Problem Set 1

January 27, 2020

1 General Instructions

Complete the following assignments in a sweave file that shows your code, output, and discussion. Hand in the compiled (pdf) version. You can use this file to get you started. You can work with others, but please hand in your own version. ***Please make sure your random number seed is set to "37" in her first R chunk.*** *Please report any glitches as soon as you discover them - thanks!*

Question 1: Omitting regressors under independence and dependence

In this question you will examine the effect of omitting an explanatory variable from your regression model when it is independent of the included variable, and when it is correlated with the included variable.

Part A: Independence, Full Model

1. *Generate two independent, normally distributed regressors (= explanatory variables), one with mean 2 and std 1, the other with mean 3 and std 1. Set the sample size to 1000 observations in each case. Call these variables x1 and x2*
2. *Create a scatterplot to examine the relationship between x1 and x2.*
3. *Create a table of sample statistics, including the correlation coefficient*
4. *Draw a normal(0,1) error term, define a vector of true parameters for the constant, x1, and x2 of* $[1, 1, -1]$*, and build your dependent variable.*
5. *Run an OLS regression on the full model. Show the output table. Call this model "Independent, full"*

Part B: Independence, Omitted

1. *Next, drop the last column in X (your x2). Update your "k" value accordingly.*
2. *Re-run the regression and capture the output. Call this model "Independent, Omit".*
3. *Comment on the estimated coefficient for x1 (with x2 omitted). Therefore, what can you conclude regarding the effects of an omitted variable that is independent from all included variables on the remaining coefficients?*

Part C: Correlation, full model
Continue with you original sweave file - **do NOT re-set the random number seed!**

1. *Generate two correlated regressors (= explanatory variables), one with mean 2 and std 1, the other with mean 3 and std 1, and with covariance (correlation in this case) of 0.8. Set the sample size to 1000 as before. Use the "mvrnorm" function in the MASS package to obtain the correlated draws (some help with this is given below)*
2. *Generate a scatter plot and a table with sample statistics, including correlation*
3. ***Use the same betas and error draws from before** and compute a new y variable. Run the full model. Call it "Correlated, full". Are there any noteworthy changes compared to the original model ("Independent, full")?*
4. *Omit x2, and estimate the model on the full sample. Call this model "Correlated, Omit"*
5. *Comment on the estimated coefficient for x1 for each partial regression (with x2 omitted). Therefore, what can you conclude regarding the effects of an omitted variable that is correlated with some included variables on the remaining coefficients?*

```
R> m<-c(2,3)
R> V<-matrix(c(1,0.8,0.8,1),nrow=2)
R> X<- mvrnorm(n=n,m,V)
R> x1<-X[,1]
R> x2<-X[,2]
```

Question 2: Omitting a variable in the wage regression

Continue with you original sweave file - **do NOT re-set the random number seed!**

Consider our wage regression from `mod1_2b`*.*

1. *Load in the data and specify your dependent variable and your regression matrix. As before, drop "age".*
2. *Capture the sample correlation across regressors (without the constant term). Show the resulting correlation matrix in your output.*
3. *Run the full regression model and capture your output in a table.*
4. *Re-run the model without "experience" (and keep "age" out as well). How do the results change? What do your findings suggest regarding the correlation of "experience" with the remaining regressors? Is the correlation strong enough to induce noticeable omitted variable bias?*

Question 3: Orthogonality and Projection

Consider the "residual maker matrix" $\mathbf{M}$ and the projection matrix $\mathbf{P}$. Show formally that the following hold (please type all Math in LaTeX):

1. $\mathbf{MX} = \mathbf{0}$ (Provide intuition).

2. $\mathbf{PX} = \mathbf{X}$ (Provide intuition).

3. $\mathbf{y} = \mathbf{Py} + \mathbf{M} * \mathbf{y}$ (Provide intuition)

4. $\mathbf{PM} = \mathbf{0}$

5. $\mathbf{e}'\mathbf{e} = \mathbf{e}'\mathbf{y}$
6. $\mathbf{y}'\mathbf{y} = \hat{\mathbf{y}}'\hat{\mathbf{y}} + \mathbf{e}'\mathbf{e}$

```
R> proc.time()-tic

   user  system elapsed
   0.05    0.00    0.05
```