# Class 11 – Multiple Regression Model Estimation (Part I)


Pedram Jahangiry

JON M. HUNTSMAN SCHOOL OF BUSINESS
UtahStateUniversity

# Motivation for multiple regression

1. Incorporate more explanatory factors into the model
2. Explicitly hold fixed other factors that otherwise would be in $u$
3. Allow for more flexible functional forms

- Example: Wage equation

Now measures effect of education <u>explicitly holding experience fixed</u>

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + u$$

All other factors…

Hourly wage

Years of education

Years of labor market experience

# Definition of the multiple linear regression model

"Explains variable $y$ in terms of variables $x_1, x_2, \ldots, x_k$"

Intercept

Slope parameters

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$$

Dependent variable,
Explained variable,
Response variable,…

Independent variables,
Explanatory variables,
Regressors,…

Error term,
disturbance,
unobservables,…

# Interpretation of the multiple regression model

$$\beta_j = \frac{\Delta y}{\Delta x_j}$$

By how much does the dependent variable change if the j-th independent variable is increased by one unit,
holding all other independent variables and the error term constant

- ✓ The multiple linear regression model manages to hold the values of other explanatory variables fixed even if, in reality, they are correlated with the explanatory variable under consideration
- ✓ "Ceteris paribus"-interpretation
- ✓ It has still to be assumed that unobserved factors do not change if the explanatory variables are changed

# Example: Predict your college GPA!

$$\widehat{colGPA} = 1.29 + .453\, hsGPA + .0094\, ACT$$

Grade point average at college

High school grade point average

Achievement test score

## Interpretation

- ✓ Holding ACT fixed, another point on high school GPA is associated with another 0.453 points college GPA
- ✓ Or: If we compare two students with the same ACT, but the hsGPA of student A is one point higher, we predict student A to have a colGPA that is **0.453** higher than that of student B
- ✓ Holding high school grade point average fixed, another 10 points on ACT are associated with less than one point on college GPA

# Example: Average test scores and per student spending

$$avgscore = \beta_0 + \beta_1 expend + \beta_2 avginc + u$$

Other factors

Average standardized test score of school

Per student spending at this school

Average family income of students at this school

- ✓ Per student spending is likely to be correlated with average family income at a given high school because of school financing
- ✓ Omitting average family income in regression would lead to biased estimate of the effect of spending on average test scores
- ✓ In a **simple regression model**, effect of per student spending would **partly** include the effect of family income on test scores

# Example: Family income and family consumption

- Meaning of “linear” regression: The model has to be linear in the parameters (not in the variables)

$$cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u$$

Family consumption

Family income

Family income squared

Other factors

✓ Model has two explanatory variables: inome and income squared
✓ Consumption is explained as a quadratic function of income
✓ One has to be very careful when interpreting the coefficients:

By how much does consumption increase if income is increased by one unit?

$$\frac{\Delta cons}{\Delta inc} \approx \beta_1 + 2\beta_2 inc$$

Depends on how much income is already there

# Example: CEO salary, sales, and CEO tenure

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + \beta_2 ceoten + \beta_3 ceoten^2 + u$$

Log of CEO salary

Log sales

Quadratic function of CEO tenure with the firm

- Model assumes a constant elasticity relationship between CEO salary and the sales of his or her firm
- Model assumes a quadratic relationship between CEO salary and his or her tenure with the firm
- Remember! The model has to be linear in the parameters (not in the variables)

# OLS Estimation of the multiple regression model

- Random sample $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i) : \; i = 1, \ldots, n\}$

- Regression residuals $\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \ldots - \hat{\beta}_k x_{ik}$

- Minimize sum of squared residuals

$$\min \sum_{i=1}^{n} \hat{u}_i^2 \quad \rightarrow \quad \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_k$$

Minimization will be carried out by computer

# Properties of OLS on any sample of data

**Fitted values** and **residuals**

$$\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_{i1} + \widehat{\beta}_2 x_{i2} + \ldots + \widehat{\beta}_k x_{ik}$$

Fitted or predicted values

$$\widehat{u}_i = y_i - \widehat{y}_i$$

Residuals

Algebraic properties of OLS regression:

$$\sum_{i=1}^{n} \widehat{u}_i = 0$$

Deviations from regression line sum up to zero

$$\sum_{i=1}^{n} x_i \widehat{u}_i = 0$$

Covariance between deviations and regressors are zero

$$\bar{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \bar{x}_1 + \ldots + \widehat{\beta}_k \bar{x}_k$$

Sample averages of y and x lie on regression line

# MRM in R

```r
library(wooldridge)


MRM <- lm(wage ~ educ + exper , wage1)
summary(MRM)
```

```
Call:
lm(formula = wage ~ educ + exper, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max 
-5.5532 -1.9801 -0.7071  1.2030 15.8370 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.39054    0.76657  -4.423 1.18e-05 ***
educ         0.64427    0.05381  11.974  < 2e-16 ***
exper        0.07010    0.01098   6.385 3.78e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.257 on 523 degrees of freedom
Multiple R-squared:  0.2252,	Adjusted R-squared:  0.2222 
F-statistic: 75.99 on 2 and 523 DF,  p-value: < 2.2e-16
```

```r
library(wooldridge)
library(stargazer)


MRM <- lm(wage ~ educ + exper , wage1)
stargazer(MRM, type = "text")
```

```
===============================================
                        Dependent variable:    
                    ---------------------------
                               wage            
-----------------------------------------------
educ                         0.644***          
                              (0.054)          
                                               
exper                        0.070***          
                              (0.011)          
                                               
Constant                     -3.391***         
                              (0.767)          
                                               
-----------------------------------------------
Observations                    526            
R2                             0.225           
Adjusted R2                    0.222           
Residual Std. Error       3.257 (df = 523)     
F Statistic            75.990*** (df = 2; 523) 
===============================================
Note:               *p<0.1; **p<0.05; ***p<0.01
```