Covariance Estimation for Multivariate Linear Models

Charles Zheng

October 29, 2015

1 Intro

Let

$$Y = XB + E$$

where $E \sim N(0, \Sigma)$. Y and E are $n \times q$ matrices, X is $n \times p$.

How can we estimate Σ?

If $n \leq p$, we can use

$$\hat{\Sigma} = (Y - X\hat{B})^T (I/n)(Y - \hat{B})$$

where $\hat{B}$ is the OLS estimator.

However, in high-dimensional settings, $\hat{B}$ will not be unbiased. Hence

$$\mathbf{E}[(Y - X\hat{B})^T (I/n)(Y - \hat{B})] = \Sigma + \delta^T X^T X \delta$$

where $\delta = B - \hat{B}$. Hence we have a lot of extra bias if we use the residual covariance.

2 Debiased estimator

Rather than the sample covariance of the residuals, we propose computing

$$\hat{\Sigma}_0 = (Y - X\hat{B})^T \Xi (Y - \hat{B})$$

where Ξ is some $n \times n$ contrast matrix. $\hat{\Sigma}_0$ is a "debiased" estimator, where the matrix Ξ is chosen to minimize the effect of the bias from the error coming from the signal $X(B - \hat{B})$. Our final estimator may take the form

$$\hat{\Sigma} = \gamma((1-\alpha)\hat{\Sigma}_0 + \alpha \mathrm{diag}(\hat{\Sigma}_0))$$

where α controls the shrinkage of off-diagonals and γ controls overall shrinkage.

For now let us outline some heuristics for choosing Ξ. We have

$$\mathbf{E}[\hat{\Sigma}_0] = \mathbf{E}[E^T \Xi E] + \mathbf{E}[\delta^T X^T \Xi X \delta]$$

since the cross-term has zero expectation. Let us deal with the first term. Letting $\Xi = V^T \Lambda V$, We have

$$\mathbf{E}[E^T \Xi E] = \mathbf{E}[\Sigma^{1/2} Z^T \Lambda Z \Sigma^{1/2}] = \Sigma^{1/2} \mathbf{E}[Z^T \Lambda Z] \Sigma^{1/2}$$

where Z is a $q \times n$ matrix with iid normal elements. We have

$$\mathbf{E}[Z^T \Lambda Z] = \sum_{i=1}^{n} \lambda_i \mathbf{E}[Z_i Z_i^T] = \sum_{i=1}^{n} \lambda_i I = \mathrm{tr}[\Lambda] I$$

where Z_i is the ith column of Z. Since $\mathrm{tr}[\Lambda] = \mathrm{tr}[\Xi]$ we therefore get

$$\mathbf{E}[E^T \Xi E] = \mathrm{tr}[\Xi] \Sigma$$

Hence it makes sense to require $\mathrm{tr}[\Xi] = 1$ so that $\mathbf{E}[E^T \Xi E] = \Sigma$ andtherefore $\hat{\Sigma}_0$ is unbiased in a limiting case where we can estimate B perfectly.

Meanwhile, let us evaluate the trace of the bias term

$$\mathrm{bias} = \mathrm{tr}\mathbf{E}[\delta^T X^T \Xi X \delta] = \mathrm{tr}[\Xi X X^T \mathbf{E}[\delta \delta^T]]$$

If we make the assumption that $\mathbf{E}[\delta\delta^T]$ is some multiple of the identity, then we get

$$\mathrm{bias} = \mathrm{tr}[\Xi X X^T]$$

Finally, let us consider the variance of $\hat{\Sigma}_0$. The full expression of the variance is quite complex. Hence, we instead look only at the variance

$$\mathrm{Var}[E^T \Xi E] = \mathrm{Var}[\Sigma^{1/2} Z^T \Lambda Z \Sigma^{1/2}]$$

This is still difficult to analyze, so we neglect the $\Sigma^{1/2}$ terms and consider

$$\mathrm{tr}\mathrm{Var}[Z^T \Lambda Z] = 2\mathrm{tr}[\Lambda^2] = 2\mathrm{tr}[\Xi^2].$$

Now we propose the method for choosing Ξ. We would like to require $\mathrm{tr}[\Xi] = 1$, so that $\hat{\Sigma}$ is unbiased in the case of perfect signal estimation, and we would like to minimize some combination of the bias and variance. This leads to the convex program

$$\text{minimize}\mathrm{tr}[\Xi X X^T] + \eta \mathrm{tr}[\Xi^2] \text{ subject to } \mathrm{tr}[\Xi] = 1.$$

Let $XX^T = \Gamma W \Gamma^T$ where W is diagonal. We claim that the minimizing Ξ also takes the form of $\Gamma D \Gamma^T$ for some diagonal D. If we accept the claim, then the above program is rewritten as

$$\text{minimize} \sum_{i=1}^n d_i \lambda_i + \eta \sum_{i=1}^n \lambda_i^2 \text{ subject to } \sum_{i=1}^n \lambda_i = 1.$$

We can solve it by forming the Lagrangian

$$\sum_{i=1}^n d_i \lambda_i + \frac{\eta}{2} \sum_{i=1}^n \lambda_i^2 + \gamma \left(1 - \sum_{i=1}^n \lambda_i \right) - \sum_i \mu_i \lambda_i$$

and from the KKT conditions we get

$$\begin{cases} 0 = \lambda_i = \frac{\mu + \gamma - d_i}{\eta} & \text{or} \\ 0 < \lambda_i = \frac{\gamma - d_i}{\eta} & \end{cases}$$

Hence, we obtain $\lambda_i = [\frac{\gamma - d_i}{\eta}]_+$, and we can determine γ by the condition

$$\sum_{i=1}^n \lambda_i = \sum_{i=1}^n \left[\frac{\gamma - d_i}{\eta} \right]_+ = 1$$

Let $\gamma(\eta)$ denote the solution to the above. We therefore form

$$\Xi(\eta) = V \mathrm{diag} \left(\left[\frac{\gamma - d_i}{\eta} \right]_+ \right) V^T$$