San Francisco State University **Michael Bar**
ECON 312 **Fall 2018**

Problem set 4

Due Thursday, October 25, in class

Names of collaborators (type below):

Name 1 MINJIN PARK

Name 2 NISHANLANG KHONGLAH

Assignment Rules

1. Homework assignments must be typed. For instruction on how to type equations and math objects please see notes "Typing Math in MS Word".
2. Homework assignments must be prepared within this template. Save this file on your computer and type your answers following each question. Do not delete the questions.
3. Your assignments must be stapled.
4. No attachments are allowed. This means that all your work must be done within this word document and attaching graphs, questions or other material is prohibited.
5. Homework assignments must be submitted at the end of the lecture, in class, on the listed dates.
6. Late homework assignments will not be accepted under any circumstances, but the lowest homework score will be dropped.
7. The first homework assignment cannot be dropped.
8. You are encouraged to work on this homework assignment in groups of up to 3 people, and submit one assignment with up to 3 names typed on this page. Sharing the electronic version of your assignment with other teams is absolutely prohibited.
9. All the graphs should be fully labeled, i.e. with a title, labeled axis and labeled curves.
10. In all the questions that involve calculations, you are required to show all your work. That is, you need to write the steps that you made in order to get to the solution.
11. This page must be part of the submitted homework.

Nonlinearities in variables

1. (10 points). Consider the following model:

$$\frac{100}{100-Y} = \beta_1 + \frac{\beta_2}{X} + u$$

   a. Show that this model is **linear in parameters** (β_1, β_2).

   This is non-linear model in variables (Y and X) but still it is linear model in parameters(β_1, β_2) because the right-hand side is a weighted average of the βs.

   b. Demonstrate how you can use OLS to estimate the unknown parameters of the model.

   In order to estimate the unknown parameters with OLS, we need to transform the non-linear variables as follows:

$$Y_{new} = \frac{100}{100-Y},\ Z_2 = \frac{1}{X}$$

   And estimate the linear model:

$$Y_{new} = \beta_1 + \beta_2\, Z_2 + u$$

   The fitted model is given by:

$$\widehat{Y_\iota} = \ b_1 + b_2 Z_2$$

2. (10 points). Consider the following model:

$$\frac{1}{Y^2} = \beta_1 + \frac{\beta_2}{X_2} + \beta_3\sqrt{X_3} + \beta_4 \log(X_4) + u$$

   a. Show that this model is **linear in parameters** $(\beta_1, \beta_2, \beta_3, \beta_4)$.

   This is non-linear model in variables (Y and Xs) but it is still linear model in parameters$(\beta_1, \beta_2, \beta_3, \beta_4)$ because the right-hand side is a weighted average of βs.

   b. Demonstrate how you can use OLS to estimate the unknown parameters of the model.

   In order to estimate the unknown parameters with OLS, we need to transform the non-linear variables as follows:

$$Y_{new} = \frac{1}{Y^2},\ Z_2 = \frac{1}{X^2},\ Z_3 = \sqrt{X_3},\ Z_4 = \log\,(X_4)$$

   And estimate the linear model:

$$Y_{new} = \beta_1 + \beta_2\, Z_2 + \beta_3\, Z_3 + \beta_4\, Z_4 + u$$

   The fitted model is given by:

$$\widehat{Y_i} = b_1 + b_2 Z_2 + b_3 Z_3 + b_4 Z_4$$

Nonlinearities in parameters

3. (10 points). Consider the following model:

$$Y = \beta_1 X_2^{\beta_2} X_3^{\beta_3} X_4^{\beta_4} e^u$$

   a. Show how you can transform this model into a linear model in parameters $(\beta_2, \beta_3, \beta_4)$.

      Taking log on both sides:

$$\ln(Y) = \ln(\beta_1) + \beta_2 \ln(X) + \beta_3 \ln(X_3) + \beta_4 \ln(X_4) + u$$

   b. Demonstrate how you can use OLS to estimate the unknown parameters of the model.

      In order to estimate the unknown parameters with OLS, we need to transform the non-linear variables as follows:

$$Y_{new} = \ln(Y)\,,\ Z_2 = \ln(X_2)\,,\ Z_3 = \ln(X_3)\,,\ Z_4 = \ln(X_4)$$

      And estimate the linear model:

$$Y_{new} = \ln\,(\beta_1) + \beta_2\, Z_2 + \beta_3\, Z_3 + \beta_4\, Z_4 + u$$

      The fitted model is given by:

$$\hat{Y} = b_1 + b_2 Z_2 + b_3 Z_3 + b_4 Z_4$$

4. (10 points). Consider the following model:

$$Y = \beta_1 e^{\beta_2 X_2} X_3^{\beta_3} e^u$$

   a. Show how you can transform this model into a linear model in parameters (β_2, β_3).

      Taking log on both sides:

$$\ln(Y) = \ln(\beta_1) + \beta_2\, X_2 \ln(e) + \beta_3 \ln(X_3) + u \qquad (\ln(e) = 1)$$

$$\ln(Y) = \ln(\beta_1) + \beta_2 X_2 + \beta_3 \ln(X_3) + u$$

   b. Demonstrate how you can use OLS to estimate the unknown parameters of the model.
      In order to estimate the unknown parameters with OLS, we need to transform the non-linear variables as follows:

$$Y_{new} = \ln(Y)\,,\ Z_2 = X_2,\ Z_3 = \ln(X_3),$$

And estimate the linear model:

$$Y_{new} = \ln(\beta_1) + \beta_2 Z_2 + \beta_3 Z_3 + u$$

The fitted model is given by:

$$\hat{Y} = b_1 + b_2 Z_2 + b_3 Z_3$$

Elasticity

5. (5 points). Consider the following model (**log-log**):

$$\log(Y) = \beta_1 + \beta_2 \log(X) + u$$

Prove that the elasticity of Y with respect to X is equal to β_2.
We see that the slope coefficient

$$\beta_2 = \frac{\partial \ln(Y)}{\partial \ln(X)} = \frac{\frac{1}{Y}\partial Y}{\frac{1}{X}\partial X} = \frac{\partial Y}{\partial X}\frac{X}{Y}$$

This expression is the familiar point elasticity formula for the elasticity of Y with respect to X.
Recall that the definition of elasticity is percentage change in Y divided by the percentage change in X:

$$\eta_{Y,X} = \frac{\%\Delta Y}{\%\Delta X},$$

which is the percentage change in Y resulting from a 1% increase in X. Recall from rules of derivatives that

$\frac{\partial Y}{Y} = \frac{\Delta Y}{Y}$ is rate of change in Y.

6. (5 points). Consider the following model (**lin-log**):

$$Y = \beta_1 + \beta_2 \log(X) + u$$

Prove that the elasticity of Y with respect to X is equal to $\frac{\beta_2}{Y}$.

$$Y = \beta_1 + \beta_2 \log(X) + u$$

$$\beta_2 = \frac{\partial Y}{\partial \log(X)} = \frac{\partial Y}{\frac{1}{X}\partial X} = \frac{\Delta Y}{\frac{\Delta X}{X}}$$

$$\beta_2 = \frac{\partial Y}{\partial X} X$$

$$\eta_{Y,X} = \frac{\beta_2}{Y}$$

7. (5 points). Consider the following model (**log-lin**):

$$\log(Y) = \beta_1 + \beta_2 X + u$$

Prove that the elasticity of Y with respect to X is equal to $\beta_2 X$.

$$\log(Y) = \beta_1 + \beta_2 X + u$$

$$\beta_2 = \frac{\partial \log(Y)}{\partial x} = \frac{\frac{1}{Y}\partial Y}{\partial X} = \frac{\frac{\Delta Y}{Y}}{\Delta X} = \frac{\partial Y}{\partial X}\frac{1}{Y}$$

$$\beta_2 = \frac{\partial Y}{\partial X}\frac{1}{Y}$$

$$\eta_{Y,X} = \beta_2 X$$

8. (5 points). Consider the following model (**linear**):

$$Y = \beta_1 + \beta_2 X + u$$

Prove that the elasticity of Y with respect to X is equal to $\beta_2 \frac{X}{Y}$.

$$Y = \beta_1 + \beta_2 X + u$$

$$\beta_2 = \frac{\partial Y}{\partial X}$$

$$\eta_{Y,X} = \beta_2 \frac{X}{Y}$$

9. (5 points). Consider the following model (**reciprocal**):

$$Y = \beta_1 + \beta_2 \frac{1}{X} + u$$

Prove that the elasticity of Y with respect to X is equal to $-\frac{\beta_2}{XY}$.

$$Y = \beta_1 + \beta_2 \frac{1}{X} + u$$

Differentiating with respect to X we get

$$\frac{\partial Y}{\partial X} = -\frac{\beta_2}{X^2}$$

$$\frac{\partial Y}{\partial X} \cdot X = -\frac{\beta_2}{X}$$

$$\eta_{Y,X} = -\frac{\beta_2}{XY}$$

Applications

Create an R script, which performs all the analysis in for questions 10 and 11. You can name the script `HW4.R`. You can either print out the script and attach it as a separate page at the end of this assignment, or copy and paste its content at the end of your assignment. Make sure to add comments explaining which question you are solving, and every command in your script.

10. (35 points). For this exercise use the `wage21` data posted on the course website. The key variables for this exercise are:
`WEIGHT85` – person's weight, in pounds in year 1985

`HEIGHT` – person’s height, in inches.

a. Regress `WEIGHT85` on `HEIGHT` and present the R estimation command and output.

   The R command is given below.

   model1 <- lm(WEIGHT85 ~ HEIGHT, data = wage) #OLS estimation
   b <- coef(model1) #Storing the OLS coefficients in vector b
   b #Displaying the estimated coefficients

   The output is given below.
   (Intercept)   HEIGHT
   -208.086159   5.395399

b. Interpret the estimated regression coefficients.

   Height = 5.40 means that in 1985, for every 1 inch increase in a person’s height, the person’s weight increased by 5.40 pounds.

   The intercept value of -208.09 is the predicted weight of a person in 1985 with 0 inches in height, which does not make any real-world sense.

c. Regress the log of `WEIGHT85` on the log of `HEIGHT` and present the R command and output.

One way would be to create new variables first, and then use them in OLS:

```
wage$LWEIGHT85 <- log(wage$WEIGHT85)
wage$LHEIGHT <- log(wage$HEIGHT)
model2 <- lm(LWEIGHT85 ~ LHEIGHT, data=wage)
```

Alternatively, the next command creates transformed variables inside the `lm()` function.

```
model2 <-lm(log(WEIGHT85) ~ log(HEIGHT), data=wage)
```

   The R command is given below.

   #Creating new variables for the log of WEIGHT85 and the log of HEIGHT
   wage$LWEIGHT85 <- log(wage$WEIGHT85)
   wage$LHEIGHT <- log(wage$HEIGHT)
   model2 <- lm(LWEIGHT85 ~ LHEIGHT, data=wage)#OLS Estimation

   #Alternative to creating new variables for log of WEIGHT85 and the log of HEIGHT
   #model2 <-lm(log(WEIGHT85) ~ log(HEIGHT), data=wage)

   b2<- coef(model2) #Storing the OLS coefficients in vector b
   b2 #Displaying the estimated coefficients

The output is given below.

(Intercept) log(HEIGHT)
 -4.788605   2.331665

d. Interpret the estimated slope coefficient.

   The slope coefficient log(HEIGHT) = 2.33 means that in 1985, for every 1 inch increase in a person's height the weight of the person increased by 2.33 pounds.

e. Interpret the reported p-value for the slope coefficient.

   The reported p-value gives the minimal significance level at which a null hypothesis can be reject. The p-value < 2.2e-16 is a very small value, and so we can reject a null hypothesis at a significance level of p-value or higher, and conclude that at least one regressor is relevant in predicting the weight of a person in 1985.

f. Report the 95% confidence interval for the slope coefficient , and (based on the CI) give your conclusion about the test:

$$H_0: \beta_2 = 2$$
$$H_1: \beta_2 \neq 2$$

   As always, you are required to report the R command and output.

   The R command is given below.

   #Obtaining confidence intervals
   confint(model1,level=0.95)

   The reported 95% confidence interval is given below.

	2.5 %	97.5 %
(Intercept)	-241.534839	-174.637479
HEIGHT	4.900164	5.890634

   This means that any null hypothesis $H_0: \beta_2 \in [4.900164,\ 5.890634]$ will not be rejected at α = 5% confidence level, while any null hypothesis $H_0: \beta_2 \notin [4.900164,\ 5.890634]$ will be rejected in favor of the two-sided alternative.

g. Calculate the test statistic for the hypothesis test in the last section, and explain why it is different from the one reported in the last regression output.

   The test statistic for the hypothesis is 21.4. The test statistic in this case measures the distance between the estimated value of the height of an individual in 1985 and the hypothesized value of the height of an individual in 1985 in units of standard errors.

11. (20 points). This question illustrates how to summarize the important regression output of several models in a single, publication-quality table. First, install the `stargazer` package in R:
`install.packages("stargazer")`
You only need to do this once. Use the same data set as in the previous question. The key variables that you need to know for this assignment are:
`EARNINGS` – hourly earnings, in $ per hour.
`S` – schooling, in years,
`EXP` – years of experience,
`SM` – schooling years of mother,
`SF` – schooling years of father.

    a. Estimate the following models,
    [model1]: $\log(EARNINGS) = \beta_1 + \beta_2 S + u$
    [model2]: $\log(EARNINGS) = \beta_1 + \beta_2 S + \beta_3 EXP + u$
    [model3]: $\log(EARNINGS) = \beta_1 + \beta_2 S + \beta_3 EXP + \beta_4 SM + \beta_5 SF + u$
    and present a summary table using `stargazer` package. The R commands:

```
model1 <- lm(log(EARNINGS) ~ S, data = wage)
model2 <- lm(log(EARNINGS) ~ S + EXP, data = wage)
model3 <- lm(log(EARNINGS) ~ S + EXP + SM + SF, data = wage)

library(stargazer)
stargazer(model1, model2, model3, type="text",
          title="Earnings Models in R",
          dep.var.labels="log(EARNINGS)",
          out="models_EARNINGS.htm", digits=4)
```

The last option `out="models_EARNINGS.htm"` saves the table in html document format, which you can copy and paste into MS Word. Make sure that you set the working directory to be the folder containing table: `setwd("C:/folder")`.

Earnings Models in R

	Dependent variable:		
	log(EARNINGS)		
	(1)	(2)	(3)
S	0.1121***	0.1274***	0.1140***
	(0.0094)	(0.0092)	(0.0104)
EXP		0.0381***	0.0373***
		(0.0051)	(0.0051)
SM			0.0019
			(0.0098)
SF			0.0187**
			(0.0081)
Constant	1.2657***	0.4122**	0.3654**
	(0.1307)	(0.1694)	(0.1743)
Observations	540	540	540
R^2	0.2086	0.2824	0.2930
Adjusted R^2	0.2071	0.2798	0.2877
Residual Std. Error	0.5238 (df = 538)	0.4992 (df = 537)	0.4965 (df = 535)
F Statistic	141.8033*** (df = 1; 538)	105.6870*** (df = 2; 537)	55.4173*** (df = 4; 535)
Note:			*p<0.1; **p<0.05; ***p<0.01

b. Interpret the slope coefficients in model3, and explain the meaning of the stars next to the estimated coefficients.

S = 0.114 means that for every additional year of schooling, an individual's hourly earnings are predicted to rise by $0.11/hour. The *** next to the estimated coefficient S means that we can reject $H_0: \beta_2 = 0$ against the alternative $H_0: \beta_2 \neq 0$ at 1% significance level.

EXP = 0.04 means that for every additional year of experience, an individual's hourly earnings are predicted to rise by $0.04/hour. The *** next to the estimated coefficient EXP means that we can reject $H_0: \beta_3 = 0$ against the alternative $H_0: \beta_3 \neq 0$ at 1% significance level.

SM = 0.002 means that for every additional year that an individual's mother spends in school, the individual's hourly earnings are predicted to rise by $0.002/hour.

SF = 0.02 means that for every additional year that an individual's father spends in school, the individual's hourly earnings are predicted to rise by \$0.02/hour. The ** next to the estimated coefficient SF means that we can reject $H_0: \beta_5 = 0$ against the alternative $H_0: \beta_5 \neq 0$ at 5% significance level.

For the next question you are required to use Stata. You must attach the do-file (similar to R-script) – a program that performs the analysis performed in this question.

12. (10 points). Using Stata, repeat the estimation of the 3 models in the last question, and create a publication-quality table using the Stata `estout` package. To install the estout package on your computer (you need to do this only once on every computer), type in Stata's command window: `ssc install estout, replace`. Open a do-file (the Stata analog of R-script) by clicking on the  icon. Save the do-file as HW4.do. Comments in Stata do-file start with // or with *. Start your do-file with the usual description and listing the names of your collaborators. To clear the workspace, use:

```
clear all
set more off
```

To read the data into Stata, the command is

```
use http://online.sfsu.edu/mbar/ECON312_files/wage21.dta
```

The dependent variable in all the models is ln(EARNINGS). To create this variable, the command is

```
generate LEARNINGS = log(EARNINGS)
```

To estimate the 3 models, and to store them, the commands are

```
regress LEARNINGS S
estimates store model1

regress LEARNINGS S EXP
estimates store model2

regress LEARNINGS S EXP SM SF
estimates store model3
```

To present the 3 models in a publication-quality table, we use the `estout` package (make sure you installed it, otherwise the following commands will not work). The commands are:

```
esttab model1 model2 model3 ///
     using models_EARNINGS.rtf, ///
     title("Earnings Models in Stata") ///
     se stat(N r2 r2_a F p) replace
```

The `///` at the end of a line break the long command into several lines. The main command, `esttab`, generates the table with the 3 models.

The `using models_EARNINGS.rtf` saves the table in rich text format file, which open with Microsoft word. The option `se` option puts standard errors for each estimate in brackets, instead of the t-statistic, which is the default. The option `stat(N r2 r2_a F p)` adds to the table the number of obesrvations used in every model (N), the R^2, the adjusted R^2, the F-statistic, and the p-value for the F-test of overall fit. All you need to do is open the file `models_EARNINGS.rtf` and copy the table into your assignment.

Earnings Models in STATA

	Dependent variable:		
	log(EARNINGS)		
	(1)	(2)	(3)
S	0.1121***	0.1274***	0.1140***
	(0.0094)	(0.0092)	(0.0104)
EXP		0.0381***	0.0373***
		(0.0051)	(0.0051)
SM			0.0019
			(0.0098)
SF			0.0187**
			(0.0081)
Constant	1.2657***	0.4122**	0.3654**
	(0.1307)	(0.1694)	(0.1743)
Observations	540	540	540
R^2	0.2086	0.2824	0.2930
Adjusted R^2	0.2071	0.2798	0.2877
Residual Std. Error	0.5238 (df = 538)	0.4992 (df = 537)	0.4965 (df = 535)
F Statistic	141.8033*** (df = 1; 538)	105.6870*** (df = 2; 537)	55.4173*** (df = 4; 535)
Note:			*p<0.1; **p<0.05; ***p<0.01