

## S1 Appendix: Details about the simulated data

This supplementary appendix describes the origin of the simulated data used for presenting the PCADSC plots. Dataset A consisted of 1000 independent simulated realizations from the  $N(0, \Sigma_1)$ -distribution, while dataset B consisted of 500 independent realizations from  $N(0, \Sigma_1)$  and 500 independent realizations from  $N(0, \Sigma_2)$ . Dataset A was furthermore appended with a grouping variable, randomly dividing the observations into two groups. Dataset B also included a grouping variable, and this variable contained information about which of the two normal distributions each observation was simulated from. The covariance matrices,  $\Sigma_1$  and  $\Sigma_2$ , were defined by

$$\Sigma_1 = \begin{pmatrix} 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 1.0 & 0.7 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.7 & 1.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.4 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.4 & 1.0 \end{pmatrix}$$

and

$$\Sigma_2 = \begin{pmatrix} 1.0 & 0.2 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.2 & 1.0 & 0.1 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.1 & 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.3 & 0.1 \\ 0.0 & 0.0 & 0.0 & 0.3 & 1.0 & 0.2 \\ 0.0 & 0.0 & 0.0 & 0.1 & 0.2 & 1.0 \end{pmatrix},$$

respectively.