DATA2002: Week 10
Interpreting log models

THE UNIVERSITY OF SYDNEY

Transformations

It is not always the case that there exists a linear relationship between two variables. Two common other relationships that occur frequently in science are the allometric and exponential relationships.

Allometric

An allometric relationship between y and x is specified as:

$$y = \theta x^{\beta}.$$

We can linearise this transformation by taking the (natural) log of both sides as follows:

$$\begin{aligned}\log(y) &= \log(\theta x^{\beta}) \\ &= \log(\theta) + \log(x^{\beta}) \\ &= \alpha + \beta \log(x)\end{aligned}$$

The resulting linear relationship has the log of y as the dependent variable and the log of x as the explanatory variable. For this reason it is sometimes referred to as a double log or log-log model.

It will be assumed throughout that we are working with log base e, that is, $\log(x) = \ln(x)$.

Note that the intercept, $\alpha = \log(\theta)$ so if we want to recover the original parameter, $\theta = e^{\alpha}$.

The interpretation of the coefficient is based on a change in x. Start with the linearised relationship:

$$\log(y) = \alpha + \beta \log(x) \tag{1}$$

Now consider what would happen if we change x by a smallish amount, $x \mapsto x + \Delta x$ then y will also change by a smallish amount $y \mapsto y + \Delta y$ in this way:

$$\log(y + \Delta y) = \alpha + \beta \log(x + \Delta x) \tag{2}$$

Note that,

$$\begin{aligned}\log(x + \Delta x) - \log(x) &= \log\left(\frac{x + \Delta x}{x}\right) \\ &= \log\left(1 + \frac{\Delta x}{x}\right) \\ &\approx \frac{\Delta x}{x}.\end{aligned}$$

Here we have used the fact that

$$\log(1 + a) \approx a$$

for small a. To see that this, consider a series expansion of $\log(1 + a)$.

If we subtract equation (1) from equation (2) we get:

$$\begin{aligned}\log(y + \Delta y) - \log(y) &= \alpha + \beta \log(x + \Delta x) - [\alpha + \beta \log(x)] \\ &= \beta \left[\log(x + \Delta x) - \log(x)\right] \\ \frac{\Delta y}{y} &\approx \beta \frac{\Delta x}{x}\end{aligned}$$

On the LHS we have the relative change in y and on the RHS we have the relative change in x multiplied by β. These relative changes in y and x can be thought of as percentage changes if we multiply them by 100. That is,

$$100 \times \frac{\Delta y}{y} = \text{percentage change in } y,$$

and

$$100 \times \frac{\Delta y}{x} = \text{percentage change in } x.$$

So we can write this:

$$\begin{gathered} 100 \times \frac{\Delta y}{y} \approx \beta \times 100 \times \frac{\Delta x}{x} \\ \% \text{ change in } y \approx \beta \times \% \text{ change in } x \end{gathered}$$

Therefore we can interpret the model as: on average, a 1% increase in x, results in a β% change in y.

Exponential

An exponential relationship between y and x is specified as:

$$y = \theta\gamma^x.$$

We can linearise this transformation by taking the (natural) log of both sides as follows:

$$\begin{aligned} \log(y) &= \log(\theta\gamma^x) \\ &= \log(\theta) + \log(\gamma^x) \\ &= \alpha + \beta x \end{aligned}$$

The resulting linear relationship has the log of y as the dependent variable and the original x as the explanatory variable. For this reason it is sometimes referred to as a semi-log or log-linear model.

Note that the intercept, $\alpha = \log(\theta)$ so if we want to recover the original parameter, $\theta = e^\alpha$, and similarly the slope coefficient is $\beta = \log(\gamma)$ so to recover the original parameter we use $\gamma = e^\beta$.

The interpretation of the coefficient is based on a change in x. Start with the linearised relationship:

$$\log(y) = \alpha + \beta x \tag{3}$$

Now consider what would happen if we change x by a smallish amount, $x \mapsto x + \Delta x$ then y will also change by a smallish amount $y \mapsto y + \Delta y$ in this way:

$$\log(y + \Delta y) = \alpha + \beta(x + \Delta x) \tag{4}$$

If we subtract equation (3) from equation (4) we get:

$$\begin{aligned}
\log(y + \Delta y) - \log(y) &= \alpha + \beta(x + \Delta x) - [\alpha + \beta x] \\
&= \beta \times \Delta x \\
\frac{\Delta y}{y} &\approx \beta \times \Delta x
\end{aligned}$$

On the RHS we have the absolute change in x multiplied by β and on the LHS we have the relative change in y, which can be interpreted as a percentage change if we multiply it by 100, i.e.,

$$\% \text{ change in } y = 100 \times \frac{\Delta y}{y} \approx 100 \times \beta \times \Delta x.$$

Therefore, on average, a 1 unit increase in x (i.e. $\Delta x = 1$) results in a $100 \times \beta\%$ change in y.