ECN 102: Analysis of Economics Data

Chapter 9: Log Transformations

Remy Beauregard   Presented by Mitali Mathur

Natural Log: Proportional vs. Level Changes

Previously, we employed a (natural) log transformation to make right-skewed data appear more symmetric, but this is not all we can do with logs. We use logs when we want to consider a *proportional* change in one or more variables instead of a *level* change. We call the proportional change in Y for a proportional change in X the **elasticity** of Y with respect to X. We call the proportional change in Y for a level change in X the **semi-elasticity** of Y with respect to X.

For example, we may care more about the % change in demand rather than the unit amount change. The % change in demand for a 1% change in price would be the *price elasticity of demand*. The % change in demand for a $1 change in price would be the *price semi-elasticity of demand*.

Natural Log: Approximating Proportional Change

A proportional change in X is given by $\frac{x_1 - x_0}{x_0} = \frac{\Delta x}{x}$. We can use the natural log to approximate a proportional change in X due to the function's derivative when dx is small:

$$\frac{d \ln x}{dx} = \frac{1}{x}$$

$$d \ln x = \frac{dx}{x}$$

$$\Rightarrow \Delta \ln x \approx \frac{\Delta x}{x}$$

Thus, a 0.01 change in the log of X is roughly equivalent to a 0.01 proportional change in X: a 1% change in X.

Natural Log: Elasticity and Semi-Elasticity

As we said before, we can compute elasticities and semi-elasticities between two variables. With our proof that $\Delta \ln x \approx \frac{\Delta x}{x}$ above:

$$\text{elasticity: } \frac{\Delta \ln y}{\Delta \ln x}$$

$$\text{semi-elasticity: } \frac{\Delta \ln y}{\Delta x} \text{ or } \frac{\Delta y}{\Delta \ln x}$$

We can estimate these [semi-]elasticities with bivariate OLS after log transforming one or both of our variables.

Natural Log: Log-Linear, Linear-Log, and Log-Log Models

We have three types of bivariate OLS models that incorporate logs. We name the model in order of dependent then independent variable transformation:

$$\text{log-linear: } \widehat{\ln y} = b_1 + b_2 x$$

$$\text{linear-log: } \hat{y} = b_1 + b_2 \ln x$$

$$\text{log-log: } \widehat{\ln y} = b_1 + b_2 \ln x$$

Obviously, we cannot apply log transformations to negative numbers. Thus, any variable must not contain any negative values to be eligible for log transformation.

Natural Log: Slope Interpretation

In each case, the interpretation of b_2 is different:

- log-linear model: a 1-unit change in X is associated with a $(b_2 \times 100)$% change in Y
- log-log model: a 1% change in X is associated with a b_2% change in Y
- linear-log model: a 1% change in X is associated with a $(b_2/100)$-unit change in Y

Natural Log: Practice Problems

How would we interpret the following results:

- $b_2 = 4$ from a log-log regression
- $b_2 = 0.2$ from a log-linear regression
- $b_2 = 0.04$ from a log-log regression
- $b_2 = 3$ from a linear-log regression
- $b_2 = 40$ from a log-linear regression

"A 1-[something] change in X is associated with a ..."

Natural Log: Stata Example

```
sysuse auto, clear // use system dataset
gen ln_price = ln(price) // log transform price
reg price mpg // linear regression
reg ln_price mpg // log-linear regression
```

We are first log-transforming the price variable and then running two regressions. The first regression is a linear model, while the second is a log-linear model. The first regression estimates the change in price for a 1-unit change in miles per gallon (mpg), while the second regression estimates the % change in price for a 1-unit change in mpg.

Natural Log: Exponential Growth

We may also want to use the natural log to linearize economic quantities that grow based on a power rule or *exponentially*, such as returns on financial assets.

If we think our returns on a principal investment x_0 grow exponentially each period t based on some positive interest rate r, we can model these returns for any period by:

$$x_t = x_0(1 + r)^t$$

We may have data on the asset value each period, but not know the principal x_0 or interest rate r. How could we use log transformations to estimate these?

Natural Log: Deriving the Log-Linear Time Model

We start with our asset value for any period $x_t = x_0(1+r)^t$ and take the natural log of both sides:

$$
\begin{aligned}
\ln x_t &= \ln(x_0(1+r)^t) \\
&= \ln x_0 + \ln((1+r)^t) \text{ (because } \ln(ab) = \ln a + \ln b\text{)} \\
&= \ln x_0 + \ln(1+r) \times t \text{ (because } \ln(a^b) = b \times \ln a\text{)} \\
&= \ln x_0 + r \times t \text{ (because } \ln(1+r) \approx r \text{ for small } r\text{)}
\end{aligned}
$$

With our new equation $\mathbf{ln}\, x_t = \mathbf{ln}\, x_0 + r \times t$, we can log-transform our outcome variable and estimate x_0 and r using OLS:

`reg ln_x t` $\rightarrow r = b_2, x_0 = \exp(b_1)$.

Natural Log: Why Base e?

Why use natural log?

We could choose to use log base 10, base 5, or any other base, but we almost always use the natural log or base e. This is because $\frac{d}{dx}e^x = e^x$ (the exponential[1]) and $\ln(e) = 1$. Thus, when we take

$$\frac{d}{dx}\log_a(x)\left(= \frac{1}{x\ln(a)}\right) \text{ with } a = e$$

we get a clean $\frac{dx}{x} \approx \frac{\Delta x}{x}$.

[1] e^x or $\exp(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!}$

End of Lecture Material

Knowledge Check 9

Describe what we are doing here. What kind of regression is this and how would you write it? What do these results say? Interpret the coefficient, t-stat, p-value, and confidence interval for b_2.

```
      Source |       SS           df       MS      Number of obs   =       248
-------------+----------------------------------   F(1, 246)       =    291.11
       Model |  .715537701         1  .715537701   Prob > F        =    0.0000
    Residual |  .604652438       246  .002457937   R-squared       =    0.5420
-------------+----------------------------------   Adj R-squared   =    0.5401
       Total |  1.32019014       247  .005344899   Root MSE        =    .04958

------------------------------------------------------------------------------
    ln_close | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       month |  -.0156221   .0009156   -17.06   0.000    -.0174255   -.0138187
       _cons |   7.183856   .0067197  1069.07   0.000     7.170621    7.197092
------------------------------------------------------------------------------
```