FIE401: Financial Econometrics

Non-linear Functional Form: Log Transformations

Darya Yuferova

Non-Linear Functional Form (i)

- The regression functions so far have been linear in the xs.
- But the linear approximation is not always a good one.
- The multiple regression model can handle regression functions that are non-linear in one or more xs.
- Linear regression $=$ Linear in parameters!

Non-Linear Functional Form (ii)

If a relation between y and x is non-linear:

- The effect on y of a unit change in x depends on the value of x – that is, the marginal effect of x is not constant.
- A linear regression is misspecified: the functional form is wrong.
- The estimator of the effect on y of x is biased.
- The solution is to estimate a regression function that is non-linear in x.

Example: Test Scores

Is the relationship between test scores and independent variables linear?

- testscr: test scores
- avginc: average family income

Logarithmic Functions of y and x

$$\ln(x) = \text{natural logarithm of } x$$

- Logarithmic transforms permit modeling relations in "percentage" terms (like elasticities).
- For small $\triangle x$:
  $\ln(x + \triangle x) - \ln(x) = \ln(\frac{x+\triangle x}{x}) = \ln(1 + \frac{\triangle x}{x}) \approx \frac{\triangle x}{x}$

Three Cases of Log Specifications

- Linear-Log: $y = \beta_0 + \beta_1 \ln(x) + u$
- Log-Linear: $\ln(y) = \beta_0 + \beta_1 x + u$
- Log-Log: $\ln(y) = \beta_0 + \beta_1 \ln(x) + u$
- The interpretation of the slope coefficient differs in each case!

Linear-Log Population Regression Function

$$y = \beta_0 + \beta_1 \ln(x) + u$$

- Compute y before and after change in x
- Before: $y = \beta_0 + \beta_1 \ln(x)$
- After: $y + \triangle y = \beta_0 + \beta_1 \ln(x + \triangle x)$
- Substract "before" from "after"
- $\triangle y = \beta_1[\ln(x + \triangle x) - \ln(x)] \approx \beta_1 \frac{\triangle x}{x}$, for small $\triangle x$
- $\beta_1 \approx \frac{\triangle y}{\triangle x / x}$

Linear-Log Case

$$y = \beta_0 + \beta_1 \ln(x) + u$$

- For small $\triangle x$: $\beta_1 \approx \frac{\triangle y}{\triangle x / x}$
- $100 \times \frac{\triangle x}{x}$ is a percentage change in x
- 1% increase in x (i.e., $\triangle x = 0.01 \times x$) is associated with $0.01 \times \beta_1$ increase in y
- 1% increase in $x \approx 0.01$ units increase in $\ln(x)$

Example: Test Scores and Income

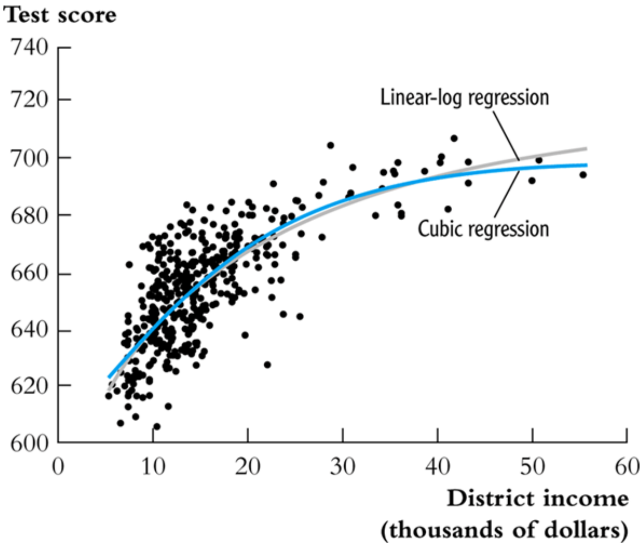
$$\widehat{testscr} = 557.83 + 36.42 \ln(avginc)$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#run regression
fit<-lm(testscr~log(avginc),data=School);
#summary of the regession
#robust to heteroskedasticity
coeftest(fit,vcov=hccm);
```

Example: Test Scores and Income

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 557.8323     3.8622 144.433 < 2.2e-16 ***
## log(avginc)  36.4197     1.4058  25.906 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Test Scores and Income



Log-Linear Population Regression Function

$$\ln(y) = \beta_0 + \beta_1 x + u$$

- Compute y before and after change in x.
- Before: $\ln(y) = \beta_0 + \beta_1 x$.
- After: $\ln(y + \triangle y) = \beta_0 + \beta_1(x + \triangle x)$.
- Substract "before" from "after".
- $\ln(y + \triangle y) - \ln(y) \approx \frac{\triangle y}{y} \approx \beta_1 \triangle x$.
- $\beta_1 \approx \frac{\triangle y / y}{\triangle x}$, for small $\triangle x$.

Log-Linear Case

$$\ln(y) = \beta_0 + \beta_1 x + u$$

- For small $\triangle x$: $\beta_1 \approx \frac{\triangle y/y}{\triangle x}$.
- $100 \times \frac{\triangle y}{y}$ is a percentage change in y.
- 1 unit increase in x (i.e., $\triangle x = 1$) is associated with $100 \times \beta_1\%$ increase in y.
- 1 unit increase in $x = \beta_1$ increase in $\ln(y)$.

Log-Log Population Regression Function

$$\ln(y) = \beta_0 + \beta_1 \ln(x) + u$$

- Compute y before and after change in x.
- Before: $\ln(y) = \beta_0 + \beta_1 \ln(x)$.
- After: $\ln(y + \triangle y) = \beta_0 + \beta_1 \ln(x + \triangle x)$.
- Substract "before" from "after".
- $\ln(y + \triangle y) - \ln(y) \approx \frac{\triangle y}{y} \approx \beta_1(\ln(x + \triangle x) - \ln(x)) \approx \beta_1 \frac{\triangle x}{x}$.
- $\beta_1 \approx \frac{\triangle y / y}{\triangle x / x}$, for small $\triangle x$.

Log-Log Case

$$\ln(y) = \beta_0 + \beta_1 \ln(x) + u$$

- For small $\triangle x$: $\beta_1 \approx \frac{\triangle y/y}{\triangle x/x}$.
- $100 \times \frac{\triangle y}{y}$ is a percentage change in y.
- $100 \times \frac{\triangle x}{x}$ is a percentage change in x.
- 1% increase in x is associated with β_1% increase in y.
- In the log-log specification, β_1 has the interpretation of an elasticity.

Example: Test Scores and Income

$$\ln(\widehat{testscr}) = 6.34 + 0.06 \ln(avginc)$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#run regression
fit<-lm(log(testscr)~log(avginc),data=School);
#summary of the regession
#robust to heteroskedasticity
coeftest(fit,vcov=hccm);
```

Example: Test Scores and Income

```
##
## t test of coefficients:
##
##              Estimate Std. Error  t value  Pr(>|t|)
## (Intercept) 6.3363494  0.0059587 1063.374 < 2.2e-16 ***
## log(avginc) 0.0554190  0.0021582   25.679 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary for Log Transformations

- Three cases depending on whether y and/or x is transformed by taking logarithms.
- The regression is linear in the new variable(s) $\ln(y)$ and/or $\ln(x)$, and the coefficients can be estimated by OLS.
- Hypothesis tests and confidence intervals are now implemented and interpreted as usual.
- The interpretation of β_1 differs from case to case.
- The choice of specification (functional form) should be guided by judgement (which interpretation makes the most sense in your application?), tests, and plotting predicted values.

Textbook

- Non-linear regression functions: Chapter 8, Stock and Watson, Introduction to Econometrics, Global Edition, 4th edition