FIE401: Financial Econometrics

Non-linear Functional Form: Interactions

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

Are test scores comlpeletely determined by class size?

Is the relationship between test scores and independent variables linear?

- testscr: test scores
- str: student-to-teacher ratio
- el_pct: proportion of English learners

Interactions Between Independent Variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2 + u$$

- Interaction between two binary variables.
- Interaction between binary and continuous variable.
- Interaction between two continuous variables.

Interactions Between Two Binary Variables

$$y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + u$$

- D_1 and D_2 are binary variables (dummy variables).
- β_1 is the effect of changing $D_1 = 0$ to $D_1 = 1$.
- In this specification, this effect does not depend on the value of D_2.

To allow the effect of changing D_1 to depend on D_2, include the "interaction term" $D_1 \times D_2$ as a regressor:

$$y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_1 \times D_2 + u$$

Interpreting Coefficients

$$y = \beta_0 + \beta_1 D_1 + \beta_2 D_2 + \beta_3 D_1 \times D_2 + u$$

- Compare the various combinations of values D_1 and D_2.
- $E(y|D_1 = 0, D_2 = 0) = \beta_0$.
- $E(y|D_1 = 0, D_2 = 1) = \beta_0 + \beta_2$.
- $E(y|D_1 = 1, D_2 = 0) = \beta_0 + \beta_1$.
- $E(y|D_1 = 1, D_2 = 1) = \beta_0 + \beta_1 + \beta_2 + \beta_3$.
- β_3 is an incremental effect of changing D_1 from 0 to 1 when $D_2 = 1$.

Example: Test Scores and Class Size

$$\widehat{testscr} = 664.14 - 1.91 large\_class - 18.16 high\_el - 3.49 large\_class \times high\_el$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#define dummies
School$large_class<-(School$str>=20)*1;
School$high_el<-(School$el_pct>=10)*1;
#run regression
fit<-lm(testscr~large_class+high_el+high_el*large_class,
        data=School);
#summary of the regession
#robust to heteroskedasticity
coeftest(fit,vcov=hccm);
```

Example: Test Scores and Class Size

```
##
## t test of coefficients:
##
##                       Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)           664.1433     1.3908 477.5271 < 2.2e-16 ***
## large_class            -1.9078     1.9416  -0.9826    0.3264
## high_el               -18.1629     2.3575  -7.7042 9.739e-14 ***
## large_class:high_el    -3.4943     3.1384  -1.1134    0.2662
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Test Scores and Class Size

- Generate cross-tabulation for large class and high percentage of English learners.

```
#load necessary packages
require(dplyr);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#define small class
School$large_class<-(School$str>=20)*1;
School$high_el<-(School$el_pct>=10)*1;
Summary<-School %>%
        group_by(large_class,high_el) %>%
        summarise(Mean = mean(testscr, na.rm=TRUE));
print(Summary);
```

Example: Test Scores and Class Size

```
## # A tibble: 4 x 3
## # Groups:   large_class [2]
##   large_class high_el  Mean
##         <dbl>   <dbl> <dbl>
## 1           0       0  664.
## 2           0       1  646.
## 3           1       0  662.
## 4           1       1  641.
```

Interactions Between Binary and Continuous Variables

$$y = \beta_0 + \beta_1 x + \beta_2 D + u$$

- D is binary variables (dummy variables).
- x is continuous variable.
- β_1 is the effect on y of a unit increase in x.
- In this specification, this effect does not depend on the value of D.

To allow the effect of changing x to depend on D, include the "interaction term" $D \times x$ as a regressor:

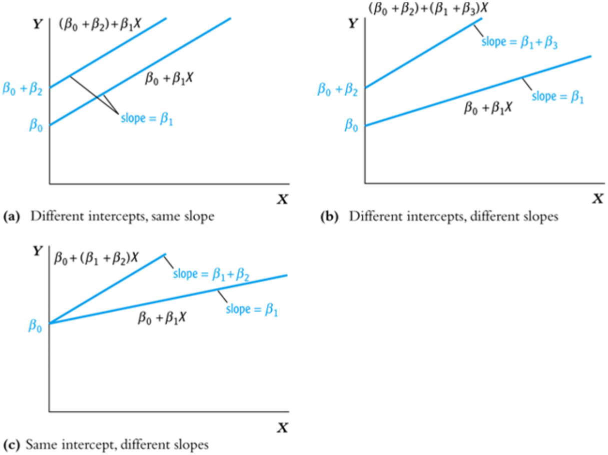
$$y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 D \times x + u$$

Binary and Continuous Interactions: Two Regression Lines

$$y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 D \times x + u$$

- When $D_i = 0$: $y = \beta_0 + \beta_1 x + u$.
- When $D_i = 1$: $y = (\beta_0 + \beta_2) + (\beta_1 + \beta_3)x + u$.
- By playing with specification of the regression equation we can have regression lines that are different:
  - Both in slopes and intercepts (current specification).
  - In slopes only (drop $\beta_2 D$ from regression equation).
  - In intercepts only (drop $\beta_3 D \times x$ from regression equation).

Binary and Continuous Interactions



Interpreting Coefficients

$$y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 D \times x + u$$

- Compute y before and after change in x.
- Before: $y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 D \times x$.
- After: $y + \triangle y = \beta_0 + \beta_1 (x + \triangle x) + \beta_2 D + \beta_3 D \times (x + \triangle x)$.
- Substract "before" from "after".
- $\triangle y = \beta_1 \triangle x + \beta_3 D \times \triangle x$.
- $\frac{\triangle y}{\triangle x} = \beta_1 + \beta_3 D$.
- The effect of x depends on D.
- $\beta_3 =$ increment to the effect of x, when $D = 1$.

Hypothesis Tests

$$y = \beta_0 + \beta_1 x + \beta_2 D + \beta_3 D \times x + u$$

- The two regression lines have the same slope H0: $\beta_3 = 0$.
- The two regression lines have the same intercept H0: $\beta_2 = 0$.
- The two regression lines are the same H0: $\beta_2 = \beta_3 = 0$.

Example: Test Scores and Class Size

$$\widehat{testscr} = 682.25 - 0.97 str + 5.64 high\_el - 1.28 str \times high\_el$$

```r
#load necessary packages
require(car);
require(lmtest);
#load data
load("M:/Projects/FIE401/data/data_lectures/School.Rdata");
#define dummies
School$high_el<-(School$el_pct>=10)*1;
#run regression
fit<-lm(testscr~str+high_el+high_el*str,
        data=School);
#summary of the regession
#robust to heteroskedasticity
coeftest(fit,vcov=hccm);
```

Example: Test Scores and Class Size

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 682.24584   12.07126 56.5182   <2e-16 ***
## str          -0.96846    0.59943 -1.6156   0.1069
## high_el       5.63914   19.88866  0.2835   0.7769
## str:high_el  -1.27661    0.98557 -1.2953   0.1959
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example: Test Scores and Class Size

- Regression for high and low percentage of English learners have the same slope H0: $\beta_3 = 0$ ($high\_el \times str$).
  - Cannot be rejected.
- Regression for high and low percentage of English learners have the same intercept H0: $\beta_2 = 0$ ($high\_el$).
  - Cannot be rejected.
- Regression for for high and low percentage of English learners are the same H0: $\beta_2 = \beta_3 = 0$ ($high\_el \times str$ and $high\_el$).

```r
myH0<-c("high_el=0","str:high_el=0");
linearHypothesis(fit,myH0,vcov=hccm);
```

  - p-value$<$ 2.2e-16
  - Rejected.

Interactions Between Two Continuous Variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

- x_1 and x_2 are continuous variables.
- As specified, the effect of x_1 does not depend on x_2.
- As specified, the effect of x_2 does not depend on x_1.
- To allow the effect of x_1 to depend on x_2, include the "interaction term" $x_1 \times x_2$ as a regressor:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2 + u$$

Interpreting Coefficients

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2 + u$$

- Compute y before and after change in x_1.
- Before: $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 \times x_2$.
- After:
  $y + \triangle y = \beta_0 + \beta_1 (x_1 + \triangle x_1) + \beta_2 x_2 + \beta_3 (x_1 + \triangle x_1) \times x_2$.
- Substract "before" from "after".
- $\triangle y = \beta_1 \triangle x_1 + \beta_3 \triangle x_1 \times x_2$.
- $\frac{\triangle y}{\triangle x} = \beta_1 + \beta_3 x_2$.
- The effect of x_1 depends on x_2.
- $\beta_3 =$ increment to the effect of x_1 from a unit change in x_2.
- Rarely used in practice due to complicated interpretation.

Summary Interactions

- Three cases depending on whether we consider interactions between two binary variables, between one binary variable and one continuous varaible, and between two continuous variables.
- Estimation and inference proceed in the same way as in the linear multiple regression model.
- The interpretation of coefficients differs from case to case.
- Interpretation of the coefficients is model-specific, but the general rule is to compute effects by comparing "before" and "after" cases.
- Many non-linear specifications are possible, so you must use judgment:
  - What non-linear effect you want to analyze?
  - What makes sense in your application?

Textbook

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