Principle of Econometrics

Gianfranco Piras

THE SIMPLE REGRESSION MODEL: DEFINITION AND ESTIMATION

Definition of the Simple Regression Model I

- We begin with cross-sectional analysis and we assume to collect a **random sample** from the population of interest.
- There are two variables, x and y, and we would like to "**study how** y **varies with changes in** x."
- Some examples: x is amount of fertilizer, and y is soybean yield; x is years of schooling, y is hourly wage.

Definition of the Simple Regression Model II

- We must confront three issues:
  1. How do we allow factors **other than** x to affect y?
  2. What is the **functional relationship** between y and x?
  3. How can we be sure we are capturing a **ceteris paribus** relationship between y and x ?

Definition of the Simple Regression Model III

Consider the following equation relating y to x:

$$y = \beta_0 + \beta_1 x + u,$$

which is assumed to hold in the **population** of interest.

This equation is referred to as the **simple linear regression model**.

Definition of the Simple Regression Model IV

One important point to note:

We want to explain y in terms of x. For example:

- It makes no sense to "explain" **past educational attainment** in terms of **future labor earnings**.
- We want to explain **student performance** (y) in terms of **class size** (x), not the other way around.

Definition of the Simple Regression Model V

y	x
Dependent Variable	Independent Variable
Explained Variable	**Explanatory Variable**
Response Variable	Control Variable
Predicted Variable	Predictor Variable
Regressand	**Regressor**

Definition of the Simple Regression Model VI

- We mentioned the **error term** or **disturbance**, u, before. The equation

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

  explicitly allows for other factors, contained in u, to affect y.
- This equation also addresses the functional form issue: y is assumed to be **linearly** related to x.
- We call β_0 the **intercept parameter** and β_1 the **slope parameter**.
- The equation describes a **population**, and our ultimate goal is to **estimate** the parameters β_0 and β_1.

Definition of the Simple Regression Model VII

The equation also addresses the *ceteris paribus* issue:

$$y = \beta_0 + \beta_1 x + u,$$

all other factors that affect y are in u. We want to know how y changes when x changes, **holding** u **fixed**.

Definition of the Simple Regression Model VIII

Let Δ denote "change." Then holding u fixed means $\Delta u = 0$. So

$$\begin{aligned} \Delta y &= \beta_1 \Delta x + \Delta u \\ &= \beta_1 \Delta x \text{ when } \Delta u = 0. \end{aligned}$$

This equation effectively defines β_1 as a **slope**, with the restriction $\Delta u = 0$.

Definition of the Simple Regression Model IX

EXAMPLE: Yield and Fertilizer

A model to explain crop yield to fertilizer use is

$$yield = \beta_0 + \beta_1 fertilizer + u,$$

- u contains land quality, rainfall, etc.
- β_1 tells us how **yield changes** when the amount of **fertilizer changes**, holding all else fixed.
- The linear function is not realistic since the **effect of fertilizer is likely to diminish** at large amounts of fertilizer.

Definition of the Simple Regression Model X

EXAMPLE: Wage and Education

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

- where u contains, among other things, ability as well as past workforce experience and tenure on the current job.

$$\Delta wage = \beta_1 \Delta educ \text{ when } \Delta u = 0$$

- Is each year of education really worth the same dollar amount no matter how much education one starts with?

Definition of the Simple Regression Model XI

We mentioned three issues to address:

1. How do we allow factors other than x to affect y?
2. What is the functional relationship between y and x?
3. How can we assure to be capturing a *ceteris paribus* relationship between y and x?

We have argued that the simple regression model

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

addresses each of them.

Definition of the Simple Regression Model XII

This seems too easy!

- How can we hope to generally estimate the **ceteris paribus** effect of y on x when we have assumed all other factors affecting y are unobserved and merged into u?
- We have to make assumptions that **restricts** how u and x are related to each other in the population.

Definition of the Simple Regression Model XIII

- The average, or **expected value** of u is zero in the population:

$$E(u) = 0$$

- Normalizing "land quality," or "ability," to be zero in the population should be no problem!

Definition of the Simple Regression Model XIV

- The presence of β_0 in

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

  allows us to assume $E(u) = 0$.

  Assume that the average of u is different (e.g, $E(u) = \alpha_0$). Then we can write

$$y = (\beta_0 + \alpha_0) + \beta_1 x + (u - \alpha_0),$$

  where the new error, $u - \alpha_0$, has a zero mean.

- The new intercept is $\beta_0 + \alpha_0$. The important point is that the slope, β_1, has not changed.

Definition of the Simple Regression Model XV

KEY QUESTION:

How do we need to restrict the dependence between u and x?

- We could assume u and x **uncorrelated** in the population:

$$Corr(x, u) = 0$$

- This implies that u and x are not **linearly** related.
- Ruling out only linear dependence can cause problems with interpretation and makes statistical analysis more difficult.

Definition of the Simple Regression Model XVI

- A different assumption involves the mean of the error term for each portion of the population determined by different values of x:

$$E(u|x) = E(u), \text{ all values } x,$$

  where $E(u|x)$ means "**the expected value of** u **given** x."

- We say that u is **mean independent** of x.

Definition of the Simple Regression Model XVII

- Let u be "ability" and x years of schooling.

$$E(ability|x = 8) = E(ability|x = 12) = E(ability|x = 16)$$

so that the average ability is the same in the different portions of the population with an 8^{th} grade education, a 12^{th} grade education, and a four-year college education.

Definition of the Simple Regression Model XVIII

- Combining $E(u|x) = E(u)$ with $E(u) = 0$ gives

$$E(u|x) = 0, \text{ all values } x$$

- Called the **zero conditional mean assumption**.

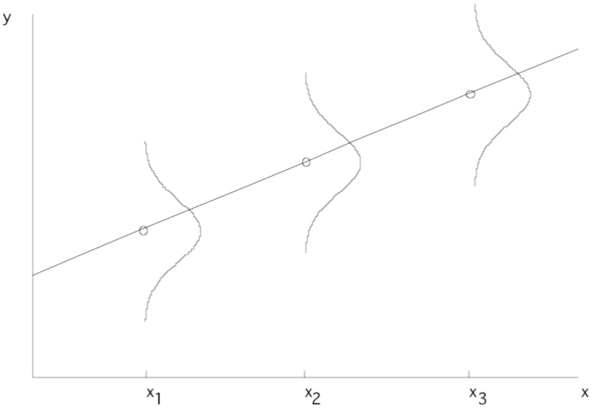
Definition of the Simple Regression Model XIX

- Keep in mind that the expected value is a linear operator, therefore $E(u|x) = 0$ implies

$$E(y|x) = \beta_0 + \beta_1 x + E(u|x) = \beta_0 + \beta_1 x,$$

  which shows the **population regression function** as a linear function of x.
- A one-unit increase in x changes the **expected value** of y by the amount β_1.

Definition of the Simple Regression Model XX



Definition of the Simple Regression Model XXI

- The straight line in the previous graph is the **Population regression function**, $E(y|x) = \beta_0 + \beta_1 x$.
- The figure also shows the **conditional distribution** of y at three different values of x.
- For a given value of x, we see a range of values for y: remember, $y = \beta_0 + \beta_1 x + u$, and u has a distribution in the population.

Definition of the Simple Regression Model XXII

EXAMPLE: College versus High School GPA.

Suppose for the population of students attending a university, we know

$$E(colGPA|hsGPA) = 1.5 + 0.5\ hsGPA,$$

If $hsGPA = 3.6$ then the *average* of $colGPA$ among students with this particular high school GPA is

$$1.5 + 0.5(3.6) = 3.3$$

Definition of the Simple Regression Model XXIII

- Anyone with $hsGPA = 3.6$ most likely will *not* have $colGPA = 3.3$. The value 3.3 is the *average* value of $colGPA$ within the portion of the population with $hsGPA = 3.6$.
- This shows that regression analysis is essentially trying to explain effects of explanatory variables on **average outcomes** of y.

Deriving the Ordinary Least Squares Estimates I

- Given data on x and y, **how can we estimate the population parameters**, β_0 and β_1?
- Let $\{(x_i, y_i) : i = 1, 2, ..., n\}$ be a sample of size n from the population.

Deriving the Ordinary Least Squares Estimates II

- We use the two restrictions

$$\begin{aligned} E(u) &= 0 \\ Cov(x, u) &= 0 \end{aligned}$$

- The first condition essentially defines the intercept.
- The second condition, stated in terms of the covariance, means that x and u are uncorrelated.

Deriving the Ordinary Least Squares Estimates III

- Remember that $Cov(x, u) = E(xu) - E(x)E(u)$.
- Since $E(u) = 0$ and $Cov(x, u) = 0$, then also $E(xu) = 0$

Deriving the Ordinary Least Squares Estimates IV

- Next we replace u in terms of the observable variables and the parameters

$$\begin{aligned} E(y - \beta_0 - \beta_1 x) &= 0 \\ E[x(y - \beta_0 - \beta_1 x)] &= 0 \end{aligned}$$

- These are the two conditions in the population that determine β_0 and β_1.

Deriving the Ordinary Least Squares Estimates V

For the **method of moments** approach to estimation, we use their sample analogs:

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\
n^{-1}\sum_{i=1}^{n}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0
\end{aligned}
$$

These are two linear equations in the two unknowns $\hat{\beta}_0$ and $\hat{\beta}_1$.

Deriving the Ordinary Least Squares Estimates VI

To solve the equations, pass the summation operator through the first equation:

$$
\begin{aligned}
n^{-1}\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= n^{-1}\sum_{i=1}^{n} y_i - n^{-1}\sum_{i=1}^{n}\hat{\beta}_0 - n^{-1}\sum_{i=1}^{n}\hat{\beta}_1 x_i \\
&= n^{-1}\sum_{i=1}^{n} y_i - \hat{\beta}_0 - \hat{\beta}_1\left(n^{-1}\sum_{i=1}^{n} x_i\right) \\
&= \bar{y} - \hat{\beta}_0 - \hat{\beta}_1\bar{x}
\end{aligned}
$$

Deriving the Ordinary Least Squares Estimates VII

We have shown that the first equation,

$$n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

implies

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Deriving the Ordinary Least Squares Estimates VIII

Rewrite this equation so that the intercept is in terms of the slope (and the sample averages):

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

and plug this into the second equation (and drop the division by n):

$$\sum_{i=1}^{n} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

so

$$\sum_{i=1}^{n} x_i [y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i] = 0$$

Deriving the Ordinary Least Squares Estimates IX

Simple algebra gives

$$\sum_{i=1}^{n} x_i(y_i - \bar{y}) = \hat{\beta}_1 \left[\sum_{i=1}^{n} x_i(x_i - \bar{x}) \right]$$

and so we have one linear equation in the one unknown $\hat{\beta}_1$

Deriving the Ordinary Least Squares Estimates X

Showing the solution for $\hat{\beta}_1$ uses three useful facts about the summation operator:

$$
\begin{aligned}
\sum_{i=1}^{n}(x_i - \bar{x}) &= 0 \\
\sum_{i=1}^{n} x_i(y_i - \bar{y}) &= \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{n}(x_i - \bar{x})y_i \\
\sum_{i=1}^{n} x_i(x_i - \bar{x}) &= \sum_{i=1}^{n}(x_i - \bar{x})^2
\end{aligned}
$$

Deriving the Ordinary Least Squares Estimates XI

We can write the equation to solve is

$$\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y}) = \hat{\beta}_1 \left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

If $\sum_{i=1}^{n}(x_i - \bar{x})^2 > 0$, we can write

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\text{Sample Covariance}(x_i, y_i)}{\text{Sample Variance}(x_i)}$$

Deriving the Ordinary Least Squares Estimates XII

- The previous formula for $\hat{\beta}_1$ is important! It shows us how to take the data we have and compute the slope estimate.
- $\hat{\beta}_1$ is called the **ordinary least squares (OLS)** slope estimate.
- It can be computed whenever the **sample variance** of the x_i is not zero, which only rules out the case where each x_i is the same value.

Deriving the Ordinary Least Squares Estimates XIII

Deriving the Ordinary Least Squares Estimates XIV

Once we have $\hat{\beta}_1$, we compute $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$. This is the OLS **intercept estimate**.

Deriving the Ordinary Least Squares Estimates XV

- Where does the name "ordinary least squares" come from?
- For any $\hat{\beta}_0$ and $\hat{\beta}_1$, we can define a **fitted value** as

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- The fitted values are the value we predict for y_i given that x has taken on the value x_i.
- The "mistake" we make is the **residual**:

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i,$$

Deriving the Ordinary Least Squares Estimates XVI

- Suppose we measure the size of the mistake, for each i, by squaring the residual: $\hat{u}_i^2$. Then we add them all up:

$$\sum_{i=1}^{n} \hat{u}_i^2 = \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

- This quantity is called the **sum of squared residuals**.
- If we choose $\hat{\beta}_0$ and $\hat{\beta}_1$ to *minimize* the sum of squared residuals it can be shown that the solutions are the slope and intercept estimates we obtained before!

Deriving the Ordinary Least Squares Estimates XVII

- Once we have the numbers $\hat{\beta}_0$ and $\hat{\beta}_1$ for a given data set, we write the **OLS regression line** as a function of x:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- The OLS regression line allows us to predict y for any value of x. It is also called the **sample regression function**.
- The intercept, $\hat{\beta}_0$, is the predicted y when $x = 0$.

Deriving the Ordinary Least Squares Estimates XVIII

- The slope, $\hat{\beta}_1$, allows us to predict changes in y for any change in x:

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x$$

- If $\Delta x = 1$, so that x increases by one unit, then $\Delta \hat{y} = \hat{\beta}_1$.

Deriving the Ordinary Least Squares Estimates XIX

EXAMPLE: Effects of Education on Hourly Wage (wage1)

- *wage* is reported in dollars per hour, *educ* is highest grade completed.
- The estimated equation is

$$\begin{aligned} \widehat{wage} &= -0.90 + 0.54\ educ \\ n &= 526 \end{aligned}$$

- The negative intercept says that *wage* is predicted to be $-\$0.90$ when $educ = 0$!
- Each additional year of schooling is estimated to be worth \$0.54.

Deriving the Ordinary Least Squares Estimates XX

```
R> library(wooldridge)
R> data("wage1")
```

Deriving the Ordinary Least Squares Estimates XXI

```
R> str(wage1)
'data.frame':        526 obs. of  24 variables:
 $ wage    : num  3.1 3.24 3 6 5.3 ...
 $ educ    : int  11 12 11 8 12 16 18 12 12 17 ...
 $ exper   : int  2 22 2 44 7 9 15 5 26 22 ...
 $ tenure  : int  0 2 0 28 2 8 7 3 4 21 ...
 $ nonwhite: int  0 0 0 0 0 0 0 0 0 0 ...
 $ female  : int  1 1 0 0 0 0 1 1 0 ...
 $ married : int  0 1 0 1 1 1 0 0 0 1 ...
 $ numdep  : int  2 3 2 0 1 0 0 0 2 0 ...
 $ smsa    : int  1 1 0 1 0 1 1 1 1 1 ...
 $ northcen: int  0 0 0 0 0 0 0 0 0 0 ...
 $ south   : int  0 0 0 0 0 0 0 0 0 0 ...
 $ west    : int  1 1 1 1 1 1 1 1 1 1 ...
```

Deriving the Ordinary Least Squares Estimates XXII

```
 $ construc: int  0 0 0 0 0 0 0 0 0 0 ...
 $ ndurman : int  0 0 0 0 0 0 0 0 0 0 ...
 $ trcommpu: int  0 0 0 0 0 0 0 0 0 0 ...
 $ trade   : int  0 0 1 0 0 0 1 0 1 0 ...
 $ services: int  0 1 0 0 0 0 0 0 0 0 ...
 $ profserv: int  0 0 0 0 0 1 0 0 0 0 ...
 $ profocc : int  0 0 0 0 0 1 1 1 1 1 ...
 $ clerocc : int  0 0 0 1 0 0 0 0 0 0 ...
 $ servocc : int  0 1 0 0 0 0 0 0 0 0 ...
 $ lwage   : num  1.13 1.18 1.1 1.79 1.67 ...
 $ expersq : int  4 484 4 1936 49 81 225 25 676 484 ...
 $ tenursq : int  0 4 0 784 4 64 49 9 16 441 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

Deriving the Ordinary Least Squares Estimates XXIII

```
R> summary(wage1)[,1:3]
      wage             educ            exper      
 Min.   : 0.530   Min.   : 0.00   Min.   : 1.00  
 1st Qu.: 3.330   1st Qu.:12.00   1st Qu.: 5.00  
 Median : 4.650   Median :12.00   Median :13.50  
 Mean   : 5.896   Mean   :12.56   Mean   :17.02  
 3rd Qu.: 6.880   3rd Qu.:14.00   3rd Qu.:26.00  
 Max.   :24.980   Max.   :18.00   Max.   :51.00  
```

Deriving the Ordinary Least Squares Estimates XXIV

```
R> mod1 <- lm(wage ~ educ, data = wage1)
R> summary(mod1)
Call:
lm(formula = wage ~ educ, data = wage1)

Residuals:
    Min      1Q  Median      3Q     Max
-5.3396 -2.1501 -0.9674  1.1921 16.6085

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 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

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
```

Deriving the Ordinary Least Squares Estimates XXV

- When we write the population regression line,

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

  we do not know β_0 and β_1.
- $\hat{\beta}_0 = -0.90$ and $\hat{\beta}_1 = 0.54$ are our *estimates* from this particular sample.
- These estimates may or may not be close to the population values.
- If we obtain another sample of the same size the estimates would almost certainly be different!

Deriving the Ordinary Least Squares Estimates XXVI

The function

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

is the OLS (or sample) regression line.

Plugging in $educ = 0$ gives the silly prediction $\widehat{wage} = -0.90$.

Extrapolating outside the range of the data can produce strange predictions.

Deriving the Ordinary Least Squares Estimates XXVII

```
R> wage1$educ[wage1$educ == 0]
[1] 0 0
R> length(which(wage1$educ == 0))
[1] 2
```

Deriving the Ordinary Least Squares Estimates XXVIII

When $educ = 8$,

$$\widehat{wage} = -0.90 + 0.54(8) = 3.42$$

The predicted hourly wage at eight years of education is \$3.42, which we can think of as our estimate of the average wage in the population when $educ = 8$. But may be no one in the sample earns exactly \$3.42!

```
R> coefficients(mod1)[1] + coefficients(mod1)[2] *8
(Intercept)
   3.426022
```

EXAMPLE: Effects of Campaign Spending on Voting Outcomes (vote1)

Two-party vote for the Congressional elections.

- *voteA* is the percentage of votes received by candidate A;
- *shareA* is the percentage share of campaign spending.

Deriving the Ordinary Least Squares Estimates XXX

```
R> data("vote1")
R> str(vote1)
'data.frame':       173 obs. of  10 variables:
 $ state    : chr  "AL" "AK" "AZ" "AZ" ...
 $ district: int  7 1 2 3 3 4 2 3 5 6 ...
 $ democA   : int  1 0 1 0 0 1 0 1 1 1 ...
 $ voteA    : int  68 62 73 69 75 69 59 71 76 73 ...
 $ expendA  : num  328.3 626.4 99.6 319.7 159.2 ...
 $ expendB  : num  8.74 402.48 3.07 26.28 60.05 ...
 $ prtystrA: int  41 60 55 64 66 46 58 49 71 64 ...
 $ lexpendA: num  5.79 6.44 4.6 5.77 5.07 ...
 $ lexpendB: num  2.17 6 1.12 3.27 4.1 ...
 $ shareA   : num  97.4 60.9 97 92.4 72.6 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
```

Deriving the Ordinary Least Squares Estimates XXXI

```
R> summary(vote1)[,1:3]
    state             district         democA      
 Length:173         Min.   : 1.000   Min.   :0.0000  
 Class :character   1st Qu.: 3.000   1st Qu.:0.0000  
 Mode  :character   Median : 6.000   Median :1.0000  
                    Mean   : 8.838   Mean   :0.5549  
                    3rd Qu.:11.000   3rd Qu.:1.0000  
                    Max.   :42.000   Max.   :1.0000  
```

Deriving the Ordinary Least Squares Estimates XXXII

```
R> summary(lm(voteA ~ shareA, data = vote1))
Call:
lm(formula = voteA ~ shareA, data = vote1)

Residuals:
     Min       1Q   Median       3Q      Max
-16.8919  -4.0660  -0.1682   3.4965  29.9772

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 26.81221    0.88721   30.22   <2e-16 ***
shareA       0.46383    0.01454   31.90   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.385 on 171 degrees of freedom
Multiple R-squared:  0.8561,	Adjusted R-squared:  0.8553
F-statistic:  1018 on 1 and 171 DF,  p-value: < 2.2e-16
```

Deriving the Ordinary Least Squares Estimates XXXIII

- We write the estimated equation as

$$\begin{aligned}\widehat{voteA} &= 26.81 + 0.464\ shareA \\ n &= 173\end{aligned}$$

- If $shareA = 50$, $\widehat{voteA} = 26.81 + .464(50) = 50.01$, or basically 50.
- If *shareA* increases by 10 (which means 10 percentage points), *voteA* is predicted to increase by 4.64 percentage points.

Properties of OLS on any Sample of Data I

- Once we have

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

we get the OLS fitted values by plugging the x_i into the equation:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i,\ i = 1, 2, ..., n$$

- The OLS residuals are

$$\hat{u}_i = y_i - \hat{y}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i,\ i = 1, 2, ..., n$$

Properties of OLS on any Sample of Data II

```
R> weq <- lm(wage ~ educ, data = wage1)
R> what <- fitted(weq)
R> uhat <- residuals(weq)
R> comp <- cbind(wage1$wage, wage1$educ, what, uhat)
R> comp[1:5,]
```

```
              what       uhat
1 3.10 11 5.050100 -1.9501003
2 3.24 12 5.591459 -2.3514594
3 3.00 11 5.050100 -2.0501002
4 6.00  8 3.426022  2.5739776
5 5.30 12 5.591459 -0.2914593
```

Properties of OLS on any Sample of Data III

- Some residuals are positive, others are negative.
- Just the fifth one is close to zero.
- Years of schooling, by itself, need not be a very good predictor of *wage*.

Properties of OLS on any Sample of Data IV

Algebraic Properties of OLS Statistics

1. The OLS residuals **always** add up to zero:

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

Because $y_i = \hat{y}_i + \hat{u}_i$ by definition,

$$n^{-1} \sum_{i=1}^{n} y_i = n^{-1} \sum_{i=1}^{n} \hat{y}_i + n^{-1} \sum_{i=1}^{n} \hat{u}_i$$

and so $\bar{y} = \overline{\hat{y}}$. In other words, **the sample average of the actual** y_i **is the same as the sample average of the fitted values.**

Properties of OLS on any Sample of Data V

```
R> sum(uhat)
[1] -6.253331e-14
R> mean(wage1$wage)
[1] 5.896103
R> mean(what)
[1] 5.896103
```

Properties of OLS on any Sample of Data VI

2. The **sample covariance** between the explanatory variables and the residuals is always zero:

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

Because the $\hat{y}_i$ are **linear functions** of the x_i, the fitted values and residuals are also uncorrelated:

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

Both of these properties hold by construction. $\hat{\beta}_0$ and $\hat{\beta}_1$ were chosen to make them true.

Properties of OLS on any Sample of Data VII

```
R> sum(uhat*wage1$educ)
[1] 1.437295e-12
R> sum(uhat*what)
[1] -4.35596e-13
```

Properties of OLS on any Sample of Data VIII

3. The point $(\bar{x}, \bar{y})$ is **always on the OLS regression line**. That is, if we plug in the average for x, we predict the sample average for y:

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$$

Again, we chose the estimates to make this true.

Goodness-of-Fit

For each observation, write

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

Define the **total sum of squares** (SST), **explained sum of squares** (SSE) and **residual sum of squares** (or sum of squared residuals) as

Properties of OLS on any Sample of Data X

$$
\begin{aligned}
SST &= \sum_{i=1}^{n}(y_i - \bar{y})^2 \\
SSE &= \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 \\
SSR &= \sum_{i=1}^{n}\hat{u}_i^2
\end{aligned}
$$

Properties of OLS on any Sample of Data XI

- Note that

$$\begin{aligned} SST &= \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}[(y_i - \hat{y}_i) + (\hat{y}_i - \bar{y})]^2 \\ &= \sum_{i=1}^{n}[\hat{u}_i + (\hat{y}_i - \bar{y})]^2 \end{aligned}$$

  and using that the fitted values and residuals are uncorrelated, can show

$$SST = SSE + SSR$$

Properties of OLS on any Sample of Data XII

- Assuming $SST > 0$, we can define the fraction of the total variation in y_i that is explained by x_i (or the OLS regression line) as

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

- Called the **R-squared** of the regression.
- It can be shown to equal the *square* of the correlation between y_i and $\hat{y}_i$. Therefore,

$$0 \leq R^2 \leq 1$$

Properties of OLS on any Sample of Data XIII

In particular:

- $R^2 = 0$ means no linear relationship between y_i and x_i.
- $R^2 = 1$ means a perfect linear relationship.
- As R^2 increases, the y_i are closer and closer to falling on the OLS regression line.

Unfortunately, having a "high" R-squared is neither necessary nor sufficient to infer causality.

Properties of OLS on any Sample of Data XIV

Years of education explains only about 16.5% of the variation in hourly wage:

$$\begin{aligned} \widehat{wage} &= -0.90 + 0.54\ educ \\ n &= 526,\ R^2 = .165 \end{aligned}$$

Properties of OLS on any Sample of Data XV

The share of the vote explains 85.6% of the variation in the vote outcome:

$$\begin{aligned} \widehat{voteA} &= 26.81 + 0.464\ shareA \\ n &= 173,\ R^2 = .856 \end{aligned}$$

Units of Measurement and Functional Form I

Units of Measurement

It is very important to know how y and x are measured in order to interpret regression functions. Consider an equation estimated from CEOSAL1 dataset, where annual CEO salary is in thousands of dollars and the return on equity is a percent

Units of Measurement and Functional Form II

```
R> data("ceosal1")
R> str(ceosal1)
```

```
'data.frame':       209 obs. of  12 variables:
 $ salary  : int  1095 1001 1122 578 1368 1145 1078 1094 1237
 $ pcsalary: int  20 32 9 -9 7 5 10 7 16 5 ...
 $ sales   : num  27595 9958 6126 16246 21783 ...
 $ roe     : num  14.1 10.9 23.5 5.9 13.8 ...
 $ pcroe   : num  106.4 -30.6 -16.3 -25.7 -3 ...
 $ ros     : int  191 13 14 -21 56 55 62 44 37 37 ...
 $ indus   : int  1 1 1 1 1 1 1 1 1 1 ...
 $ finance : int  0 0 0 0 0 0 0 0 0 0 ...
 $ consprod: int  0 0 0 0 0 0 0 0 0 0 ...
 $ utility : int  0 0 0 0 0 0 0 0 0 0 ...
 $ lsalary : num  7 6.91 7.02 6.36 7.22 ...
```

Units of Measurement and Functional Form III

```
 $ lsales  : num  10.23 9.21 8.72 9.7 9.99 ...
 - attr(*, "time.stamp")= chr "25 Jun 2011 23:03"
R> summary(ceosal1[,c(1,4)])
     salary             roe
 Min.   :  223   Min.   : 0.50
 1st Qu.:  736   1st Qu.:12.40
 Median : 1039   Median :15.50
 Mean   : 1281   Mean   :17.18
 3rd Qu.: 1407   3rd Qu.:20.00
 Max.   :14822   Max.   :56.30
```

Units of Measurement and Functional Form IV

```
R> summary(lm(salary ~ roe, data = ceosal1))
Call:
lm(formula = salary ~ roe, data = ceosal1)

Residuals:
    Min      1Q  Median      3Q     Max
-1160.2  -526.0  -254.0   138.8 13499.9

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   963.19     213.24   4.517 1.05e-05 ***
roe            18.50      11.12   1.663   0.0978 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1367 on 207 degrees of freedom
Multiple R-squared:  0.01319,        Adjusted R-squared:  0.008421
F-statistic: 2.767 on 1 and 207 DF,  p-value: 0.09777
```

Units of Measurement and Functional Form V

$$\begin{aligned}\widehat{salary} &= 963.191 + 18.501\ roe \\ n &= 209,\ R^2 = .0132\end{aligned}$$

- When $roe = 0$ (it never is in the data), $\widehat{salary} = 963.191$. But salary is in thousands of dollars, so $963,191.
- A one percentage point increase in *roe* increases predicted salary by 18.501, or $18,501 (since salary is in thousands of dollars).

Units of Measurement and Functional Form VI

- What if we measure *roe* as a decimal, rather than a percent?

$$roedec = roe/100$$

- What will happen to the intercept, slope, and R^2 when we regress

$$salary \text{ on } roedec?$$

Units of Measurement and Functional Form VII

The new regression is

$$\begin{aligned}\widehat{salary} &= 963.191 + 1,850.1\ roedec \\ n &= 209,\ \ R^2 = .0132\end{aligned}$$

Now a one percentage point change in *roe* is the same as $\Delta roedec = .01$. Then $1,850.1(.01) = 18.501$ or $\$18,501$ and so we get the same effect as before.

Units of Measurement and Functional Form VIII

What if we measure salary in dollars, rather than thousands of dollars? $salarydol = 1,000 \cdot salary$.

Both the intercept and slope get multiplied by 1,000:

$$\begin{aligned} \widehat{salarydol} &= 963,191 + 18,501 \; roe \\ n &= 209, \; R^2 = .0132 \end{aligned}$$

Question: What happens in the voting example if both *voteA* and *shareA* are measured as decimals rather than percents?

Units of Measurement and Functional Form X

The original equation is

$$\begin{aligned} \widehat{voteA} &= 26.81 + 0.464\ shareA \\ n &= 173,\ R^2 = .856 \end{aligned}$$

If we divide both *voteA* and *shareA* by 100, nothing happens to the slope and nothing happens to R^2. But the intercept gets divided by 100 and becomes .2681.

Units of Measurement and Functional Form XI

Using the Natural Logarithm in Simple Regression

Recall the wage example:

$$
\begin{aligned}
\widehat{wage} &= -0.90 + 0.54\ educ \\
n &= 526,\ R^2 = .165
\end{aligned}
$$

Units of Measurement and Functional Form XII

- Might be an okay approximation, but unsatisfying for a couple of reasons.
    1. First, **the negative intercept** is a bit strange (even though the equation gives sensible predictions for education ranging from 8 to 20).
    2. Second reason is more important: **the dollar value of another year of schooling is constant**: 54 cents is the increase for either the first year of education or the twentieth year. But the expectation is that additional years of schooling are worth more, in dollar terms, than previous years.

Units of Measurement and Functional Form XIII

How can we incorporate an increasing effect?
One way is to assume a **constant percentage** effect. We can approximate percentage changes using the natural log.

Units of Measurement and Functional Form XIV

To see this, let the dependent variable be $\log(wage)$:

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

Holding u fixed,

$$\Delta \log(wage) = \beta_1 \Delta educ$$

so

$$\beta_1 = \frac{\Delta \log(wage)}{\Delta educ}$$

Units of Measurement and Functional Form XV

An useful result from calculus tells us that:

$$\log(x_1) - \log(x_0) \approx \frac{(x_1 - x_0)}{x_0} = \frac{\Delta x}{x_0}$$

Therefore, if we replace $\Delta \log(x) = \log(x_1) - \log(x_0)$ and multiply both sides by 100, we have:

$$100 \cdot \Delta \log(x) \approx \%\Delta x$$

which in the context of the wage equation reads as

$$100 \cdot \Delta \log(wage) \approx \%\Delta wage$$

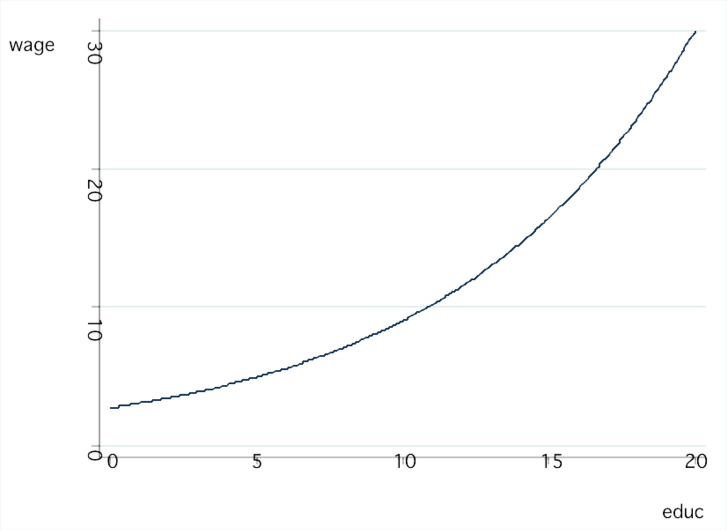
Units of Measurement and Functional Form XVI

This means when $\log(wage) = \beta_0 + \beta_1 educ + u$, we have a simple interpretation of β_1:

$$100\beta_1 \approx \%\Delta wage \text{ when } \Delta educ = 1$$

- In this example, $100\beta_1$ is often called the *return to education*.
- This measure is free of units of measurement of wage (currency, price level).
- The next graph shows the relationship between *wage* and *educ* when $u = 0$.

Units of Measurement and Functional Form XVII



The results of the model specified in terms of log(wage) are:

$$\begin{aligned}\widehat{(lwage)} &= 0.583 + 0.083\ educ \\ n &= 526,\ R^2 = .186\end{aligned}$$

```
R> summary(lm(lwage ~ educ, data = wage1))
Call:
lm(formula = lwage ~ educ, data = wage1)

Residuals:
     Min       1Q   Median       3Q      Max
-2.21158 -0.36393 -0.07263  0.29712  1.52339

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 0.583773   0.097336   5.998 3.74e-09 ***
educ        0.082744   0.007567  10.935  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4801 on 524 degrees of freedom
Multiple R-squared:  0.1858,        Adjusted R-squared:  0.1843
F-statistic: 119.6 on 1 and 524 DF,  p-value: < 2.2e-16
```

Units of Measurement and Functional Form XX

- The estimated return to each year of education is about 8.3%.
- This R-squared is not directly comparable to the R-squared when *wage* is the dependent variable. The total variation (SSTs) in $wage_i$ and $lwage_i$ that we must explain are completely different.

We can use the log on both sides of the equation to get **constant elasticity models**. For example, if

$$\log(salary) = \beta_0 + \beta_1 \log(sales) + u$$

then

$$\beta_1 \approx \frac{\%\Delta salary}{\%\Delta sales}$$

The elasticity is free of units of *salary* and *sales*.

A constant elasticity model for salary and sales makes more sense than a constant dollar effect.

Units of Measurement and Functional Form XXII

Using CEOSAL1:

```
R> summary(lm(lsalary ~ lsales, data = ceosal1))

Call:
lm(formula = lsalary ~ lsales, data = ceosal1)

Residuals:
     Min       1Q   Median       3Q      Max
-1.01038 -0.28140 -0.02723  0.21222  2.81128

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.82200    0.28834  16.723  < 2e-16 ***
lsales       0.25667    0.03452   7.436  2.7e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5044 on 207 degrees of freedom
Multiple R-squared:  0.2108,        Adjusted R-squared:  0.207
F-statistic:  55.3 on 1 and 207 DF,  p-value: 2.703e-12
```

Units of Measurement and Functional Form XXIII

The estimated elasticity of CEO salary with respect to firms sales is about .257.

A 10 percent increase in sales

$$.257(10) = 2.57$$

is associated with percent increase in salary.

Units of Measurement and Functional Form XXIV

Model	Dep. Var.	Indep. Var.	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	$\log(y)$	$\log(x)$	$\%\Delta y = \beta_1 \%\Delta x$

Units of Measurement and Functional Form XXV

The possibility of using the natural log to get nonlinear relationships between y and x raises a question: What do we mean now by "linear" regression? The answer is that the model is linear in the *parameters*, β_0 and β_1. We can use any transformations of the dependent and independent variables to get interesting interpretations for the parameters.

An appendix on Maximum Likelihood I

- We argued that there are mostly three ways to estimate a linear regression model.
- In fact, we have seen that OLS and MM gave the same results for the parameters of the model.
- Now we briefly introduce the third estimation method and we will observe that also this one will lead to estimators that are identical to the previous.

An appendix on Maximum Likelihood II

- Let us start from the simple regression model:

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

  and let us assume that the errors u_i are independent and identically distributed following a normal with mean 0 and variance σ^2, that is $u_i \sim N(0, \sigma^2)$.
- For a reason that we will discuss later, if u_i are normal, also y_i are normal.
- Hence the density function of the y_i can be expressed as:

$$f(y; \beta_0, \beta_1, \sigma) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp^{-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$$

- This is called the Likelihood function.

An appendix on Maximum Likelihood III

- The idea is to find the values of the (unknown) parameters that maximize the likelihood function
- In practice, we generally consider the log-likelihood function which is easier to deal with:

$$
\begin{aligned}
L(y;\beta_0,\beta_1,\sigma) &= \log\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi\sigma^2}}\exp^{-\frac{(y_i-\beta_0-\beta_1x_i)^2}{2\sigma^2}} \\
&= \sum_{i=1}^{n}\log(\frac{1}{\sqrt{2\pi\sigma^2}}\exp^{-\frac{(y_i-\beta_0-\beta_1x_i)^2}{2\sigma^2}}) \\
&= -\frac{n}{2}\log 2\pi - n\log\sigma - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2
\end{aligned}
$$

An appendix on Maximum Likelihood IV

- To find the maximum we derive the likelihood function with respect to the parameters
- After some calculus, we obtain the

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$
$$\hat{\beta}_1 = \frac{Cov(x, y)}{var(x)}$$

An appendix on Maximum Likelihood V

```
R> set.seed(3)
R> x <- rnorm(100, mean = 3, sd = 5)
R> u <- rnorm(100, mean = 0, sd = 1)
R> y <- 2 + 3*x + u
```

An appendix on Maximum Likelihood VI

```
R> lik <- function(par, n, y, x){
+    -((-n/2) *log(pi) -
+        n*log(par[3]) -
+        (1/(2*par[3]^2))*
+        sum((y-par[1]-par[2]*x)^2))
+    }
```

An appendix on Maximum Likelihood VII

```
R> par <- c(1.5, 3.5, 1.3)
R> ML <- optim(par, lik,
+              method = "L-BFGS-B",
+              n = 100,
+              y = y,
+              x = x)
R> b0b1 <- ML$par[1:2]
R> names(b0b1) <- c("(Intercept)", "x")
R> b0b1
(Intercept)           x
   2.074021    2.981968
R> coefficients(lm(y~x))
(Intercept)           x
   2.074030    2.981967
```