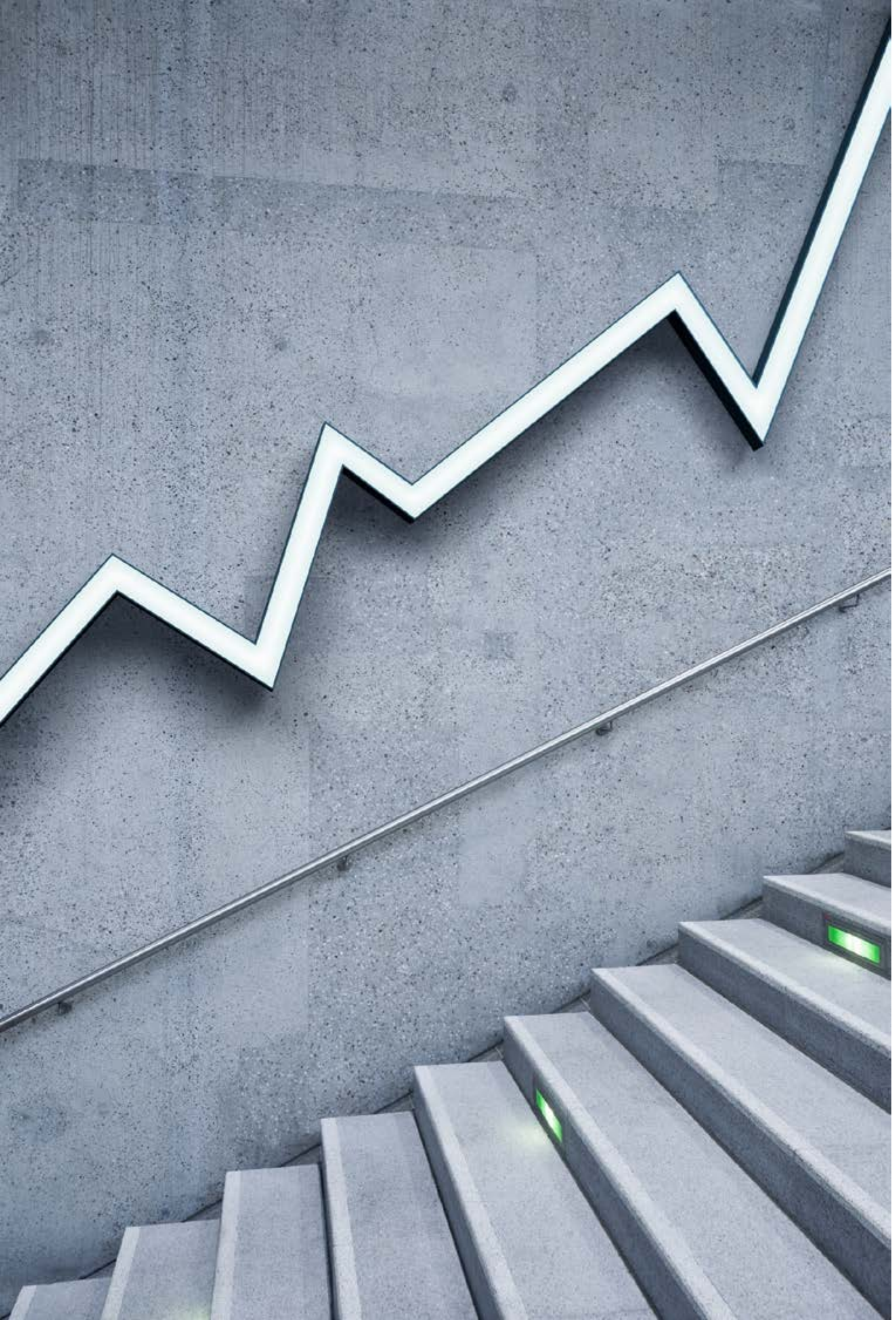


Lecture 2 The Multiple Regression Model: Estimation

LECTURED BY WENYA WANG

Outline

1. Motivation for multiple regression
2. Mechanics and interpretation of ordinary least squares
3. The expected value of the OLS estimators
4. The variance of the OLS estimators
5. Efficiency of OLS: the Gauss-Markov theorem

Introduction

The simple regression model

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

- Caveat: problematic ceteris paribus analysis
- (Very) strong assumption $E(\mu|x) = 0$

Example of Mincer equation: suppose in the population

$$Hourly\ Wage = \beta_0 + \beta_1 Education + \beta_2 Ability + \epsilon$$

- Let's say we run the simple regression $Hourly\ Wage = \beta_0 + \beta_1 Education + \mu$ in our randomly selected sample
- We know that one's education level is positively associated with its ability, i.e., $E(\mu|x) \neq 0$
- Failure of zero conditional mean

In multiple regression model, we could control for ability, e.g., proxied by SAT scores

Introduction: two explanatory variables

General form of multiple regression model with two explanatory variables

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

β_0: intercept

β_1: change in y with respect to x_1, holding other factors fixed (ceteris paribus)

β_2: change in y with respect to x_2, , holding other factors fixed (ceteris paribus)

Introduction: two explanatory variables

In a special case for capturing the non-linear relation between y and x_1,

$$x_2 = x_1^2$$

Interpretations of β_1 and β_2 are not on ceteris paribus

- Cannot hold x_1^2 constant while there's a change in x_1 ; vice versa.
- Rather, change in y with respect to x_1 is

$$\frac{\Delta y}{\Delta x_1} = \beta_1 + \beta_2 x_1$$

which increases with x_1 .

We also discuss a non-linear model in Lecture 1: $\log(y) = \beta_0 + \beta_1 x_1 + \mu$. How do they differ?

Introduction: k explanatory variables

Multiple linear regression (MLR) model:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \ldots + \beta_k x_k + u,$$

where

β_0 is the **intercept**.
β_1 is the parameter associated with x_1.
β_2 is the parameter associated with x_2, and so on.

Again, we assume zero conditional mean

$$E(\mu | x_1, x_2, \ldots, x_k) = 0$$

Estimation: Ordinary-Least-Squares

As in Lecture 1, minimize the distance between y and $\hat{y}$

$$min \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \dots - \hat{\beta}_k x_k)^2$$

Take derivative with respect to $\hat{\beta}_j$, j=0,1,...,k, first order conditions:

$$\sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

$$\sum_{i=1}^{n} x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

$$\sum_{i=1}^{n} x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0.$$

Estimation: Ordinary-Least-Squares

Hard to solve by hands, although easier if we re-write them in forms of matrix operations.

Often times, we resort to computer software

Example: how do one's income, age and self-employment status affect one's credit card limit.

```
. reg logcc_limit logincome age selfemp

      Source |       SS           df       MS      Number of obs   =    16,075
-------------+----------------------------------   F(3, 16071)     =   1753.17
       Model |  6431.63886         3  2143.87962   Prob > F        =    0.0000
    Residual |  19652.5217    16,071  1.22285618   R-squared       =    0.2466
-------------+----------------------------------   Adj R-squared   =    0.2464
       Total |  26084.1606    16,074   1.6227548   Root MSE        =    1.1058

------------------------------------------------------------------------------
 logcc_limit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   logincome |   .3580465   .0064211    55.76   0.000     .3454603    .3706326
         age |   .0092951   .0005886    15.79   0.000     .0081413    .0104489
     selfemp |   .2117563    .020649    10.26   0.000     .1712819    .2522306
       _cons |   4.818577   .0720057    66.92   0.000     4.677438    4.959716
------------------------------------------------------------------------------
```

Interpreting OLS estimates

Partial effect or ceteris paribus

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2,$$

- Holding x_2 constant, i.e., $\Delta x_2 = 0$, changes in y corresponding to Δx_1 units change is $\hat{\beta}_1 \Delta x_1$
- Vice versa for changes in x_1, ceteris paribus

Why this is a powerful perspective?

- Separate the effect of one factor alone, out of many confounding factors
- Ideal case: experiments with treatment, e.g., low x_1, and control, e.g., high x_1, groups
- Not feasible, or morally unacceptable for most economic applications

Simultaneous changes in multiple variables

Changes in y corresponding to Δx_1 units change in x_1 and Δx_2 units change in x_2

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2,$$

Properties of OLS estimates

1. Fitted or predicted values $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \ldots + \hat{\beta}_k x_{ik},$

Residuals $\hat{u}_i = y_i - \hat{y}_i.$

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0$$
$$\sum_{i=1}^{n}x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0$$
$$\sum_{i=1}^{n}x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0$$
$$\vdots$$
$$\sum_{i=1}^{n}x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \ldots - \hat{\beta}_k x_{ik}) = 0.$$

$$\Rightarrow \sum_{i=1}^{n}\hat{u}_i = 0 \Rightarrow \bar{y} = \bar{\hat{y}}$$
$$= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \ldots + \hat{\beta}_k \bar{x}_k$$
$$\Downarrow$$

The point $(\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_k, \bar{y})$ is always on the OLS regression line: $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \hat{\beta}_2 \bar{x}_2 + \ldots + \hat{\beta}_k \bar{x}_k$.

Properties of OLS estimates

2. "Partialing out" interpretation of multiple regression

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

Is there a way to solve OLS estimates of $\hat{\beta}_1$ and $\hat{\beta}_2$ by hand, not using the above first order conditions? Yes!

Frisch-Waugh Theorem the OLS estimate of $\hat{\beta}_1$ in multiple regression model could also be obtained by the following three steps:

a, regress y on x_2, and obtain the residual $\hat{r}_y$ (typo in the textbook)

b, regress x_1 on x_2, and obtain the residual $\hat{r}_{x1}$

c. regress $\hat{r}_y$ on $\hat{r}_{x1}$, and obtain the same estimate of $\hat{\beta}_1$ =$(\sum_i \hat{r}_{x1}\hat{r}_y)/(\sum_i \hat{r}_{x1}^2)$

Properties of OLS estimates

a, regress y on x_2, and obtain the residual $\hat{r}_y$

b, regress x_1 on x_2, and obtain the residual $\hat{r}_{x1}$

Partialing out the variations of y and x_1 that are correlated with x_2 *(another form of ceteris paribus)*

c. regress $\hat{r}_y$ on $\hat{r}_{x1}$ $\Rightarrow$ The remaining: variation of the residual y that could be explained by the variation of residual x_1

The theorem can be generalized into k explanatory variables case. First two steps are now:

a, regress y on $x_2, x_3, \dots, x_k$,....

b, regress x_1 on $x_2, x_3, \dots, x_k$,....

Properties of OLS estimates

Use the credit card limit as the partialing-out example.

Interested in the coefficient of logincome.

```
. quietly reg logcc_limit age selfemp

. predict r_logcc_limit,resid

. quietly reg logincome age selfemp

. predict rlogincome,resid

. reg r_logcc_limit rlogincome

      Source |       SS           df       MS      Number of obs   =    16,075
-------------+----------------------------------   F(1, 16073)     =   3109.63
       Model |  3802.15411         1  3802.15411   Prob > F        =    0.0000
    Residual |  19652.5217    16,073  1.22270402   R-squared       =    0.1621
-------------+----------------------------------   Adj R-squared   =    0.1621
       Total |  23454.6758    16,074  1.45916858   Root MSE        =    1.1058

------------------------------------------------------------------------------
r_logcc_li~t |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  rlogincome |   .3580465   .0064207    55.76   0.000     .3454611    .3706319
       _cons |  -1.55e-09   .0087214    -0.00   1.000    -.0170949    .0170949
------------------------------------------------------------------------------
```

Properties of OLS estimates

3. Comparison of simple and multiple regression estimates

Suppose the theory (i.e., the true data generating process) is

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \mu$$

When do we get the same estimates of β_1 by

(i) regressing y on x_1;

(ii) regressing y on x_1 and x_2?

One obvious scenario is β_2=0, or we get a very small estimate, $\hat{\beta}_2$

Properties of OLS estimates

The other scenario is when x_1 and x_2 are uncorrelated.

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$

To see this, note that regressing y on x_1 means

- the relation between $\tilde{\beta}_1$ and $\hat{\beta}_1$ in the multiple regression model $y = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$ is

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$$

where $\tilde{\delta}_1$ is the estimate in $\mathrm{x}_1 = \hat{\delta}_0 + \hat{\delta}_2 x_2$, and =0 when x_1 and x_2 are uncorrelated.

Properties of OLS estimates

```
. reg logcc_limit logincome age selfemp

      Source |       SS           df       MS      Number of obs   =    16,075
-------------+----------------------------------   F(3, 16071)     =   1753.17
       Model |  6431.63886         3  2143.87962   Prob > F        =    0.0000
    Residual |  19652.5217    16,071  1.22285618   R-squared       =    0.2466
-------------+----------------------------------   Adj R-squared   =    0.2464
       Total |  26084.1606    16,074   1.6227548   Root MSE        =    1.1058

------------------------------------------------------------------------------
 logcc_limit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   logincome |   .3580465   .0064211    55.76   0.000     .3454603    .3706326
         age |   .0092951   .0005886    15.79   0.000     .0081413    .0104489
     selfemp |   .2117563    .020649    10.26   0.000     .1712819    .2522306
       _cons |   4.818577   .0720057    66.92   0.000     4.677438    4.959716
------------------------------------------------------------------------------
```

Example: (i) guess changes of the coefficients of logincome and selfemp if we remove age.

(ii) can you name another example that fits into the second scenario, given the following correlation matrix?

```
 corr logincome age selfemp
(obs=16,075)

             | loginc~e      age  selfemp
-------------+---------------------------
   logincome |   1.0000
         age |   0.2085   1.0000
     selfemp |   0.4044   0.0767   1.0000
```

Properties of OLS estimates

4. Goodness-of-fit

As with simple regression, we can define the **total sum of squares (SST)**, the **explained sum of squares (SSE)**, and the **residual sum of squares** or **sum of squared residuals (SSR)** as

$$\text{SST} \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \quad [3.24]$$

$$\text{SSE} \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \quad [3.25]$$

$$\text{SSR} \equiv \sum_{i=1}^{n} \hat{u}_i^2. \quad [3.26]$$

$$\Rightarrow \quad R^2 \equiv \text{SSE/SST} = 1 - \text{SSR/SST},$$

Important property (caveat) of R-squared: always increasing when we add more explanatory variables

*may make the model unnecessarily complex.

*Solution: adjusted R-squared that punish more when the number of variables increases.

Properties of OLS estimates

Example: find the R-squared and interpret the number.

```
. reg logcc_limit logincome age selfemp

      Source |       SS           df       MS      Number of obs   =    16,075
-------------+----------------------------------   F(3, 16071)     =   1753.17
       Model |  6431.63886         3  2143.87962   Prob > F        =    0.0000
    Residual |  19652.5217    16,071  1.22285618   R-squared       =    0.2466
-------------+----------------------------------   Adj R-squared   =    0.2464
       Total |  26084.1606    16,074   1.6227548   Root MSE        =    1.1058

------------------------------------------------------------------------------
 logcc_limit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   logincome |   .3580465   .0064211    55.76   0.000     .3454603    .3706326
         age |   .0092951   .0005886    15.79   0.000     .0081413    .0104489
     selfemp |   .2117563    .020649    10.26   0.000     .1712819    .2522306
       _cons |   4.818577   .0720057    66.92   0.000     4.677438    4.959716
------------------------------------------------------------------------------
```

Expected value of the OLS estimators

To derive unbiasedness of OLS estimators as in the simple regression model, we need four assumptions.

1. (MLR.1) linear in parameters

The model in the population can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u, \qquad [3.31]$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown parameters (constants) of interest and u is an unobserved random error or disturbance term.

Expected value of the OLS estimators

2. (MRL.2) Random sampling

We have a random sample of n observations, $\{(x_{i1}, x_{i2}, \ldots, x_{ik}, y_i): i = 1, 2, \ldots, n\}$, following the population model in Assumption MLR.1.

3. (MRL.3) No perfect collinearity

In the sample (and therefore in the population), none of the independent variables is constant, and there are no *exact linear* relationships among the independent variables.

Example of perfect collinearity: $x_k = \delta_0 + \delta_1 x_1 + \delta_2 x_2$

Expected value of the OLS estimators

Note that how $x_k = \delta_0 + \delta_1 x_1 + \delta_2 x_2$ differs from $x_k = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \epsilon$

- the error term ϵ is not the error term μ in the original equation
- Intuitively, $x_k = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \epsilon$ says x_k should contain some information other than x_1 and x_2
- Otherwise, x_k is simply repeating information from x_1 and x_2, and hence perfectly correlates with them.

More rigorously speaking, the matrix of $(x_1, x_2, \ldots, x_k)$ should have a full rank.

Expected value of the OLS estimators

4. (MRL.4) Zero conditional mean

The error u has an expected value of zero given any values of the independent variables. In other words,

$$E(u|x_1, x_2, \ldots, x_k) = 0. \quad [3.36]$$

Theorem of unbiasedness

Theorem 3.1 Unbiasedness of OLS

Under assumptions MLR. 1 through MLR. 4,

$$E(\hat{\beta}_j) = \beta_j, j = 0,1, \ldots, k$$

for any values of the population parameter β_j. In other words, the OLS estimators are unbiased estimators of the population parameters.

Discussions on the unbiasedness

We assumed that we know the true model in the population

- a fairly strong assumption: truth is yet to discover, but not assumed known

The model in the population can be written as

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u, \qquad [3.31]$$

where $\beta_0, \beta_1, \ldots, \beta_k$ are the unknown parameters (constants) of interest and u is an unobserved random error or disturbance term.

What if

(i) I add another irrelevant explanatory variable x_{k+1}?

(ii) I miss some explanatory variable(s)?

Discussions on the unbiasedness

(i) Inclusion of an irrelevant variable or overspecifying the model

By meaning irrelevant, we should immediately have $E(\mu|x_{k+1}) = 0$

- Otherwise, equation 3.31 is not the population model

Thus, $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k + \beta_{k+1} x_{k+1} + \mu$

- β_{k+1}=0
- Recall from the previous discussion: estimates of β_1, …., β_k would be exactly the same as we get from

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

Discussions on the unbiasedness

(ii) Excluding a relevant variable or underspecifying the model

Let's use the Mincer equation as the example.

Suppose the true population model:

$$Hourly\ Wage = \beta_0 + \beta_1 Education + \beta_2 Ability + \mu \quad \text{(A)}$$

Misspecification:

$$Hourly\ Wage = \beta_0 + \beta_1 Education + \nu \qquad (B)$$

From earlier discussion, we know that the estimate $\tilde{\beta}_1$ in equation (B) relates to that of (A) $\hat{\beta}_1$

$$\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1$$

Discussions on the unbiasedness

Thus $E(\tilde{\beta}_1) = E(\hat{\beta}_1) + E(\hat{\beta}_2\tilde{\delta}_1) = \beta_1 \boxed{+ \beta_2\ \tilde{\delta}_1}$ $\Rightarrow$ Omitted variable bias

Positive or negative bias?

TABLE 3.2 Summary of Bias in $\tilde{\beta}_1$ when x_2 Is Omitted in Estimating Eqution (3.40)

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

Terminology often used in applied economics

- Positive bias: upward bias
- Negative bias: downward bias

Same idea if the population model has k variables

Discussions on the unbiasedness

Example: missing self-employment variable in the credit card limit example

```
. reg logcc_limit logincome age

      Source |       SS           df       MS      Number of obs   =    16,075
-------------+----------------------------------   F(2, 16072)     =   2560.58
       Model |  6303.03626         2  3151.51813   Prob > F        =    0.0000
    Residual |  19781.1243    16,072  1.23078175   R-squared       =    0.2416
-------------+----------------------------------   Adj R-squared   =    0.2415
       Total |  26084.1606    16,074   1.6227548   Root MSE        =    1.1094

------------------------------------------------------------------------------
 logcc_limit |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   logincome |   .3842712    .005909    65.03   0.000     .3726889    .3958535
         age |    .009244   .0005905    15.65   0.000     .0080865    .0104015
       _cons |    4.58641   .0685753    66.88   0.000     4.451995    4.720826
------------------------------------------------------------------------------
```

```
. corr logincome selfemp
(obs=16,075)

             | loginc~e  selfemp
-------------+------------------
   logincome |   1.0000
     selfemp |   0.4044   1.0000
```

The variance of the OLS estimators

As in the simple regression model, additional assumption on homoscedasticity

5. (MLR. 5) The error u has the same variance given any values of the explanatory variables. In other words, $\mathrm{Var}(u|x_1, \ldots, x_k) = \sigma^2$.

THEOREM 3.2

SAMPLING VARIANCES OF THE OLS SLOPE ESTIMATORS

Under Assumptions MLR.1 through MLR.5, conditional on the sample values of the independent variables,

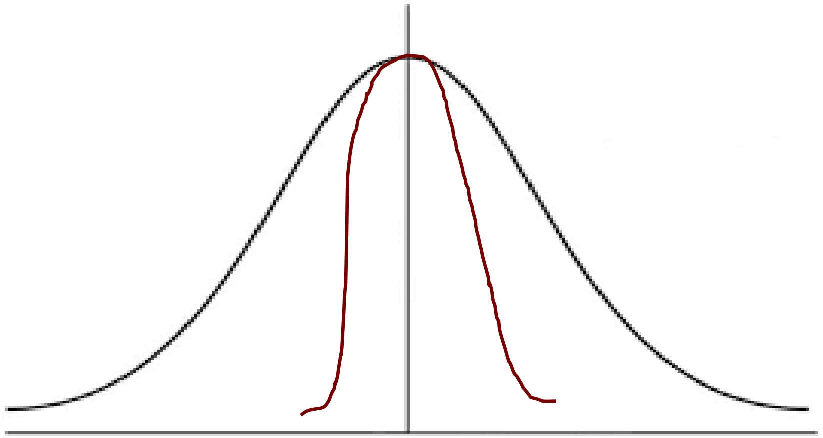
$$\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\mathrm{SST}_j(1 - R_j^2)}, \qquad [3.51]$$

for $j = 1, 2, \ldots, k$, where $\mathrm{SST}_j = \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)^2$ is the total sample variation in x_j, and R_j^2 is the R-squared from regressing x_j on all other independent variables (and including an intercept).

The variance of the OLS estimators

Some intuitions of theorem 3.2: A higher variance $\mathrm{Var}(\hat{\beta}_j)$

1. when σ^2 increases, i.e., too much noise in the error term
2. when $\mathrm{SST}_j = \sum_i (x_j - \bar{x})^2$ *decreases*, i.e., there is little variation in x_j
3. When R_j^2 increases, i.e., when x_j could be better explained by other explanatory variables, or equivalently speaking, when x_j tells us little additional information given other variables (extreme case: multicollinearity)



Ideally, $\mathrm{Var}(\hat{\beta}_j)$ is infinitely small, such that $\hat{\beta}_j$ is infinitely close to the true value β_j

The variance of the OLS estimators

But that may not be achievable.

You may come up with any estimators of β_j, but the $Var(\hat{\beta}_j)$ reaches its minimum when it's OLS estimator.

Gauss-Markov Theorem

Under assumptions MLR. 1 through MLR. 5, the OLS estimator of β_j, $\hat{\beta}_j$, is the best linear unbiased estimator (BLUE, or equivalently speaking most efficient), i.e.,

$$0 < Var(\hat{\beta}_j) < Var(\tilde{\beta}_j)$$

for any other linear unbiased estimator $\tilde{\beta}_j$.

The variance of the OLS estimators

Clarifications on linear & unbiased

1. Linear estimator. We call an estimator as linear if it's a linear function of the sampled y

$$\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i,$$

where w_{ij} is a function of the sample values of all the independent variables.

- Easier to see this in matrix form

$$\tilde{\beta} = W(X)Y$$

and the OLS estimator is $(X'X)^{-1}X'Y$

The variance of the OLS estimators

2. Unbiased

We've already show (i) definition of unbiasedness; (ii) OLS estimators are unbiased

Question is: are there unbiased estimators other than OLS? Yes!

Example in the simple regression model $y = \beta_0 + \beta_1 x_1 + \mu$

- Check the estimator $\tilde{\beta} = \frac{\sum_{i=2}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=2}^{n}(x_i - \bar{x})^2}$ is unbiased

- And its variance is greater than that of OLS estimator

The variance of the OLS estimators

Note that σ^2 remained unknown in this equation $\mathrm{Var}(\hat{\beta}_j) = \frac{\sigma^2}{\mathrm{SST}_j(1 - R_j^2)}$,

Like in the simple regression model, we could use the sample analog of

$$E(\mu^2) = \sigma^2$$

which is

$$n^{-1} \sum_{i=1}^{n} \hat{\mu}_i^2$$

$$= n^{-1} \sum_{i=1}^{n} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_1 - \hat{\beta}_2 x_2 - \ldots - \hat{\beta}_k x_k)^2$$

The variance of the OLS estimators

Again, by the OLS estimation process: the n residuals $\hat{\mu}$ are not independent

- k+1 first order conditions
- Know $\hat{\mu}_{n-k}$, … , $\hat{\mu}_n$ when we know $\hat{\mu}_1, \hat{\mu}_2, \dots, \hat{\mu}_{n-k-1}$
- Degree of freedom n-k-1

$$\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

$$\sum_{i=1}^{n}x_{i1}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

$$\sum_{i=1}^{n}x_{i2}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0$$

$$\vdots$$

$$\sum_{i=1}^{n}x_{ik}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \dots - \hat{\beta}_k x_{ik}) = 0.$$

Hence the unbiased estimator of σ^2 is

$$\sum_{i=1}^{n}\hat{\mu}_i^2/(n-k-1)$$

The variance of the OLS estimators

The standard error of OLS estimator of β_j is then

$$\mathrm{se}(\hat{\beta}_j) = \hat{\sigma}/[\mathrm{SST}_j(1 - R_j^2)]^{1/2}.$$

Note this differs from the term standard deviation

$$\mathrm{sd}(\hat{\beta}_j) = \sigma/[\mathrm{SST}_j(1 - R_j^2)]^{1/2}.$$

Summary

1. Motivation for multiple regression
2. Mechanics and interpretation of ordinary least squares
3. The expected value of the OLS estimators
4. The variance of the OLS estimators
5. Efficiency of OLS: the Gauss-Markov theorem

Next: statistical inference in the multiple regression model.