}_{ij})+1}(\hat{Z}_{ij} - Z_{ij}) \\ &\quad + 4\frac{\exp(2\xi)(\exp(2\xi)-1)}{(\exp(2\xi)+1)^3}(\hat{Z}_{ij} - Z_{ij})^2 \end{aligned} \tag{23}$$

where ξ is a value between Z_{ij} and $\hat{Z}_{ij}$. We can place an upper bound on the coefficient of the last term in (23):

$$\left|\frac{\exp(2\xi)(\exp(2\xi)-1)}{(\exp(2\xi)+1)^3}\right| \le \frac{1}{6\sqrt{3}}. \tag{24}$$

Using Equations (22), (23) and (24), we can write

$$\begin{aligned}|\hat{R}_{ij} - R_{ij}| \le& 4\frac{\exp(2\hat{Z}_{ij})}{(\exp(2\hat{Z}_{ij})+1)^2}|\hat{Z}_{ij} - Z_{ij}| \\ &+ \frac{2}{3\sqrt{3}}|\hat{Z}_{ij} - Z_{ij}|^2\end{aligned} \tag{25}$$

Using the definition of $\hat{Z}_{ij}$ in Equation (11), we get

$$\frac{\exp(2\hat{Z}_{ij})}{(\exp(2\hat{Z}_{ij})+1)^2} = \frac{(1-\hat{R}^2_{ij})}{4}. \tag{26}$$

Using the above expression in (25), we get:

$$|\hat{R}_{ij} - R_{ij}| \le (1-\hat{R}^2_{ij})|\hat{Z}_{ij} - Z_{ij}| + \frac{2}{3\sqrt{3}}|\hat{Z}_{ij} - Z_{ij}|^2 \tag{27}$$

Using (21) and (27), we have:

$$\Pr\left(|\hat{R}_{ij} - R_{ij}| < (1-\hat{R}^2_{ij})M(\epsilon)\eta(n_{ij}) + \frac{2}{3\sqrt{3}}M^2(\epsilon)\eta^2(n_{ij}) \middle| R_{ij}\right) > (1-\epsilon).$$

Since, $\hat{R}_{ij}$ and R_{ij} are both correlation terms, they lie between -1 and $+1$ and hence with probability one:

$$|\hat{R}_{ij} - R_{ij}| \le 2 \tag{28}$$

Combining Equations (27) and (28), we get

$$\Pr\left(|\hat{R}_{ij} - R_{ij}| < \min\{2, B\} \middle| R_{ij}\right) > (1-\epsilon) \tag{29}$$

where,

$$B = (1-\hat{R}^2_{ij})M(\epsilon)\eta(n_{ij}) + \frac{2}{3\sqrt{3}}M^2(\epsilon)\eta^2(n_{ij})$$

which completes the proof of Lemma 1.

In (20), if we choose $\epsilon = 0.001$, we have $M(\epsilon) \approx 3$—hence, (29) leads to:

$$\Pr\left(|\hat{R}_{ij} - R_{ij}| < \min\left\{2, 3(1-\hat{R}^2_{ij})\eta(n_{ij}) + 2\sqrt{3}\eta^2(n_{ij})\right\} \middle| R_{ij}\right) > (1-\epsilon) \tag{30}$$

which proves Corollary 1. Usually this result holds good [22] for any $n_{ij} > 3$. If however $n_{ij} \to \infty$ for all (i,j) pairs, then the bound on $|\hat{R}_{ij} - R_{ij}|$ in (29) approaches 0 and $\hat{R}_{ij}$ would be close to R_{ij}.

A General Likelihood Framework for Robocov Covariance Matrix Estimation

We propose a generalization of the Robocov covariance matrix estimation framework presented in Section 2.1 – the loss function presented here is directly motivated by the Fisher's Z-score framework discussed above, but differs from that appearing in Section 2.1.

Recall that the estimators in Section 2.1 are special cases of the following regularized loss minimization framework:

$$\min_{\Sigma \succeq 0} \; \mathcal{L}(\Sigma) + \lambda \xi(\Sigma) \tag{31}$$

where $\mathcal{L}$ is the data fidelity function and ξ is the penalty function. and λ is a tuning parameter that controls the trade-off between data-fidelity and regularization. We can choose $\mathcal{L}(\Sigma; \hat{\Sigma}) = \sum_{ij} \mathcal{L}_{ij}(\Sigma_{ij}, \hat{\Sigma}_{ij})$ with $\mathcal{L}_{ij}(\Sigma_{ij}, \hat{\Sigma}_{ij}) = \max\{|\hat{\Sigma}_{ij} - \Sigma_{ij}| - C_{ij}, 0\}$ for all i, j. This leads to a regularized convex optimization problem of the form:

$$\min \; \frac{1}{\lambda}\mathcal{L}(\Sigma; \hat{\Sigma}) + \sum_{i<j} |\Sigma_{ij}|. \tag{32}$$

In the limiting case, $\lambda \to 0+$ i.e., $1/\lambda \to \infty$, estimator obtained from Problem (32) will reduce to the estimator available from (2). This is because, for sufficiently large values of $1/\lambda$, an optimal solution to (32) will lead to a zero loss—$\mathcal{L}(\Sigma; \hat{\Sigma}) = 0$ which implies that $\mathcal{L}_{ij}(\Sigma_{ij}; \hat{\Sigma}_{ij}) = 0$ for all i, j — these are the data-fidelity constraints in (2). We note that in our numerical experiments, estimator (2) had a performance which was roughly similar to that of the general estimator (32).

A quadratic loss alternative to covariance estimation problem

We present below (See (33)) a convex quadratic loss function $\mathcal{L}(\Sigma)$. While this differs from the loss function considered in (2), in practice, the performances of these two estimators were found to be similar (at least on the datasets we experimented on).

To derive the loss function, we make use of Lemma 2 — which presents the (conditional) mean and variance of $\hat{R}_{ij}$ (given R_{ij}). This leads to a loss function of the form:

$$\sum_{ij} \frac{(\hat{R}_{ij} - E(\hat{R}_{ij}|R_{ij}))^2}{\text{var}(\hat{R}_{ij}|R_{ij})}$$

Using the expressions for conditional mean/variances from Lemma 2 (see below), in the above expression, we get:

$$\sum_{ij} \left(\hat{R}_{ij} - \left(R_{ij} + R_{ij}(1 - R_{ij}^2)\eta^2(n_{ij})\right)\right)^2 / ((1 - R_{ij}^2)^2 \eta^2(n_{ij})).$$

We set $\hat{R}_{ij} = \hat{\Sigma}_{ij}/(\hat{\sigma}_i \hat{\sigma}_j)$ above, and obtain

$$\sum_{ij} \left\{ \frac{\left(\hat{\sigma}_i \hat{\sigma}_j R_{ij} + \hat{\sigma}_i \hat{\sigma}_j R_{ij}(1 - R_{ij}^2)\eta^2(n_{ij}) - \hat{\Sigma}_{ij}\right)}{\hat{\sigma}_i \hat{\sigma}_j (1 - R_{ij}^2)\eta(n_{ij})} \right\}^2.$$

The loss function above is a highly nonconvex function in R_{ij} or Σ_{ij}. To this end, we approximate the above by replacing some unknown population quantities by their sample analogues. This results in a loss function:

$$\mathcal{L}(\Sigma) = \sum_{ij} \left\{ \frac{\left(\Sigma_{ij} + \Sigma_{ij}(1 - \hat{R}_{ij}^2)\eta^2(n_{ij}) - \hat{\Sigma}_{ij}\right)}{\hat{\sigma}_i \hat{\sigma}_j (1 - \hat{R}_{ij}^2)\eta(n_{ij})} \right\}^2, \tag{33}$$

which is convex in Σ. In words, $\mathcal{L}(\Sigma)$ above, is a measure of how close Σ_{ij}s are to the pairwise covariance terms $\hat{\Sigma}_{ij}$s—this critically depends upon the number of observed samples n_{ij} for every pair (i, j).

We now present Lemma 2 and its proof:

Lemma 2. *Assume that all conditions of Lemma 1 hold. If n_{ij} is large so that Cn_{ij}^{-4} is negligible for a constant C, we have:*

$$E\left(\hat{R}_{ij}|R_{ij}\right) \approx R_{ij} + R_{ij}(1 - R_{ij}^2)\eta^2(n_{ij}) \tag{34}$$

and

$$var\left(\hat{R}_{ij}|R_{ij}\right) \approx (1 - R_{ij}^2)^2 \eta^2(n_{ij}) \tag{35}$$

where $\eta(n_{ij})$ is as described in (15).

Proof of Lemma 2

We re-write $\hat{R}_{ij}$ as a function of the Fisher Z-score

$$\hat{R}_{ij} = \frac{\exp(2\hat{Z}_{ij}) - 1}{\exp(2\hat{Z}_{ij}) + 1} \tag{36}$$

We then expand $\hat{R}_{ij}$ as a function of $\hat{Z}_{ij}$ around the population Fisher Z-score Z_{ij} using the 2nd order Taylor series expansion as follows:

$$\begin{aligned}\hat{R}_{ij} \approx\; & \frac{\exp(2Z_{ij}) - 1}{\exp(2Z_{ij}) + 1} + \frac{4\exp(2Z_{ij})}{\exp(2Z_{ij}) + 1}(\hat{Z}_{ij} - Z_{ij}) + \\ & \frac{4\exp(2Z_{ij})(\exp(2Z_{ij}) - 1)}{(\exp(2Z_{ij}) + 1)^3}(\hat{Z}_{ij} - Z_{ij})^2 \\ =& R_{ij} + (1 - R_{ij}^2)(\hat{Z}_{ij} - Z_{ij}) + R_{ij}(1 - R_{ij}^2)(\hat{Z}_{ij} - Z_{ij})^2\end{aligned} \tag{37}$$

Using the fact that $E(\hat{Z}_{ij}|R_{ij}) = E(\hat{Z}_{ij}|Z_{ij}) = Z_{ij}$, we get from (37)

$$\begin{aligned}E(\hat{R}_{ij}|R_{ij}) \approx& R_{ij} + R_{ij}(1 - R_{ij}^2)E\left((\hat{Z}_{ij} - Z_{ij})^2|R_{ij}\right) \\ =& R_{ij} + R_{ij}(1 - R_{ij}^2)\eta_{ij}^2\end{aligned} \tag{38}$$

and

$$\begin{aligned}\text{var}\left(\hat{R}_{ij}|R_{ij}\right) \approx& (1 - R_{ij}^2)^2\eta^2(n_{ij}) + Cn_{ij}^{-4} \\ \approx& (1 - R_{ij}^2)^2\eta^2(n_{ij}),\end{aligned} \tag{39}$$

where (39) makes use of the fact that Cn_{ij}^{-4} is negligible as per the condition of Lemma 2; and the cross (covariance) term vanishes as it is the third moment of a Gaussian with mean zero.

Derivation of D in (7)

Here we discuss how we derive the analytical form of D in (7) in the optimization framework in (6).

Let $\tilde{\Sigma}$ be the sample covariance matrix of X^f (i.e., the fully observed version of X) We implicitly assume that the perturbation amount Δ is such that $\hat{\Sigma} + \Delta$ is a good approximation to the unobserved $\tilde{\Sigma}$. That is,

$$|\Delta_{ij}| \approx |\hat{\Sigma}_{ij} - \tilde{\Sigma}_{ij}| \leq D_{ij} \tag{40}$$

We can write

$$|\hat{\Sigma}_{ij} - \tilde{\Sigma}_{ij}| \leq |\hat{\Sigma}_{ij} - \Sigma_{ij}| + |\tilde{\Sigma}_{ij} - \Sigma_{ij}|. \tag{41}$$

We propose bounds on each of the two terms on the right using our results from the Robocov covariance matrix section. We know that the first term would be bounded by C_{ij} from Corollary 1. Note that $\tilde{\Sigma}_{ij}$ is an instance of $\hat{\Sigma}_{ij}$ when $n_{ij} = N$ — i.e., all samples are observed. Hence, the bound will be similar to C_{ij} but with n_{ij} replaced by N. We therefore define

$$Q_{ij} := \hat{\sigma}_i \hat{\sigma}_j \min\left(2, \eta(N)\left\{3(1 - \tilde{R}_{ij}^2) + 2\sqrt{3}\eta(N)\right\}\right) \tag{42}$$

where $\tilde{R}$ is the correlation matrix corresponding to $\tilde{\Sigma}$.

When N is reasonably large, $|\eta(N)\tilde{R}_{ij}^2 - \eta(N)\hat{R}_{ij}^2|$ is very small since both $\hat{R}_{ij}^2$ and $\tilde{R}_{ij}^2$ are bounded between 0 and 1 and $\eta(N) \to 0$ as $N \to \infty$.

Therefore we can effectively replace Q_{ij} by $C_{ij}^{'}$ defined as:

$$C_{ij}^{'} := \hat{\sigma}_i \hat{\sigma}_j \min\left(2, \eta(N)\left\{3(1 - \hat{R}_{ij}^2) + 2\sqrt{3}\eta(N)\right\}\right) \tag{43}$$

This provides a justification for the choice of D appearing in (7).

Arriving at the Robocov inverse covariance estimator in Section 2.2

Note that Problem (6) involves minimization of a pointwise maximum (over Δ) of convex functions $\Omega \mapsto L(\Omega; \hat{\Sigma} + \Delta) + \lambda \sum_{ij} |\Omega_{ij}|$. Hence, Problem (6) is convex [16] in Ω.

Here we explain how the min-max optimization problem in (6) leads to the optimization problem in (8).

To this end, note that:

$$\begin{aligned}
&\max_{\Delta:|\Delta_{ij}|\le D_{ij},\forall i,j} \Big\{-\log\det(\Omega) + \langle \Omega, \hat{\Sigma}+\Delta\rangle\Big\} \\
&= \max_{\Delta:\,|\Delta_{ij}|\le D_{ij},\forall i,j} \Big\{-\log\det(\Omega) + \langle \Omega, \hat{\Sigma}\rangle + \langle \Omega, \Delta\rangle\Big\} \\
&= -\log\det(\Omega) + \langle \Omega, \hat{\Sigma}\rangle + \max_{\Delta:\,|\Delta_{ij}|\le D_{ij},\forall i,j} \langle \Omega, \Delta\rangle \\
&= -\log\det(\Omega) + \langle \Omega, \hat{\Sigma}\rangle + \sum_{i,j} D_{ij}|\Omega_{ij}|
\end{aligned} \tag{44}$$

where, the last line follows by noting that

$$\langle \Omega, \Delta\rangle = \sum_{ij}\Omega_{ij}\Delta_{ij} \le \sum_{ij}|\Omega_{ij}|\cdot|\Delta_{ij}| \le \sum_{ij}|\Omega_{ij}|D_{ij}$$

and an equality above holds when $\Delta_{ij} = \text{sign}(\Omega_{ij})|D_{ij}|$ for all i, j,

Using (44), Problem (6) becomes:

$$\begin{aligned}
&\min_{\Omega\succeq 0}\ \Big\{ \max_{\substack{\Delta:\\ |\Delta_{ij}|\le D_{ij},\forall i,j}} \Big\{-\log\det(\Omega) + \langle \Omega, \hat{\Sigma}+\Delta\rangle\Big\} \\
&\qquad + \lambda\sum_{ij}|\Omega_{ij}|\Big\} \\
&= \min_{\Omega\succeq 0}\Big\{ -\log\det(\Omega) + \langle \Omega, \hat{\Sigma}\rangle + \sum_{i,j} D_{ij}|\Omega_{ij}| \\
&\qquad + \lambda\sum_{ij}|\Omega_{ij}|\Big\}
\end{aligned}$$

which is the formulation appearing in (8).

Simulation settings

The parameter models for the simulated population models in Figure 1 are as follows.

- **Hub**: The hub matrix population model for both Figure 1 and Table 1 comprised of correlation blocks of size 5. Each block had all off-diagonal entries equal to 0.7.
- **Toeplitz**: The Toeplitz matrix population model A in Figure 1 had entries of the form $A_{ij} = \max\{0, 1 - 0.1 * |i-j|\}$.
- **1-band precision**: The 1-band precision matrix population model in Figure 1 is of the form $A_{i,i+1} = 0.5$ an $A_{i,j} = 0$ for $j \neq i, i+1$ for each feature i.

Performance metrics

Three performance metrics were used to compare different correlation and partial correlation estimators for different simulation settings (Table 1). They include

- **FP2 : False Positive 2-norm**: Euclidean distance of the estimated correlation or partial correlation values for feature pairs with population correlation or partial correlation equal to 0.
- **FPR: False Positive Rate**: The proportion of feature pairs with population correlation (partial correlation) equal to 0 that have estimated correlation (partial correlation) greater than 0.1.
- **FNR: False Negative Rate**: The proportion of feature pairs with population correlation (partial correlation) greater than 0.1 that have estimated correlation (partial correlation) less than 0.01.

Stratified LD-score regression

Stratified LD score regression (S-LDSC) is a method that assesses the contribution of a genomic annotation to disease and complex trait heritability[31, 34]. S-LDSC assumes that the per-SNP heritability or variance of effect size (of standardized genotype on trait) of each SNP is equal to the linear contribution of each annotation

$$var\,(\beta_j) := \sum_c a_{cj}\tau_c, \tag{45}$$

where a_{cj} is the value of annotation c for SNP j, where a_{cj} is binary in our case, and τ_c is the contribution of annotation c to per-SNP heritability conditioned on other annotations. S-LDSC estimates the τ_c for each annotation using the following equation

$$E\left[\chi_j^2\right] = N\sum_c l(j,c)\tau_c + 1, \tag{46}$$

where $l(j,c) = \sum_k a_{ck} r_{jk}^2$ is the *stratified LD score* of SNP j with respect to annotation c and r_{jk} is the genotypic correlation between SNPs j and k computed using data from 1000 Genomes Project[30] (see URLs); N is the GWAS sample size.

We assess the informativeness of an annotation c using two metrics. The first metric is enrichment (E_c), defined as follows (for binary and probabilistic annotations only):

$$E_c = \frac{\frac{h_g^2(c)}{h_g^2}}{\frac{\sum_j a_{cj}}{M}}, \tag{47}$$

where $h_g^2(c)$ is the heritability explained by the SNPs in annotation c, weighted by the annotation values.

The second metric is standardized effect size ($\tau^\star$) defined as follows (for binary, probabilistic, and continuous-valued annotations):

$$\tau_c^\star = \frac{\tau_c sd_c}{\frac{h_g^2}{M}}, \tag{48}$$

where sd_c is the standard error of annotation c, h_g^2 the total SNP heritability and M is the total number of SNPs on which this heritability is computed (equal to $5,961,159$ in our analyses). $\tau_c^\star$ represents the proportionate change in per-SNP heritability associated to a 1 standard deviation increase in the value of the annotation.