GDP growth and investment

## Part 1: True False Questions

1, Endogeneity of regressors causes that the OLS coeffecients are biased. True. Because the estimated coeffecients are not BLUE anymore

2, Dropping an important variable from the set of regressors always creates an omitted variable bias. False. Omitted variable bias only happens when residuals correlated with explanatory variables. If dropped variable doesn???t affect to other explanatory variables, the residuals into which the impact of the dropped variable entered does not also affect to other variables.

3, A high degree of correlation between the explanatory variables causes that the OLS coefficients are biased. False. Multicollinearity causes to increase of variance of OLS coefficients, but not causes to biasnesss.

4, R^2 = Var(y^)/Var(y) = 7.8/10.3 = 0.75728??? R-squared close to 1. This means that the model explains the data well.

5, Including an irrelevant variables does not lead to biased OLS coefficients. True. It leads to multicollinearity which cause to increase of variance of OLS coefficients.

6, I reject a Null-hypothesis on a single coefficient, if the absolute value of the t-statistics is larger than the critical value. True.

7, An OLS estimator becomes more accurate, the larger the dispersion of the explanatory variables x\_it. True. If the Var(x\_it) becomes larger, the Var(B^) will decrease.

## Part2: Regression preparetion

1. Firstly, We generate the variables that we need for running Cobb-Douglas production function.

# 1 read CobbDouglas.dta

cobbdata <- read.dta13("/Users/Takuma/Google Drive/GitHub/GDP-Growth-and-Investment/CobbDouglas.dta")

# 2 generate explanatory variables

class(cobbdata)

## [1] "data.frame"

head(cobbdata)

## state st\_abb yr hwy water util pc gsp emp unemp  
## 1 ALABAMA AL 1970 7325.80 1655.68 6051.20 35793.80 28418 1010.5 4.7  
## 2 ALABAMA AL 1971 7525.94 1721.02 6254.98 37299.91 29375 1021.9 5.2  
## 3 ALABAMA AL 1972 7765.42 1764.75 6442.23 38670.30 31303 1072.3 4.7  
## 4 ALABAMA AL 1973 7907.66 1742.41 6756.19 40084.01 33430 1135.5 3.9  
## 5 ALABAMA AL 1974 8025.52 1734.85 7002.29 42057.31 33749 1169.8 5.5  
## 6 ALABAMA AL 1975 8158.23 1752.27 7405.76 43971.71 33604 1155.4 7.7

logY <- log(cobbdata$gsp)  
logK1 <- log(cobbdata$hwy + cobbdata$water + cobbdata$util)  
logK2 <- log(cobbdata$pc)  
logL <- log(cobbdata$emp)  
unemp <- cobbdata$unemp

1. Because we have to avoid omitted variable biases. If we omit these variables (private capital, employment rate), the residual may significantry correlated with explanatory variable (public capital). That leads to biasness of estimated coeficients.

## Estimating pooled OLS

We estimate the Cobb-Douglas production function below.

regression <- lm(logY ~ logK1 + logK2 + logL + unemp)  
summary(regression)

##   
## Call:  
## lm(formula = logY ~ logK1 + logK2 + logL + unemp)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -0.23176 -0.06104 -0.00010 0.05085 0.35111   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.643302 0.057587 28.536 < 2e-16 \*\*\*  
## logK1 0.155007 0.017154 9.036 < 2e-16 \*\*\*  
## logK2 0.309190 0.010272 30.100 < 2e-16 \*\*\*  
## logL 0.593935 0.013747 43.203 < 2e-16 \*\*\*  
## unemp -0.006733 0.001416 -4.754 2.36e-06 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 0.0881 on 811 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.9926, Adjusted R-squared: 0.9926   
## F-statistic: 2.717e+04 on 4 and 811 DF, p-value: < 2.2e-16

As the summary shows, the result is following

logY = B0 + B1*logK1 + B2*logK2 + B3\* logL + B4 \* unemp

logY = 1.643302 + 0.155007 \* logK1 + 0.309190 \* logK2 + 0.593925 \* logL + -0.006733 \* unemp

R-Squared = 0.9926 Adjusted R-Squared = 0.9926

1. The model fitted the data well, because the Adjusted R-Squared is close to 100%. According to the regression, the coefficient of logarismic public capital is 0.155007. That means if public capital changes 10%, growth of production changes 1.55%. I would say publi capital has significant effect on growth of production.
2. We formulize the hypothesis like following,

Null Hypothesis: B1-B2 = 0

t-statistics = (B1-B2)/(Var(B1-B2))^(1/2)

Var(B1-B2) = Var(B1) + Var(B2) +2\*covar(B1,B2)

I need covariance of B1 and B2 to calculate this t-statistics. (see <http://www.uv.es/uriel/4%20Hypothesis%20testing%20in%20the%20multiple%20regression%20model.pdf>)

vcov(regression)

## (Intercept) logK1 logK2 logL  
## (Intercept) 3.316291e-03 -5.782633e-04 -1.942134e-04 6.205055e-04  
## logK1 -5.782633e-04 2.942518e-04 -8.774564e-05 -1.913027e-04  
## logK2 -1.942134e-04 -8.774564e-05 1.055137e-04 -9.094008e-06  
## logL 6.205055e-04 -1.913027e-04 -9.094008e-06 1.889927e-04  
## unemp 1.657240e-06 -1.258420e-06 -1.088128e-06 1.256500e-06  
## unemp  
## (Intercept) 1.657240e-06  
## logK1 -1.258420e-06  
## logK2 -1.088128e-06  
## logL 1.256500e-06  
## unemp 2.006121e-06

As shown in the covariance matrices, the covar(B1,B2) eqauls to -8.774564e-05.

B1 <- 0.155007  
B2 <- 0.309190  
varB1 <- 0.017154   
varB2 <- 0.010272  
vcovB1B2 <- (-1)\*8.774564\*10^(-5)  
  
standarderror <- varB1 + varB2 +2\*vcovB1B2  
  
tstatistics <- (B1-B2)/standarderror\*(1/2)  
tstatistics

## [1] -2.828993

range(tstatistics)

## [1] -2.828993 -2.828993