Data simulation and calculation of betas

```
library(knitr)
library(kableExtra)
library(cowplot)
library(ggforce)
```

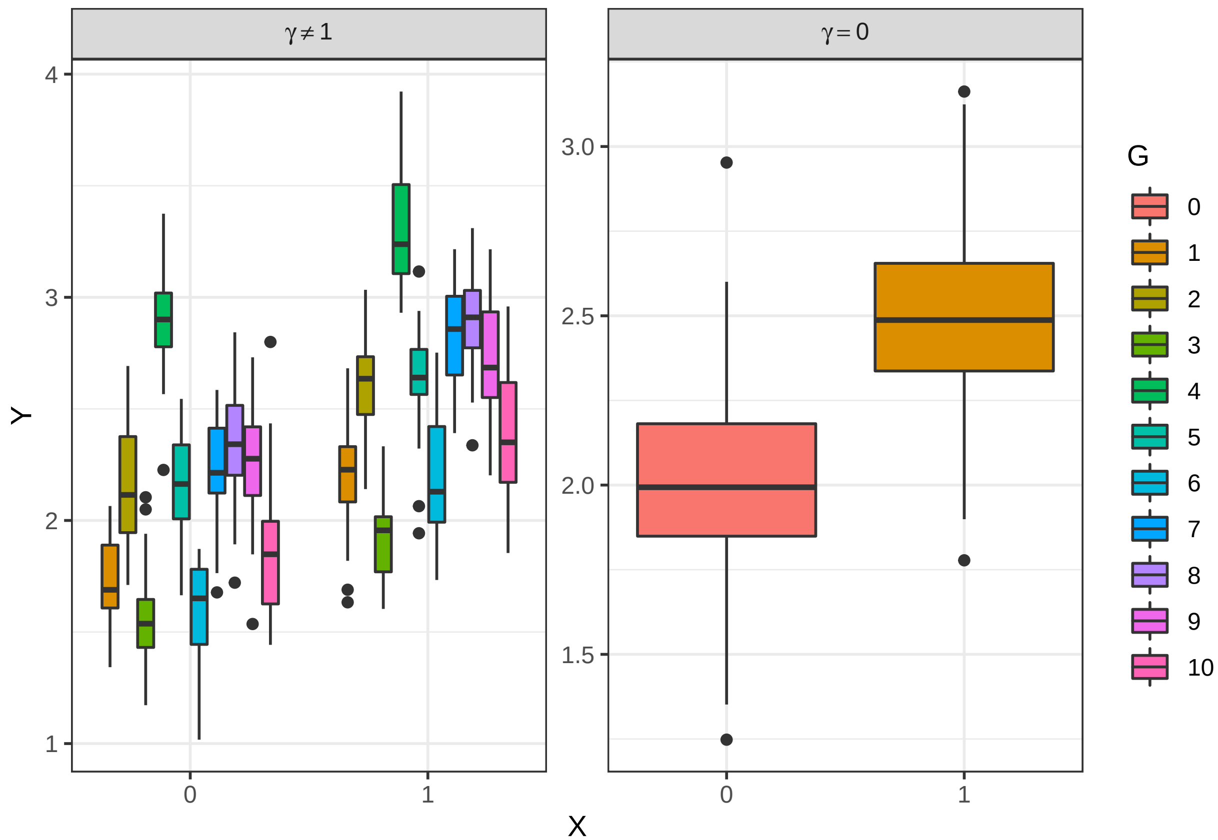
```
## Loading required package: ggplot2
```

```
library(latex2exp)
knitr::opts_chunk$set(cache = FALSE, warning = FALSE, message = FALSE, cache.lazy = FALSE)
```

Simulation data

```
set.seed(1)
K <- 10 # Nombre de groupe
nK <- 50 # Nombre d'observations par groupe
N <- K * nK # Nombre total d'observation
G <- factor(rep(1:K, each = nK))
intercept <- 2
fixefEffect <- .5
aleaEffect <- rnorm(K, sd = .5)
bias <- rnorm(N, sd = .25)

X <- rbinom(N,size=1,prob = .5)
# X_gauss <- rnorm(N,1,2)
Y_withoutAlea <- intercept + fixefEffect * X + bias
# Y_G <- intercept + fixefEffect * X_gauss + aleaEffect[G]  + bias
Y <- intercept + fixefEffect * X + aleaEffect[G] + bias
dfWith <- data.frame(X, Y, G)
dfWithout <- data.frame(X, Y=Y_withoutAlea, G)

H0 <- formula(Y ~ X)
H1 <- formula(Y ~ X + (1|G))

dataPlot = cbind.data.frame(X=rep(factor(X),2), Y = c(Y, Y_withoutAlea),
                            G = factor(c(G,X),levels = 0:10),
                            Effect = factor(rep(c("gamma != 1", "gamma == 0"),
                                                each=length(X))))

ggplot(dataPlot) +
  geom_boxplot(aes(x=X, y=Y, fill = G)) +
  scale_x_discrete(expand = c(0, 0.5)) +
  theme_bw()+
```

```
  ggforce::facet_row(vars(Effect), scales = 'free', space = 'free',
                     labeller = "label_parsed")
```



Linear model with lm function

```
options(warn=-1)
X_prime <- cbind(1, X)

# With Random Effect
lmModel <- lm(H0,data = dfWith)
Sigm <- sqrt(((summary(lmModel)$sigma)**2)*solve(t(X) %*% X))
betaLm <- lmModel$coefficients
SE <- (betaLm[-1] - confint(lmModel)[-1,][1])/1.96

A <- ggplot(dfWith, aes(x = X, y = Y, color = G) ) +
     geom_point() +
     geom_smooth(formula = as.formula(y ~ x), method = "lm", se=.3,aes(fill = G))+
  theme_bw()+theme(legend.position="none")+
  annotate(geom='text', label=TeX("$\\gamma \\neq 0$"), y=3.7, x=.5)


# Without Random Effect
```

```
lmModel <- lm(H0,data = dfWithout)
Sigm.1 <- sqrt(((summary(lmModel)$sigma)**2)*solve(t(X) %*% X))
betaLm.1 <- lmModel$coefficients
SE.1 <- (betaLm.1[-1] - confint(lmModel)[-1,][1])/1.96

B <- ggplot(dfWithout, aes(x = X, y = Y, color = G) ) +
     geom_point() +
     geom_smooth(formula = as.formula(y ~ x), method = "lm",se = .3, aes(fill = G))+
  theme_bw()+theme(axis.title.y = element_blank())+
  annotate(geom='text', label=TeX("$\\gamma=0$"), y=3, x=.5)


cowplot::plot_grid(A, B, labels = c('',''))
```

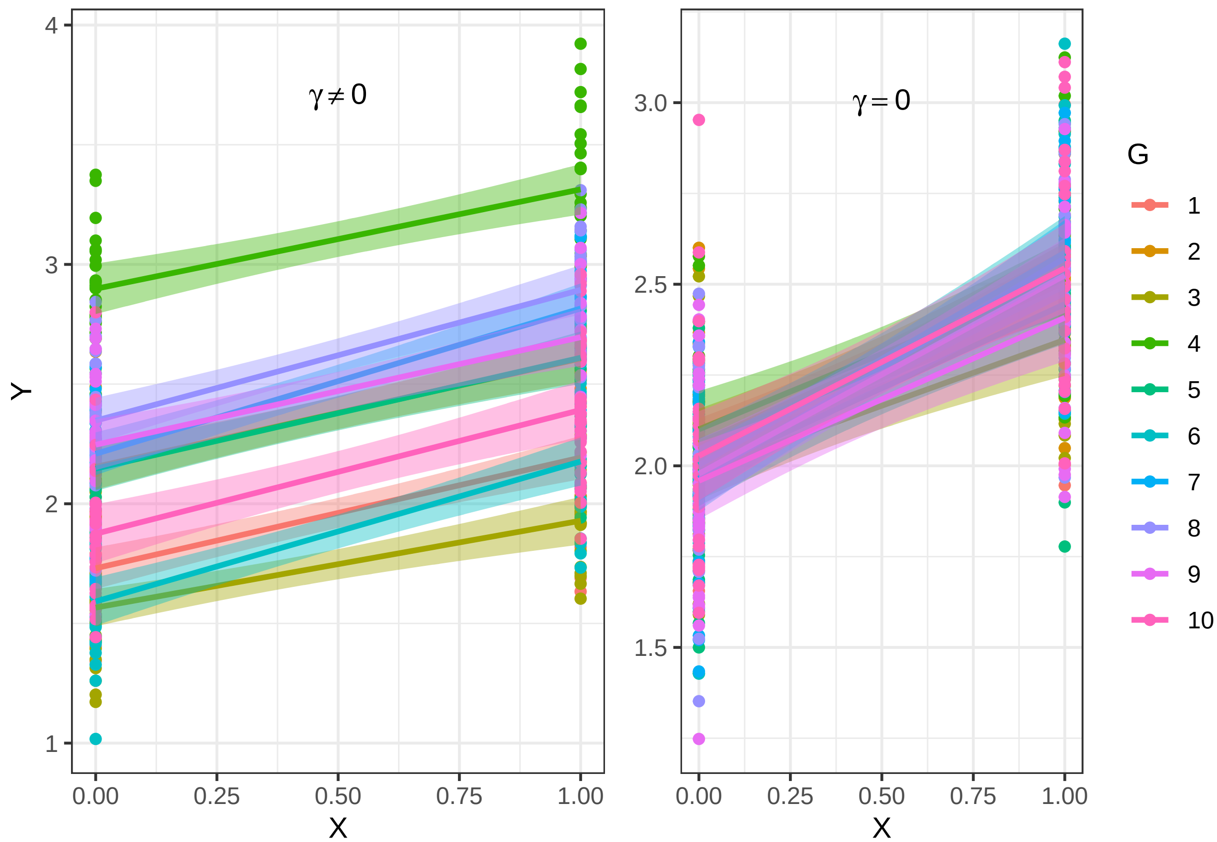


Table 1: Linear regression estimates with `lm()`

	$\hat{\beta}$	$SE(\hat{\beta})$	$\sigma(\hat{\beta})$
$\gamma \neq 0$	0.4996	0.0409	0.03
$\gamma = 0$	0.4925	0.0229	0.0168

Mixed model with lmer function

```
options(warn=-1)
# With Random Effect
lmerModel <- lme4::lmer(H1, data=dfWith)
betaLmer <- lme4::fixef(lmerModel)
SE <- sqrt(diag(as.matrix(vcov(lmerModel))))[-1]
Sigm <- sqrt(stats::var(as.vector(betaLmer %*% t(X))))
A <- ggplot(dfWith, aes(x=X, y=Y, colour=G)) +
    geom_point(size=1.5) +
    geom_line(aes(y=predict(lmerModel), group=G), size=1.3) +
      theme_bw()+theme(legend.position="none")+
  annotate(geom='text', label=TeX("$\\gamma \\neq 0$"), y=3.7, x=.5)

# Without Random Effect
lmerModel.1 <- lme4::lmer(H1, data=dfWithout)
betaLmer.1 <- lme4::fixef(lmerModel.1)
SE.1 <- sqrt(diag(as.matrix(vcov(lmerModel.1))))[-1]
Sigm.1 <- sqrt(stats::var(as.vector(betaLmer.1 %*% t(X))))

B <- ggplot(dfWithout, aes(x=X, y=Y, colour=G)) +
    geom_point(size=1.5) +
    geom_line(aes(y=predict(lmerModel.1), group=G), size=1.3) +
  theme_bw()+theme(axis.title.y = element_blank())+
  annotate(geom='text', label=TeX("$\\gamma=0$"), y=3, x=.5)
cowplot::plot_grid(A, B, labels = c('',''))
```

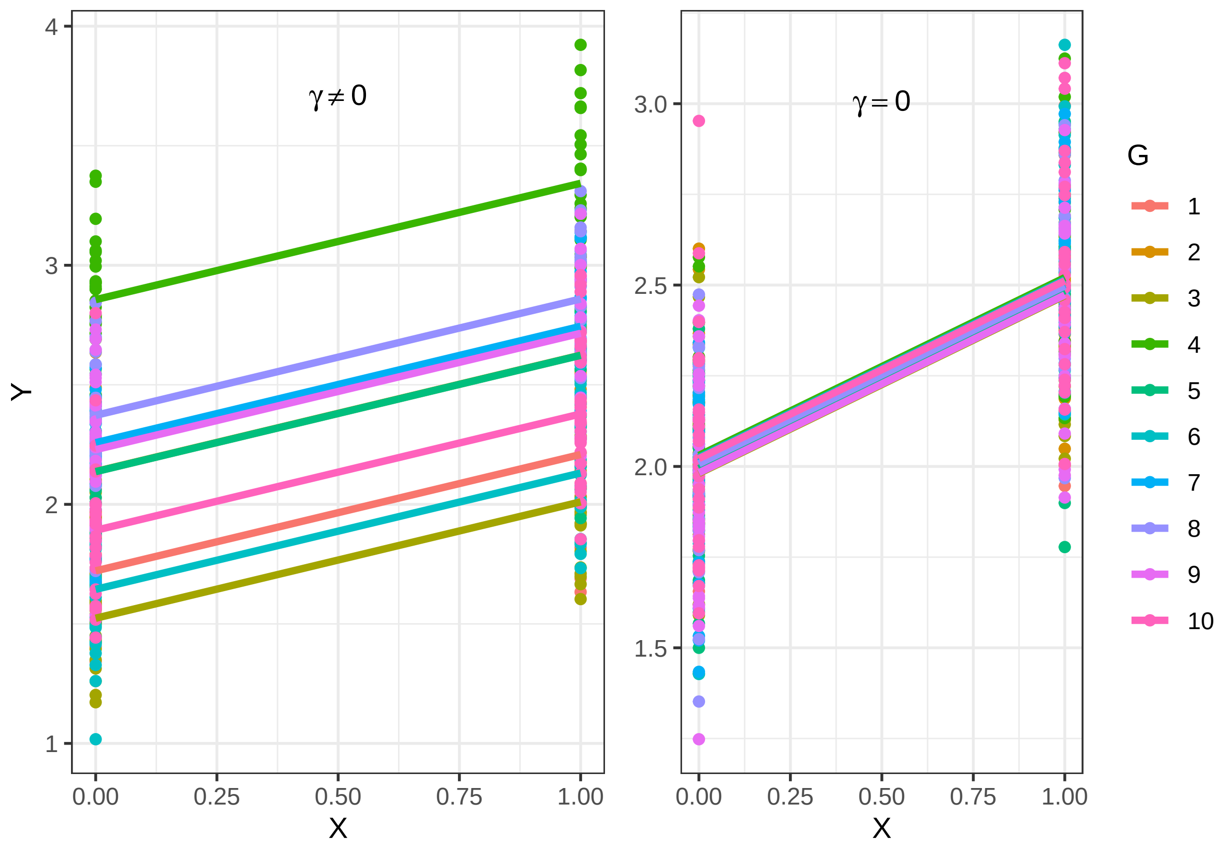


Table 2: Mixed Model estimates with `lmer()`

	$\hat{\beta}$	$SE(\hat{\beta})$	$\sigma(\hat{\beta})$
$\gamma \neq 0$	0.4867	0.0228	0.8356
$\gamma = 0$	0.4902	0.0227	0.8078

OLS

```
OLS <- function(Y, X){
  modelmat <- model.matrix(~.,cbind.data.frame(X=X))
  indexes_X <- which(substring(colnames(modelmat), 1, 1) == "X")
  modX_OLS <- modelmat[, c(1, indexes_X), drop = FALSE]
  Y <- as.numeric(Y)
  betaOLS <- solve(crossprod(modX_OLS))%*%(t(modX_OLS)%*%Y)
  k <- ncol(modX_OLS)
  n <- nrow(modX_OLS)

  residuals <- as.matrix(Y - (betaOLS[1, , drop=FALSE]) - X * betaOLS[indexes_X, ,
                                                                  drop=FALSE])


  RSS <- as.numeric(t(residuals)%*%residuals)
```

```
  Sigma2 <- as.numeric(RSS/(n-k))
  Vb <- Sigma2*solve(t(X)%*% X)
  Sigm <- sqrt(Vb)

  OLSCOV <- 1/(n-k) * as.numeric(t(residuals)%*%residuals) * solve(t(modX_OLS)%*%modX_OLS)
  SE <- sqrt(diag(OLSCOV))[-1]

  return(list('betaOLS'=betaOLS[indexes_X, ,drop=FALSE], 'SE'=SE,'Sigm'=Sigm))
}
options(warn=-1)

# With Random Effect
res <- OLS(dfWith$Y,X)

# Without Random Effect
res.1 <- OLS(dfWithout$Y,X)
```

Table 3: OLS estimates

	$\hat{\beta}$	$SE(\hat{\beta})$	$\sigma(\hat{\beta})$
$\gamma \neq 0$	0.4996	0.0408	0.8356
$\gamma = 0$	0.4925	0.0228	0.8078

Monte Carlo

Simulating a mixed model

The following represents a simulate a random intercepts model obtenaid 500 Monte-Carlo replicates. The method follows the following simulation framework from gaussian distributions with or without the presence of random effects:

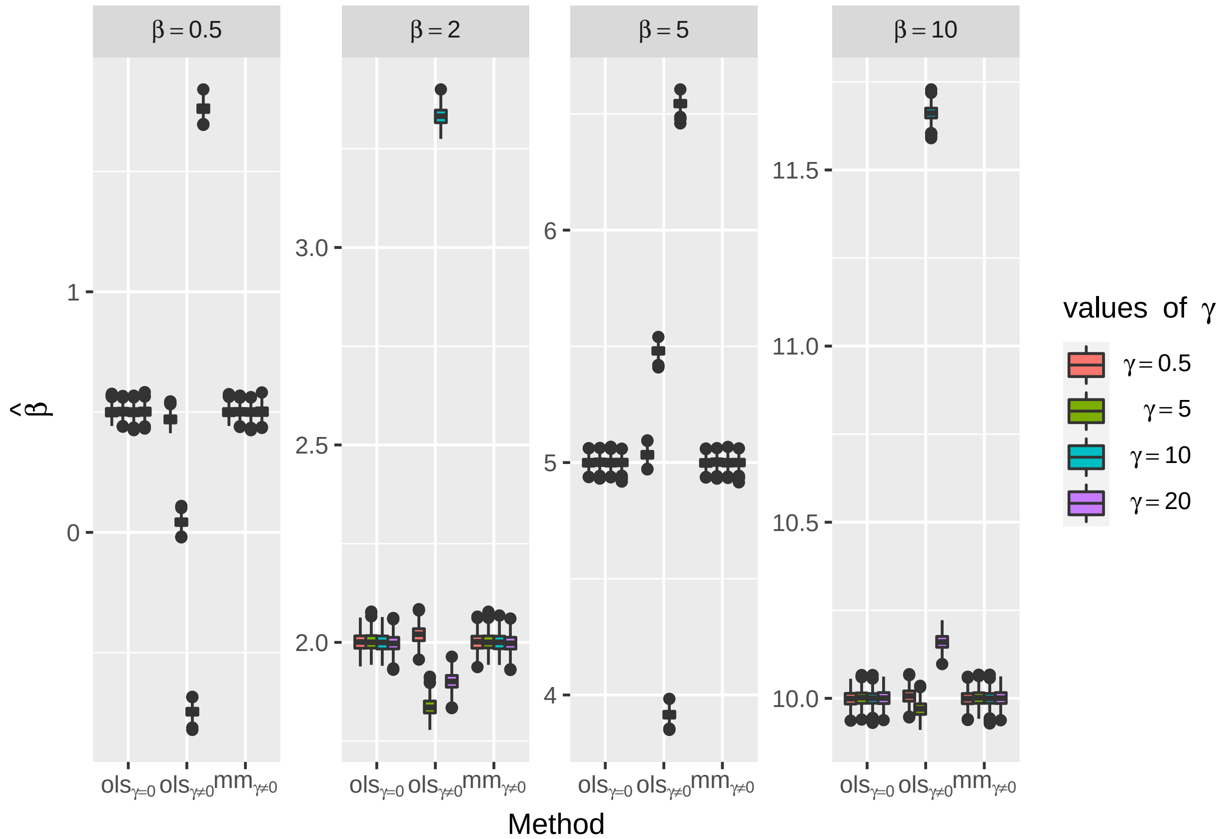
$$Y = \beta_0 + \beta X + \gamma G + \epsilon \tag{1}$$

with $\beta_0 = 2$, $\beta \in \{0.5, 2, 5, 10\}$ the fixed effect of $X \sim \mathcal{B}\binom{n}{0.5}$ with $n = 500$ the number total of samples, $\epsilon \sim \mathcal{N}(0, 0.25)$ the bias associated, the random effect of group $\gamma \sim \mathcal{N}(0, \sigma_\gamma)$ if simulated with random effect with $\sigma_\gamma \in \{0.5, 5, 10, 20\}$) and $\gamma = 0$ if not random effect and the group $G \in \{1, \ldots, K\}$ with $K = 10$.

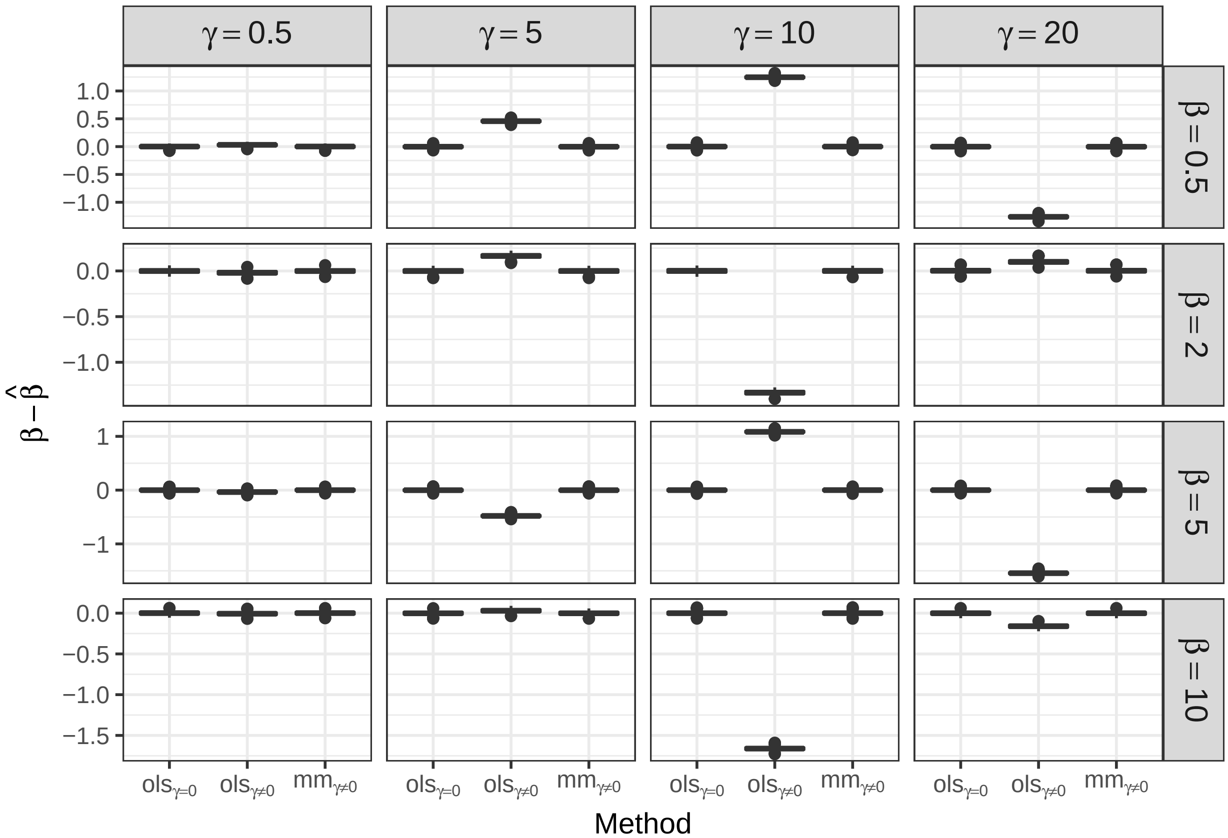
```
simOLS <- function (intercept=2, fixefEffects = c(.5,2,5,10), sds = c(.5,5,10,20),
                    K=10, nK=50, sims=5000){
  resAllOLS.1 <- list()
  resAllOLS.2 <- list()
  resAllMM.1 <- list()
  estimateAll <- matrix(0, length(fixefEffects)*length(sds)*sims, 12)
  n <- K * nK
  G <- factor(rep(1:K, each = nK))
  X <- rbinom(n,size=1,prob = .5)
  k = 1

  for (fixefEffect in fixefEffects) {
```

```
  for (sdUnit in sds) {
      aleaEffect <- rnorm(K, sd = sdUnit)
      resOLS.1 <- matrix(0, sims, 3)
      resOLS.2 <- matrix(0, sims, 3)
      resMM <- matrix(0, sims, 3)


      for(i in 1:sims){
        options(warn=-1)
        bias <- rnorm(n, sd = .25)
        Y_with <- intercept + fixefEffect * X + aleaEffect[G] + bias
        Y_without <- intercept + fixefEffect * X + bias
        modOLS.1 <- OLS(Y_with,X)
        modOLS.2 <- OLS(Y_without,X)

        resOLS.1[i,] <- c(modOLS.1$betaOLS, modOLS.1$SE, modOLS.1$Sigm)
        resOLS.2[i,] <- c(modOLS.2$betaOLS, modOLS.2$SE, modOLS.2$Sigm)

        lmerModel.1 <- try({ lme4::lmer(Y_with ~ X + (1|G), REML = TRUE)}, silent = T)

        beta <- lme4::fixef(lmerModel.1)[-1]
        SE_b <- sqrt(diag(as.matrix(vcov(lmerModel.1))))[-1]
        Sigm <- sqrt(stats::var(as.vector(beta %*% t(X))))
        resMM[i,] <- c(beta, SE_b,Sigm)
        estimateAll[k,] <- c(fixefEffect, modOLS.1$betaOLS,modOLS.2$betaOLS,beta,
                             modOLS.1$SE,modOLS.2$SE,SE_b,
                             modOLS.1$Sigm,modOLS.2$Sigm,Sigm,sdUnit,i)
        k <- k+1
      }

      M <- apply(resOLS.1, 2, mean)
      S <- apply(resOLS.1, 2, sd)
      H_ols.1 <- matrix(c(M[1], S[1], M[2], S[2], M[3], S[3]), ncol = 2, byrow = TRUE)

      M <- apply(resOLS.2, 2, mean)
      S <- apply(resOLS.2, 2, sd)
      H_ols.2 <- matrix(c(M[1], S[1], M[2], S[2], M[3], S[3]), ncol = 2, byrow = TRUE)

      M <- apply(resMM, 2, mean)
      S <- apply(resMM, 2, sd)
      H_mm <- matrix(c(M[1], S[1], M[2], S[2], M[3], S[3]), ncol = 2, byrow = TRUE)

      dimnames(H_ols.1) <- dimnames(H_mm) <-
        dimnames(H_ols.2)  <- list( c('Beta','Standard Error', 'Sigm'), c('mean', 'se'))

      resAllOLS.1[[as.character(paste0(fixefEffect,"_",sdUnit))]] <- H_ols.1
      resAllOLS.2[[as.character(paste0(fixefEffect,"_",sdUnit))]] <- H_ols.2
      resAllMM.1[[as.character(paste0(fixefEffect,"_",sdUnit))]] <- H_mm
  }

}
estimateAll <- as.data.frame(estimateAll)
colnames(estimateAll) = c('fixefEffect','betaOLS.1','betaOLS.2','betaMM',
```

```
                            'seOLS.1','seOLS.2','seMM','sigmOLS.1',
                            'sigmOLS.2','sigmMM','sdUnit','simId')
  return (list('resAllOLS'=resAllOLS.1, 'resAllMM'=resAllMM.1,
               'resAllOLSWithoutAleaEffect'=resAllOLS.2, "estimateAll"=estimateAll))
}

resSim <- simOLS(sims = 500)
dataPLOT <- cbind.data.frame(betaEs = c(resSim$estimateAll[[2]],resSim$estimateAll[[3]],
                                        resSim$estimateAll[[4]]),
                             betas = factor(rep(resSim$estimateAll[[1]],3)),
                             Sigm = factor(rep(resSim$estimateAll[[11]],3)),
                             Method = as.factor(rep(c('OLS.1','OLS.2','MM'),
                                                    each=length(resSim$estimateAll[[1]]))))

dataPLOT$Bias = rep(resSim$estimateAll[[1]],3)-dataPLOT$betaEs

dataPLOT$betas <- gsub(0.5, "beta == 0.5",
           gsub(2, "beta == 2",
           gsub(10, "beta == 10",dataPLOT$betas)))
dataPLOT[dataPLOT$betas==5,'betas']="beta == 5"
dataPLOT$betas = factor(dataPLOT$betas,levels = c("beta == 0.5","beta == 2",
                                                  "beta == 5","beta == 10"))

dataPLOT$Sigm <- gsub(0.5, "gamma == 0.5",
           gsub(10, "gamma == 10",
           gsub(20, "gamma == 20",dataPLOT$Sigm)))
dataPLOT[dataPLOT$Sigm==5,'Sigm']="gamma == 5"
dataPLOT$Sigm = factor(dataPLOT$Sigm,levels = c("gamma == 0.5","gamma == 5",
                                                "gamma == 10","gamma == 20"))

ggplot(dataPLOT, aes(Method,betaEs))+
geom_boxplot(position="dodge",aes(fill=Sigm))+
facet_wrap(~betas, scales = "free_y",labeller = label_parsed,ncol = 4)+
  ylab(TeX("$\\hat{\\beta}$"))+
  scale_x_discrete(labels=c(TeX("$ols_{\\gamma=0}$"),TeX("$ols_{\\gamma\\neq 0}$"),
                            TeX("$mm_{\\gamma\\neq 0}$")))+
  scale_fill_discrete(name=TeX("$values\\ of\\ \\gamma$"),
                         labels=c(TeX("$\\gamma=.5$"),TeX("$\\gamma=5$"),
                            TeX("$\\gamma=10$"),TeX("$\\gamma=20$")))
```




```r
ggplot(data=dataPLOT, aes(x=Method, y=Bias))+
  geom_boxplot()+
  theme_bw()+
  facet_grid(betas~Sigm, scales = "free_y",labeller = label_parsed)+
  theme(strip.text.x = element_text(size=12, face="bold"),
        strip.text.y = element_text(size=12, face="bold",)) +
  ylab(TeX("$\\beta-\\hat{\\beta}$"))+
  scale_x_discrete(labels=c(TeX("$ols_{\\gamma=0}$"),TeX("$ols_{\\gamma\\neq 0}$"),
                            TeX("$mm_{\\gamma\\neq 0}$")))
```



$\gamma = 0.5$
$\gamma = 5$
$\gamma = 10$
$\gamma = 20$
$\beta = 0.5$
$\beta = 2$
$\beta = 5$
$\beta = 10$
$\beta - \hat{\beta}$
1.0
0.5
0.0
−0.5
−1.0
0.0
−0.5
−1.0
1
0
−1
0.0
−0.5
−1.0
−1.5
$ols_{\gamma=0}$
$ols_{\gamma\neq0}$
$mm_{\gamma\neq0}$
Method

Table 4: OLS and Mixed Model estimates with 500 Monte-Carlo replicates

				$\boldsymbol{\Sigma = 0.5}$		$\boldsymbol{\Sigma = 5}$		$\boldsymbol{\Sigma = 10}$		$\boldsymbol{\Sigma = 20}$	
β	Model	γ	Estimate	μ	SE	μ	SE	μ	SE	μ	SE
$\boldsymbol{\beta = 0.5}$	OLS	$\gamma \neq 0$	$\hat{\beta}$	0.469	0.0221	0.0432	0.021	-0.7455	0.0221	1.7612	0.023
			$\boldsymbol{SE}(\hat{\beta})$	0.0511	0.001	0.511	9×10^{-4}	0.9147	0.001	1.6608	0.001
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0362	7×10^{-4}	0.362	7×10^{-4}	0.6481	7×10^{-4}	1.1767	7×10^{-4}
		$\gamma = 0$	$\hat{\beta}$	0.4991	0.0221	0.5018	0.021	0.5003	0.0221	0.5016	0.023
			$\boldsymbol{SE}(\hat{\beta})$	0.0224	7×10^{-4}	0.0223	7×10^{-4}	0.0223	7×10^{-4}	0.0224	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0159	5×10^{-4}	0.0158	5×10^{-4}	0.0158	5×10^{-4}	0.0159	5×10^{-4}
	MM	$\gamma \neq 0$	$\hat{\beta}$	0.499	0.0224	0.5018	0.0213	0.5002	0.0224	0.5017	0.023
			$\boldsymbol{SE}(\hat{\beta})$	0.0225	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	0.2497	0.0112	0.2512	0.0106	0.2503	0.0112	0.2511	0.0115
$\boldsymbol{\beta = 2}$	OLS	$\gamma \neq 0$	$\hat{\beta}$	2.0201	0.0221	1.8366	0.0227	3.332	0.0224	1.9012	0.0228
			$\boldsymbol{SE}(\hat{\beta})$	0.0404	0.001	0.5625	0.001	0.7208	0.0011	0.9151	0.001
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0286	7×10^{-4}	0.3985	7×10^{-4}	0.5107	8×10^{-4}	0.6483	7×10^{-4}
		$\gamma = 0$	$\hat{\beta}$	2.0012	0.0221	2.0014	0.0227	1.9996	0.0224	1.9979	0.0228
			$\boldsymbol{SE}(\hat{\beta})$	0.0224	7×10^{-4}	0.0224	7×10^{-4}	0.0224	7×10^{-4}	0.0223	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0159	5×10^{-4}	0.0159	5×10^{-4}	0.0158	5×10^{-4}	0.0158	5×10^{-4}
	MM	$\gamma \neq 0$	$\hat{\beta}$	2.0014	0.0223	2.0015	0.0229	1.9997	0.0225	1.9979	0.0228
			$\boldsymbol{SE}(\hat{\beta})$	0.0226	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	1.0017	0.0112	1.0017	0.0115	1.0008	0.0113	0.9999	0.0114
$\boldsymbol{\beta = 5}$	OLS	$\gamma \neq 0$	$\hat{\beta}$	5.0345	0.0226	5.4794	0.0217	3.9153	0.0218	6.5439	0.0221
			$\boldsymbol{SE}(\hat{\beta})$	0.0517	0.001	0.576	0.001	1.0475	0.001	0.9771	0.001
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0366	7×10^{-4}	0.4081	7×10^{-4}	0.7422	7×10^{-4}	0.6923	7×10^{-4}
		$\gamma = 0$	$\hat{\beta}$	4.9995	0.0226	5.0011	0.0217	4.9994	0.0218	4.9995	0.0221
			$\boldsymbol{SE}(\hat{\beta})$	0.0224	7×10^{-4}	0.0224	7×10^{-4}	0.0223	7×10^{-4}	0.0224	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0159	5×10^{-4}	0.0159	5×10^{-4}	0.0158	5×10^{-4}	0.0158	5×10^{-4}
	MM	$\gamma \neq 0$	$\hat{\beta}$	4.9998	0.0227	5.0012	0.0219	4.9994	0.022	4.9997	0.0221
			$\boldsymbol{SE}(\hat{\beta})$	0.0226	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	2.5024	0.0114	2.5031	0.0109	2.5022	0.011	2.5023	0.0111
$\boldsymbol{\beta = 10}$	OLS	$\gamma \neq 0$	$\hat{\beta}$	10.006	0.023	9.9696	0.0223	11.661	0.0213	10.1606	0.0219
			$\boldsymbol{SE}(\hat{\beta})$	0.0468	0.001	0.282	0.001	0.9186	0.001	1.5969	0.001
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0332	7×10^{-4}	0.1998	7×10^{-4}	0.6509	7×10^{-4}	1.1314	7×10^{-4}
		$\gamma = 0$	$\hat{\beta}$	9.9987	0.023	10.0011	0.0223	9.9994	0.0213	10.0011	0.0219
			$\boldsymbol{SE}(\hat{\beta})$	0.0224	7×10^{-4}	0.0223	7×10^{-4}	0.0224	7×10^{-4}	0.0224	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	0.0159	5×10^{-4}	0.0158	5×10^{-4}	0.0159	5×10^{-4}	0.0158	5×10^{-4}
	MM	$\gamma \neq 0$	$\hat{\beta}$	9.9986	0.0234	10.001	0.0224	9.9994	0.0215	10.0012	0.0222
			$\boldsymbol{SE}(\hat{\beta})$	0.0225	8×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}	0.0225	7×10^{-4}
			$\boldsymbol{\sigma}(\hat{\beta})$	5.0043	0.0117	5.0055	0.0112	5.0047	0.0108	5.0056	0.0111