Natural Log in Regression Models

Daniel Cullen

October 4, 2017

Economists often utilize logs (in particular natural logs: ln) because they are a nice way of linearizing variables with exponential growth (eg. wages). There are three types of log-transformed regression models: log-log, log-linear, linear-log

Examples:

$$ln(salary_i) = \alpha + \beta hours_i + \varepsilon_i \qquad \text{(log-linear)}$$
$$ln(wage_i) = \alpha + \beta ln(parentinc_i) + \varepsilon_i \qquad \text{(log-log)}$$
$$deposits_t = \alpha + \beta ln(GDP_t) + \varepsilon_t \qquad \text{(linear-log)}$$

Another (very) useful thing about logs is that they are a great way of approximating percentage changes. Say, for example, that we wanted to approximate the percentage change of going from 10 to 11 (10%). We can do the following:

$$ln(11) - ln(10) = 0.0953 \approx 0.10$$

Recall, also, some useful log rules:

- $ln(AB) = ln(A) + ln(B)$
- $ln(A/B) = ln(A) - ln(B)$

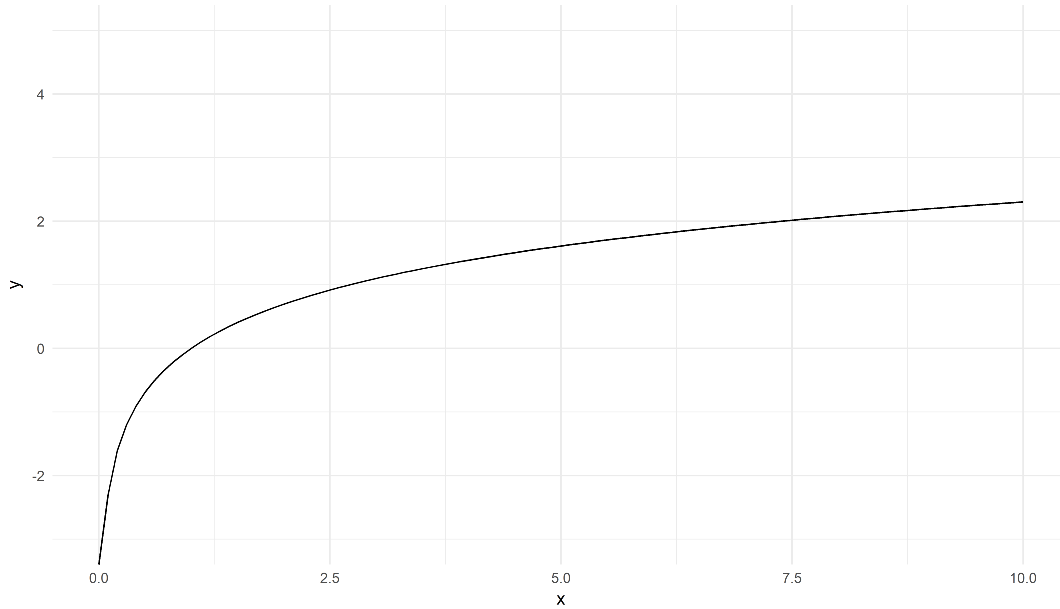
That is, we could have written our approximation as $ln(11/10)$. Further, the approximation gets better the closer are A and B. Now, let's return to our examples and interpret the β's.

$$\frac{dln(salary_i)}{dhours_i} = \beta \quad \longrightarrow \quad \text{a 1 unit inc. in hours increases wage by } 100 \times \beta\%$$

$$\frac{dln(wage_i)}{dln(parentinc_i)} = \beta \quad \longrightarrow \quad \text{a 1\% increase in parent income increases wages by } \beta\%$$

$$\frac{ddeposits_t}{dln(GDP_t)} = \beta \quad \longrightarrow \quad \text{a 1\% increase in GDP increases deposits by } \beta/100 \text{ dollars}$$

A common way to specify a non-linear relationship is the natural logarithm.



The logarithm is useful because of the following relationship. When Δx is small

$$\ln(x + \Delta x) - \ln(x) \approx \frac{\Delta x}{x}$$

or

$$\ln\left(\frac{x + \Delta x}{x}\right) = \ln\left(1 + \frac{\Delta x}{x}\right) \approx \frac{\Delta x}{x}.$$

This means that a small change in x leads to a percent change in the $\ln(x)$.

To see why this is true begin with the infinite series expansion of $\ln(1 + x)$:

$$\ln(1 + x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$
$$\ln(1 + x) = x + O(x^2)$$

$O(x^2)$ means that the remainder is bounded by Ax^2 as $x \to 0$ for some $A < \infty$. So,

$$\ln(1 + x) \approx x.$$

If, y^* is $c\%$ greater than y, then

$$y^* = y + \frac{c}{100}y = \left(1 + \frac{c}{100}\right)y.$$

Taking the log,

$$\ln(y^*) = \ln\left(y\left(1 + \frac{c}{100}\right)\right) = \ln(y) + \ln\left(1 + \frac{c}{100}\right)$$

so,

$$\ln(y^*) - \ln(y) = \ln\left(1 + \frac{c}{100}\right) \approx \frac{c}{100}.$$

The figure below shows that x and $\ln(1+x)$ are approximately equal for values close to $|x| \leq 0.1$.