BS2280 – Econometrics I
Homework 10: Nonlinear Models and Transformation of Variables II - Solution

1

The output shows the result of regression of $WEIGHT04$ (in pounds) on $HEIGHT$ (in inches) and its square, $HEIGHTSQ$. Provide an interpretation of the regression results.

```
> EAWE21$HEIGHTSQ <- EAWE21$HEIGHT^2
> WEIGHTfit <- lm(WEIGHT04~HEIGHT+HEIGHTSQ, data=EAWE21)
> summary(WEIGHTfit)

Call:
lm(formula = WEIGHT04 ~ HEIGHT + HEIGHTSQ, data = EAWE21)

Residuals:
    Min      1Q  Median      3Q     Max
-62.986 -22.986  -8.206  16.909 132.379

Coefficients:
               Estimate  Std. Error t value Pr(>|t|)
(Intercept) -132.556566  388.924367  -0.341    0.733
HEIGHT         3.758453   11.446696   0.328    0.743
HEIGHTSQ       0.009659    0.084018   0.115    0.909

Residual standard error: 34.61 on 497 degrees of freedom
Multiple R-squared:  0.262,     Adjusted R-squared:  0.259
F-statistic:  88.2 on 2 and 497 DF,  p-value: < 2.2e-16
```

The estimated regression model is

$$\widehat{Weight04_i} = -132.556 + 3.758 Height_i + 0.0096 Height_i^2$$

Here we have a quadratic model. To be able to interpret the marginal effect of height on weight, we have to take the first derivative of weight with respect to height (Hint: Make use of the power rule, derivative of X^2 is $2X^{2-1} = 2X$)

$$\frac{\partial \widehat{Weight04_i}}{\partial Height_i} = 3.758 + 2 \times 0.0096 \times Height_i$$

Plug in various values of height to get the marginal effect.
For example, if a person with a height of 70 inches,

$$\frac{\partial Weight04_i}{\partial Height_i} = 3.758 + 2 \times 0.0096 \times 70 = 5.102$$

A one inch increase in height of this person will increase weight on average by 5.102 pounds, ceteris paribus. Make sure that you use reasonable height values when calculating the marginal effect.
There is a positive sign of the coefficient of $HEIGHTSQ$, however, the square term has no statistical significance (p-value is 0.909, which is higher than even 10% significance level), therefore we cannot say that there is a non-linear effect of height on weight.

2

Why do economists usually stick with quadratic models, but do not consider cubic, quartic, or a polynomial of even higher order?

1. Diminishing marginal effects are standard in economic theory, justifying quadratic specifications.
2. There will be an improvement in fit as higher-order terms are added, but because these terms are not theoretically justified, the improvement will be sample-specific.
3. Unless the sample is very small, the fits of higher-order polynomials are unlikely to be very different from those of a quadratic over the main part of the data range

3

The output shows the results of regressing the logarithm of hourly earnings in USD on S (educational attainment, in years), EXP (work experience, in years), AGE (in years), and $SAGE$, an interactive term defined as the product of S and AGE. Derive the marginal effects of the coefficients of S and AGE and calculate their sizes at the mean values for S and AGE. The mean of S is 14.866 and the mean of AGE was 28.932.

```
> EAWE21$LGEARN <- log(EAWE21$EARNINGS)
> EAWE21$SAGE <- EAWE21$S * EAWE21$AGE
> EARNfit <- lm(LGEARN~S+EXP+AGE+SAGE, data=EAWE21)
> summary(EARNfit)

Call:
lm(formula = LGEARN ~ S + EXP + AGE + SAGE, data = EAWE21)

Residuals:
     Min       1Q   Median       3Q      Max
-1.94986 -0.27769  0.01489  0.29884  1.59737

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  6.962076   2.587967   2.690 0.007383 **
S           -0.290998   0.171287  -1.699 0.089969 .
EXP          0.043710   0.011416   3.829 0.000145 ***
AGE         -0.200335   0.090096  -2.224 0.026629 *
SAGE         0.013263   0.005916   2.242 0.025416 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5126 on 495 degrees of freedom
Multiple R-squared:  0.1477,    Adjusted R-squared:  0.1408
F-statistic: 21.44 on 4 and 495 DF,  p-value: 2.498e-16
```

The estimated regression model is

$$\widehat{LGEARN_i} = 6.962 - 0.290S_i + 0.043EXP_i - 0.200AGE_i + 0.013SAGE_i$$

Deriving and calculating the marginal effects of S:

$$\frac{\partial \widehat{LGEARN_i}}{\partial S_i} = -0.290 + 0.013AGE_i$$

For a 28.932 year old person,

$$\frac{\partial \widehat{LGEARN_i}}{\partial S_i} = -0.290 + 0.013AGE_i = -0.290 + 0.013 \times 28.932 = 0.086$$

one more year of education will increase hourly wages on averages by 8.6%, ceteris paribus (Don't forget that we have a semi-elasticity here!).
S is statistically significant at the 10% significance level (p-value is 0.089, which is only smaller than 10% significance level) and the interaction term is statistically significant at the 5% significance level (p-value is 0.025, which is smaller than 5% significance level), showing that education is likely to have an impact on hourly earnings.
Deriving and calculating the marginal effects of S:

$$\frac{\partial \widehat{LGEARN_i}}{\partial AGE_i} = -0.200 + 0.013S_i$$

For a person with 14.866 years of education,

$$\frac{\partial \widehat{LGEARN_i}}{\partial AGE_i} = -0.200 + 0.013 S_i = -0.200 + 0.013 \times 14.866 = -0.0067$$

one more year of age will decrease hourly wages on averages by 0.67%, ceteris paribus (Don't forget that we have a semi-elasticity here!). Both Age (p-value is 0.026, which is smaller than 5% significance level) and the interaction term (p-value is 0.025, which is smaller than 5% significance level) are statistically significant at the 5% significance level.