8 – Logarithmic functions of Y and/or X

Another way to approximate the nonlinear relationship between Y and X is by using logarithms.

- In some rare cases, taking logarithms is **not** an approximation – it is an exact way to linearize a relationship. Examples in macro/time-series
- In other cases, we can exploit a property of logs – small changes in $\log(x)$ are *approximately* percentage changes.
- How do percentage changes help us? It is a type of nonlinear effect. Example – wages and gender. (Regressions almost always use $\log(wage)$ on the LHS instead of just *wage*).

Log-approximation

Percentage change:

$$\frac{\Delta X}{X} \times 100 = \frac{x_2 - x_1}{x_1} \times 100$$

x_1 is the initial value of X, x_2 is the final value of X.

The approximation:

$$[\ln(X + \Delta X) - \ln X] \times 100 \cong \frac{\Delta X}{X} \times 100$$

$$(\ln x_2 - \ln x_1) \times 100 \cong \frac{x_2 - x_1}{x_1} \times 100$$

The approximation is better the smaller Δx.

Change in x	Percentage change: $\frac{x_2 - x_1}{x_1} \times 100$	Approximated percentage change $(\ln x_2 - \ln x_1) \times 100$
$x_1 = 1, x_2 = 2$	100%	69.32%
$x_1 = 1, x_2 = 1.1$	10%	9.53%
$x_1 = 1, x_2 = 1.01$	1%	0.995%
$x_1 = 5, x_2 = 6$	20%	18.23%
$x_1 = 11, x_2 = 12$	9.09%	8.70%
$x_1 = 11, x_2 = 11.1$	0.91%	0.91%

So how is this helpful?

<u>Three log regression specifications</u>

Case	**Population regression function**
I. linear-log	$Y = \beta_0 + \beta_1 \ln(X) + \epsilon$
II. log-linear	$\ln(Y) = \beta_0 + \beta_1 X + \epsilon$
III. log-log	$\ln(Y) = \beta_0 + \beta_1 \ln(X) + \epsilon$

- The interpretation of the slope coefficient differs in each case.
- The interpretation can be found by figuring out the change in Y for a given change in X.

Interpretation of coefficients

lin-log: $Y = \beta_0 + \beta_1 \ln(X) + \epsilon$

- A 1% change in X is associated with a $0.01\beta_1$ change in Y

log-lin: $\ln(Y) = \beta_0 + \beta_1 X + \epsilon$

- A change in X of 1 is associated with a $100\beta_1$% change in Y

log-log: $\ln(Y) = \beta_0 + \beta_1 \ln(X) + \epsilon$

- A 1% change in X is associated with a β_1% change in Y
- β_1 can be interpreted as an *elasticity*

<u>A note on R^2</u>

R^2 measures the proportion of variation in the dependent (*Y*) variable that can be explained using the *X* variables.

- When we take log(*Y*), the variance of the dependent variable changes (it tends to get smaller)
- We cannot use R^2 to compare models with different dependent variables! That is, we should not use R^2 to decide between two models, where the LHS variable is *wage* in one, and log(*wage*) in the other.

Example: CPS wages

```
install.packages("AER")
library(AER)
data("CPS1985")
attach(CPS1985)
```

Estimate log-lin model:

```
summary(lm(log(wage) ~ education + gender + age +
     experience))
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.15357    0.69387   1.663    0.097 .
education     0.17746    0.11371   1.561    0.119
genderfemale -0.25736    0.03948  -6.519 1.66e-10 ***
age          -0.07961    0.11365  -0.700    0.484
experience    0.09234    0.11375   0.812    0.417
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation: 1 more year of *education* → 17.7% increase in *wage*, etc.

Dummy variables are a bit tricky: women earn 25.7% less than men (but it's actually 100×(exp(-0.257) -1) = -22.7%)