# Chapter 6: Multiple Regression Analysis: Further Issues

Susumu Shikano

Last compiled at 18. Juli 2022

### Adding regressors which are uncorrelated with the independent variables of interest

We generate 5000 datasets with n=25 under the GM-assumptions. The number of independent variables is 2. The true regression line has the intercept of 1 and the slope of 5, -2.5. The independent variables are generated with the mean 2, -1, variances 3, 5 and covariance 0.

```
CLM.samples <- data.generation(sample.size=sample.size,
                               n.sim=num.datasets,
                               n.iv=n.iv,
                               x.mu=x.mu,
                               x.Sigma=x.Sigma,
                               para=c(true.intercept,true.slope),
                               err.dist = "normal",
                               err.disp = true.err.var)
```

As stated above, the data above was generated based on two independent variables whose correlation is zero in the population.

We can estimate the following two regression models

$$\hat{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

and

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$

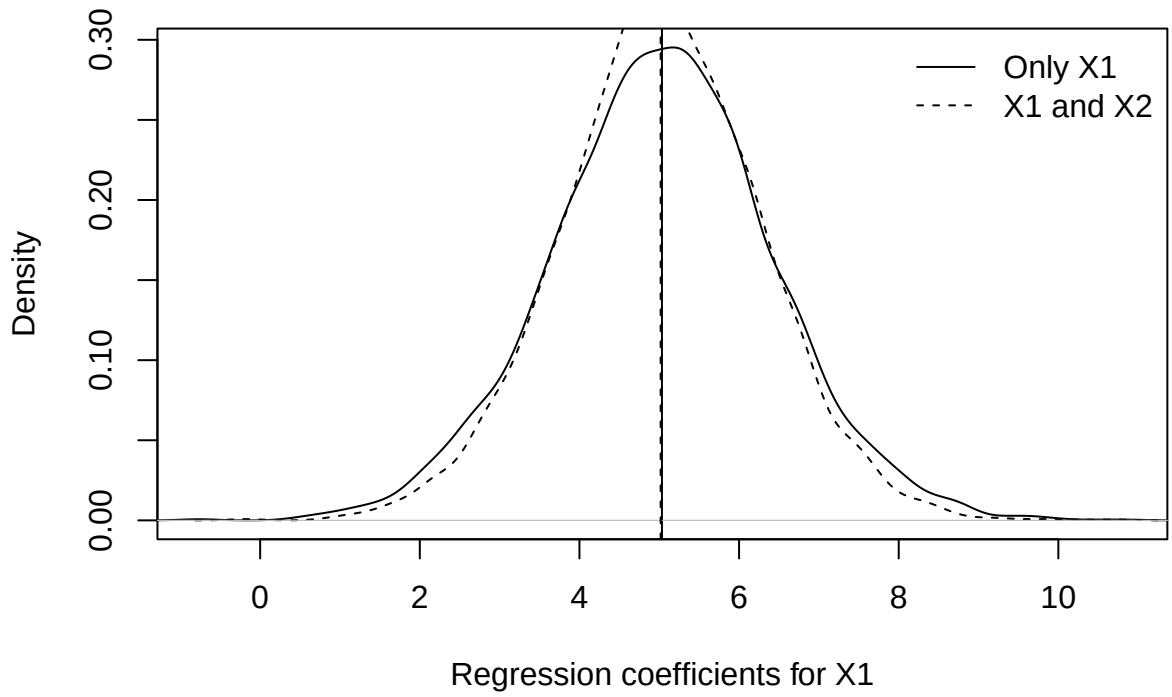
and compare both slope estimates.

```
all.coef <- array(NA,dim=c(num.datasets,4,2))
all.coef.se <- array(NA,dim=c(num.datasets,3,2))

for (i in 1:num.datasets){
    this.data <- CLM.samples$generated.data[[i]]

    lm.out <- lm(y ~ X1  ,data=  this.data)

    all.coef[i,1:2,1] <- coef(lm.out)
    all.coef[i,4,1] <- summary(lm.out)$sigma
    all.coef.se[i,c(1:2),1] <- coef(summary(lm.out))[,2]

    lm.out <- lm(y ~ X1 + X2 ,data=  this.data)

    all.coef[i,1:3,2] <- coef(lm.out)
    all.coef[i,4,2] <- summary(lm.out)$sigma
    all.coef.se[i,,2] <- coef(summary(lm.out))[,2]
```

```
    }
dimnames(all.coef)[[2]] <- c("b0","b1","b2","sigma")
dimnames(all.coef.se)[[2]] <- c("b0","b1","b2")
```

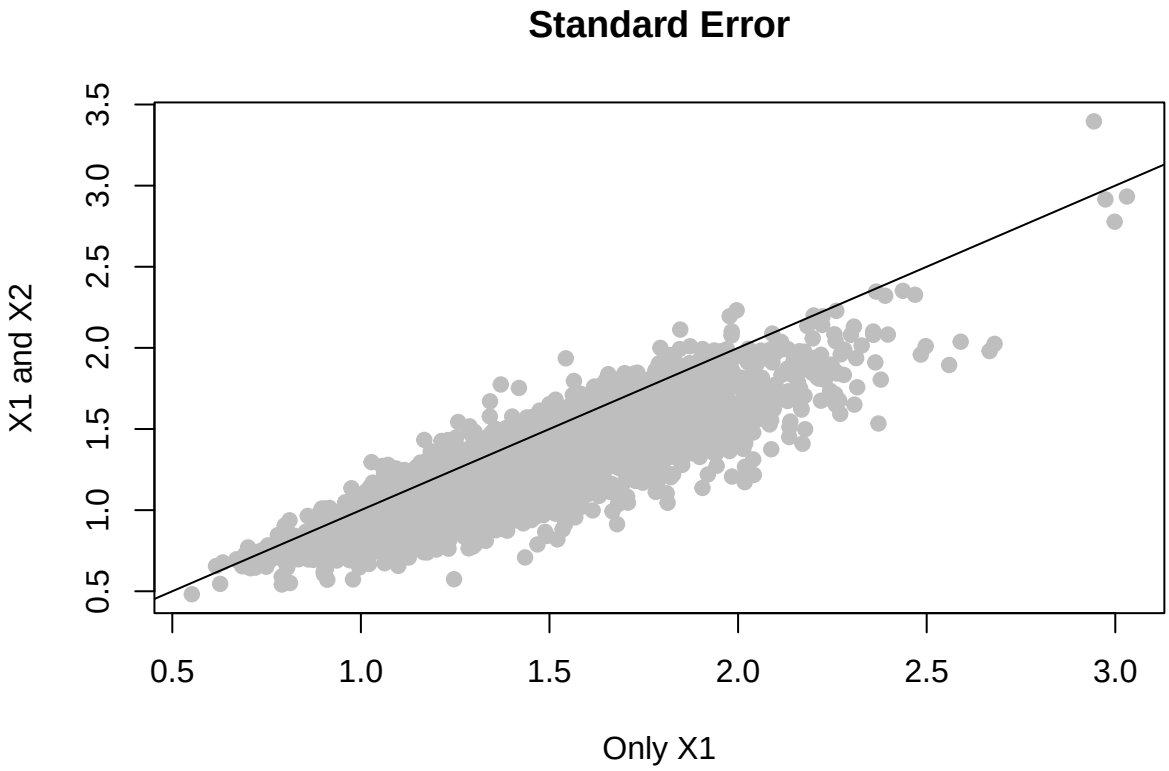
Below you will find the distribution of both estimated regression coefficients:

```
    x.range <- range(c(all.coef[,"b1",]))
    plot(density.out <- density(all.coef[,"b1",1]),
         main="",xlab=paste0("Regression coefficients for X1"),
         xlim=x.range)
    abline(v=mean(all.coef[,"b1",1]))
    par(new=T)
    plot(density(all.coef[,"b1",2]),ann=F,xlab="",ylab="",
         axes=F,
         xlim=x.range,lty=2,
         ylim=c(0,max(density.out$y)))
    abline(v=mean(all.coef[,"b1",2]),lty=2)
    legend("topright",lty=c(1,2),c("Only X1","X1 and X2"),bty="n")
```




Both estimators are unbiased, that is, $\tilde{\beta}$ has no omitted variable bias. But $\hat{\beta}$ has smaller variance since by considering $X_2$ the estimated error variance becomes smaller.

```
plot(all.coef.se[,2,],pch=19,col="grey",xlab="Only X1",ylab="X1 and X2",main="Standard Error")
abline(coef=c(0,1))
```




Comparing standard errors in both models, the second model with both independent variables has smaller standard errors in most cases.