Most of this is optional; if you just want the bare necessities, jump to Table 1. But if you are a big nerd like me and want to know a bit more about why log forms have the interpretations we use, then continue on.

1 Logs as Percentage Changes

Suppose $y = \log(x)$. You hopefully recall from calculus that $dy/dx = 1/x$. Multiplying both sides by dx gives $dy = dx/x$. The interpretation of dx is a really, really, really, ridiculously small change in x. Ergo dx/x is the change in x as a proportion of the level of x.

While the preceding is *exactly* true for really, really, really, ridiculously small changes in x, it is *approximately* true for larger (but still small) changes in x. (The approximation typically becomes worse as the change in x gets bigger, but it's arbitrary when the approximation should be considered "bad.") So as long as we're talking about fairly small changes in x, we can replace dx with Δx and write $\Delta y \approx \Delta x/x$. Since changes in economics tend to be small — think of inflation rates and interest rates, changes usually less than 0.10 annually — this approximation is widely used in economics.

So in words, the change in y is approximately equal to the proportional change in x. If x increases by 1% (that is, if $\Delta x/x = 0.01$), then y increases by approximately 0.01. If x decreases by 3%, then y decreases by approximately 0.03. And so on and so forth.

Let me use some numbers to illustrate more explicitly. Suppose $x = 100$, so that $y = \log(100) = 4.6052$. Now we increase x to 101, from which it follows that $\Delta x = 1$ and $\Delta x/x = 0.01$. Okay, so x has increased by 1%. And now $y = \log(101) = 4.6151$, from which it follows that $\Delta y = 0.00995$. Yeah, that's pretty close to $\Delta x/x$, as expected.[1]

Okay, now back to regressions.

2 Linearity in Parameters

The OLS estimation technique requires that our model be *linear in parameters*. What this means is, each β term must appear essentially as a constant: we cannot have β_1^2 or $\log(\beta_1)$

[1] Suppose x_1 is the value of x in period 1 and x_2 is the value of x in period 2. Therefore $y_1 = \log(x_1)$ and $y_2 = \log(x_2)$, so we can write $\Delta y = \log(x_2) - \log(x_1)$. But we just showed that $\Delta y \approx \Delta x/x$. So we can conclude that

$$\frac{x_2 - x_1}{x_1} \approx \log(x_2) - \log(x_1).$$

In words, we can approximate a proportional change in x by subtracting logs. (And of course, multiply both sides by 100 to get the percentage change.) This is also widely used in economics.

or $\beta_1\beta_2$, for instance. This is because the OLS technique is only able to solve explicitly for each β if they appear in a linear fashion.

However, this does not necessitate that the model be linear in *variables*. There is no reason why we can't specify a model of the form

$$y = \beta_1 + \beta_2 \log(x) + u$$

if we think it is useful to do so.[2] And it certainly might be useful to do so. Consider the relationship between healthcare expenditure and life expectancy. We would expect more healthcare expenditure to be correlated with higher life expectancy, but at a diminishing rate (since there is a natural limit to life expectancy that medical treatment cannot overcome). So it wouldn't make sense to impose a linear relationship on healthcare expenditure and life expectancy; instead, use a log to capture diminishing returns to healthcare expenditure.

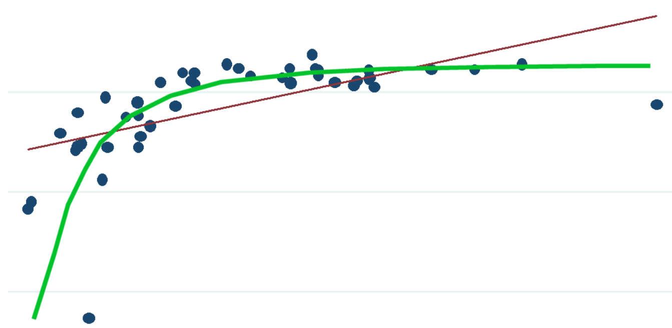

FIGURE 1: Specifying a logarithmic relationship (green) generates a better fit of the data compared to a linear relationship (red).

3 Functional Forms

There are an infinite number of ways we could specify such a model. I focus on three due to their salient economic interpretations. It will be assumed throughout that the zero conditional mean assumption holds; this implies that x and u are uncorrelated. In practice, this means that when we change x, there is no change in u on average. This is useful because we take derivatives, for which the zero conditional mean assumption implies that $du/dx = 0$.

[2] If you'd like, you can define $v \equiv \log(x)$ and rewrite the model as $y = \beta_1 + \beta_2 v + u$, from which it is obvious that the model exhibits the same form as that with which we are familiar.

3.1 Linear-Linear Regression

A **linear-linear** regression is of the form

$$y = \beta_1 + \beta_2 x + u. \tag{1}$$

It is named as such because the dependent variable is linear (it's simply y) and the regressor variable is also linear (it's simply x). The interpretation is that an increase in x by one unit is associated with a change in y of β units. This is the regression we've been focusing on.

3.2 Linear-Log Regression

A **linear-log** regression is of the form

$$y = \beta_1 + \beta_2 \log(x) + u. \tag{2}$$

Although $\log(x)$ may do a better job of capturing the data or a salient economic phenomenon, we are not interested in how a change in $\log(x)$ will affect y; we are interested in how a change in x will affect y. We'll have to do a little work to squeeze out that information, but it's not so bad. First, we can take the derivative of both sides with respect to x, which yields

$$\frac{dy}{dx} = \frac{\beta_2}{x}.$$

Now multiply both sides by dx, and then multiply the right-hand size by $100/100$. Moving around the 100 factors, we end up with

$$dy = \frac{\beta_2}{100}\left(\frac{dx}{x} \times 100\right).$$

Notice that the term in parentheses is the percentage change in x.

Although calculus operations are in terms of infinitesimally small changes dy and dx, the equation still constitutes a valid approximation for small changes Δy and Δx. In practice, we will typically consider $\%\Delta x = 1$. The interpretation in words is: the change in the level of y equals $\beta_2/100$ times the percentage change in x,

$$\Delta y \approx \frac{\beta_2}{100} \times \%\Delta x. \tag{3}$$

3.3 Log-Linear Regression (Semi-Elasticity)

A **log-linear** regression is of the form

$$\log(y) = \beta_1 + \beta_2 x + u. \tag{4}$$

To interpret, start by taking the derivative of both sides with respect to x, which yields

$$\frac{d\log(y)}{dx} = \beta_2.$$

In order to squeeze y out of the left-hand side, we will appeal to the chain rule of calculus. In particular, we can write

$$\frac{d\log(y)}{dy}\frac{dy}{dx} = \beta_2.$$

We know that $d\log(y)/dy = 1/y$, so let's make that substitution. Let's also multiply both sides by $100 \times dx$, which yields

$$\frac{dy}{y} \times 100 = (100 \times \beta_2)dx.$$

In words: the percentage change in y equals $100 \times \beta_2$ times the change in the level of x,

$$\%\Delta y \approx 100\beta_2 \times \Delta x. \tag{5}$$

In this form, coefficient β_2 is referred to as the **semi-elasticity** of y with respect to x.

3.4 Log-Log Regression (Elasticity)

A **log-log** regression is of the form

$$\log(y) = \beta_1 + \beta_2 \log(x) + u. \tag{6}$$

To interpret, take the derivative of both sides with respect to x, which gives

$$\frac{d\log(y)}{dx} = \frac{\beta_2}{x}.$$

Use the chain rule again on the right-hand side so that

$$\frac{d\log(y)}{dy}\frac{dy}{dx} = \frac{\beta_2}{x}.$$

We know that $d\log(y)/dy = 1/y$, so let's make that substitution. Also multiply both sides by dx and both sides by 100. Doing so yields

$$\frac{dy}{y} \times 100 = \beta_2 \left(\frac{dx}{x} \times 100 \right).$$

In words: the percentage change in y is equal to β_2 times the percentage change in x,

$$\%\Delta y \approx \beta_2 \times \%\Delta x. \tag{7}$$

In this form, coefficient β_2 is referred to as the **elasticity** of y with respect to x, which you hopefully remember from a microeconomics course.

4 Summary

Again, there are a multitude of other functional forms we could consider, e.g. quadratic forms, that are useful in certain contexts. Those will be discussed later as they become pertinent. But for now, the following table summarizes the four functional forms introduced here.

Model	Dependent Variable	Regressor	Interpretation of β_2
linear	y	x	$\Delta y = \beta_2 \times \Delta x$
linear-log	y	$\log(x)$	$\Delta y \approx \frac{\beta_2}{100} \times \%\Delta x$
log-linear (semi-elasticity)	$\log(y)$	x	$\%\Delta y \approx 100\beta_2 \times \Delta x$
log-log (elasticity)	$\log(y)$	$\log(x)$	$\%\Delta y \approx \beta_2 \times \%\Delta x$

TABLE 1: Common functional forms and their interpretations.

5 Life Expectancy and Healthcare Expenditure

Consider data with three variables: country, life expectancy at birth, and healthcare spending per-capita in 2015. As can be ascertained from the Stata output that follows, the linear-

linear regression

$$\text{lifeexpect} = \beta_1 + \beta_2 \text{hcspending} + u$$

gives goodness-of-fit measure $R^2 = 0.363$. The linear-log regression as suggested earlier,

$$\text{lifeexpect} = \beta_1 + \beta_2 \log(\text{hcspending}) + u,$$

gives goodness-of-fit measure $R^2 = 0.542$, implying a better fit. The interpretation of the linear-log model is that a 1% increase in healthcare spending is associated with, on average, an increase in life expectancy by about 0.0465 years.

```
. regress lifeexpect hcspending

      Source |       SS           df       MS      Number of obs   =        44
-------------+----------------------------------   F(1, 42)        =     23.93
       Model |  394.963176         1  394.963176   Prob > F        =    0.0000
    Residual |  693.098579        42  16.5023471   R-squared       =    0.3630
-------------+----------------------------------   Adj R-squared   =    0.3478
       Total |  1088.06175        43  25.3037617   Root MSE        =    4.0623

------------------------------------------------------------------------------
  lifeexpect |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  hcspending |   .0014559   .0002976     4.89   0.000     .0008553    .0020565
       _cons |   73.98101   1.148863    64.39   0.000     71.66251    76.29951
------------------------------------------------------------------------------

. gen loghc = log(hcspending)

. regress lifeexpect loghc

      Source |       SS           df       MS      Number of obs   =        44
-------------+----------------------------------   F(1, 42)        =     49.67
       Model |  589.550424         1  589.550424   Prob > F        =    0.0000
    Residual |  498.511331        42  11.8693174   R-squared       =    0.5418
-------------+----------------------------------   Adj R-squared   =    0.5309
       Total |  1088.06175        43  25.3037617   Root MSE        =    3.4452

------------------------------------------------------------------------------
  lifeexpect |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       loghc |   4.646454   .6592863     7.05   0.000      3.31596    5.976947
       _cons |   42.30245    5.19564     8.14   0.000     31.81723    52.78768
------------------------------------------------------------------------------
```

There's an **important caveat** in comparing the R^2 of different models: it is *not* meaningful to compare the R^2 of models that have different dependent variables! If the regressors are different but the dependent variables are the same, then it is fine to compare R^2.

6 Retransformation Bias

Suppose you run a log-linear regression and estimate

$$\widehat{\log(y)} = b_1 + b_2 x + e.$$

This tells you what the fitted value of $\log(y)$ is. But what if you want to know what the fitted value of y is? In other words, what if you have $\widehat{\log(y)}$ but you want $\widehat{y}$?

The first instinct for most is to exponentiate both sides because $e^{\log(y)} = y$, and therefore $e^{\widehat{\log(y)}} = \widehat{y}$. ***This is wrong!!!!11!*** Don't do it. Your value for $\widehat{y}$ will be biased, and this is called the **retransformation bias**.

The following explanation is very much optional (but the conclusion is not). Assuming the zero conditional mean holds, $E[u|x] = 0$ implies that u and x are uncorrelated. Transforming the estimated log-linear form implies that

$$y = e^{\beta_1 + \beta_2 x + u}.$$

Now taking the conditional mean gives

$$\begin{aligned} E[y|x] =& E[e^{\beta_1 + \beta_2 x + u}|x] \\ & E[e^{\beta_1 + \beta_2 x} e^{u}|x] \\ & e^{\beta_1 + \beta_2 x} E[e^{u}|x]. \end{aligned}$$

However, $E[u|x] = 0$ does not impy that $E[e^u|x] = 1$, so in general it is the case that

$$E[y|x] \neq e^{\beta_1 + \beta_2 x}.$$

Not good, and such is the source of the retransformation bias.

Okay, so then how can we get $\widehat{y}$ from $\widehat{\log(y)}$? Use a *bias correction* term, $e^{s_e^2/2}$, so that

$$\begin{aligned} \widehat{y} &= e^{s_e^2/2} e^{\widehat{\log(y)}} \\ &= e^{s_e^2/2} e^{b_1 + b_2 x}, \end{aligned}$$

where s_e is the standard error of the log-linear regression (i.e. its root MSE in Stata). Do note that this correction requires normally distributed errors and homoskedasticity to be valid.