PBHL-B574\_simulation\_code

Yao Chen

2023-04-24

## Simulation example 1 code

### correlated data when .

library(PGEE) # requiring `mvtnorm` and `MASS` packages   
set.seed(12345) # for reproducibility  
# correlated normal responses   
#True parameter vector  
  
# covariate matrices  
ar1\_cor <- function(n, rho) {  
 exponent <- abs(matrix(1:n - 1, nrow = n, ncol = n, byrow = TRUE) -   
 (1:n - 1))  
 rho^exponent  
}  
exch\_cor <- function(n, rho) {  
 mat <- matrix(rho, nrow = n, ncol = n)  
 diag(mat) <- 1  
 return(mat)  
}  
  
# strength of within cluster correlation   
rho\_values <- c(0.5, 0.8)  
  
# Create the function to simulate data set   
# consisting of N=200 study participants  
# with 4 observations per participant  
# p = 200 covariates   
simulate<-function(N, rho){  
 # The number of repeated measurements is 4  
 timepoints=4  
 # The number of covariates is 200   
 p <- 200  
 # b\_0  
 beta0 <- c(2, 3, 1.5, 2, rep(0, p - 4))  
 # Simulate errors from MVN  
 e\_i <- c(sapply(1, function(x) t(mvrnorm(n = N, mu = rep(0,timepoints), Sigma = exch\_cor(4,rho)))) )  
 # Generate ID variable  
 id <- c(sapply(1:N, function(x) rep(x, timepoints)))  
 # Simulate x\_ij  
 # x\_ij,1 from Bernoulli(0.5)   
 x\_1 <- c(sapply(1: N, function(x) rep(t(rbinom(1, 1, 0.5)),  
 timepoints)))   
 # x\_ij,2-200 from MVN   
 x\_2 <- matrix(sapply(1:(p-1), function(x) t(mvrnorm(n = N, mu = rep(0,timepoints),   
 Sigma = ar1\_cor(4,0.5)))), ncol=c(p-1))   
 colnames(x\_2) <- paste0("x", 2:200)  
 X <- cbind(x\_1,x\_2)   
 # Simulate response   
 y<- X %\*% beta0 +e\_i  
 # Return data frame  
 data.frame(id,y,X)  
}  
# generate simulation data with correlated normal responses   
data\_sim1 <- simulate(N=200, rho=rho\_values[1])  
data\_sim2 <- simulate(N=200, rho=rho\_values[2])  
  
colnames(data\_sim1)[1:10]  
head(data\_sim1,5)[1:10]  
  
formula <- "y ~.-id"  
family <- gaussian(link = "identity")  
lambda.vec <- seq(0.1,0.9,0.1)  
# find the optimum lambda   
cv\_sim1 <- CVfit(formula = formula, id = id, data = data\_sim1, family = family, scale.fix = TRUE,  
 scale.value = 1, fold = 4, lambda.vec = lambda.vec, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-3)  
print(cv\_sim1) # 0.4 is the best lambda   
names(cv\_sim1)  
cv\_sim1$lam.opt  
cv\_sim2 <- CVfit(formula = formula, id = id, data = data\_sim2, family = family, scale.fix = TRUE,  
 scale.value = 1, fold = 4, lambda.vec = lambda.vec, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-3)  
print(cv\_sim2) # 0.5 is the best lambda   
names(cv\_sim2)  
cv\_sim2$lam.opt   
  
# number of generated data set for each set up = 100   
nsim=100   
b0 <- c(2, 3, 1.5, 2, rep(0, 200 - 4))  
  
# Perform the simulation - PGEE and rho=0.5   
# Perform the simulation - rho=0.5   
for(i in 1:nsim){  
 print(i)   
 N=200   
 rho=rho\_values[1] # rho=0.5, 0.8   
 data<-simulate(N, rho)  
 myfit1\_indp <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit1\_exch <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit1\_ar1 <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 pgee\_b1 <- coef(myfit1\_indp)  
 pgee\_b2 <- coef(myfit1\_exch)  
 pgee\_b3 <- coef(myfit1\_ar1)  
 myfit2\_indp <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit2\_exch <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit2\_ar1 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 gee\_b1 <- coef(myfit2\_indp)  
 gee\_b2 <- coef(myfit2\_exch)  
 gee\_b3 <- coef(myfit2\_ar1)  
 myfit3\_indp <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(0, b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit3\_exch <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(0, b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit3\_ar1 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(0, b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 oracle\_b1 <- coef(myfit3\_indp)  
 oracle\_b2 <- coef(myfit3\_exch)  
 oracle\_b3 <- coef(myfit3\_ar1)  
 if(i==1){  
 pgee1\_beta1<-pgee\_b1  
 pgee1\_beta2<-pgee\_b2  
 pgee1\_beta3<-pgee\_b3  
 gee1\_beta1<-gee\_b1  
 gee1\_beta2<-gee\_b2  
 gee1\_beta3<-gee\_b3  
 oracle1\_beta1<-oracle\_b1  
 oracle1\_beta2<-oracle\_b2   
 oracle1\_beta3<-oracle\_b3  
 } else {  
 pgee1\_beta1<-rbind(pgee1\_beta1,pgee\_b1)  
 pgee1\_beta2<-rbind(pgee1\_beta2,pgee\_b2)  
 pgee1\_beta3<-rbind(pgee1\_beta3,pgee\_b3)  
 gee1\_beta1<-rbind(gee1\_beta1,gee\_b1)  
 gee1\_beta2<-rbind(gee1\_beta2,gee\_b2)  
 gee1\_beta3<-rbind(gee1\_beta3,gee\_b3)  
 oracle1\_beta1<-rbind(oracle1\_beta1,oracle\_b1)  
 oracle1\_beta2<-rbind(oracle1\_beta2,oracle\_b2)  
 oracle1\_beta3<-rbind(oracle1\_beta3,oracle\_b3)  
 }}  
  
# MSE when rho=0.5  
pgee1\_mse <- c(mean(colSums((pgee1\_beta1[,-1] - b0)^2))/nrow(pgee1\_beta1),   
 mean(colSums((pgee1\_beta2[,-1] - b0)^2))/nrow(pgee1\_beta2),   
 mean(colSums((pgee1\_beta3[,-1] - b0)^2))/nrow(pgee1\_beta3))   
names(pgee1\_mse) <- c("indp", "exch", "ar1")  
gee1\_mse <- c(mean(colSums((gee1\_beta1[,-1] - b0)^2))/nrow(gee1\_beta1),   
 mean(colSums((gee1\_beta2[,-1] - b0)^2))/nrow(gee1\_beta2),   
 mean(colSums((gee1\_beta3[,-1] - b0)^2))/nrow(gee1\_beta3))   
names(gee1\_mse) <- c("indp", "exch", "ar1")  
oracle1\_mse <- c(mean(colSums((oracle1\_beta1[,-1] - b0)^2))/nrow(oracle1\_beta1),   
 mean(colSums((oracle1\_beta2[,-1] - b0)^2))/nrow(oracle1\_beta2),   
 mean(colSums((oracle1\_beta3[,-1] - b0)^2))/nrow(oracle1\_beta3))   
names(oracle1\_mse) <- c("indp", "exch", "ar1")  
  
# U/O/EXACT when rho=0.5  
num\_indp\_pgee1 <- apply(pgee1\_beta1[,-1], 1, function(x) sum(abs(x) > 10^-3))  
num\_exch\_pgee1 <- apply(pgee1\_beta2[,-1], 1, function(x) sum(abs(x) > 10^-3))  
num\_ar1\_pgee1 <- apply(pgee1\_beta3[,-1], 1, function(x) sum(abs(x) > 10^-3))  
prop\_pgee1\_indp <- c(mean(num\_indp\_pgee1 < 4), mean(num\_indp\_pgee1 > 4), mean(num\_indp\_pgee1 == 4))  
prop\_pgee1\_exch <- c(mean(num\_exch\_pgee1 < 4), mean(num\_exch\_pgee1 > 4), mean(num\_exch\_pgee1 == 4))  
prop\_pgee1\_ar1 <- c(mean(num\_ar1\_pgee1 < 4), mean(num\_ar1\_pgee1 > 4), mean(num\_ar1\_pgee1 == 4))  
prop\_pgee1 <- rbind(prop\_pgee1\_indp, prop\_pgee1\_exch, prop\_pgee1\_ar1)  
colnames(prop\_pgee1) <- c("U", "O", "EXACT")  
rownames(prop\_pgee1) <- c("pgee1\_indp", "pgee1\_exch", "pgee1\_ar1")  
prop\_pgee1  
num\_indp\_gee1 <- apply(gee1\_beta1[,-1], 1, function(x) sum(abs(x) > 10^-3))  
num\_exch\_gee1 <- apply(gee1\_beta2[,-1], 1, function(x) sum(abs(x) > 10^-3))  
num\_ar1\_gee1 <- apply(gee1\_beta3[,-1], 1, function(x) sum(abs(x) > 10^-3))  
prop\_gee1\_indp <- c(mean(num\_indp\_gee1 < 4), mean(num\_indp\_gee1 > 4), mean(num\_indp\_gee1 == 4))  
prop\_gee1\_exch <- c(mean(num\_exch\_gee1 < 4), mean(num\_exch\_gee1 > 4), mean(num\_exch\_gee1 == 4))  
prop\_gee1\_ar1 <- c(mean(num\_ar1\_gee1 < 4), mean(num\_ar1\_gee1 > 4), mean(num\_ar1\_gee1 == 4))  
prop\_gee1 <- rbind(prop\_gee1\_indp, prop\_gee1\_exch, prop\_gee1\_ar1)  
colnames(prop\_gee1) <- c("U", "O", "EXACT")  
rownames(prop\_gee1) <- c("gee1\_indp", "gee1\_exch", "gee1\_ar1")  
prop\_gee1  
# TP when rho=0.5  
mean(apply(pgee1\_beta1[,2:5], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(pgee1\_beta2[,2:5], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(pgee1\_beta3[,2:5], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(gee1\_beta1[,2:5], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(gee1\_beta2[,2:5], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(gee1\_beta3[,2:5], 1, function(x) sum(abs(x) > 10^-3)))  
# FP when rho=0.5  
mean(apply(pgee1\_beta1[,-c(1:5)], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(pgee1\_beta2[,-c(1:5)], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(pgee1\_beta3[,-c(1:5)], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(gee1\_beta1[,-c(1:5)], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(gee1\_beta2[,-c(1:5)], 1, function(x) sum(abs(x) > 10^-3)))  
mean(apply(gee1\_beta3[,-c(1:5)], 1, function(x) sum(abs(x) > 10^-3)))

### correlated data when .

# Perform the simulation - rho=0.8  
for(i in 1:nsim){  
 print(i)   
 N=200   
 rho=rho\_values[2] # rho=0.5, 0.8   
 data<-simulate(N, rho)  
 myfit1\_indp <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit1\_exch <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit1\_ar1 <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 pgee\_b1 <- coef(myfit1\_indp)  
 pgee\_b2 <- coef(myfit1\_exch)  
 pgee\_b3 <- coef(myfit1\_ar1)  
 myfit2\_indp <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit2\_exch <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit2\_ar1 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 gee\_b1 <- coef(myfit2\_indp)  
 gee\_b2 <- coef(myfit2\_exch)  
 gee\_b3 <- coef(myfit2\_ar1)  
 myfit3\_indp <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(0, b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit3\_exch <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(0, b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit3\_ar1 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(0, b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 oracle\_b1 <- coef(myfit3\_indp)  
 oracle\_b2 <- coef(myfit3\_exch)  
 oracle\_b3 <- coef(myfit3\_ar1)  
 if(i==1){  
 pgee2\_beta1<-pgee\_b1  
 pgee2\_beta2<-pgee\_b2  
 pgee2\_beta3<-pgee\_b3  
 gee2\_beta1<-gee\_b1  
 gee2\_beta2<-gee\_b2  
 gee2\_beta3<-gee\_b3  
 oracle2\_beta1<-oracle\_b1  
 oracle2\_beta2<-oracle\_b2   
 oracle2\_beta3<-oracle\_b3  
 } else {  
 pgee2\_beta1<-rbind(pgee2\_beta1,pgee\_b1)  
 pgee2\_beta2<-rbind(pgee2\_beta2,pgee\_b2)  
 pgee2\_beta3<-rbind(pgee2\_beta3,pgee\_b3)  
 gee2\_beta1<-rbind(gee2\_beta1,gee\_b1)  
 gee2\_beta2<-rbind(gee2\_beta2,gee\_b2)  
 gee2\_beta3<-rbind(gee2\_beta3,gee\_b3)  
 oracle2\_beta1<-rbind(oracle2\_beta1,oracle\_b1)  
 oracle2\_beta2<-rbind(oracle2\_beta2,oracle\_b2)  
 oracle2\_beta3<-rbind(oracle2\_beta3,oracle\_b3)  
 }}

####################### test: initial beta values = 0 ####################  
############################################################################  
############################################################################  
############################################################################  
############################################################################  
for(i in 1:2){  
 print(i)   
 N=200   
 rho=rho\_values[1] # rho=0.5, 0.8   
 data<-simulate(N, rho)  
 myfit1\_indp <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,   
 beta\_int = NULL, R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit1\_exch <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,   
 beta\_int = NULL, R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit1\_ar1 <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,   
 beta\_int = NULL, R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv\_sim1$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 pgee\_b1 <- coef(myfit1\_indp)  
 pgee\_b2 <- coef(myfit1\_exch)  
 pgee\_b3 <- coef(myfit1\_ar1)  
 myfit2\_indp <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,   
 beta\_int = NULL, R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit2\_exch <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,   
 beta\_int = NULL, R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit2\_ar1 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,   
 beta\_int = NULL,   
 R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 gee\_b1 <- coef(myfit2\_indp)  
 gee\_b2 <- coef(myfit2\_exch)  
 gee\_b3 <- coef(myfit2\_ar1)  
 myfit3\_indp <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit3\_exch <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "exchangeable", Mv = NULL,  
 beta\_int = c(b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 myfit3\_ar1 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "AR-1", Mv = NULL,  
 beta\_int = c(b0), R = NULL, scale.fix = TRUE,  
 scale.value = 1,   
 maxiter = 30, tol = 10^-6, silent = TRUE)  
 oracle\_b1 <- coef(myfit3\_indp)  
 oracle\_b2 <- coef(myfit3\_exch)  
 oracle\_b3 <- coef(myfit3\_ar1)  
 if(i==1){  
 pgee1\_beta1<-pgee\_b1  
 pgee1\_beta2<-pgee\_b2  
 pgee1\_beta3<-pgee\_b3  
 gee1\_beta1<-gee\_b1  
 gee1\_beta2<-gee\_b2  
 gee1\_beta3<-gee\_b3  
 oracle1\_beta1<-oracle\_b1  
 oracle1\_beta2<-oracle\_b2   
 oracle1\_beta3<-oracle\_b3  
 } else {  
 pgee1\_beta1<-rbind(pgee1\_beta1,pgee\_b1)  
 pgee1\_beta2<-rbind(pgee1\_beta2,pgee\_b2)  
 pgee1\_beta3<-rbind(pgee1\_beta3,pgee\_b3)  
 gee1\_beta1<-rbind(gee1\_beta1,gee\_b1)  
 gee1\_beta2<-rbind(gee1\_beta2,gee\_b2)  
 gee1\_beta3<-rbind(gee1\_beta3,gee\_b3)  
 oracle1\_beta1<-rbind(oracle1\_beta1,oracle\_b1)  
 oracle1\_beta2<-rbind(oracle1\_beta2,oracle\_b2)  
 oracle1\_beta3<-rbind(oracle1\_beta3,oracle\_b3)  
 }}

### real-world data

library(PGEE) # requiring `mvtnorm` and `MASS` packages   
# load data  
data(yeastG1)  
data = yeastG1  
# get the column names  
colnames(data)[1:9]  
# see some portion of yeast G1 data  
head(data,5)[1:9]  
# define the input arguments  
formula <- "y ~.-id"  
family <- gaussian(link = "identity")  
lambda.vec <- seq(0.01,0.2,0.01)  
# find the optimum lambda  
cv <- CVfit(formula = formula, id = id, data = data, family = family, scale.fix = TRUE,  
 scale.value = 1, fold = 4, lambda.vec = lambda.vec, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6)  
# print the results  
print(cv)  
# see the returned values by CVfit  
names(cv)  
# get the optimum lambda  
cv$lam.opt  
# fit the PGEE model  
myfit1 <- PGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, lambda = cv$lam.opt, pindex = c(1,2), eps = 10^-6,  
 maxiter = 30, tol = 10^-6, silent = TRUE)  
# get the values returned by myfit object  
names(myfit1)  
  
# see a portion of the results returned by coef(summary(myfit1))  
head(coef(summary(myfit1)),7)  
  
# see the variables which have non-zero coefficients  
index1 <- which(abs(coef(summary(myfit1))[,"Estimate"]) > 10^-3)  
names(abs(coef(summary(myfit1))[index1,"Estimate"]))  
# see the PGEE summary statistics of these non-zero variables  
coef(summary(myfit1))[index1,]  
# fit the GEE model  
myfit2 <- MGEE(formula = formula, id = id, data = data, na.action = NULL,  
 family = family, corstr = "independence", Mv = NULL,  
 beta\_int = c(rep(0,dim(data)[2]-1)), R = NULL, scale.fix = TRUE,  
 scale.value = 1, maxiter = 30, tol = 10^-6, silent = TRUE)  
# see the significantly associated TFs in PGEE analysis  
names(which(abs(coef(summary(myfit1))[index1,"Robust z"]) > 1.96))  
  
# see the significantly associated TFs in GEE analysis  
names(which(abs(coef(summary(myfit2))[,"Robust z"]) > 1.96))