ECO 5350-701 T. Fomby
Intro. Econometrics Spring 2021

EXERCISE 7
KEY

Purpose: We are going to be focusing on (1) **interpreting the coefficients in level-level, log-level, level-log, and log-log regression models**, (2) **conducting a Chow Test of the difference in the regression functions of two groups**, (3) **the analysis of a regression equation with a quadratic part of the model**, and (4) **interpreting the coefficients of a standardized regression**. This exercise is to be handed in on **Tuesday, March 9 at 5:00 pm CT on Canvas**. The **Lecture Notes 9.pdf** should contain all the information your need to complete this exercise.

When I refer to reporting a regression model in **standard form** I mean something like:

$$\widehat{lsalary} = \underset{(0.257)}{4.504} + \underset{(0.039)}{0.163}lsales + \underset{(0.050)}{0.109}lmktval + \underset{(0.0053)}{0.0117}ceoten$$

where the standard errors of the coefficient estimates are placed in parentheses below the coefficient estimates.

Use the **Hitters data set** that we have analyzed in class that is part of the **ISLR library in R**. **Do all of your work in this exercise using R and RStudio.**

(a) Using **Salary** as the dependent variable and **CHits** as the sole independent variable, report the following regressions in standard form and for each model interpret the coefficient on the sole independent variable, CHits.

(i) level-level regression

Answer: $\boldsymbol{\widehat{Salary} = \underset{(34.91386)}{260.0385} + \underset{(0.03601)}{0.38202}CHits}$

For each additional hit, the player's salary increases, on average, by 0.382 units. The units of measure for the Salary variable is thousands of dollars. Therefore, the increase is 0.382*$1,000 = $382.

(ii) log-level regression

Answer: $\boldsymbol{\widehat{lSalary} = \underset{(0.0646)}{5.313} + \underset{(6.663e\text{-}05)}{0.00085046}CHits}$

For each additional hit, the player's salary increases, on average, by (100*0.0008504) = 0.08%. That is 8 one-hundreds of one percent.

(iii) level-log regression

Answer: $\widehat{Salary} = -839.739 + 224.59 lCHits$
(132.55) (21.35)

For each 1% increase in hits, the player's salary increases, on average, by (224.59/100) = 2.2459 units or $2,245.90.

(iv) log-log regression

Answer: $\widehat{lSalary} = 2.35569 + 0.58452 lCHits$
(0.21623) (0.03483)

For each 1% increase in hits, the player's salary increases, on average, by 0.58452%.

(b) Here we are going to compare the level-level salary equations of the National League versus the American League using **the Chow Test** applied to the Additive/Multiplicative Dummy variable model. The sole independent variable we are going to be using is CHits while the Dummy variable is the variable "League" in the Hitters data set. Report your estimated Additive/Multiplicative Dummy variable model in standard form. Separately, write out the fitted equation of the National League salaries and the fitted equation of the American League salaries (no standard errors needed). Using the F-statistic, test the null hypothesis that the regression model of the National League is the same as the regression model of the American League. Report your calculated F-statistic and its p-value. In EXCEL you can use the "F.DIST.RT()" function to get the p-value of your calculated F statistic. What is your conclusion?

Answer: The fitted additive/multiplicative dummy variable model is:

$$\widehat{Salary} = 185.552 + 0.48420 CHits + 137.21428 LeagueN -$$
(49.79263) (0.05213) (69.291)

$$0.19214 LeagueN * CHits$$
(0.07147)

American League Model: $\widehat{Salary} = 185.55292 + 0.48420 CHits$

National League Model: $\widehat{Salary} = (185.55292 + 137.21428) + (0.48420 - 0.19214 CHits = 322.7672 + 0.29206 CHits$

The National League has higher starting salaries but slower increase thereafter.

Now for the Chow Test F-statistic:

$$R_{ur}^2 = 0.3203, R_r^2 = 0.3013, n = 263, k = 3, q = 2, n - k - 1 = 259$$

$$F = \frac{(0.3203-0.3013)/2}{(1-0.3203)/(263-3-1)} = 3.619979$$

F.DIST.RT(3.619979,2,259) = 0.028147

This p-value is less than the conventional type-one error of 0.05 so we reject the null hypothesis that the regression equations of the two leagues are the same. They are apparently different.

(c) Now add to your original level-level equation the **quadratic terms** involving the explanatory variable "Years." Write out your estimated model in standard form. At what year do major league baseball salaries reach a peak, on average. Show your work. Suppose that you are a player with 7 years of experience. How would expect your salary to change in the transition to the eighth year, other factors held constant? Show your work.

Answer:

$$\widehat{Salary} = \underset{(55.41871)}{-64.13971} + \underset{(0.06917)}{0.87219CHits} + \underset{(14.90025)}{79.25133Years} - \underset{(0.76357)}{7.97982Years\text{^}2}$$

The point where the major league baseball players' salaries peak is at approximately 5 years.

$$\frac{-79.25133}{2(-7.97982)} = 4.966\ years.$$

You can answer the last question of this part in <u>either of two ways</u>.

(i) You can use the <u>continuous approximation</u> offered by the calculus:

$$\frac{dSalary}{dYears} = 79.25133 - 2(7.97982Years) = 79.25133 - 2(7.97982 * 7)$$

= -32.4662 which translates to -$32,466

That is, in going from 7 years in the league to 8 years in the league means, on average, that a player could expect to lose -$32,466 in salary.

(ii) Alternatively, one can get a <u>discrete approximation</u> by evaluating the quadratic part of the above equation at Years = 7 and Years = 8 and see how much they differ. (Remember that we are holding other things are held constant while we change the Years variable.)

Years = 8: 79.25133(8) – 7.97982(64) = 123.3022
Years = 7: 79.25133(7) – 7.97982(49) = 163.7481
(Years = 8) – (Years = 7) = -40.446 which amounts to a loss of salary of $40,446.

(d) Run a **standardized regression** of Salary on CHits. Be sure and drop the intercept of your regression in this case as in lm((y) ~ -1 + (x)). Report your regression in standard form. Interpret the coefficient on the standardized CHits variable. Compare the t-ratio on the standardized CHits variable with the CHits variable in your first level-level regression. What do you conclude from this?

Answer:

$$\widehat{scale(Salary)} = \underset{(0.05214)}{0.55419}\, scale(CHits)$$

Notice that the intercept coefficient has been set to zero because the means of scale(Salary) and scale(CHits) are both zero and, by necessity, the intercept must be zero.

The interpretation of the coefficient on the scale(CHits) is with a one standard deviation increase in CHits one would expect, on average, a 0.55419 deviation increase in Salary.

The standard deviation in CHits is 654.47.
The standard deviation in Salary is 451.12 which equals $451,120.

Extra Comments:

 # Standardized regression: Properly run, with only standardized
 # variables in the regression, we should expect the intercept to
 # be zero since the standardized variables have zero means by
 # construction. Imposing the zero-intercept restriction in theory
 # provides for more efficient estimation of the coefficients
of the standardized regression. The t-ratio for CHits in this
 # zero-intercept standardized regression is 10.63 which is
 # almost identical to the t-ratio of 10.609 in the non-standardized
 # regression. In model7A below, the intercept is not suppressed
 # and the t-ratio of the scale(CHits) variable exactly matches the
 # 10.609 number in the non-standardized regression. Still,
 # in running a standardized regression with only standardized
 # variables in the regression, it is preferable, given estimation
 # theory, to suppress the intercept.
#
 # The BOTTOM LINE is that the statistical significance of variables
 # in the standardized regression will match the statistical significance
# of variables in the non-standardized regression. Of course, we
 # know that the advantage of the standardized regression is to
 # make the effects of explanatory variables with very different
 # units of measurement more comparable since with standardization the

 # the scales of the variables are the same in terms of standard
 # deviation measures.
#

(e) Report the R program that you used to complete this exercise.

```
library(ISLR)
attach(Hitters)
?Hitters
summary(Hitters)
model1 <- lm(Salary~CHits,data=Hitters)
summary(model1)
lSalary <- log(Hitters$Salary)
lCHits <- log(Hitters$CHits)
model2 <- lm(lSalary~CHits,data=Hitters)
summary(model2)
model3 <- lm(Salary~lCHits,data=Hitters)
summary(model3)
model4 <- lm(lSalary~lCHits,data=Hitters)
summary(model4)
model5 <- lm(Salary~CHits+League+League*CHits)
summary(model5)
anova(model5,model1)
Years2 <- Years^2
model6 <- lm(Salary~CHits+Years+Years2,data=Hitters)
summary(model6)
model7 <- lm(scale(Salary) ~ -1 + scale(CHits))
summary(model7)
model7A <- lm(scale(Salary) ~ scale(CHits))
summary(model7A)

# To get the standard deviations for CHits and Salary.
install.packages("psych")
library(psych)
describe(CHits)
describe(Salary)
```