Introductory Econometrics

Lecture 7: Variable transformations and interpretation

Alex Alekseev

November 28, 2022

University of Regensburg, Department of Economics

Previously on *Introductory Econometrics*...

- Variance of OLS
- Efficiency of OLS
- Model selection

Linear models and non-linear relationships

Introduction

- Let's start with the following linear regression

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + U.$$

- We call this model **linear** because the outcome is written as a linear function of the predictors
- Sometimes it is said the the model is linear in coefficients
- Essentially, what makes a model linear is that the only way you can combine predictors is to multiply them by a constant and add together
- However, it does **not** mean that a linear model cannot model non-linear relationships

Example

- Consider the **Cobb-Douglas** production function

$$P = AK^{\beta_1}L^{\beta^2}.$$

- Taking the natural logs on both sides yields

$$\ln P = \ln A + \beta_1 \ln K + \beta_2 \ln L.$$

- Let's denote $\beta_0 \equiv \ln A$ and add an error term U to make it an econometric model

$$\ln P = \beta_0 + \beta_1 \ln K + \beta_2 \ln L + U.$$

- Does it look similar to the model we started with?

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + U.$$

What is going on?

- We started with a **non-linear** model (Cobb-Douglas) and ended up with a **linear** regression
- **Transforming** the outcome and predictor variables in a linear regression allows one to model non-linear relationships between variables
- However, remember that while you can transform the outcome and predictors using whatever functions you want, after all these transformations you can only **multiply** the variables by a **constant** and **add** them together to preserve the **linear** structure of the regression.


YOU CAN'T USE LOGS,
IT'S A LINEAR MODEL!
imgflip.com
IT'S STILL
LINEAR IN COEFFICIENTS

More examples of linear models

- Logs:
$$Y = \beta_0 + \beta_1 \ln(X) + \beta_2 Z + U.$$
- Polynomials:
$$Y = \beta_0 + \beta_1 X + \beta_2 X^2 + U.$$
- Interaction terms:
$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + U.$$

Non-linear models

- However, there are many more models that are not linear and cannot be transformed to be linear, e.g.,

$$Y = \beta_0 + \beta_1 X^{\beta_2} + U$$

- or

$$Y = \beta_0 + \frac{1}{\beta_1 X + \beta_2 Z} + U$$

Variable transformations

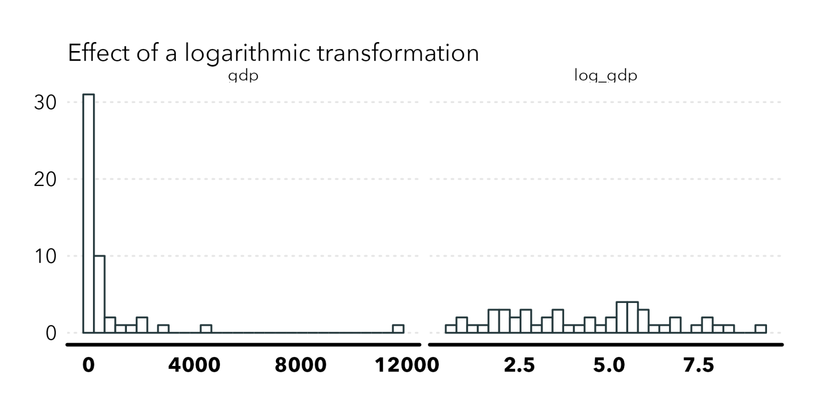
Logarithms: why 1

- There are several reasons for using a logarithmic transformation on your variables
- Logs arise when we want to transform a **multiplicative** relationship into an **additive** one, because logs transform products into sums
- The Cobb-Douglas production function we considered is an example of using logs to linearize a model

Logarithms: why 2

- The logarithmic function is **concave**, it squishes down the value of its argument
- Thus if your predictor has a lot of small values and a few very large values, using the log will make all the values closer together
- As a consequence, very large values (often called **outliers**) will not have such a big effect on the estimates
- The log transformation works well for variables that are heavily **right skewed**, like the income distribution or GDP

Example


Effect of a logarithmic transformation
gdp
log_gdp
30
20
10
0
0
4000
8000
12000
2.5
5.0
7.5

Quick recap

- Let's start with the model

$$Y = \beta_0 + \beta_1 X + U.$$

- The CEF of Y is

$$\mathbb{E}[Y \mid X] = \beta_0 + \beta_1 X.$$

- If we increase X by c units, we get

$$\mathbb{E}[Y \mid X + c] = \beta_0 + \beta_1 (X + c) = \beta_0 + \beta_1 X + \beta_1 c.$$

- Subtracting from this expression the previous one, we get

$$\mathbb{E}[Y \mid X + c] - \mathbb{E}[Y \mid X] = c\beta_1.$$

- Thus, if we increase X by c units, the conditional expectation of Y changes by $c\beta_1$ units. For example, if X increases by 1 unit, the CEF of Y changes by β_1 units.

Log on the right

- Now let's consider a model where we transform X as $\ln X$:

$$Y = \beta_0 + \beta_1 \ln X + U.$$

- As before, we have that

$$\mathbb{E}[Y \mid \ln X + c] - \mathbb{E}[Y \mid \ln X] = c\beta_1,$$

- i.e., if $\ln X$ increases by c units, the CEF of Y changes by $c\beta_1$ units

But what does it mean that $\ln X$ increases by c units?

Properties of logs

- For small c, we have

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

- Therefore

$$\ln X + c \approx \ln X + \ln(1+c) = \ln(X(1+c)).$$

- This means that increasing $\ln X$ by c units is approximately equivalent to increasing X by $100c\%$
- For example, increasing $\ln X$ by 0.01 is equivalent to increasing X by 1%

Back to the model

- We have
$$\mathbb{E}[Y \mid \ln X + c] - \mathbb{E}[Y \mid \ln X] = c\beta_1,$$
- Increasing X by $100c\%$ changes the CEF of Y by $c\beta_1$ units
- For example, increasing X by 1% changes the CEF of Y by $0.01\beta_1$ units.

Logs on the left

- We can apply a similar logic to the model in which we transform Y instead:
$$\ln Y = \beta_0 + \beta_1 X + U.$$
- In this case, we have that
$$\mathbb{E}[\ln Y \mid X + c] - \mathbb{E}[\ln Y \mid X] = c\beta_1.$$
- Increasing X by c units changes the CEF of $\ln Y$ by $c\beta_1$ units
- And from the previous analysis we know that changing $\ln Y$ by $c\beta_1$ units is approximately equivalent to changing Y by $100c\beta_1\%$
- For example, increasing X by 1 unit changes Y by $100\beta_1\%$.

Logs on both sides

- Finally, let's consider the case when both the outcome and the predictor are log-transformed:

$$\ln Y = \beta_0 + \beta_1 \ln X + U.$$

- As before, we have that

$$\mathbb{E}[\ln Y \mid \ln X + c] - \mathbb{E}[\ln Y \mid \ln X] = c\beta_1.$$

- Increasing $\ln X$ by c units changes the CEF of $\ln Y$ by $c\beta_1$ units

Logs on both sides

- Finally, let's consider the case when both the outcome and the predictor are log-transformed:

$$\ln Y = \beta_0 + \beta_1 \ln X + U.$$

- As before, we have that

$$\mathbb{E}[\ln Y \mid \ln X + c] - \mathbb{E}[\ln Y \mid \ln X] = c\beta_1.$$

- Increasing $\ln X$ by c units changes the CEF of $\ln Y$ by $c\beta_1$ units
- This means that increasing X by $100c\%$ changes Y by $100c\beta_1\%$
- For example, increasing X by 1% changes Y by $\beta_1\%$

Logs on both sides

- Finally, let's consider the case when both the outcome and the predictor are log-transformed:

$$\ln Y = \beta_0 + \beta_1 \ln X + U.$$

- As before, we have that

$$\mathbb{E}[\ln Y \mid \ln X + c] - \mathbb{E}[\ln Y \mid \ln X] = c\beta_1.$$

- Increasing $\ln X$ by c units changes the CEF of $\ln Y$ by $c\beta_1$ units
- This means that increasing X by $100c\%$ changes Y by $100c\beta_1\%$
- For example, increasing X by 1% changes Y by $\beta_1\%$
- In this case β_1 represents the **elasticity** of Y with respect to X
- Another way to see this is by taking the derivative

$$\frac{d \ln Y}{d \ln X} = \frac{dY/Y}{dX/X} = \beta_1.$$

Alternatives to logarithms

- An issue with using a logarithmic transformation is that it is **not defined** for zeros
- If the variable you plan to transform has zero values, you cannot take the log of it
- There are several practical ways of dealing with zeros
- One commonly used option is to simply **add a small value** to all of the values of your variable, so that none of the values is exactly zero
- It is an easy fix, although somewhat crude and arbitrary.

Square root

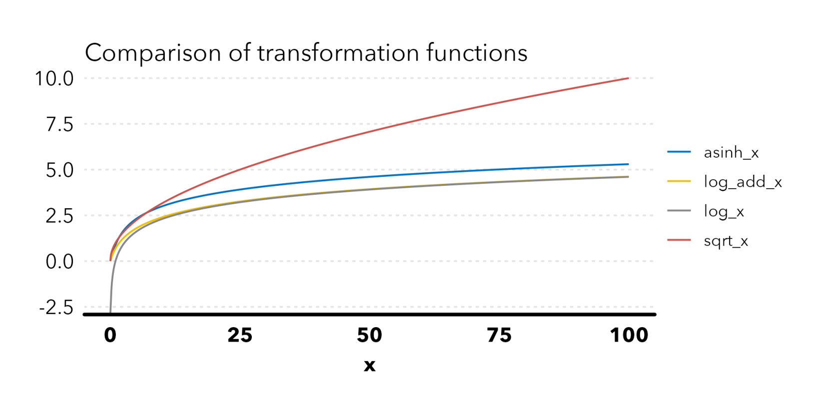
- Another option is to use a **square root** transformation
- It is also concave (squishes down large values) and it can handle zeros
- However, it does not squish down large values as much as the logarithm does
- It also becomes harder to interpret a unit change in the transformed variable

Inverse hyperbolic sine

- A more sophisticated approach to dealing with zeros is to use the **inverse hyperbolic sine** (or **asinh**) transformation defined as

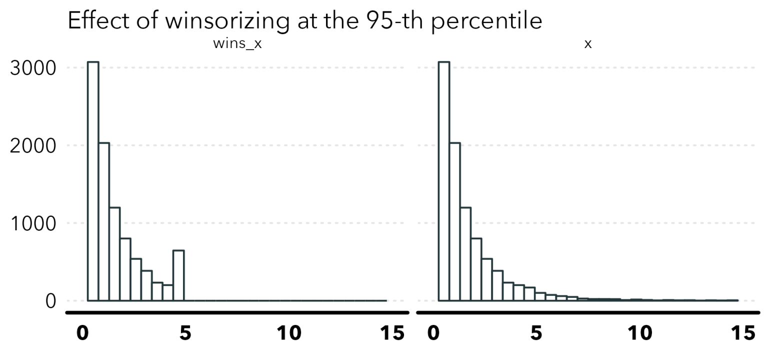
$$\mathrm{asinh}(x) = \ln(x + \sqrt{x^2 + 1}).$$

- This transformation handles zeros and behaves almost like the logarithm for large values of x
- However, it still does not have an interpretation of percentage changes


Comparison of transformation functions
10.0
7.5
5.0
2.5
0.0
-2.5
0
25
50
75
100
x
asinh_x
log_add_x
log_x
sqrt_x

Winsorizing

- An alternative way of dealing with outliers is **winsorizing**
- Instead of squishing down all the values of a variable using a concave function, winsorizing works by limiting extreme values
- A fist step in winsorizing is to define **outliers**
- For example, we can say that the values that are above the 95-th percentile of the distribution of our variable are outliers
- To winsorize, we simply take all the outliers, as defined above, and replace them with the value of our variable at the 95-th percentile
- Note that winsorizing is different from **truncating** (or **trimming**)
- Winsorizing does not discard outliers, it replaces them with other values


Effect of winsorizing at the 95-th percentile
wins_x
x
3000
2000
1000
0
0
5
10
15
0
5
10
15

Standardizing

- To standardize a variable means to subtract its mean and divide by the standard deviation:
$$\frac{X - \mathbb{E}X}{sd(X)}.$$
- Standardizing simply rescales a variable by making it have a zero mean and a standard deviation of one (compare to a **standard normal distribution**)
- The reason for standardizing variables is to make their effects on an outcome **comparable**
- If all the predictors are standardized, then each slope coefficient tells us by how many units an outcome changes if a given predictor increases by **one standard deviation**
- In this case we do not have to worry about the units of measurement of individual predictors

Standardizing explained

- Let's call the standardizing transformation $std(X)$
- Our model is then

$$Y = \beta_0 + \beta_1 std(X) + U.$$

- As before

$$\mathbb{E}[Y \mid std(X) + c] - \mathbb{E}[Y \mid std(X)] = c\beta_1.$$

- Then increasing $std(X)$ by c units means

$$std(X)+c = \frac{X - \mathbb{E}X}{sd(X)}+c = \frac{X + c \cdot sd(X) - \mathbb{E}X}{sd(X)} = std(X+c \cdot sd(X)).$$

- Increasing the standardized X by one unit is equivalent to increasing X by one standard deviation
- And increasing X by one standard deviation will change the CEF of Y by β_1 units

Standardizing outcomes

- In addition to standardizing predictors, we can also standardize an **outcome**
- In this case, the interpretation of each slope coefficient will be by how many standard deviations the outcome changes if a given predictor increases by one standard deviation
- The coefficients in a linear regression in which all the variables (outcome and predictors) are standardized are called **standardized** (or **beta**) coefficients

Next Time on *Introductory Econometrics*...

Polynomials, categorical predictors, interaction terms