

ECON-320-Lab-8

Sonan Memon

- While the OLS is a linear model, it is not linear in the extremely restrictive sense. It has the capacity to incorporate many non-linear relationships.
- More precisely, OLS is linear in terms of the impact of x on y but x itself can be a non-linear variable such as square or an interaction-term.
- Loops are fundamental building blocks for any programming exercise; they allow for automation and efficiency when there exists a recursion in the model.

Packages

```
library(tinytex)
```

```
library(tidyverse)
```

```
library(dslabs)
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
library(tibble)
```

```
library(broom)
```

```
library(modelsummary)
```

- While the OLS is a linear model, it is not linear in the extremely restrictive sense. It has the capacity to incorporate many non-linear relationships.
- More precisely, OLS is linear in terms of the impact of x on y but x itself can be a non-linear variable such as square or an interaction-term.
- Loops are fundamental building blocks for any programming exercise; they allow for automation and efficiency when there exists a recursion in the model.

Data Cleaning

```
data <- read.csv("nlsy79-2020.csv")

data <- data %>%
  rename(
    logWealth = log_wealth,
    logEduc = log_educ
  )

View(data)
```

Non-Linear Relations

- Some common form of non-linear relationships in economics are linear, log-liner, linear-log and log-log models.

```
linear = lm(data = data, wealth ~ educ)

model1 <- modelsummary(
  list("Linear Model" = linear),
  stars = TRUE,
  statistic = "std.error",
  output = "tinytable",
  title = "Results From NLS-1979",
  gof_omit = ".*" # remove AIC, BIC, etc.
)
```

Non-Linear Relations

model1

Table 1: Results From NLS-1979

| | Linear Model |
|-------------|---------------------------------|
| (Intercept) | -2837577.048*** (236970.310) |
| educ | 270430.004*** (16790.823) |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Non-Linear Relations

- Log-Linear Case:

```
log_lin = lm(data = data, logWealth ~ educ)

model2 <- modelsummary(
  list("Log-Linear" = log_lin),
  stars = TRUE,
  statistic = "std.error",
  output = "tinytable",
  title = "Results From NLS-1979",
  gof_omit = ".*" # remove AIC, BIC, etc.
)
```


Non-Linear Relations

Table 2: Results From
NLS-1979

| | Log-Linear |
|-------------|---------------------|
| (Intercept) | 7.213*** (0.180) |
| educ | 0.343*** (0.013) |

+ $p < 0.1$, * $p < 0.05$, **
 $p < 0.01$, *** $p < 0.001$

Non-Linear Relations

- Linear-Log Case:

```
lin_log = lm(data = data, wealth ~ logEduc)

model3<- modelsummary(
  list("Linear-Log" = lin_log),
  stars = TRUE,
  statistic = "std.error",
  output = "tinytable",
  title = "Results From NLS-1979",
  gof_omit = ".*" # remove AIC, BIC, etc.
)
```

Table 3: Results From NLS-1979

| | Linear-Log |
|-------------|---------------------------------|
| (Intercept) | -8640087.816*** (617063.317) |
| logEduc | 3655431.205*** (235443.611) |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Non-Linear Relations

- Log-Log case

```
log_log = lm(data = data, logWealth ~ logEduc)

model4 <- modelsummary(
  list("Log-Log" = log_log),
  stars = TRUE,
  statistic = "std.error",
  output = "tinytable",
  title = "Results From NLS-1979",
  gof_omit = ".*" # remove AIC, BIC, etc.
)
```

Table 4: Results From
NLS-1979

| | Log-Log |
|-------------|---------------------|
| (Intercept) | -0.538 (0.467) |
| logEduc | 4.787*** (0.178) |

+ $p < 0.1$, * $p < 0.05$,
** $p < 0.01$, *** $p < 0.001$

Function

```
fun = function(y, x){  
  
  df = tibble(y,x)  
  df = df %>%mutate(log_y = log(y),log_x = log(x))  
  
  lin = lm(data = df, y ~ x)%>%broom::tidy()  
  log_lin = lm(data = df, log_y ~ x)%>%broom::tidy()  
  lin_log = lm(data = df, y ~ log_x)%>%broom::tidy()  
  log_log = lm(data = df, log_y ~ log_x)%>%broom::tidy()  
  bind_rows(lin, log_lin, lin_log, log_log) %>%  
  filter(term == 'x' | term == 'log_x') %>%  
  mutate(interpret =c(estimate[1],estimate[2]*100, estimate[3]/100,  
  type = c("lin", "log_lin", "lin_log", "log_log"),  
  x_change = c("unit","unit","percent","percent"),  
  y_change = c("unit", "percent", "unit", "percent")  
  )  
}
```

Function: Check

```
results = fun(data$wealth, data$educ)
results[,6:9]
```

```
# A tibble: 4 x 4
```

| | interpret | type | x_change | y_change |
|---|-----------|---------|----------|----------|
| | <dbl> | <chr> | <chr> | <chr> |
| 1 | 270430. | lin | unit | unit |
| 2 | 34.3 | log_lin | unit | percent |
| 3 | 36554. | lin_log | percent | unit |
| 4 | 4.79 | log_log | percent | percent |

Quadratic Model

```
data <- data %>% mutate(EducSqr = educ^2)

quad = lm(data = data, wealth ~ educ + EducSqr)

modelsquare <- modelsummary(
  list("OLS with Quadratic Term" = quad),
  stars = TRUE,
  statistic = "std.error",
  output = "tinytable",
  title = "Results From NLS-1979",
  gof_omit = ".*" # remove AIC, BIC, etc.
)
```


Quadratic Model

Table 5: Results From NLS-1979

| OLS with Quadratic Term | |
|-------------------------|-----------------------------|
| (Intercept) | 70951.081 (1113209.272) |
| educ | -142317.210 (155269.011) |
| EducSqr | 14172.191** (5300.141) |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, ***
 $p < 0.001$

Quadratic Model: Interpretation

- $wealth = \beta_0 + \beta_1 * educ + \beta_2 * educ^2$
- $\frac{\partial wealth}{\partial educ} = \beta_1 + 2 * \beta_2 * educ$

Quadratic Model: Visualisation

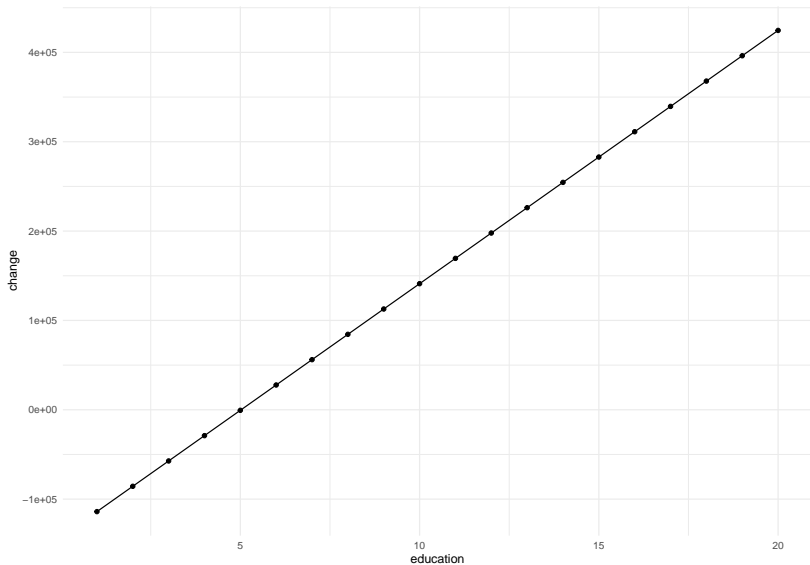
```
quad = lm(data = data, wealth ~ educ + EducSqr) %>% tidy()

education = seq(1, 20, 1)
change = c()

for(i in education){
  change[i] = quad$estimate[2] + 2*quad$estimate[3]*education[i]
}

df = tibble(education, change)
```

Quadratic Model: Visualisation



Interaction Terms

- When the effect of x on y varies based on whether z is high or low or other continuous or categorical variable, we use interaction terms.
- For instance, effect of education on income can vary across genders, racial groups, parent's socioeconomic status etc.
- Interaction terms with dummy variables have specially elegant interpretation.

Interaction Terms

```
int = lm(data = data, logWealth ~ logEduc +  
         race + logEduc:race)  
  
modelint <- modelsummary(  
  list("OLS with Interaction Term" = int),  
  stars = TRUE,  
  statistic = "std.error",  
  output = "tinytable",  
  title = "Results From NLS-1979",  
  gof_omit = ".*" # remove AIC, BIC, etc.  
)
```

Model with Interaction Term

- Effect of education on wealth is slightly dampened by race; in other words, African Americans benefit more from one additional year of education relative to white population.

Table 6: Results From NLS-1979

| OLS with Interaction Term | |
|---------------------------|---------------------|
| (Intercept) | -1.120 (0.830) |
| logEduc | 4.654*** (0.320) |
| race | 2.167* (0.983) |
| logEduc \times race | -0.309 (0.378) |

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$