FIE401: Financial Econometrics

Non-linear Functional Form: Polynomials

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:

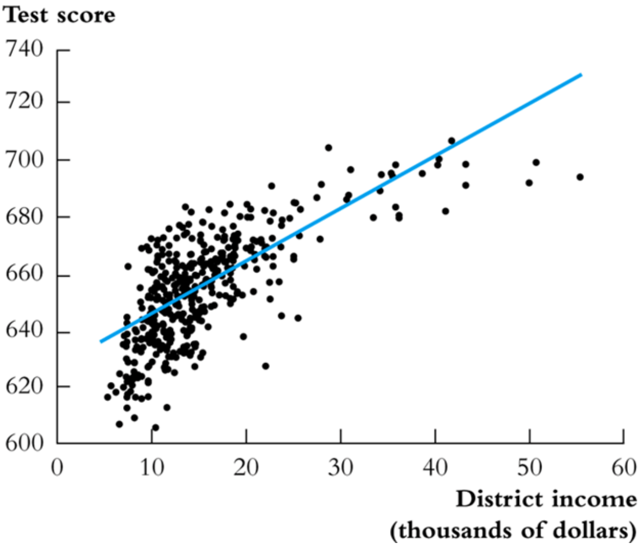
- The effect on y of a unit change in x depends on the value of x – that is, the marginal effect of x is not constant.
- A linear regression is misspecified: the functional form is wrong.
- The estimator of the effect on y of x is biased.
- The solution is to estimate a regression function that is non-linear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

Example: Test Scores

Is the relationship between test scores and independent variables linear?

- testscr: test scores
- avginc: average family income

Example: Test Scores and Income



Polynomials in x

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_r x^r + u$$

- This is just the linear multiple regression model – except that the regressors are powers of x!
- Estimation and hypotheses testing proceeds as in the multiple regression model using OLS.
- The coefficients are difficult to interpret, but the regression function itself is interpretable.

Example: Test Scores and Income

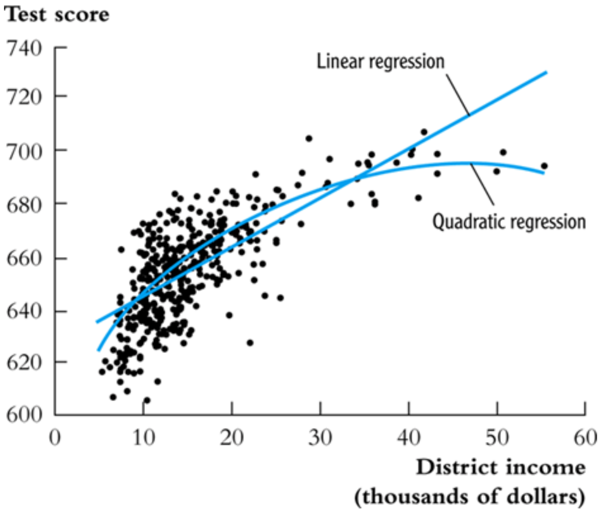
$$\widehat{testcr} = 607.30 + 3.85avginc - 0.04avginc^2$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#run quadratic regression
fit<-lm(testscr~avginc+I(avginc^2),
         data=School);
#summary of quadratic 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) 607.3017350   2.9242231 207.6797 < 2.2e-16 ***
## avginc        3.8509947   0.2711044  14.2048 < 2.2e-16 ***
## I(avginc^2)  -0.0423085   0.0048809  -8.6681 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Test Scores and Income



Example: Test Scores and Income

$$\widehat{testcr} = 607.30 + 3.85 avginc - 0.04 avginc^2$$

Marginal effect of additional USD 1,000 of income:
$\triangle testscr = 3.85 - 2 \times 0.04 \times avginc$

- $avginc = 5$ and $\triangle avginc = 1 => \triangle testscr = 3.45$
- $avginc = 20$ and $\triangle avginc = 1 => \triangle testscr = 2.25$
- $avginc = 50$ and $\triangle avginc = 1 => \triangle testscr = -0.15$

Example: Test Scores and Income

$$\widehat{testcr} = 607.30 + 3.85avginc - 0.04avginc^2$$

- Test scores maximum with respect to the income:

$$\tfrac{3.85}{2\times 0.04} \approx 48.1$$

- Does this mean test score decrease after income passing 48.1?
- Not necessarily. It depends on how many observations in the sample lie right to the turnaround point.
- In the given example, these are 3 observation or about 0.7% of the observations.
- Note: do not extrapolate outside the range of the data!

Example: Test Scores and Income

$$\widehat{testcr} = 600.10 + 5.02 avginc - 0.10 avginc^2 + 0.0007 avginc^3$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#run cubic regression
fit<-lm(testscr~avginc+I(avginc^2)+I(avginc^3),
         data=School);
#summary of the cubic 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)  6.0008e+02  5.4623e+00 109.8581 < 2.2e-16 ***
## avginc       5.0187e+00  7.8729e-01   6.3746 4.878e-10 ***
## I(avginc^2) -9.5805e-02  3.4052e-02  -2.8135  0.005133 **
## I(avginc^3)  6.8548e-04  4.3692e-04   1.5689  0.117430
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Test Scores and Income

$$\widehat{testcr} = 600.10 + 5.02 avginc - 0.10 avginc^2 + 0.0007 avginc^3$$

$$F = 29.678 \text{ and } p\text{-value}=8.944\text{e-}13$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#run cubic regression
fit<-lm(testscr~avginc+I(avginc^2)+I(avginc^3),
         data=School);
#define H0
myH0<-c("I(avginc^2)=0","I(avginc^3)=0");
#F-test
#robust to heteroskedasticity
linearHypothesis(fit,myH0,vcov=hccm);
```

Example: Test Scores and Income

```
#define H0
myH0<-c("I(avginc^2)=0","I(avginc^3)=0");
#F-test
#robust to heteroskedasticity
linearHypothesis(fit,myH0,vcov=hccm);
```

```
## Linear hypothesis test
##
## Hypothesis:
## I(avginc^2) = 0
## I(avginc^3) = 0
##
## Model 1: restricted model
## Model 2: testscr ~ avginc + I(avginc^2) + I(avginc^3)
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1    418
## 2    416  2 29.678 8.944e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Summary for Polynomials in x

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + ... + \beta_r x^r + u$$

- Estimation: by OLS after defining new regressors.
- Coefficients have complicated interpretations.
- To interpret the estimated regression function:
  - plot predicted values as a function of x.
  - compute predicted $\frac{\triangle y}{\triangle x}$ for different values of x.
- Hypotheses concerning degree r can be tested by t- and F-tests on the appropriate (blocks of) variable(s).

Textbook

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