CHAPTER 13

Making Regression More Flexible

13.0 What We Need to Know When We Finish This Chapter

For our purposes, regression has to be a linear function of the constant and coefficients so that their estimators can be linear functions of the dependent variable. However, explanatory variables can appear in discrete and nonlinear form. These forms give us the opportunity to represent a wide and varied range of possible relationships between the explanatory and dependent variables.

1. **Section 13.2:** Dummy variables identify the absence or presence of an indivisible characteristic. The intercept for observations that don't have this characteristic is *a*. The effective intercept for observations

that do have this characteristic is $a + b_2$, where b_2 is the slope associated with the dummy variable. The slope b_2 estimates the fixed difference in y_i between those that do and those that do not have the characteristic at issue.

2. **Section 13.2:** We fall into the dummy variable trap when we enter one dummy variable for a particular characteristic, and another dummy variable for the opposite or absence of that characteristic. These dummy variables are perfectly correlated, so slopes and their variances are undefined. We only need one dummy variable, because its slope measures the difference between the values of y_i for observations that have the characteristic and those that don't or have its opposite.
3. **Equations (13.13) and (13.18), section 13.3:** The quadratic specification is

   $$y_i = \alpha + \beta_1 x_i + \beta_2 x_i^2 + \varepsilon_i.$$

   $\beta_1 x_i$ is the linear term. $\beta_2 x_i^2$ is the quadratic term. If $\beta_1 > 0$ and $\beta_2 < 0$, small changes in x_i increase $E(y_i)$ when $x_i < -\beta_1/2\beta_2$ and reduce it when $x_i > -\beta_1/2\beta_2$.
4. **Equations (13.35) and (13.36), section 13.4:** The log-linear or semi-log specification is

   $$\ln y_i = \alpha + \beta x_i + \varepsilon_i.$$

   The coefficient is the expected relative change in the dependent variable in response to an absolute change in the explanatory variable:

   $$\beta = \frac{E[\Delta y / y_i]}{\Delta x}.$$

5. **Equations (13.38) and (13.39), section 13.4:** The log-log specification is

   $$\ln y_i = \alpha + \beta \ln x_i + \varepsilon_i.$$

   The coefficient is the elasticity of the expected change in the dependent variable with respect to the change in the explanatory variable:

   $$\beta = \frac{E[\Delta y / y_i]}{\Delta x / x_i} = \eta_{yx}.$$

6. **Equations (13.48) and (13.52), section 13.5:** Interactions allow the effect of one variable to depend on the value of another. The population relationship with an interaction is

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{1i} x_{2i} + \varepsilon_i .$$

The change in the expected value of y_i with a change in x_{1i} is

$$\frac{\Delta y}{\Delta x_1} = \beta_1 + \beta_3 x_{2i} .$$