RDPG with Covariates

Cong Mu

1 Notes

Consider RDPG with covariates as

$$P_{ij} = X_i^\top X_j + \beta 1_{\{Z_i = Z_j\}}. \tag{1}$$

Given an adjacency matrix A with observed covariates Z, we use the following procedure to estimate β.

1. Estimate $\hat{X}$ using Adjacency Spectral Embedding (ASE).
2. Cluster $\hat{X}$ using Gaussian Mixture Model (GMM) and compute the means of clusters $\hat{\mu}$.
3. Compute matrix $B_{\hat{\mu}} = \hat{\mu}\hat{\mu}^\top$ and cluster the diagonal of $B_{\hat{\mu}}$.
4. Estimate β by subtracting all corresponding (based on clusters in last step) terms in $B_{\hat{\mu}}$ and taking the mean.
5. Post analysis such as removing the effects of covariates.

We do some simulations to test whether this procedure could work well if we change

- Number of Blocks
- Dimension of Latent Position
- Size of Each Block and Each Gender (Binary Covariate)

2 Simulation

2.1 Number of Blocks

Here we fix number of nodes $n = 2000$, dimension of latent position $d = 1$, the size of each block and each gender (binary covariate) to be balanced, and consider number of blocks $K = 2, 4$.

2.1.1 K = 2

We consider latent position to be $[0.1, 0.3]$, i.e. $p = 0.1$, $q = 0.3$ and $\beta = 0.3$. Then we have the block probability matrix as

$$B_{cov} = \begin{array}{c} \\ male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \begin{array}{c} \begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array} \\ \begin{pmatrix} 0.31 & 0.01 & 0.33 & 0.03 \\ 0.01 & 0.31 & 0.03 & 0.33 \\ 0.33 & 0.03 & 0.39 & 0.09 \\ 0.03 & 0.33 & 0.09 & 0.39 \end{pmatrix} \end{array} \tag{2}$$

With our procedure, we compute the following matrix

$$B_{\hat{\mu}} = \hat{\mu}\hat{\mu}^\top = \begin{pmatrix} 0.2991 & 0.0006 & 0.3377 & 0.0368 \\ 0.0006 & 0.2986 & 0.0369 & 0.3370 \\ 0.3377 & 0.0369 & 0.3857 & 0.0825 \\ 0.0368 & 0.3370 & 0.0825 & 0.3848 \end{pmatrix} \tag{3}$$

And estimate β as $\hat{\beta} = 0.3005$.

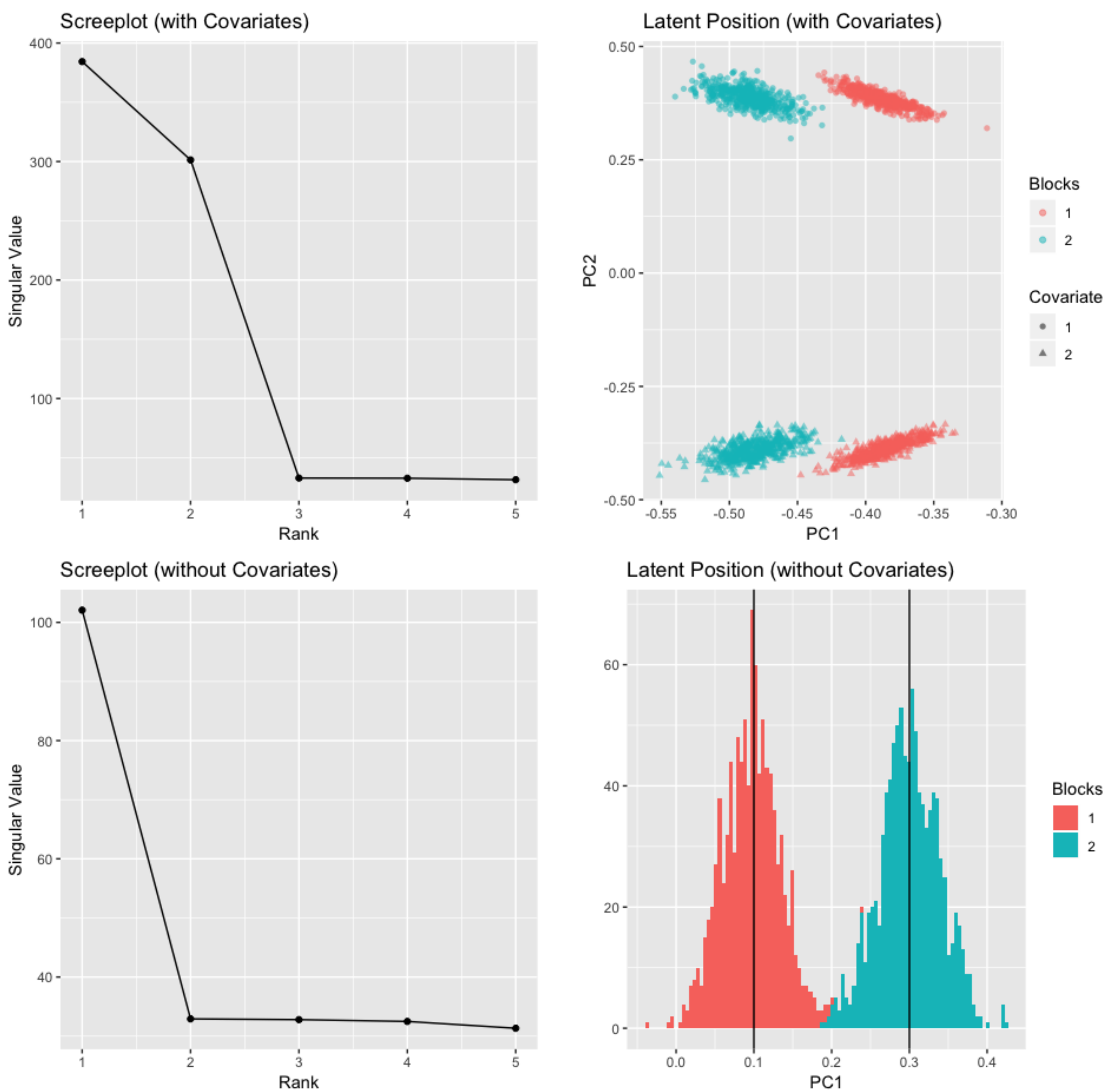

Figure 1: $K = 2,\ d = 1,\ n = 2000,\ p = 0.1,\ q = 0.3,\ \beta = 0.3$, Balanced

2.1.2 K = 4

We consider latent position to be $[0.1, 0.3, 0.5, 0.7]$, and $\beta = 0.3$. Then we have the block probability matrix as

$$B_{cov} = \begin{array}{c|cccccccc} & male_1 & female_1 & male_2 & female_2 & male_3 & female_3 & male_4 & female_4 \\ \hline male_1 & 0.31 & 0.01 & 0.33 & 0.03 & 0.35 & 0.05 & 0.37 & 0.07 \\ female_1 & 0.01 & 0.31 & 0.03 & 0.33 & 0.05 & 0.35 & 0.07 & 0.37 \\ male_2 & 0.33 & 0.03 & 0.39 & 0.09 & 0.45 & 0.15 & 0.51 & 0.21 \\ female_2 & 0.03 & 0.33 & 0.09 & 0.39 & 0.15 & 0.45 & 0.21 & 0.51 \\ male_3 & 0.35 & 0.05 & 0.45 & 0.15 & 0.55 & 0.25 & 0.65 & 0.35 \\ female_3 & 0.05 & 0.35 & 0.15 & 0.45 & 0.25 & 0.55 & 0.35 & 0.65 \\ male_4 & 0.37 & 0.07 & 0.51 & 0.21 & 0.65 & 0.35 & 0.79 & 0.49 \\ female_4 & 0.07 & 0.37 & 0.21 & 0.51 & 0.35 & 0.65 & 0.49 & 0.79 \end{array} \tag{4}$$

With our procedure, we compute the following matrix

$$B_{\hat{\mu}} = \hat{\mu}\hat{\mu}^\top = \begin{pmatrix} 0.2547 & 0.3534 & -0.0451 & 0.3037 & 0.0041 & 0.0545 & 0.4045 & 0.1057 \\ 0.3534 & 0.5473 & 0.0558 & 0.4497 & 0.1534 & 0.2502 & 0.6452 & 0.3499 \\ -0.0451 & 0.0558 & 0.2543 & 0.0052 & 0.3064 & 0.3533 & 0.1030 & 0.4040 \\ 0.3037 & 0.4497 & 0.0052 & 0.3762 & 0.0785 & 0.1520 & 0.5241 & 0.2273 \\ 0.0041 & 0.1534 & 0.3064 & 0.0785 & 0.3831 & 0.4535 & 0.2244 & 0.5289 \\ 0.0545 & 0.2502 & 0.3533 & 0.1520 & 0.4535 & 0.5466 & 0.3440 & 0.6457 \\ 0.4045 & 0.6452 & 0.1030 & 0.5241 & 0.2244 & 0.3440 & 0.7663 & 0.4676 \\ 0.1057 & 0.3499 & 0.4040 & 0.2273 & 0.5289 & 0.6457 & 0.4676 & 0.7695 \end{pmatrix} \tag{5}$$

And estimate β as $\hat{\beta} = 0.2994$.

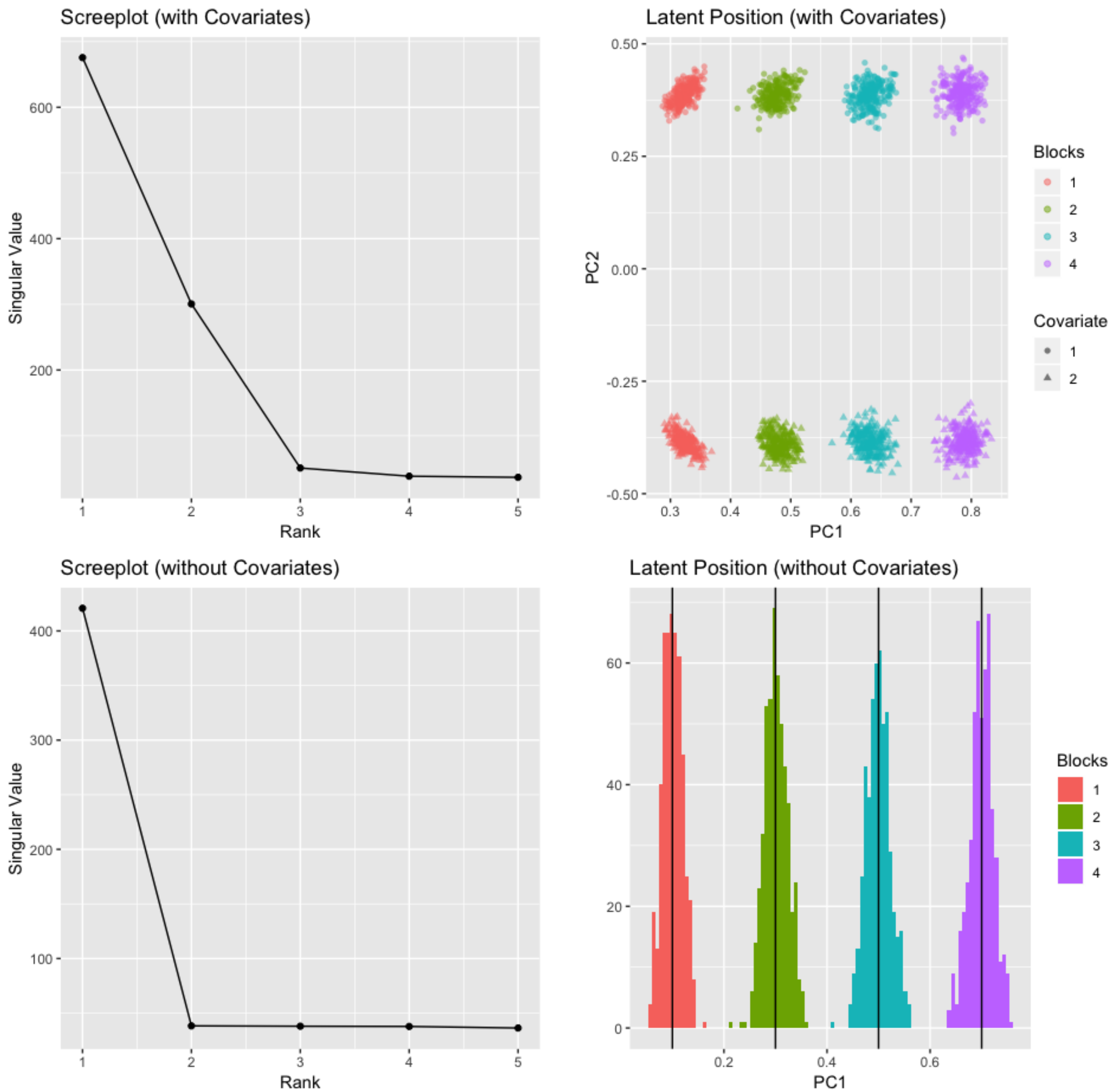

Figure 2: $K = 4,\ d = 1,\ n = 2000,\ p = 0.1,\ q = 0.3,\ r = 0.5,\ s = 0.7,\ \beta = 0.3$, Balanced

2.2 Dimension of Latent Posistion

Here we fix number of nodes $n = 2000$, number of blocks $K = 2$, the size of each block and each gender (binary covariate) to be balanced, and consider dimension of latent position $d = 1, 2$.

2.2.1 d = 1

We consider latent position to be $[0.3, 0.7]$, i.e. $p = 0.3$, $q = 0.7$ and $\beta = 0.2$. Then we have the block probability matrix as

$$B_{cov} = \begin{array}{c} \\ male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \begin{array}{c} \begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array} \\ \begin{pmatrix} 0.29 & 0.09 & 0.41 & 0.21 \\ 0.09 & 0.29 & 0.21 & 0.41 \\ 0.41 & 0.21 & 0.69 & 0.49 \\ 0.21 & 0.41 & 0.49 & 0.69 \end{pmatrix} \end{array} \tag{6}$$

With our procedure, we compute the following matrix

$$B_{\hat{\mu}} = \hat{\mu}\hat{\mu}^{\top} = \begin{pmatrix} 0.2725 & 0.0739 & 0.4180 & 0.2177 \\ 0.0739 & 0.2740 & 0.2195 & 0.4187 \\ 0.4180 & 0.2195 & 0.6854 & 0.4843 \\ 0.2177 & 0.4187 & 0.4843 & 0.6834 \end{pmatrix} \tag{7}$$

And estimate β as $\hat{\beta} = 0.1997$.

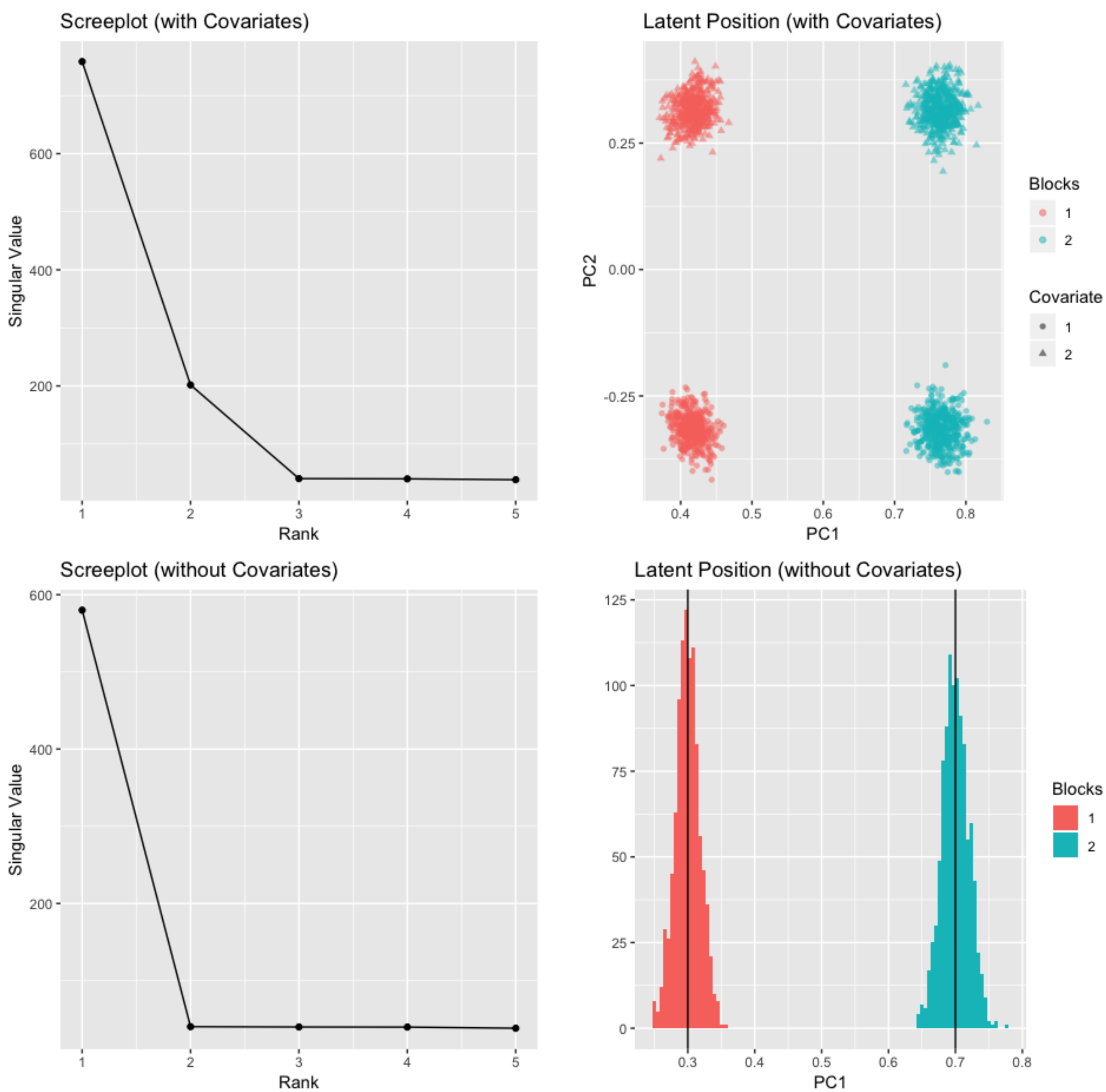

Figure 3: $K = 2,\ d = 1,\ n = 2000,\ p = 0.3,\ q = 0.7,\ \beta = 0.2$, Balanced

2.2.2 d = 2

We consider latent position to be $x_1 = [0.63, -0.14]$, $x_2 = [0.69, 0.13]$, and $\beta = 0.2$. Then we have the block probability matrix as

$$B_{cov} = \begin{array}{c} \\ male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \begin{array}{c} \begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array} \\ \begin{pmatrix} 0.6165 & 0.4165 & 0.6165 & 0.4165 \\ 0.4165 & 0.6165 & 0.4165 & 0.6165 \\ 0.6165 & 0.4165 & 0.6930 & 0.4930 \\ 0.4165 & 0.6165 & 0.4930 & 0.6930 \end{pmatrix} \end{array} \tag{8}$$

With our procedure, we compute the following matrix

$$B_{\hat{\mu}} = \hat{\mu}\hat{\mu}^\top = \begin{pmatrix} 0.5957 & 0.6345 & 0.3963 & 0.4350 \\ 0.6345 & 0.6761 & 0.4332 & 0.4749 \\ 0.3963 & 0.4332 & 0.5960 & 0.6329 \\ 0.4350 & 0.4749 & 0.6329 & 0.6727 \end{pmatrix} \tag{9}$$

And estimate β as $\hat{\beta} = 0.1995$.

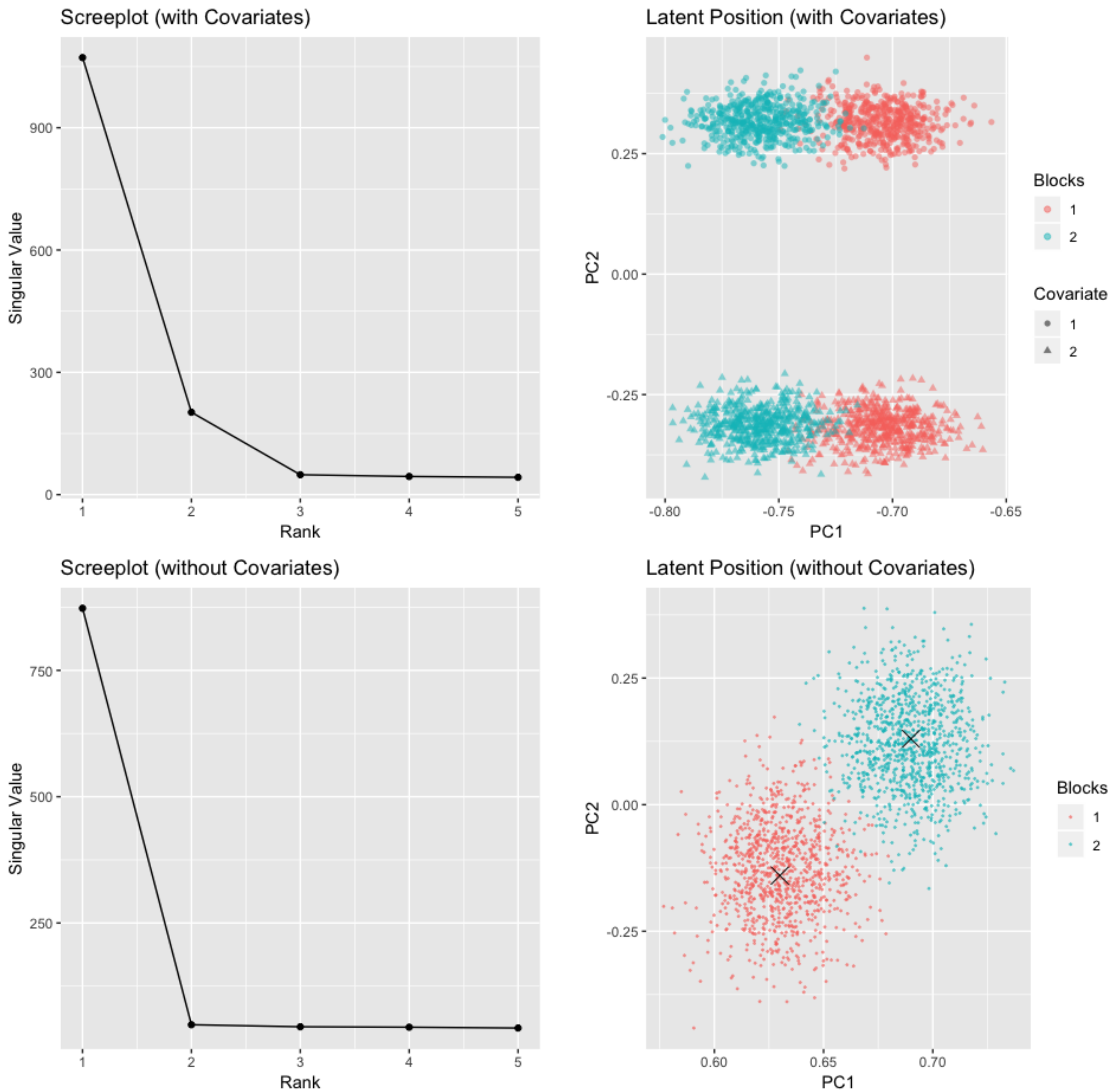

Figure 4: $K = 2,\ d = 2,\ n = 2000,\ x_1 = [0.63, -0.14],\ x_2 = [0.69, 0.13],\ \beta = 0.2$, Balanced

2.3 Size of Each Block and Each Gender (Binary Covariate)

Here we fix number of nodes $n = 2000$, number of blocks $K = 2$, dimension of latent position $d = 1$, and consider the size of each block to be $(0.2, 0.8)$, the size of each gender to be $(0.3, 0.7)$. We consider latent position to be $[0.2, 0.6]$, i.e. $p = 0.2$, $q = 0.6$ and $\beta = 0.4$. Then we have the block probability matrix as

$$B_{cov} = \begin{array}{c} \\ male_1 \\ female_1 \\ male_2 \\ female_2 \end{array} \begin{array}{c} \begin{array}{cccc} male_1 & female_1 & male_2 & female_2 \end{array} \\ \begin{pmatrix} 0.44 & 0.04 & 0.52 & 0.12 \\ 0.04 & 0.44 & 0.12 & 0.52 \\ 0.52 & 0.12 & 0.76 & 0.36 \\ 0.12 & 0.52 & 0.36 & 0.76 \end{pmatrix} \end{array} \tag{10}$$

With our procedure, we compute the following matrix

$$B_{\hat{\mu}} = \hat{\mu}\hat{\mu}^\top = \begin{pmatrix} 0.3858 & -0.0102 & 0.5235 & 0.1289 \\ -0.0102 & 0.3936 & 0.1235 & 0.5265 \\ 0.5235 & 0.1235 & 0.7583 & 0.3601 \\ 0.1289 & 0.5265 & 0.3601 & 0.7570 \end{pmatrix} \tag{11}$$

And estimate β as $\hat{\beta} = 0.3987$.

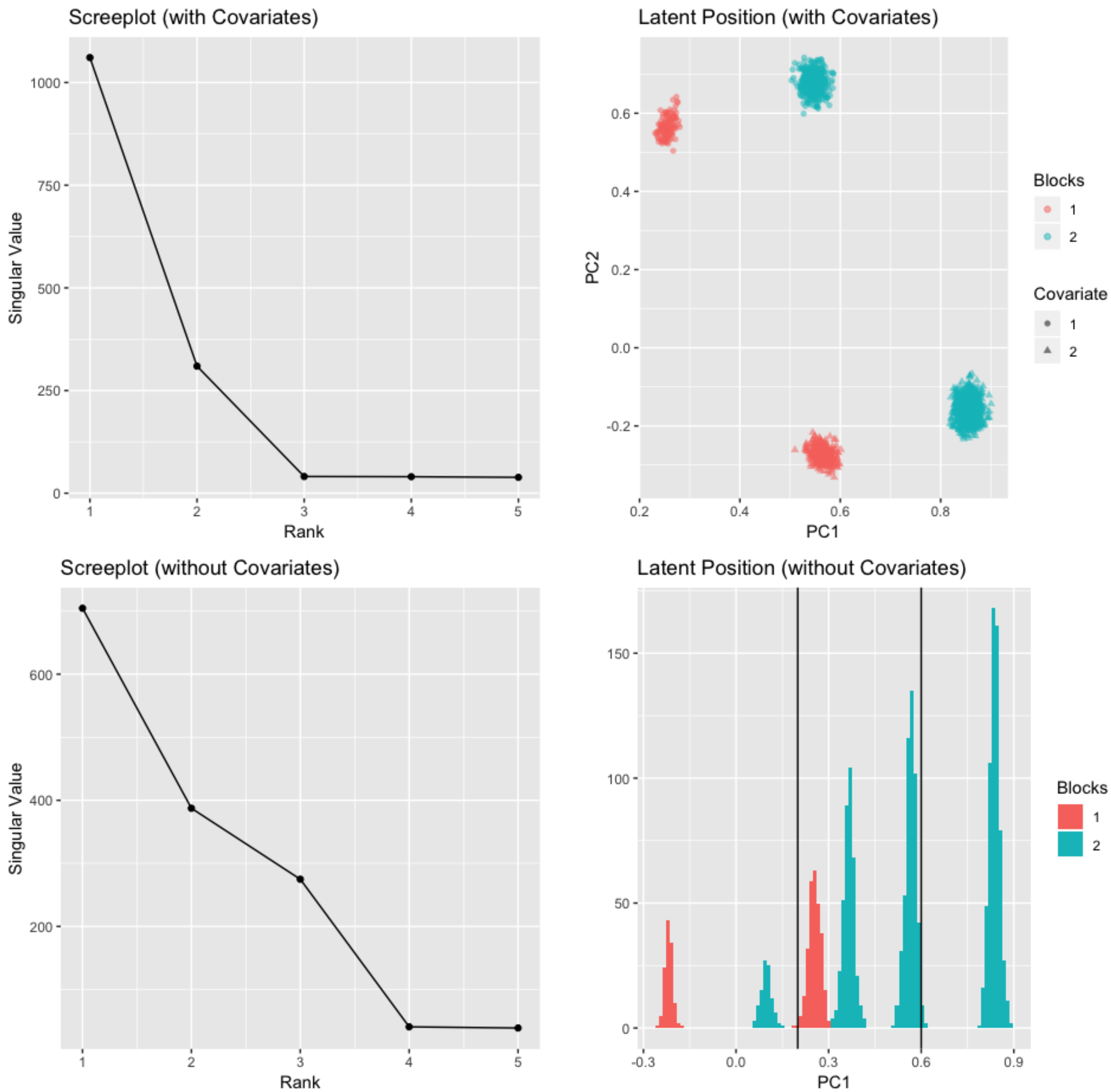

Figure 5: $K = 2,\ d = 1,\ n = 2000,\ p = 0.2,\ q = 0.6,\ \beta = 0.4$, Size of Block = (0.2, 0.8), Size of Gender = (0.3, 0.7)

2.4 Summary

With respect to the estimation of β, we summarize our simulations as follows.

n	K	d	**Size of Blocks**	**Size of Gender**	β	$\hat{\beta}$
2000	2	1	Balanced	Balanced	0.3	0.3005
2000	4	1	Balanced	Balanced	0.3	0.2994
2000	2	1	Balanced	Balanced	0.2	0.1997
2000	2	2	Balanced	Balanced	0.2	0.1995
2000	2	1	(0.2, 0.8)	(0.3, 0.7)	0.4	0.3987

Table 1: Summary of Simulation

In general, the procedure seems to work well under different setting. Note that here we only consider $\beta > 0$, when $\beta < 0$, the matrix $B_{\hat{\mu}}$ seems to be a little different. But we think we could still recover β, just with some additional steps and we are working on that.