BS2280 - Econometrics 1

Lecture 11 - Part 1: Identifying Nonlinearities and Multicollinearity

Dr. Yichen Zhu

Structure of today's lecture

1. Interactive Explanatory Variables
2. Ramsey's Test of Functional Misspecification

Intended Learning Outcomes

- Interpreting the coefficients of interactive explanatory variables
- Understanding Ramsey RESET test

Interactive Explanatory Variables

- Assume the following model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- This model is linear in parameter, so it can be estimated with OLS
- However, it is nonlinear in variables, so it has impact on the interpretation of the parameters.

Interactive Explanatory Variables

- Assume the following model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- This model is linear in parameter, so it can be estimated with OLS
- However, it is nonlinear in variables, so it has impact on the interpretation of the parameters.

Interactive Explanatory Variables

- Assume the following model:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- This model is linear in parameter, so it can be estimated with OLS
- However, it is nonlinear in variables, so it has impact on the interpretation of the parameters.

Interactive Explanatory Variables

- Until now the coefficients measured the marginal effect holding everything else constant, for example

$$\widehat{EARNINGS}_i = -14.668 + 1.877S_i + 0.983EXP_i$$

On average, every additional schooling year increases hourly earnings by \$1.877, controlling for the effects of other *EXP* variables

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- In this model, this interpretation is not possible.
- You cannot interpret β_2 as the effect of X_2 on Y, holding X_3 and $X_2 X_3$ constant!!!
- It is not possible to hold both X_3 and $X_2 X_3$ constant if X_2 changes.

Interactive Explanatory Variables

- Until now the coefficients measured the marginal effect holding everything else constant, for example

$$\widehat{EARNINGS}_i = -14.668 + 1.877 S_i + 0.983 EXP_i$$

On average, every additional schooling year increases hourly earnings by \$1.877, controlling for the effects of other *EXP* variables

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- In this model, this interpretation is not possible.
- You cannot interpret β_2 as the effect of X_2 on Y, holding X_3 and $X_2 X_3$ constant!!!
- It is not possible to hold both X_3 and $X_2 X_3$ constant if X_2 changes.

Interactive Explanatory Variables

- Until now the coefficients measured the marginal effect holding everything else constant, for example

$$\widehat{EARNINGS}_i = -14.668 + 1.877S_i + 0.983EXP_i$$

On average, every additional schooling year increases hourly earnings by \$1.877, controlling for the effects of other *EXP* variables

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- In this model, this interpretation is not possible.
- You cannot interpret β_2 as the effect of X_2 on Y, holding X_3 and X_2X_3 constant!!!
- It is not possible to hold both X_3 and X_2X_3 constant if X_2 changes.

Interactive Explanatory Variables

- Until now the coefficients measured the marginal effect holding everything else constant, for example

$$\widehat{EARNINGS}_i = -14.668 + 1.877S_i + 0.983EXP_i$$

On average, every additional schooling year increases hourly earnings by \$1.877, controlling for the effects of other *EXP* variables

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- In this model, this interpretation is not possible.
- You cannot interpret β_2 as the effect of X_2 on Y, holding X_3 and $X_2 X_3$ constant!!!
- It is not possible to hold both X_3 and $X_2 X_3$ constant if X_2 changes.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \quad (1)$$

- Rearrange (1) to:

$$Y_i = \beta_1 + \boxed{(\beta_2 + \beta_4 X_{3i}) X_{2i}} + \beta_3 X_{3i} + u_i$$

- Marginal effect of X_2 on Y is

$$\frac{dY_i}{dX_{2i}} = \beta_2 + \beta_4 X_{3i}$$

- The marginal effect of X_2 depends on X_3!
- β_4 is the change in the coefficient of X_2 when X_3 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \quad (1)$$

- Rearrange (1) to:

$$Y_i = \beta_1 + \boxed{(\beta_2 + \beta_4 X_{3i}) X_{2i}} + \beta_3 X_{3i} + u_i$$

- Marginal effect of X_2 on Y is

$$\frac{dY_i}{dX_{2i}} = \beta_2 + \beta_4 X_{3i}$$

- The marginal effect of X_2 depends on X_3!
- β_4 is the change in the coefficient of X_2 when X_3 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \tag{1}$$

- Rearrange (1) to:

$$Y_i = \beta_1 + (\beta_2 + \beta_4 X_{3i}) X_{2i} + \beta_3 X_{3i} + u_i$$

- Marginal effect of X_2 on Y is

$$\frac{dY_i}{dX_{2i}} = \beta_2 + \beta_4 X_{3i}$$

- The marginal effect of X_2 depends on X_3!
- β_4 is the change in the coefficient of X_2 when X_3 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \tag{1}$$

- Rearrange (1) to:

$$Y_i = \beta_1 + \boxed{(\beta_2 + \beta_4 X_{3i}) X_{2i}} + \beta_3 X_{3i} + u_i$$

- Marginal effect of X_2 on Y is

$$\frac{dY_i}{dX_{2i}} = \beta_2 + \beta_4 X_{3i}$$

- The marginal effect of X_2 depends on X_3!
- β_4 is the change in the coefficient of X_2 when X_3 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \tag{2}$$

- Rearrange (2) to:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \boxed{(\beta_3 + \beta_4 X_{2i}) X_{3i}} + u_i$$

- Marginal effect of X_3 on Y is

$$\frac{dY_i}{dX_{3i}} = \beta_3 + \beta_4 X_{2i}$$

- The marginal effect of X_3 depends on X_2!
- β_4 is the change in the coefficient of X_3 when X_2 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \tag{2}$$

- Rearrange (2) to:

$$Y_i = \beta_1 + \beta_2 X_{2i} + (\beta_3 + \beta_4 X_{2i}) X_{3i} + u_i$$

- Marginal effect of X_3 on Y is

$$\frac{dY_i}{dX_{3i}} = \beta_3 + \beta_4 X_{2i}$$

- The marginal effect of X_3 depends on X_2!
- β_4 is the change in the coefficient of X_3 when X_2 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \qquad (2)$$

- Rearrange (2) to:

$$Y_i = \beta_1 + \beta_2 X_{2i} + (\beta_3 + \beta_4 X_{2i}) X_{3i} + u_i$$

- Marginal effect of X_3 on Y is

$$\frac{dY_i}{dX_{3i}} = \beta_3 + \beta_4 X_{2i}$$

- The marginal effect of X_3 depends on X_2!
- β_4 is the change in the coefficient of X_3 when X_2 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i \tag{2}$$

- Rearrange (2) to:

$$Y_i = \beta_1 + \beta_2 X_{2i} + (\beta_3 + \beta_4 X_{2i}) X_{3i} + u_i$$

- Marginal effect of X_3 on Y is

$$\frac{dY_i}{dX_{3i}} = \beta_3 + \beta_4 X_{2i}$$

- The marginal effect of X_3 depends on X_2!
- β_4 is the change in the coefficient of X_3 when X_2 changes by one unit.

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- **Special case**: When X_2 and X_3 have value 0 in the data set:
- If $X_3 = 0$, β_2 is the marginal effect of X_2

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

$$\frac{dY_i}{dX_{2i}} = \beta_2$$

- If $X_2 = 0$, β_3 is the marginal effect of X_3

$$Y_i = \beta_1 + \beta_3 X_{3i} + u_i$$

$$\frac{dY_i}{dX_{3i}} = \beta_3$$

- This interpretation only makes sense when 0 is in the sample range of X_2 and X_3 which may not always be the case

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- **Special case**: When X_2 and X_3 have value 0 in the data set:
- If $X_3 = 0$, β_2 is the marginal effect of X_2

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

$$\frac{dY_i}{dX_{2i}} = \beta_2$$

- If $X_2 = 0$, β_3 is the marginal effect of X_3

$$Y_i = \beta_1 + \beta_3 X_{3i} + u_i$$

$$\frac{dY_i}{dX_{3i}} = \beta_3$$

- This interpretation only makes sense when 0 is in the sample range of X_2 and X_3 which may not always be the case

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- **Special case**: When X_2 and X_3 have value 0 in the data set:
- If $X_3 = 0$, β_2 is the marginal effect of X_2

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

$$\frac{dY_i}{dX_{2i}} = \beta_2$$

- If $X_2 = 0$, β_3 is the marginal effect of X_3

$$Y_i = \beta_1 + \beta_3 X_{3i} + u_i$$

$$\frac{dY_i}{dX_{3i}} = \beta_3$$

- This interpretation only makes sense when 0 is in the sample range of X_2 and X_3 which may not always be the case

Interactive Explanatory Variables: Interpretations

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{2i} X_{3i} + u_i$$

- **Special case**: When X_2 and X_3 have value 0 in the data set:
- If $X_3 = 0$, β_2 is the marginal effect of X_2

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i$$

$$\frac{dY_i}{dX_{2i}} = \beta_2$$

- If $X_2 = 0$, β_3 is the marginal effect of X_3

$$Y_i = \beta_1 + \beta_3 X_{3i} + u_i$$

$$\frac{dY_i}{dX_{3i}} = \beta_3$$

- This interpretation only makes sense when 0 is in the sample range of X_2 and X_3 which may not always be the case

Interactive Explanatory Variables: Example

- We estimate a semi-log model where we regress log of hourly earnings (*LGEARN*) in USD, on
  - years of education (S)
  - years of work experience (EXP)
  - and an interaction term of education and experience ($SEXP = S \times EXP$)
- The interaction term is the product of education and work experience

$$LGEARN_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + \beta_4 SEXP_i + u_i$$

Interactive Explanatory Variables: Example

- We estimate a semi-log model where we regress log of hourly earnings ($LGEARN$) in USD, on
  - years of education (S)
  - years of work experience (EXP)
  - and an interaction term of education and experience ($SEXP = S \times EXP$)
- The interaction term is the product of education and work experience

$$LGEARN_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + \beta_4 SEXP_i + u_i$$

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = \hat{\beta}_1 + \hat{\beta}_2 S_i + \hat{\beta}_3 EXP_i + \hat{\beta}_4 SEXP_i$$

```
Call:
lm(formula = LGEARN ~ S + EXP + SEXP, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.308507   0.333209   3.927 9.82e-05 ***
S           0.084342   0.020859   4.043 6.11e-05 ***
EXP         0.023414   0.043323   0.540    0.589
SEXP        0.001218   0.003002   0.406    0.685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5146 on 496 degrees of freedom
Multiple R-squared:  0.1393,  Adjusted R-squared:  0.1341
F-statistic: 26.75 on 3 and 496 DF,  p-value: 4.713e-16
```

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- If an individual has no work experience ($EXP = 0$), an additional year of education will lead to an increase in earnings by $100 \times 0.084 = 8.4\%$, ceteris paribus.
- If an individual has no education ($S = 0$), an additional year of work experience will lead to and increase in earnings by $100 \times 0.023 = 2.3\%$, ceteris paribus.
- However, especially no education is very unlikely!!!!

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = \hat{\beta}_1 + \hat{\beta}_2 S_i + \hat{\beta}_3 EXP_i + \hat{\beta}_4 SEXP_i$$

```
Call:
lm(formula = LGEARN ~ S + EXP + SEXP, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.308507   0.333209   3.927 9.82e-05 ***
S           0.084342   0.020859   4.043 6.11e-05 ***
EXP         0.023414   0.043323   0.540    0.589
SEXP        0.001218   0.003002   0.406    0.685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5146 on 496 degrees of freedom
Multiple R-squared:  0.1393,  Adjusted R-squared:  0.1341
F-statistic: 26.75 on 3 and 496 DF,  p-value: 4.713e-16
```

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- If an individual has no work experience ($EXP = 0$), an additional year of education will lead to an increase in earnings by $100 \times 0.084 = 8.4\%$, ceteris paribus.
- If an individual has no education ($S = 0$), an additional year of work experience will lead to and increase in earnings by $100 \times 0.023 = 2.3\%$, ceteris paribus.
- However, especially no education is very unlikely!!!!

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = \hat{\beta}_1 + \hat{\beta}_2 S_i + \hat{\beta}_3 EXP_i + \hat{\beta}_4 SEXP_i$$

```
Call:
lm(formula = LGEARN ~ S + EXP + SEXP, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.308507   0.333209   3.927 9.82e-05 ***
S           0.084342   0.020859   4.043 6.11e-05 ***
EXP         0.023414   0.043323   0.540    0.589
SEXP        0.001218   0.003002   0.406    0.685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5146 on 496 degrees of freedom
Multiple R-squared:  0.1393,  Adjusted R-squared:  0.1341
F-statistic: 26.75 on 3 and 496 DF,  p-value: 4.713e-16
```

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- If an individual has no work experience ($EXP = 0$), an additional year of education will lead to an increase in earnings by $100 \times 0.084 = 8.4\%$, ceteris paribus.
- If an individual has no education ($S = 0$), an additional year of work experience will lead to and increase in earnings by $100 \times 0.023 = 2.3\%$, ceteris paribus.
- However, especially no education is very unlikely!!!!

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = \hat{\beta}_1 + \hat{\beta}_2 S_i + \hat{\beta}_3 EXP_i + \hat{\beta}_4 SEXP_i$$

```
Call:
lm(formula = LGEARN ~ S + EXP + SEXP, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.308507   0.333209   3.927 9.82e-05 ***
S           0.084342   0.020859   4.043 6.11e-05 ***
EXP         0.023414   0.043323   0.540    0.589
SEXP        0.001218   0.003002   0.406    0.685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5146 on 496 degrees of freedom
Multiple R-squared:  0.1393,  Adjusted R-squared:  0.1341
F-statistic: 26.75 on 3 and 496 DF,  p-value: 4.713e-16
```

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i$$

- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i$$

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = \hat{\beta}_1 + \hat{\beta}_2 S_i + \hat{\beta}_3 EXP_i + \hat{\beta}_4 SEXP_i$$

```
Call:
lm(formula = LGEARN ~ S + EXP + SEXP, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 1.308507   0.333209   3.927 9.82e-05 ***
S           0.084342   0.020859   4.043 6.11e-05 ***
EXP         0.023414   0.043323   0.540    0.589
SEXP        0.001218   0.003002   0.406    0.685
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5146 on 496 degrees of freedom
Multiple R-squared:  0.1393,  Adjusted R-squared:  0.1341
F-statistic: 26.75 on 3 and 496 DF,  p-value: 4.713e-16
```

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i$$

- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i$$

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i$$

- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i$$

- To get a better idea of the marginal effects, it is common to calculate the marginal effects based on different levels of S and EXP
- If you only want to select one value, then it is common to select the mean values of EXP or S

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i$$

- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i$$

- To get a better idea of the marginal effects, it is common to calculate the marginal effects based on different levels of S and EXP
- If you only want to select one value, then it is common to select the mean values of EXP or S

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = 1.308 + 0.084 S_i + 0.023 EXP_i + 0.001 SEXP_i$$

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i$$

- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i$$

- To get a better idea of the marginal effects, it is common to calculate the marginal effects based on different levels of S and EXP
- If you only want to select one value, then it is common to select the mean values of EXP or S

Interactive Explanatory Variables: Example

$$\widehat{LGEARN}_i = 1.308 + 0.084S_i + 0.023EXP_i + 0.001SEXP_i$$

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001EXP_i$$

- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001S_i$$

- To get a better idea of the marginal effects, it is common to calculate the marginal effects based on different levels of S and EXP
- If you only want to select one value, then it is common to select the mean values of EXP or S

Interactive Explanatory Variables: Example

EAWE21.S	EAWE21.EXP
Min. : 8.00	Min. : 0.000
1st Qu.:12.00	1st Qu.: 4.356
Median :15.00	Median : 6.212
Mean :14.87	Mean : 6.445
3rd Qu.:17.00	3rd Qu.: 8.548
Max. :20.00	Max. :13.923

- Marginal effect of S

$$\frac{d\widehat{LGEARN_i}}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i = 0.084 + 0.001 \times 6.445 = 0.090$$

- For an individual with average years of work experience, an additional year of education will lead to an increase in earnings by 9%, ceteris paribus.
- Marginal effect of EXP

$$\frac{d\widehat{LGEARN_i}}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i = 0.023 + 0.001 \times 14.87 = 0.037$$

- For an individual with average years of education, an additional year of work experience will lead to an increase in earnings by 3.8%, ceteris paribus.

Interactive Explanatory Variables: Example

EAWE21.S	EAWE21.EXP
Min. : 8.00	Min. : 0.000
1st Qu.:12.00	1st Qu.: 4.356
Median :15.00	Median : 6.212
Mean :14.87	Mean : 6.445
3rd Qu.:17.00	3rd Qu.: 8.548
Max. :20.00	Max. :13.923

- Marginal effect of S

$$\frac{d\widehat{LGEARN_i}}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i = 0.084 + 0.001 \times 6.445 = 0.090$$

- For an individual with average years of work experience, an additional year of education will lead to an increase in earnings by 9%, ceteris paribus.
- Marginal effect of EXP

$$\frac{d\widehat{LGEARN_i}}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i = 0.023 + 0.001 \times 14.87 = 0.037$$

- For an individual with average years of education, an additional year of work experience will lead to an increase in earnings by 3.8%, ceteris paribus.

Interactive Explanatory Variables: Example

EAWE21.S	EAWE21.EXP
Min. : 8.00	Min. : 0.000
1st Qu.:12.00	1st Qu.: 4.356
Median :15.00	Median : 6.212
Mean :14.87	Mean : 6.445
3rd Qu.:17.00	3rd Qu.: 8.548
Max. :20.00	Max. :13.923

- Marginal effect of S

$$\frac{d\widehat{LGEARN_i}}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i = 0.084 + 0.001 \times 6.445 = 0.090$$

- For an individual with average years of work experience, an additional year of education will lead to an increase in earnings by 9%, ceteris paribus.
- Marginal effect of EXP

$$\frac{d\widehat{LGEARN_i}}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i = 0.023 + 0.001 \times 14.87 = 0.037$$

- For an individual with average years of education, an additional year of work experience will lead to an increase in earnings by 3.8%, ceteris paribus.

Interactive Explanatory Variables: Example

EAWE21.S	EAWE21.EXP
Min. : 8.00	Min. : 0.000
1st Qu.:12.00	1st Qu.: 4.356
Median :15.00	Median : 6.212
Mean :14.87	Mean : 6.445
3rd Qu.:17.00	3rd Qu.: 8.548
Max. :20.00	Max. :13.923

- Marginal effect of S

$$\frac{d\widehat{LGEARN}_i}{dS_i} = \hat{\beta}_2 + \hat{\beta}_4 EXP_i = 0.084 + 0.001 EXP_i = 0.084 + 0.001 \times 6.445 = 0.090$$

- For an individual with average years of work experience, an additional year of education will lead to an increase in earnings by 9%, ceteris paribus.
- Marginal effect of EXP

$$\frac{d\widehat{LGEARN}_i}{dEXP_i} = \hat{\beta}_3 + \hat{\beta}_4 S_i = 0.023 + 0.001 S_i = 0.023 + 0.001 \times 14.87 = 0.037$$

- For an individual with average years of education, an additional year of work experience will lead to an increase in earnings by 3.8%, ceteris paribus.

Student Task

- We estimate the following model:

$$bwght_i = \beta_1 + \beta_2 mage_i + \beta_3 cigs_i + \beta_4 magecigs_i + \delta male_i + u_i$$

- Where

bwght	birthweight of babies (grams)
mage	mother's age (years)
cigs	average number of cigarettes per day
magecigs	interaction term of *mage* and *cigs*, $mage \times cigs$
male	gender of the baby (1 if male and 0 if female)

Student Task

- Calculate the marginal effects of mother's age and average number of cigarettes smoked using the tables below

```
Call:
lm(formula = bwght ~ mage + cigs + magecigs + male, data = bwght2)
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 3270.4538    91.7708  35.637  < 2e-16 ***
mage           3.5305     3.0087   1.173 0.240790
cigs         -32.6623    19.3738  -1.686 0.091997 .
magecigs       0.7408     0.6567   1.128 0.259420
male          91.2886    27.3398   3.339 0.000859 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
bwght2.mage      bwght2.cigs
Min.   :16.00    Min.   : 0.000
1st Qu.:26.00    1st Qu.: 0.000
Median :29.00    Median : 0.000
Mean   :29.56    Mean   : 1.089
3rd Qu.:33.00    3rd Qu.: 0.000
Max.   :44.00    Max.   :40.000
```

Ramsey RESET Test

- **Question**: How do we know if we should use a nonlinear model
- Ramsey Regression Equation Specification Error Test (RESET) test is a general specification test for the linear regression model.
- More specifically, it tests whether nonlinear combinations of the explanatory variables help to explain the response variable.
- This test is a simple indicator of nonlinearity

Ramsey RESET Test

- **Question**: How do we know if we should use a nonlinear model
- Ramsey Regression Equation Specification Error Test (RESET) test is a general specification test for the linear regression model.
- More specifically, it tests whether nonlinear combinations of the explanatory variables help to explain the response variable.
- This test is a simple indicator of nonlinearity

Ramsey RESET Test

- **Question**: How do we know if we should use a nonlinear model
- Ramsey Regression Equation Specification Error Test (RESET) test is a general specification test for the linear regression model.
- More specifically, it tests whether nonlinear combinations of the explanatory variables help to explain the response variable.
- This test is a simple indicator of nonlinearity

Ramsey RESET Test

- **Question**: How do we know if we should use a nonlinear model
- Ramsey Regression Equation Specification Error Test (RESET) test is a general specification test for the linear regression model.
- More specifically, it tests whether nonlinear combinations of the explanatory variables help to explain the response variable.
- This test is a simple indicator of nonlinearity

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model
$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$
  2. Save the fitted values of the dependent variable $\hat{Y}_i$
$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki}$$
  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get
$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$

- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

  2. Save the fitted values of the dependent variable $\hat{Y}_i$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki}$$

  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$

- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

  2. Save the fitted values of the dependent variable $\hat{Y}_i$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki}$$

  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$

- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + u_i$$

  2. Save the fitted values of the dependent variable $\hat{Y}_i$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + ... + \hat{\beta}_k X_{ki}$$

  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + ... + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$

- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$

  2. Save the fitted values of the dependent variable $\hat{Y}_i$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki}$$

  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$

- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model
  $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$
  2. Save the fitted values of the dependent variable $\hat{Y}_i$
  $$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki}$$
  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get
  $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$
- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement

- To implement it, we need 4 steps
  1. Run the linear regression model
  $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + u_i$$
  2. Save the fitted values of the dependent variable $\hat{Y}_i$
  $$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \ldots + \hat{\beta}_k X_{ki}$$
  3. Square $\hat{Y}_i$ and get $\hat{Y}_i^2$
  4. Add $\hat{Y}_i^2$ to the original regression model and get
  $$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \ldots + \beta_k X_{ki} + \gamma \hat{Y}_i^2 + u_i$$
- γ should pick up any nonlinearities
- If γ is statistically significant, it indicates that some nonlinearity may be present

Ramsey RESET Test: Implement in R

- We will use again the example of

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Should we consider a nonlinear model for this example?

Step 1: Run the linear regression model	EARNfit <- lm(EARNINGS $\sim$ S+EXP,data=EAWE21)
Step 2: Save the fitted values of the dependent variable $\hat{EARNINGS}_i$	EAWE21\$FITTED <- predict(EARNfit)
Step 3: Square $\hat{EARNINGS}_i$ and get $\hat{EARNINGS}_i^2$	EAWE21\$FITTEDSQ <-EAWE21\$FITTED^2
Step 4: Add $\hat{Y}_i^2$ to the original regression model	RAMSEY <- lm(EARNINGS $\sim$ S+EXP+FITTEDSQ, data=EAWE21)

Ramsey RESET Test: Implement in R

- We will use again the example of

$$EARNINGS_i = \beta_1 + \beta_2 S_i + \beta_3 EXP_i + u_i$$

- Should we consider a nonlinear model for this example?

Step 1: Run the linear regression model	EARNfit <- lm(EARNINGS $\sim$ S+EXP,data=EAWE21)
Step 2: Save the fitted values of the dependent variable $\hat{EARNINGS}_i$	EAWE21$FITTED <- predict(EARNfit)
Step 3: Square $\hat{EARNINGS}_i$ and get $\hat{EARNINGS}_i^2$	EAWE21$FITTEDSQ <-EAWE21$FITTED^2
Step 4: Add $\hat{Y}_i^2$ to the original regression model	RAMSEY <- lm(EARNINGS $\sim$ S+EXP+FITTEDSQ, data=EAWE21)

Ramsey RESET Test: Implement in R

```
Call:
lm(formula = EARNINGS ~ S + EXP + FITTEDSQ, data = EAWE21)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 25.09320   17.79509   1.410   0.1591
S           -1.33416    1.41307  -0.944   0.3455
EXP         -0.64412    0.73731  -0.874   0.3828
FITTEDSQ     0.04608    0.02002   2.302   0.0218 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.08 on 496 degrees of freedom
Multiple R-squared:  0.1335,  Adjusted R-squared:  0.1282
F-statistic: 25.46 on 3 and 496 DF,  p-value: 2.453e-15
```

- For *FITTEDSQ*, p-value is $0.0218 < 5\%$
- The coefficient of *FITTEDSQ* is significant at the 5% level p-value is $0.0218 < 5\%$
- The addition of quadratic terms may improve the specification of the model.

Ramsey RESET Test: Implement in R

```
Call:
lm(formula = EARNINGS ~ S + EXP + FITTEDSQ, data = EAWE21)

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 25.09320   17.79509   1.410   0.1591
S           -1.33416    1.41307  -0.944   0.3455
EXP         -0.64412    0.73731  -0.874   0.3828
FITTEDSQ     0.04608    0.02002   2.302   0.0218 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.08 on 496 degrees of freedom
Multiple R-squared:  0.1335,  Adjusted R-squared:  0.1282
F-statistic: 25.46 on 3 and 496 DF,  p-value: 2.453e-15
```

- For *FITTEDSQ*, *p*-value is $0.0218 < 5\%$
- The coefficient of *FITTEDSQ* is significant at the 5% level *p*-value is $0.0218 < 5\%$
- The addition of quadratic terms may improve the specification of the model.

Ramsey RESET Test: Implement in R

```
Call:
lm(formula = EARNINGS ~ S + EXP + FITTEDSQ, data = EAWE21)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 25.09320   17.79509   1.410   0.1591
S           -1.33416    1.41307  -0.944   0.3455
EXP         -0.64412    0.73731  -0.874   0.3828
FITTEDSQ     0.04608    0.02002   2.302   0.0218 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 11.08 on 496 degrees of freedom
Multiple R-squared:  0.1335,  Adjusted R-squared:  0.1282
F-statistic: 25.46 on 3 and 496 DF,  p-value: 2.453e-15
```

- For *FITTEDSQ*, *p*-value is $0.0218 < 5\%$
- The coefficient of *FITTEDSQ* is significant at the 5% level *p*-value is $0.0218 < 5\%$
- The addition of quadratic terms may improve the specification of the model.

Notes on Ramsey RESET Test

- The RESET test detects nonlinearity
- Does not provide information on the most appropriate nonlinear model
- It may fail to detect some types of nonlinearity
- However, it is very easy to implement and consumes only one degree of freedom

Notes on Ramsey RESET Test

- The RESET test detects nonlinearity
- Does not provide information on the most appropriate nonlinear model
- It may fail to detect some types of nonlinearity
- However, it is very easy to implement and consumes only one degree of freedom

Notes on Ramsey RESET Test

- The RESET test detects nonlinearity
- Does not provide information on the most appropriate nonlinear model
- It may fail to detect some types of nonlinearity
- However, it is very easy to implement and consumes only one degree of freedom

Notes on Ramsey RESET Test

- The RESET test detects nonlinearity
- Does not provide information on the most appropriate nonlinear model
- It may fail to detect some types of nonlinearity
- However, it is very easy to implement and consumes only one degree of freedom