Chapter 5
Transformations of Variables and Interactions

5.1 Basic idea

One limitation in the linear regression analysis is that the dependent variable has to be linear in the parameters:

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 X_3 + \cdots + \beta_k X_k + u. \tag{5.1}$$

However, there are equations that are not linear, for example:

$$Y = \beta_1 + \beta_2 X_2 + X_3^{\beta_3} + u. \tag{5.2}$$

This Equation 5.2 cannot be estimated using OLS. One way to estimate nonlinear models is by using Nonlinear Least Squares (NLS), which is an extension of the methods we discussed before. In this chapter, rather that focusing on NLS, we will see how transformations in the variables can allow us to use OLS on a variety of non-linear models. For example, consider the estimation of the following Cobb-Douglas production function:

$$P_i = AL_i^{\beta_2} K_i^{\beta_3} e^{\varepsilon_i}, \tag{5.3}$$

where P_i is total production or total output, A is a technology constant, K_i is the amount of capital, and L_i is labor. Taking natural logs we have:

$$\log P_i = \log A + \beta_2 \log L_i + \beta_3 \log K_i + \varepsilon_i. \tag{5.4}$$

If we simple set $Y_i = \log P_i$, $\beta_1 = \log A$, $X_2 = \log L_i$, and $X_3 = \log K_i$ we can write Equation 5.4 as:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i, \tag{5.5}$$

that can be easily estimated via OLS. β_2 and β_3 will correspond to the ones given in Equation 5.3. Another example of a model that can be estimated with OLS is:

$$Y_i = \beta_1 + \beta_2 Z_{2i}^2 + \beta_3 \sqrt{Z_{3i}} + \beta_4 \frac{1}{Z_{4i}} + \varepsilon_i. \tag{5.6}$$

We just need to replace $X_{2i} = Z_{2i}^2$, $X_{3i} = \sqrt{Z_{3i}}$, $X_{4i} = \frac{1}{Z_{4i}}$.

5.2 Logarithmic transformations

To explain the logarithm transformation let's go over one example in Gretl. If we want to estimate the following model:

$$\log \texttt{crime}_i = \beta_1 + \beta_2 \log \texttt{pop}_i + \beta_3 \texttt{unem}_i + \beta_4 \texttt{offi}_i + u_i, \tag{5.7}$$

you need to create the new variables first. Go to `Add` → `Define new variable` and type:

```
logcrime = log(crime)
```

This will generate the new variable `logcrime`. Do the same thing for log population and then estimate the model. The regression output is:

```
Model 1: OLS, using observations 1-92
Dependent variable: logcrime

             coefficient    std. error    t-ratio    p-value
  ----------------------------------------------------------
  const      -0.709735      0.807193      -0.8793   0.3817
  unem       -0.00456848    0.00903041    -0.5059   0.6142
  offi        0.000144915   6.15429e-05    2.355    0.0208    **
  logpop      0.864044      0.0662782     13.04     2.92e-022 ***

Mean dependent var   10.33774   S.D. dependent var   0.742056
Sum squared resid    6.883563   S.E. of regression   0.279683
R-squared            0.862628   Adjusted R-squared   0.857945
F(3, 88)             184.1989   P-value(F)           8.20e-38
Log-likelihood      -11.28034   Akaike criterion     30.56069
Schwarz criterion    40.64784   Hannan-Quinn         34.63195

Excluding the constant, p-value was highest for variable 3 (unem)
```

$$\widehat{\texttt{logcrime}} = \underset{(0.80719)}{-0.709735} - \underset{(0.00903)}{0.00456}\,\texttt{unem} + \underset{(6.1543\text{e–}005)}{0.000144915}\,\texttt{offi} + \underset{(0.0663)}{0.8640}\,\texttt{logpop}$$

$$N = 92 \quad \bar{R}^2 = 0.8579 \quad F(3,88) = 184.20 \quad \hat{\sigma} = 0.27968$$

(standard errors in parentheses)

First, notice how the coefficients are very different from the one obtain with no logarithm transformation. Here the interpretation is different. β_2 is interpreted as the elasticity of `crime` with respect to `pop`:

$$\beta_2 = \frac{\Delta \texttt{crime}/\texttt{crime}}{\Delta \texttt{pop}/\texttt{pop}}. \tag{5.8}$$

A one percentage increase in `pop` will increase `crime` by 0.864 percent. Δ`crime/crime` is interpreted as a percentage change in `crime`. For β_4 we have:

$$\beta_4 = \frac{\Delta \texttt{crime/crime}}{\Delta \texttt{offi}}. \tag{5.9}$$

Here, a one unit increase in `offi` is associated with a 0.014% (0.00014 $\times$ 100 percent) increase in `crime`.

5.3 Quadratic terms

So far we have bee estimating the marginal effects (βs) that are constant across all possible values of X. The simplest way to introduce nonlinearities in the marginal effect is to estimate the model with quadratic terms. For example, let the model be:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^2 + \varepsilon_i. \tag{5.10}$$

In this case the marginal effect of X on Y is given by:

$$\frac{\Delta Y}{\Delta X} = \beta_2 + 2 \cdot \beta_3 X_i. \tag{5.11}$$

If we want to estimate the marginal effect of experience of wages and in addition we allow for a nonlinear effect we can estimate:

$$\texttt{wage}_i = \beta_1 + \beta_2 \texttt{exper}_i + \beta_3 \texttt{expersq}_i + \varepsilon_i, \tag{5.12}$$

where `wage` is average hourly earnings, `exper` is years of experience and `expersq` is the number of years of experience squared. The Gletl output is the following:

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

             coefficient   std. error    t-ratio    p-value
  ---------------------------------------------------------
  const       3.72541      0.345939      10.77     1.46e-024 ***
  exper       0.298100     0.0409655      7.277    1.26e-012 ***
  expersq    -0.00612989   0.000902517   -6.792    3.02e-011 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    6496.147   S.E. of regression   3.524334
R-squared            0.092769   Adjusted R-squared   0.089300
F(2, 523)            26.73982   P-value(F)           8.77e-12
Log-likelihood      -1407.455   Akaike criterion     2820.910
Schwarz criterion    2833.706   Hannan-Quinn         2825.920
```

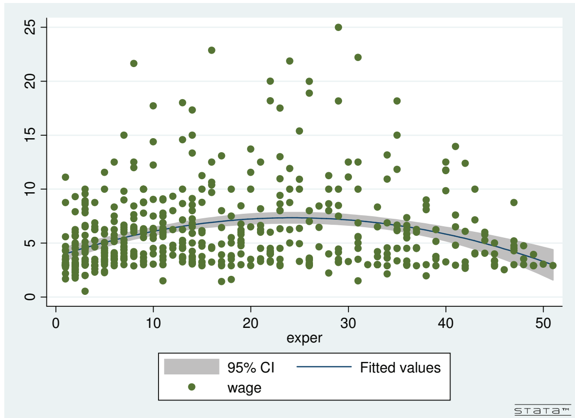

Fig. 5.1 Predicted values for Equation 5.12

$$\widehat{\texttt{wage}} = \underset{(0.34594)}{3.72541} + \underset{(0.040966)}{0.298100}\,\texttt{exper} - \underset{(0.00090252)}{0.00612989}\,\texttt{expersq}$$
$$N = 526 \quad \bar{R}^2 = 0.0893 \quad F(2,523) = 26.740 \quad \hat{\sigma} = 3.5243$$
$$\text{(standard errors in parentheses)}$$

Here, the marginal effect of experience on average hourly wage is:

$$\begin{aligned}\frac{\Delta\texttt{wage}}{\Delta\texttt{exper}} &= 0.2981 + 2\cdot(-0.006)\texttt{exper}\\ &= 0.2981 - 0.012\texttt{exper}.\end{aligned}$$

For a person with 2 years of experience, the effect of an additional year of experience on wage is 0.2741 (=0.2981 - 0.012 $\times$ 2) and for a person with 15 years of experience, the marginal effect of an additional year of experience is 0.1181 (=0.2981 - 0.012 $\times$ 15). Hence, we can say that for a reasonable range of years of experience, experience has a positive effect on wage. In addition, this effect is smaller as you accumulate more experience.

Figure 5.1 show the fitted regression line along with the 95% confidence interval for the fitted values and the actual data. This figure clearly shows the nonlinear marginal effect and innlustrates how wages increase with experience for about the first 25 years, but then wages decrease later on.

5.4 Interaction terms

A second popular approach to allow for the marginal effect to change over different values of X is to include interaction terms in the regression equation. For example,

$$Y = \beta_1 + \beta_2 X_2 + \beta_3 (X_2 \times X_3) + \varepsilon_i. \tag{5.13}$$

In this case the marginal effect of X_2 on Y depends on X_3 is given by:

$$\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_3. \tag{5.14}$$

Consider the next example with the interaction between `exper` and `educ` in a wage equation:

$$\mathtt{wage}_i = \beta_1 + \beta_2 \mathtt{exper}_i + \beta_3 (\mathtt{exper}_i \times \mathtt{educ}_i) + \varepsilon_i, \tag{5.15}$$

where the marginal effect of experience on wage depends on the level of education:

$$\frac{\Delta \mathtt{wage}}{\Delta \mathtt{exper}} = \beta_2 + \beta_3 \mathtt{educ}. \tag{5.16}$$

When estimating this equation in Gretl we have to make sure we generate the interaction term first. That is, go to `Add` $\rightarrow$ `Define new variable` and type:

```
expereduc = exper*educ
```

Then we are ready to estimate the equation via OLS. The regression output is:

```
Model 1: OLS, using observations 1-526
Dependent variable: wage

              coefficient   std. error   t-ratio    p-value
  ---------------------------------------------------------
  const        4.88993      0.242730     20.15     3.11e-067 ***
  exper       -0.188124     0.0253904    -7.409    5.13e-013 ***
  expereduc    0.0207731    0.00217625    9.545    5.17e-020 ***

Mean dependent var   5.896103   S.D. dependent var   3.693086
Sum squared resid    6020.313   S.E. of regression   3.392803
R-squared            0.159223   Adjusted R-squared   0.156008
F(2, 523)            49.52175   P-value(F)           2.01e-20
Log-likelihood      -1387.449   Akaike criterion     2780.897
Schwarz criterion    2793.693   Hannan-Quinn         2785.908
```

$$\widehat{\mathtt{wage}} = \underset{(0.24273)}{4.88993} - \underset{(0.025390)}{0.188124}\,\mathtt{exper} + \underset{(0.0021762)}{0.0207731}\,\mathtt{expereduc}$$

$$N = 526 \quad \bar{R}^2 = 0.1560 \quad F(2,523) = 49.522 \quad \hat{\sigma} = 3.3928$$

(standard errors in parentheses)

Here, the marginal effect of experience on wage is:

$$\frac{\Delta \texttt{wage}}{\Delta \texttt{exper}} = -0.1881 + 0.0208\texttt{educ}$$

For a person with twelve years education (high school), the marginal effect from an additional year of education is 0.0615 (=-0.1881+0.208×12). However, with more education the marginal effect is larger. A person with 16 years of education (high school + college) will have a marginal effect of 0.1447 (=-0.1881+0.208×16). Notice that for an important range of education the marginal effect is positive, meaning that more experience leads to higher wages. In addition, the effect if larger if you have more education. This means that going to school is not only good because it directly increases your expected wage but also makes additional years of experience more valuable.