Multiple Regression Analysis: Estimation

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

BS1802 Statistics and Econometrics

Jiahua Wu

382 Business School
j.wu@imperial.ac.uk

Imperial College Business School

Imperial means Intelligent Business

Roadmap

- Regression analysis with cross-sectional data
  - The multiple regression analysis
    - Basics: estimation, inference, analysis with dummy variables
    - More technically involved: asymptotics, heteroskedasticity, specification and data issues
- Advanced topics
  - Limited dependent variable models
  - Panel data analysis
  - Regression analysis with time series data

Outline (Wooldridge, Ch. 2.4, 3.1 - 3.5)

- Definition of the multiple regression model
- Ordinary least squares (OLS) estimates
- Units of measurement and functional form
- Statistical properties of OLS estimators
- Reporting regression results

Outline

- Definition of the multiple regression model
- Ordinary least squares (OLS) estimates
- Units of measurement and functional form
- Statistical properties of OLS estimators
- Reporting regression results

Multiple Regression Model

- Definition

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

  - y: dependent variable (observable)
  - $x_1, \ldots, x_k$: independent variables (observable)
  - $\beta_1, \ldots, \beta_k$: slope parameters (to be estimated)
  - β_0: intercept parameter (to be estimated)
  - u: error term or disturbance (unobservable)
  - k: the number of independent variables

- The disturbance u represents factors other than x's
- With the intercept β_0, the unconditional mean of u can always be set to zero: $E(u) = 0$

Zero Conditional Mean Assumption

- Consider a simple regression model: $y = \beta_0 + \beta_1 x + u$
- If other factors in u are held fixed ($\Delta u = 0$), the causal effect of x on y is β_1:

$$\Delta y = \beta_1 \Delta x$$

- But under what condition u can be held fixed while x changes?
  - As x and u are treated as random variables, "u is fixed while x varying" is described as "the mean of u for any given x is the same (zero)"
  - The required condition is

$$E(u|x) = E(u) = 0,$$

    which is known as zero-conditional-mean (ZCM) assumption

- ZCM implies that x and u are uncorrelated, i.e., $Cov(x, u) = 0$

Zero Conditional Mean Assumption: An Example

- Eg. The wage model

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

- Suppose u represents inner ability, then ZCM assumption amounts to

$$Cov(educ, ability) = 0,$$

  i.e., a person's education level is uncorrelated with her/his ability

Zero Conditional Mean Assumption: An Example

- Eg. The wage model

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

- Suppose u represents inner ability, then ZCM assumption amounts to

$$Cov(educ, ability) = 0,$$

  i.e., a person's education level is uncorrelated with her/his ability

- This is NOT true if, on average, people with higher ability choose to become more educated

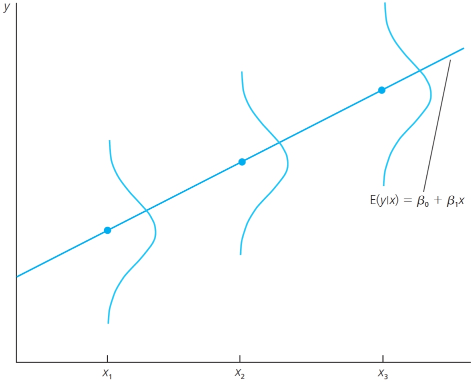
Zero Conditional Mean Assumption

- For the multiple regression model, the ZCM assumption is

$$E(u|x_1, \ldots, x_k) = 0$$

  - It requires the average of u to be the same irrespective of the values of x's
  - It implies that the factors in u are uncorrelated with $x_1, \ldots, x_k$
  - It is a key condition for the OLS estimators being unbiased
  - It defines the population regression function (PRF) $E(y|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

Population Regression Function



The distribution of $y = \beta_0 + \beta_1 x + u$ is centred around $E(y|x)$

- Systematic part of y: $E(y|x)$
- Unsystematic part of y: u

Outline

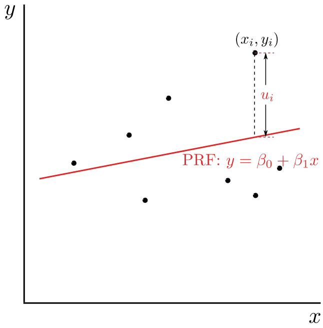
- Definition of the multiple regression model
- Ordinary least squares (OLS) estimates
- Units of measurement and functional form
- Statistical properties of OLS estimators
- Reporting regression results

Observations on $(x_1, \dots, x_k, y)$

- A random sample is a set of independent observations $\{(x_{i1}, x_{i2}, \dots, x_{ik}, y_i), i = 1, 2, \dots, n\}$
- At observation level, the model may be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i, \qquad i = 1, 2, \dots, n$$

  where i is the observation index
- For a simple regression model,



Estimate Multiple Regression

- The model

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i, \qquad i = 1, 2, \ldots, n$$

- Let $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ be estimates of $(\beta_0, \beta_1, \ldots, \beta_k)$
- Corresponding residual is

$$\hat{u}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}, \qquad i = 1, 2, \ldots, n$$

- The sum of squared residuals (SSR)

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik})^2$$

indicates the goodness of the estimates

- Good estimates should make SSR small

Ordinary Least Squares (OLS)

- The OLS estimates $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ minimize the SSR
- Detailed steps
  - The first order conditions lead to

$$\frac{\partial SSR}{\partial \hat{\beta}_0} = -2 \cdot \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) = 0$$

$$\frac{\partial SSR}{\partial \hat{\beta}_j} = -2 \cdot \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_k x_{ik}) x_{ij} = 0$$

  - Solve the system of $k+1$ equations

Ordinary Least Squares (OLS): Matrix Form

- At observation level, the model may be written as

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + u_i, \qquad i = 1, 2, \ldots, n$$

where i is the observation index

- Or collectively,

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & \cdots & x_{1,k} \\ 1 & x_{2,1} & \cdots & x_{2,k} \\ \vdots & \ddots & \vdots & \\ 1 & x_{n,1} & \cdots & x_{n,k} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

- Matrix notation: $y = X\beta + u$
- OLS estimates: $\hat{\beta} = (X'X)^{-1}X'y$

Sample Regression Function (SRF)

- Once OLS estimates are obtained,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}$$

  is the fitted value of y when $(x_1, \ldots, x_k) = (x_{i1}, \ldots, x_{ik})$

- The OLS regression line or sample regression function (SRF) is

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k,$$

  which is an estimate of the PRF

- "Run a regression of y on $x_1, \ldots, x_k$": use OLS to estimate the multiple regression model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

Interpretation of SRF

- The OLS regression line or SRF can be written in the form of changes:

$$\Delta\hat{y} = \hat{\beta}_1\Delta x_1 + \cdots + \hat{\beta}_k\Delta x_k$$

  - The coefficient on x_j is the partial effect of x_j on y holding other x's fixed: $\Delta\hat{y} = \hat{\beta}_j\Delta x_j$
  - $\hat{\beta}_j$ has a ceteris paribus interpretation when ZCM holds, i.e., factors in u are not correlated with x_j

PRF vs SRF

- The dependent variable y may be decomposed either as the sum of the SRF and the residual

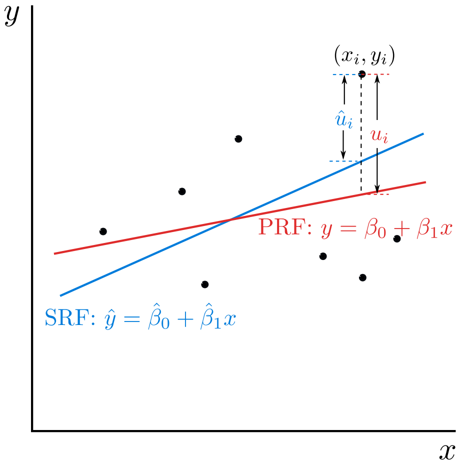
$$y = \hat{y} + \hat{u}$$

or as the sum of the PRF and the disturbance

$$y = E(y|x) + u$$

PRF vs SRF

- For a simple regression model,



- Hope: SRF = PRF "on average" or "when n goes to infinity"

Goodness-of-Fit

- How well does x explain y? Or how well does the OLS regression line fit data?
- We may use the fraction of variation in y that is explained by x's (or by the SRF) to measure the goodness-of-fit

Goodness-of-Fit: Sum of Squares

- Each y_i may be decomposed into $y_i = \hat{y}_i + \hat{u}_i$
- Measure variations from $\bar{y}$
  - Total sum of squares (total variation in y_i):

$$SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

  - Explained sum of squares (variation in $\hat{y}_i$):

$$SSE = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$$

  - Sum of squared Residuals (variation in $\hat{u}_i$):

$$SSR = \sum_{i=1}^{n} \hat{u}_i^2$$

  - It can be shown that $SST = SSE + SSR$

Goodness-of-Fit: R-Squared

- R-squared (coefficient of determination):

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

  - larger R^2, better fit
  - $0 \leq R^2 \leq 1$

OLS Example: Returns to Education

- Example 2.4.(wage1.RData) A sample of 526 people in the workforce. Regress *wage* on *educ*, and interpret the output.

OLS Example: Returns to Education

- Example 2.4.(wage1.RData) A sample of 526 people in the workforce. Regress *wage* on *educ*, and interpret the output.
    - OLS SRF: $n = 526$

$$\widehat{wage} = -0.90 + 0.54 educ$$

OLS Example: Returns to Education

- Example 2.4.(wage1.RData) A sample of 526 people in the workforce. Regress *wage* on *educ*, and interpret the output.
  - OLS SRF: $n = 526$

$$\widehat{wage} = -0.90 + 0.54 educ$$

  - Interpretation
    - Slope of 0.54: each additional year of schooling increases the wage by \$0.54
    - Intercept of -0.90: fitted wage of a person with $educ = 0$? SRF does poorly at low levels of education

OLS Example: Returns to Education

- Example 2.4.(wage1.RData) A sample of 526 people in the workforce. Regress *wage* on *educ*, and interpret the output.
  - OLS SRF: $n = 526$

$$\widehat{wage} = -0.90 + 0.54 educ$$

  - Interpretation
    - Slope of 0.54: each additional year of schooling increases the wage by $0.54
    - Intercept of -0.90: fitted wage of a person with $educ = 0$? SRF does poorly at low levels of education
  - Predicted wage for a person with $educ = 10$?

OLS Example: Returns to Education

- Example 2.4.(wage1.RData) A sample of 526 people in the workforce. Regress *wage* on *educ*, and interpret the output.
  - OLS SRF: $n = 526$

$$\widehat{wage} = -0.90 + 0.54educ$$

  - Interpretation
    - Slope of 0.54: each additional year of schooling increases the wage by $0.54
    - Intercept of -0.90: fitted wage of a person with $educ = 0$? SRF does poorly at low levels of education
  - Predicted wage for a person with $educ = 10$?
  - $R^2 = 0.165$ indicates that 16.5% of variation in wage is explained by *educ*

Outline

- Definition of the multiple regression model
- Ordinary least squares (OLS) estimates
- Units of measurement and functional form
- Statistical properties of OLS estimators
- Reporting regression results

Effects of Changing Units of Measurement

- How do OLS intercept and slope estimates change, if y is multiplied by a constant c?

Effects of Changing Units of Measurement

- How do OLS intercept and slope estimates change, if y is multiplied by a constant c?
- Example 2.4 (wage1.RData)
  - `> summary(lm(wage ~ educ, wage.data))`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
educ         0.54136    0.05325  10.167   <2e-16 ***
```

  - `> summary(lm(100 * wage ~ educ, wage.data))`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -90.485     68.497  -1.321    0.187
educ          54.136      5.325  10.167   <2e-16 ***
```

Effects of Changing Units of Measurement

- How do OLS intercept and slope estimates change, if y is multiplied by a constant c?
- Example 2.4 (wage1.RData)
  - `> summary(lm(wage ~ educ, wage.data))`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
educ         0.54136    0.05325  10.167   <2e-16 ***
```

  - `> summary(lm(100 * wage ~ educ, wage.data))`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -90.485     68.497  -1.321    0.187
educ          54.136      5.325  10.167   <2e-16 ***
```

- If y is multiplied by a constant c, all OLS intercept and slope estimates are also multiplied by c

Effects of Changing Units of Measurement

- How do OLS intercept and slope estimates change, if x_j is multiplied by a constant c?

Effects of Changing Units of Measurement

- How do OLS intercept and slope estimates change, if x_j is multiplied by a constant c?
- Example 2.4 (wage1.RData)
  - `> summary(lm(wage ~ educ, wage.data))`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
educ         0.54136    0.05325  10.167   <2e-16 ***
```

  - `> summary(lm(wage ~ I(100 * educ), wage.data))`

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.9048516  0.6849678  -1.321    0.187
I(100 * educ)  0.0054136  0.0005325  10.167   <2e-16 ***
```

Effects of Changing Units of Measurement

- How do OLS intercept and slope estimates change, if x_j is multiplied by a constant c?
- Example 2.4 (wage1.RData)
  - `> summary(lm(wage ~ educ, wage.data))`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.90485    0.68497  -1.321    0.187
educ         0.54136    0.05325  10.167   <2e-16 ***
```

  - `> summary(lm(wage ~ I(100 * educ), wage.data))`

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)   -0.9048516  0.6849678  -1.321    0.187
I(100 * educ)  0.0054136  0.0005325  10.167   <2e-16 ***
```

- If x_j is multiplied by a constant c, the slope estimate $\hat{\beta}_j$ is multiplied by $1/c$. All other OLS estimates remain unchanged

Effects of Changing Units of Measurement

- How does R^2 change when varying the units of measurement?

Effects of Changing Units of Measurement

- How does R^2 change when varying the units of measurement?
- Example 2.4 (wage1.RData)
  - `> summary(lm(wage ~ educ, wage.data))`

```
Residual standard error: 3.378 on 524 degrees of freedom
Multiple R-squared:  0.1648,	Adjusted R-squared:  0.1632 
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

  - `> summary(lm(100 * wage ~ educ, wage.data))`

```
Residual standard error: 337.8 on 524 degrees of freedom
Multiple R-squared:  0.1648,	Adjusted R-squared:  0.1632 
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

  - `> summary(lm(wage ~ I(100 * educ), wage.data))`

```
Residual standard error: 3.378 on 524 degrees of freedom
Multiple R-squared:  0.1648,	Adjusted R-squared:  0.1632 
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

Effects of Changing Units of Measurement

- How does R^2 change when varying the units of measurement?
- Example 2.4 (wage1.RData)
  - `> summary(lm(wage ~ educ, wage.data))`

```
Residual standard error: 3.378 on 524 degrees of freedom
Multiple R-squared:  0.1648,    Adjusted R-squared:  0.1632
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

  - `> summary(lm(100 * wage ~ educ, wage.data))`

```
Residual standard error: 337.8 on 524 degrees of freedom
Multiple R-squared:  0.1648,    Adjusted R-squared:  0.1632
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

  - `> summary(lm(wage ~ I(100 * educ), wage.data))`

```
Residual standard error: 3.378 on 524 degrees of freedom
Multiple R-squared:  0.1648,    Adjusted R-squared:  0.1632
F-statistic: 103.4 on 1 and 524 DF,  p-value: < 2.2e-16
```

- The R^2 does not change when varying units of measurement

Nonlinear Relationship between x and y

- The OLS only requires the regression model

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k + u$$

  to be linear in parameters $(\beta_0, \beta_1, \ldots, \beta_k)$

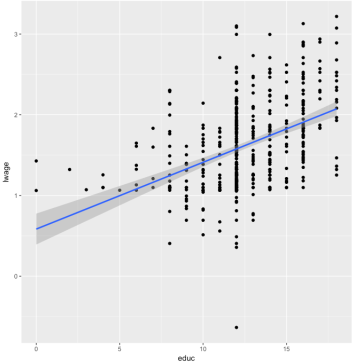
- Nonlinear relationships between y and x can be easily accommodated

Nonlinear Relationship between x and y: An Example

- Example 2.4 (wage1.RData) A better model to reflect the nonlinear relationship between *wage* and *educ* can be

$$\log(wage) = \beta_0 + \beta_1 educ + u.$$

-
```
> ggplot(data = wage.data, aes(x = educ, y = lwage)) +
    geom_point() + stat_smooth(method = "lm")
```



Nonlinear Relationship between x and y: An Example

- ```
  > linear.m1 <- lm(wage ~ educ, data = wage.data)
  > log.m1 <- lm(lwage ~ educ, data = wage.data)
  > stargazer(linear.m1, log.m1, align = TRUE)
  ```

	Dependent variable:	
	wage	lwage
	(1)	(2)
educ	0.541^{***}	0.083^{***}
	(0.053)	(0.008)
Constant	-0.905	0.584^{***}
	(0.685)	(0.097)
Observations	526	526
R^2	0.165	0.186
Adjusted R^2	0.163	0.184
Residual Std. Error (df = 524)	3.378	0.480
F Statistic (df = 1; 524)	103.363^{***}	119.582^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	
  ```

Nonlinear Relationship between x and y: An Example

- For two observations $(educ_1, wage_1)$ and $(educ_2, wage_2)$, we have

$$\begin{aligned} \log(wage_1) &= \beta_0 + \beta_1 educ_1 + u_1 & (1) \\ \log(wage_2) &= \beta_0 + \beta_1 educ_2 + u_2 & (2) \end{aligned}$$

- Subtracting (2) from (1),

$$\begin{aligned} & \log(wage_1) - \log(wage_2) = \beta_1(educ_1 - educ_2) + (u_1 - u_2) \\ \Rightarrow \quad & \Delta wage/wage_2 = \beta_1 \Delta educ + \Delta u \end{aligned}$$

- When $\Delta u = 0$, the percentage change in the wage is approximately $\%\Delta wage = (100)\beta_1 \Delta educ$.

Nonlinear Relationship between x and y

- Linear models are linear in parameters
- OLS applies to linear models no matter how x and y are defined
- But be careful about the interpretation of β

TABLE 2.3

Summary of Functional Forms Involving Logarithms

Model	Dependent Variable	Independent Variable	Interpretation of β_1
Level-level	y	x	$\Delta y = \beta_1 \Delta x$
Level-log	y	$\log(x)$	$\Delta y = (\beta_1/100)\%\Delta x$
Log-level	$\log(y)$	x	$\%\Delta y = (100\beta_1)\Delta x$
Log-log	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Example: Returns to Education

- Example 3.2.(wage1.RData) Use *educ*, *exper*, *tenure* (years with current employer) to explain hourly wage.

Example: Returns to Education

- Example 3.2.(wage1.RData) Use *educ*, *exper*, *tenure* (years with current employer) to explain hourly wage.
  OLS SRF:
  $$\widehat{\log(wage)} = .284 + .092educ + .004exper + .022tenure,$$
  where $n = 526$, $R^2 = 0.316$
  - How to interpret the coefficient on *educ*?
    - Holding *exper* and *tenure* fixed, an extra year of education is predicted to increase log(*wage*) by 0.092 (or 9.2% increase in wage)
  - Holding *educ* fixed, what is the effect of an individual staying at the same firm for an extra year on log(*wage*)?

Example: Returns to Education

- Example 3.2.(wage1.RData) Use *educ*, *exper*, *tenure* (years with current employer) to explain hourly wage.
  OLS SRF:

  $$\widehat{\log(wage)} = .284 + .092educ + .004exper + .022tenure,$$

  where $n = 526$, $R^2 = 0.316$

  - How to interpret the coefficient on *educ*?
    - Holding *exper* and *tenure* fixed, an extra year of education is predicted to increase log(*wage*) by 0.092 (or 9.2% increase in wage)
  - Holding *educ* fixed, what is the effect of an individual staying at the same firm for an extra year on log(*wage*)?

    $$\Delta\widehat{\log(wage)} = .004 + .022 = .026$$

  - $R^2 = 0.316$ indicates that 31.6% of variation in wage is explained by *educ*, *exper*, and *tenure*

Outline

- Definition of the multiple regression model
- Ordinary least squares (OLS) estimates
- Units of measurement and functional form
- Statistical properties of OLS estimators
- Reporting regression results

OLS Estimators

- A random sample, containing independent draws from the same population, is random
  - A data set is a realization of the random sample
- OLS "estimates" $(\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k)$ computed from a random sample is random, called the OLS estimators
- To make inference about the population parameters $(\beta_0, \beta_1, \ldots, \beta_k)$, we need to understand the statistical properties of the OLS estimators
- In particular, we like to know the means and variances of the OLS estimators

Unbiasedness of OLS Estimators

Theorem (3.1)

With a "good" model (where certain conditions are satisfied), the OLS estimators are unbiased, i.e., $E(\hat{\beta}_j) = \beta_j$, $j = 0, 1, \ldots, k$

Unbiased estimators $(\hat{\beta}_0, \ldots, \hat{\beta}_k)$

- They are "centred" around $(\beta_0, \ldots, \beta_k)$
- They correctly estimate $(\beta_0, \ldots, \beta_k)$ on average
- They will be "near" $(\beta_0, \ldots, \beta_k)$ for a "typical" sample

What if an irrelevant x is included?

- "Irrelevant" means the population coefficient of that variable is 0
- Eg. x_3 is irrelevant in the model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u.$$

The OLS estimators are unbiased: $E(\hat{\beta}_j) = \beta_j$, $j = 0, 1, \ldots, k$

  - In particular, $E(\hat{\beta}_3) = \beta_3 = 0$
  - The estimate of β_3 will "typically" be near 0

What if an irrelevant x is included?

- Example 3.2.(wage1.RData) Use *educ*, *exper*, *tenure* (years with current employer) to explain hourly wage. What if we include an irrelevant variable?

- ```
  > log.m2 <- lm(lwage ~ educ + exper + tenure, data = wage.
      data)
  > irr <- rnorm(nrow(wage.data), mean = 0, sd = 1)
  > log.m3 <- lm(lwage ~ educ + exper + tenure + irr, data =
      wage.data)
  > stargazer(log.m2, log.m3, align = TRUE, no.space = TRUE)
  ```
  ```

What if an irrelevant x is included?

	Dependent variable:	
	lwage	
	(1)	(2)
educ	0.092^{***}	0.092^{***}
	(0.007)	(0.007)
exper	0.004^{**}	0.004^{**}
	(0.002)	(0.002)
tenure	0.022^{***}	0.022^{***}
	(0.003)	(0.003)
irr		0.004
		(0.019)
Constant	0.284^{***}	0.285^{***}
	(0.104)	(0.104)
Observations	526	526
R^2	0.316	0.316
Adjusted R^2	0.312	0.311
Residual Std. Error	0.441 (df = 522)	0.441 (df = 521)
F Statistic	80.391^{***} (df = 3; 522)	60.193^{***} (df = 4; 521)
Note:		*p<0.1; **p<0.05; ***p<0.01

What if a relevant x is omitted?

- Eg. When x_2 is omitted from the true model

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

  - It becomes: $y = \beta_0 + \beta_1 x_1 + v$, with $v = \beta_2 x_2 + u$
  - The estimated model is $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$
  - It can be shown that OLS is biased: $E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}$, where $\beta_2 \tilde{\delta}$ is known as omitted variable bias and $\tilde{\delta}$ is the coefficient of regressing x_2 on x_1
  - The omitted variable bias is zero in two special cases
    - when $\beta_2 = 0$ or
    - when $\tilde{\delta} = 0$

Omitted Variable Bias

- The OLS estimators will generally be biased
- The direction and size of bias depend on how the omitted is related to the included

	$Corr(x_1, x_2) > 0$	$Corr(x_1, x_2) < 0$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- In practice, the knowledge about the signs of β_2 and $Corr(x_1, x_2)$ is useful for interpreting estimation result

Omitted Variable Bias: An Example

- Example 3.2.(wage1.RData) Use *educ*, *exper*, *tenure* (years with current employer) to explain hourly wage. What if we omit *exper* in the model?

- ```
  > log.m2 <- lm(lwage ~ educ + exper + tenure, data = wage.
      data)
  > log.m4 <- lm(lwage ~ educ + tenure, data = wage.data)
  > stargazer(log.m2, log.m4, align = TRUE, no.space = TRUE)
  ```
  ```

Omitted Variable Bias: An Example

	Dependent variable:	
	lwage	
	(1)	(2)
educ	0.092***	0.087***
	(0.007)	(0.007)
exper	0.004**	
	(0.002)	
tenure	0.022***	0.026***
	(0.003)	(0.003)
Constant	0.284***	0.404***
	(0.104)	(0.092)
Observations	526	526
R^2	0.316	0.309
Adjusted R^2	0.312	0.306
Residual Std. Error	0.441 (df = 522)	0.443 (df = 523)
F Statistic	80.391*** (df = 3; 522)	116.674*** (df = 2; 523)
Note:		*p<0.1; **p<0.05; ***p<0.01

Variance of OLS Estimators

Theorem (3.2)

With a "good" model, the variances of the OLS estimators are given by

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}, \qquad j = 1, \dots, k,$$

where $SST_j = \sum_{i=1}^{n}(x_{ij} - \bar{x}_j)^2$, $\bar{x}_j = n^{-1}\sum_{i=1}^{n} x_{ij}$, *and* R_j^2 *is the R-squared from regressing* x_j *on all other independent variables.*

- The larger σ^2, the greater $Var(\hat{\beta}_j)$
- The larger the variation in x_j, the smaller $Var(\hat{\beta}_j)$
- The larger R_j^2, the greater $Var(\hat{\beta}_j)$

Variance of OLS Estimators: Impact of σ^2

```
> x <- rnorm(100, mean = 1, sd = 2)
> y1 <- x + rnorm(100, mean = 1, sd = 2)
> y2 <- x + rnorm(100, mean = 1, sd = 8)
> m1 <- lm(y1 ~ x)
> m2 <- lm(y2 ~ x)
> stargazer(m1, m2, align = TRUE, no.space = TRUE)
```

Variance of OLS Estimators: Imapct of σ^2

	Dependent variable:	
	y1	y2
	(1)	(2)
x	0.948^{***}	0.947^{**}
	(0.107)	(0.427)
Constant	1.001^{***}	1.510
	(0.264)	(1.048)
Observations	100	100
R^2	0.442	0.048
Adjusted R^2	0.437	0.038
Residual Std. Error (df = 98)	2.158	8.582
F Statistic (df = 1; 98)	77.769^{***}	4.910^{**}
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$	

Variance of OLS Estimators: Imapct of SST_j

```
> x1 <- rnorm(100, mean = 1, sd = 1)
> y1 <- x1 + rnorm(100, mean = 1, sd = 8)
> x2 <- rnorm(100, mean = 1, sd = 8)
> y2 <- x2 + rnorm(100, mean = 1, sd = 8)
> m1 <- lm(y1 ~ x1)
> m2 <- lm(y2 ~ x2)
> stargazer(m1, m2, align = TRUE, no.space = TRUE)
```

Variance of OLS Estimators: Imapct of SST_j

	Dependent variable:	
	y1	y2
	(1)	(2)
x1	1.861^{**}	
	(0.711)	
x2		1.056^{***}
		(0.107)
Constant	0.016	0.906
	(1.040)	(0.818)
Observations	100	100
R^2	0.065	0.497
Adjusted R^2	0.056	0.492
Residual Std. Error (df = 98)	7.737	8.060
F Statistic (df = 1; 98)	6.851^{**}	96.798^{***}
Note:	*p<0.1; **p<0.05; ***p<0.01	

Variance of OLS Estimators: Impact of R_j^2

```
> x1 <- rnorm(100, mean = 1, sd = 2)
> x2 <- x1 + rnorm(100, mean = 1, sd = 1)
> x3 <- x1 + rnorm(100, mean = 1, sd = 8)
> y1 <- x1 + x2 + rnorm(100, mean = 1, sd = 8)
> y2 <- x1 + x3 + rnorm(100, mean = 1, sd = 8)
> m1 <- lm(y1 ~ x1 + x2)
> m2 <- lm(y2 ~ x1 + x3)
> stargazer(m1, m2, align = TRUE, no.space = TRUE)
```

Variance of OLS Estimators: Impact of R_j^2

	Dependent variable:	
	y1	y2
	(1)	(2)
x1	1.515*	1.069***
	(0.888)	(0.373)
x2	0.900	
	(0.802)	
x3		0.837***
		(0.092)
Constant	−0.439	2.039**
	(1.187)	(0.832)
Observations	100	100
R^2	0.281	0.539
Adjusted R^2	0.266	0.530
Residual Std. Error (df = 97)	7.958	7.195
F Statistic (df = 2; 97)	18.942***	56.783***
Note:	*p<0.1; **p<0.05; ***p<0.01	

Multicollinearity

- The larger R_j^2, the greater $Var(\hat{\beta}_j)$
  - R_j^2 is the R-squared from regressing x_j on all other x's
  - The more variation in x_j is explained by other x's, the larger is R_j^2
- High correlation between two or more independent variables is known as multicollinearity

Estimation of σ^2

- As the residual approximates u, the estimator of σ^2 is

$$\hat{\sigma}^2 = \frac{SSR}{n - (k+1)} = \frac{\sum_{i=1}^{n} \hat{u}_i^2}{n - (k+1)}$$

- The degrees of freedom (df) for the regression is $n - (k+1) = \#$ of observations - $\#$ of estimated coefficients
- $\hat{\sigma} = \sqrt{\hat{\sigma}^2}$ is known as the standard error of the regression

Outline

- Definition of the multiple regression model
- Ordinary least squares (OLS) estimates
- Units of measurement and functional form
- Statistical properties of OLS estimators
- Reporting regression results

Reporting Regression Results

- Good practice (minimum)
  - Report estimated coefficients AND standard errors
  - Report R-squared
  - Report sample size
  - Report in equation-form if the number of equations is small
    - Eg. Log wage model (standard errors are in brackets):

$$\widehat{\log(wage)} = \underset{(.104)}{.284} + \underset{(.007)}{.092}\, educ + \underset{(.0017)}{.0041}\, exper + \underset{(.003)}{.022}\, tenure$$

$$n = 526,\ R^2 = .316$$

  - Report in table-form

Reporting Regression Results

```
lm(formula = bwght ~ cigs, data = data)

Residuals:
    Min      1Q  Median      3Q     Max
-96.772 -11.772   0.297  13.228 151.228

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 119.77190    0.57234 209.267  < 2e-16 ***
cigs         -0.51377    0.090[illegible]  -5.678 1.[illegible]6e-08 ***
---
Signif. codes:  0 *** 0.001 ** [illegible]01 * 0.05 . 0.1   1

Residual standard error: 20.13 on [illegible] degrees of freedom
Multiple R-squared:  0.02273,  [illegible]jus[illegible] R-squared:  0.02202
F-statistic: 32.24 on 1 and 13[illegible] DF, [illegible]-value: 1.662e-08
```