Introductory Econometrics
Ch3 Multiple Regression Analysis: Estimation

LIU Chenyuan

Spring 2024

Outline

Why We Need Multiple Regression Model?

Estimation and Interpretation
- Estimation
- Interpretation

Expected Values and Variances of the OLS Estimators

A Few Practical Issues

Outline

Why We Need Multiple Regression Model?

Estimation and Interpretation

- Estimation
- Interpretation

Expected Values and Variances of the OLS Estimators

A Few Practical Issues

Multiple Regression Model

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

Why do we need a multiple regression model?

- Descriptive analysis: sometimes, we want to estimate the conditional mean of y on multiple variables.
- Causal estimation: we know that something other than x may affect y, so we explicitly control them.
- Forecasting: we want to use more variables to better predict y.

Example: Education and Wage

We are interested in the causal relationship between education and wages. Consider the two models:

$$wage = \beta_0 + \beta_1 edu + u, \tag{1}$$

$$wage = \beta_0 + \beta_1 edu + \beta_2 ability + u, \tag{2}$$

where *wage* is hourly wage, *edu* is year of education, *ability* is ability.

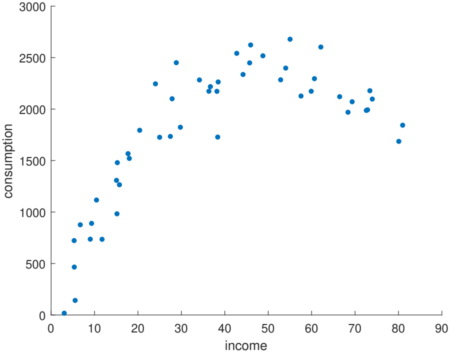
1. In (1), under what condition will β_1 represent the causal effects between *edu* and *wage*?

Example Education and Wage

- For the first model, we need $E(u|edu) = 0$ to make sure there exists a causal relationship. In reality, this is hardly true. For example, ability is a factor affecting wage other than education. However, high-ability people may also have higher education.
- In the second model, we directly add *ability* into the regression. By doing this, we could measure the causal effect of education on wage, holding fixed ability.
- Sometimes, we may also be interested in the impact of ability on wage.
- Multiple regression model can better establish causal relationships by controlling for more variables directly.

Example Consumption and Income

Suppose we are interested in forecasting consumption given income:



Because consumption and income have a non-monotonic relationship, the following model seems better:

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

Outline

Why We Need Multiple Regression Model?

Estimation and Interpretation
- Estimation
- Interpretation

Expected Values and Variances of the OLS Estimators

A Few Practical Issues

Estimation

- The population regression model is:

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

- Zero conditional mean assumption: $E(u|x_1, x_2, ..., x_k) = 0$.
- We can also prove that, $\forall j \in 1, 2, ..., k$:

$$\begin{aligned} E(x_j u) &= E(E(x_j u|x_1, x_2, ..., x_k)) \\ &= E(x_j E(u|x_1, x_2, ..., x_k)) \\ &= E(x_j 0) = 0. \end{aligned}$$

  and

$$E(u) = 0.$$

- Define a random sample $(x_{i1}, ..., x_{ik}, y_i) : i = 1, ..., N$, where i in x_{ij} means the i-th observation, j means the j-th independent variable.
- Fitted value: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_k x_{ik}$.
- Residual: $\hat{u}_i = y_i - \hat{y}_i$

Sample Analogues

Population expectations	Sample analogue
$E(u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i = 0$
$E(x_1 u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} x_{i1}\hat{u}_i = 0$
...	...
$E(x_k u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} x_{ik}\hat{u}_i = 0$

Ordinary Least Squares

- Ordinary Least Squares: minimize the sum of the residual square.
- Fitted value: $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + ... + \hat{\beta}_k x_{ik}$.
- Residual: $\hat{u}_i = y_i - \hat{y}_i$
- We choose $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ to minimize

$$H \equiv \sum_{i=1}^{N} \hat{u}_i^2 = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - ... - \hat{\beta}_k x_{ik})^2.$$

Ordinary Least Squares

The first-order conditions are:

$$\frac{\partial H}{\partial \hat{\beta}_0} = -\sum_{i=1}^{N} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - ... - \hat{\beta}_k x_{ik}) = 0,$$

$$\frac{\partial H}{\partial \hat{\beta}_j} = -\sum_{i=1}^{N} 2x_{ij}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - ... - \hat{\beta}_k x_{ik}) = 0, \forall j = 1, 2, ..., k.$$

- The sample analogue conditions are the same as the first-order condition in OLS.
- In other words, the two motivations are mathematically equivalent.
- There are $k+1$ equations, and $k+1$ unknowns. With proper assumptions discussed later, the system of equations has a unique solution.
- We can solve them by hand or simply using Stata.

```
reg y x1 x2 x3
```

OLS Fitted Values and Residuals

- We call the estimated equation **OLS regression line** or the **sample regression function**:

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

- The fitted value:

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

- The residual is:

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

Properties of the OLS Fitted Values and Residuals

1. The sample average of the residuals is zero.

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

2. The sample covariance between each independent variable and the OLS residuals is zero. Consequently, the sample covariance between the OLS fitted values and the OLS residuals is zero.

$$\sum_{i=1}^{N} \hat{u}_i x_{ij} = 0, \forall j = 1, 2, ..., k.$$

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

Properties of the OLS Fitted Values and Residuals

1. The sample average of the residuals is zero.
2. The sample covariance between each independent variable and the OLS residuals is zero. Consequently, the sample covariance between the OLS fitted values and the OLS residuals is zero.
3. The point $(\bar{x}_1, \bar{x}_2, ..., \bar{x}_k, \bar{y})$ is always on the OLS regression line.

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

Interpretation of the Multiple Regression Model

We first think from the causal perspective.

- Consider the case with two independent variables:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

- The estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ have partial effect, or "all else equal" interpretations:

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

- When x_2 is held fixed, so that $\Delta x_2 = 0$, we have

$$\Delta \hat{y} = \hat{\beta}_1 \Delta x_1.$$

- The coefficient of x_1 represents **holding fixed other factors, the change in y when x_1 increases by one unit**.

- The case with more independent variables is similar:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + ... + \hat{\beta}_k x_k.$$

- Written in terms of changes:

$$\Delta\hat{y} = \hat{\beta}_1 \Delta x_1 + \hat{\beta}_2 \Delta x_2 + ... + \hat{\beta}_k \Delta x_k.$$

- The coefficient on x_1 measures the change in $\hat{y}$ due to a one-unit increase in x_1, holding all other independent variables fixed. That is,

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

holding $x_2, x_3, ..., x_k$ fixed.

- In lab experiments, we can directly control other factors and make sure they are the same.
- Multiple regression models allow us to do something similar in a non-experimental environment: by explicitly controlling other factors, we make sure that we are studying the effect of x_1 on y when x_2 is held constant, even though in our sample, x_2 may be different.

A “Partialling Out” Interpretation

- Consider the model

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + ... + \beta_k x_{ik} + u_i.$$

- Let $\hat{\beta}_1$ denote the OLS estimator we obtained either using the sample analogue method or minimizing the sum of residual squares
- The Frisch-Waugh-Lovell Theorem (FWL) theorem states that $\hat{\beta}_1$ can be obtained by the following two alternative procedures.
- These alternative procedures provide explicit formula for $\hat{\beta}_1$, as well as a way to interpret the coefficient.

Frisch-Waugh-Lovell Theorem

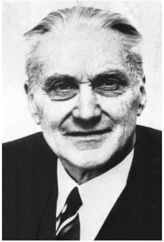


Figure: Ragnar Frisch

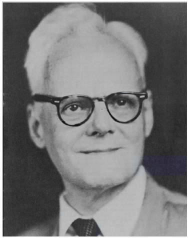


Figure: Frederick Waugh

Figure: Michael Lovell

Interlude: Note on Terminology

- The following three expressions are equivalent
  1. "regress y on x"
  2. "regress y on 1 and x"
  3. "run a regression of y on 1 and x"
- All of them means that we estimate the OLS regression line $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$
- Regression through the origin: we choose a slope parameter such that

$$\tilde{y} = \tilde{\beta}_1 x.$$

  Using the method of least squares:

$$\tilde{\beta}_1 = \frac{\sum_{i=1}^{N} x_i y_i}{\sum_{i=1}^{N} x_i^2}.$$

  $\tilde{\beta}_1 = \hat{\beta}_1$ if $\bar{x} = 0$ in the sample.

FWL Theorem: Version I

- $\hat{\beta}_1$ can be obtained by
  1. Regress x_1 on other independent variables and 1, obtain the residual $\hat{r}_{i1}$.
  2. Regress y on $\hat{r}_{i1}$. The resulting slope coefficient is $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \hat{r}_{i1} y_i}{\sum_{i=1}^{N} \hat{r}_{i1}^2}.$$

- In step 2, whether including the intercept does not matter because the sum of the residuals is zero.

FWL Theorem: Version I

How to understand?

- The residual $\hat{r}_{i1}$ are the part of x_1 that is uncorrelated with other x. It is x_1 after the effects of other x have been partialled out, or netted out.
- Thus, $\hat{\beta}_1$ measures the sample relationship between y and x_1 after other x has been partialled out.

FWL Theorem: Version II

- $\hat{\beta}_1$ can be obtained by
  1. Regress x_1 on other independent variables (including the constant), obtain the residual $\hat{r}_{i1}$.
  2. Regress y on other independent variables (including the constant), obtain the residual $\hat{r}_{iy}$.
  3. Regress $\hat{r}_{iy}$ on $\hat{r}_{i1}$. The resulting slope coefficient is $\hat{\beta}_1$.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N} \hat{r}_{i1}\hat{r}_{iy}}{\sum_{i=1}^{N} \hat{r}_{i1}^2}.$$

- In the final step, whether including the intercept does not matter because the sum of the residuals is zero.

FWL Theorem: Version II

How to understand?

- The residual $\hat{r}_{i1}$ are the part of x_1 that is uncorrelated with other x. It is x_1 after the effects of other x have been partialled out or netted out.
- The residual $\hat{r}_{iy}$ are the part of y that is uncorrelated with other x. It is y after the effects of other x have been partialled out or netted out.
- Thus, $\hat{\beta}_1$ measures the sample relationship between y and x_1 after other x has been partialled out.

FWL Theorem Proof

- The proof involves two parts
  1. Show that the two versions are equivalent
  2. Show that either is equivalent to the OLS estimator we obtained using the sample analogue or least squares method
- Let's prove part 1. We leave the proof of part 2 as an after-class exercise and present the solution in the notes (see the Web-Learning page).

Proof: Part I

- We want to show:

$$\frac{\sum_{i=1}^N y_i \hat{r}_{i1}}{\sum_{i=1}^N \hat{r}_{i1}^2} = \frac{\sum_{i=1}^N \hat{r}_{iy} \hat{r}_{i1}}{\sum_{i=1}^N \hat{r}_{i1}^2}.$$

- Plug in $y_i = \hat{\omega}_0 + \hat{\omega}_2 x_{i2} + \cdots + \hat{\omega}_k x_{ik} + \hat{r}_{iy}$:

$$\begin{aligned}\sum_{i=1}^N y_i \hat{r}_{i1} &= \sum_{i=1}^N (\hat{\omega}_0 + \hat{\omega}_2 x_{i2} + \cdots + \hat{\omega}_k x_{ik} + \hat{r}_{iy}) \hat{r}_{i1} \\ &= \hat{\omega}_0 \sum_{i=1}^N \hat{r}_{i1} + \hat{\omega}_2 \sum_{i=1}^N x_{i2} \hat{r}_{i1} + \cdots + \hat{\omega}_k \sum_{i=1}^N x_{ik} \hat{r}_{i1} + \sum_{i=1}^N \hat{r}_{iy} \hat{r}_{i1} \\ &= \sum_{i=1}^N \hat{r}_{iy} \hat{r}_{i1}.\end{aligned}$$

where we use the fact that $\sum_{i=1}^N x_{is} \hat{r}_{i1} = 0, \forall s = 2, ..., k$ and $\sum_{i=1}^N \hat{r}_{i1} = 0$.

Quiz: Factors Affecting Years of Education

Consider the following estimated model:

$$\widehat{edu} = 10.36 - 0.094 sibs + 0.131 meduc + 0.21 feduc.$$

edu means an individual's year of education, *sibs* represents the number of siblings, *feduc* and *meduc* represents father and mother's education respectively.

- Based on the model, what's my best guess for the education years of an individual who is a single child and whose parents both had 9 years of education?
- Discuss the meaning of 0.131 from a causal perspective.
- Both A and B are the only child in the family. A's father and mother had 12 years of education. B's father had 16 years of education. Mother had 9 years of education. What's the difference in their expected year of education?

Outline

Why We Need Multiple Regression Model?

Estimation and Interpretation

- Estimation
- Interpretation

Expected Values and Variances of the OLS Estimators

A Few Practical Issues

Unbiasedness of OLS

- **Unbiasedness** means the expectation of the estimator equals the true value.
- Similar to the simple linear regression model case, we discuss four assumptions under which OLS estimators are unbiased.

1. MLR.1 Linear in Parameters
2. MLR.2 Random Sampling
3. MLR.3 No Perfect Collinearity
4. MLR.4 Zero Conditional Mean: $E(u|x) = 0$

We first define the basic model:

MLR.1-Linear in Parameters

In the population model, the dependent variable, y, is related to the independent variable, x, and the error (or disturbance), u, as

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

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

MLR.2 Random Sampling

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

MLR.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.

- If an independent variable is an exact linear combination of the other independent variables, then we say the model suffers from perfect collinearity, and it cannot be estimated by OLS.
- MLR3 is different but related to the assumption that there is variation in x. If x has no variation, then it is perfectly correlated with the constant.

- Consider this example:

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

  There is no way to uncover β_1 separately from β_2.
- Similarly, we cannot separately estimate β_1 and β_2 in the following model:

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u, \\ x_2 &= \gamma_0 + \gamma_1 x_1. \end{aligned}$$

  Why?

$$\begin{aligned} y &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u \\ &= \beta_0 + \beta_1 x_1 + \beta_2(\gamma_0 + \gamma_1 x_1) + u \\ &= (\beta_0 + \beta_2\gamma_0) + (\beta_1 + \beta_2\gamma_1)x_1 + u. \end{aligned}$$

Quiz: Perfect Collinearity

Which of the following violates MLR.3?

1. $y = \beta_0 + \beta_1 female + \beta_2 male + u$
2. $y = \beta_1 female + \beta_2 male + u$
3. $q = \beta_0 + \beta_1 spring + \beta_2 summer + \beta_3 fall + \beta_4 winter + u.$
4. $cons = \beta_0 + \beta_1 inc + \beta_2 inc^2 + u.$

MLR.4 Zero Conditional Mean

The error u has an expected value of zero given any value of the explanatory variable. In other words,

$$E(u|x_1, x_2, ..., x_k) = 0.$$

Cases when MLR4 is violated:

- Misspecified functional form. e.g. the true model is $y = \beta_0 + \beta_1 log(x_1) + \beta_2 x_2 + u$, but we set $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. We'll discuss this in Chapter 9.
- Omitting an important variable correlated with $x_1, x_2, ..., x_k$.
- When Assumption MLR.4 holds, we often say that we have exogenous explanatory variables. If x_j is correlated with u for any reason, then x_j is said to be an endogenous explanatory variable.

The Unbiasedness of the OLS Estimator

Unbiasedness of OLS

Using assumptions MLR.1 through MLR.4,

$$E[\hat{\beta}_j] = \beta_j,$$

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

▸ Proof (not required)

The Variance of the OLS Estimator

We now obtain the variance of the OLS estimators.
We add another assumption:

MLR.5 Homoskedasticity

The error u has the same variance given any value of the explanatory variables. In other words,

$$Var(u|x_1, x_2, ..., x_k) = \sigma^2.$$

Assumptions MLR.1 through MLR.5 are collectively known as the **Gauss-Markov assumptions**.

Sampling variances of the OLS slope parameters

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

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}.$$

for $j = 1, 2, ..., k$, where $SST_j = \sum_{i=1}^{N}(x_{ij} - \bar{x}_j)^2$ 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).

▸ Proof (not required)

Estimating σ^2

- σ^2 is unknown, so we need the sample to estimate it. The unbiased estimator of σ^2 is:

$$\hat{\sigma}^2 = \frac{1}{N-k-1} \sum_{i=1}^{N} \hat{u}_i^2.$$

- Degree of freedom: $N-k-1$
- Thus, in obtaining the OLS estimates, $k+1$ restrictions are imposed on the OLS residuals. This means that given $N-(k+1)$ of the residuals, the remaining $k+1$ residuals are known: there are only $N-(k+1)$ degrees of freedom in the residuals.

Unbiased estimation of σ^2

Under MLR.1 through MLR.5,

$$E(\hat{\sigma}^2) = \sigma^2.$$

Standard Deviation and Standard Error of $\hat{\beta}_j$

Standard deviation	Standard error
$sd(\hat{\beta}_j)$	$se(\hat{\beta}_j)$
$\frac{\sigma}{[SST_j(1-R_j^2)]^{1/2}}$	$\frac{\hat{\sigma}}{[SST_j(1-R_j^2)]^{1/2}}$
$\sigma^2 = \text{variance of } u$	$\hat{\sigma}^2 = \frac{1}{N-k-1} \sum_{i=1}^{N} \hat{u}_i^2$
unknown	estimated using the sample

The Gauss-Markov Theorem

Why use OLS? They are the best linear unbiased estimator (BLUE).

- Linear: $\tilde{\beta}_j$ is a linear estimator of β_j if and only if it can be expressed as a linear function of the data on the dependent variable

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

  where each w_{ij} can be a function of the sample values of all the independent variables.

- Best: for the current theorem, best is defined as having the smallest variance. Given two unbiased estimators, it is logical to prefer the one with the smallest variance.

The Gauss-Markov Theorem

Why use OLS? They are the best linear unbiased estimator (BLUE).[1]

Gauss-Markov Theorem

Under Assumptions MLR.1 through MLR.5, $\hat{\beta}_0, \hat{\beta}_1, ..., \hat{\beta}_k$ are the best linear unbiased estimators of $\beta_0, \beta_1, ..., \beta_k$ respectively.

[1] There are some unsettled debates about whether OLS is the best unbiased estimator (BUE). See Hansen, Bruce. "A Modern Gauss-Markov Theorem," *Econometrica*, (2022), 90, 1283-1294.

Figure: Gauss

Figure: Markov

Outline

Why We Need Multiple Regression Model?

Estimation and Interpretation

- Estimation
- Interpretation

Expected Values and Variances of the OLS Estimators

A Few Practical Issues

A Few Practical Issues

We analyze three issues common in empirical research:

- Underspecifying the model: excluding a relevant variable
- Overspecifying the model: including irrelevant variables in a regression model
- Multicollinearity: high (but not perfect) correlation between two or more independent variables

Compare the simple regression model and the multiple regression model:

- Regress y on x_1. The estimated model is:

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

- Regress y on x_1 and x_2. The estimated model is:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2.$$

What's the relationship between coefficients in the two models?

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

where $\tilde{\delta}_1$ is the slope coefficient regressing x_{i2} on x_{i1}. Proof

Case 1: Omitted Variable Bias

- Omitted variable bias: excluding a relevant variable
- Assume the true model is $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$. Instead, we regress y on x_1 only, then this creates the omitted variable bias: we omit a relevant variable x_2.
- What's the impact on the estimation?
- $E(\tilde{\beta}_1) = \beta_1 + \beta_2 \tilde{\delta}_1$. Thus $Bias(\tilde{\beta}_1) = \beta_2 \tilde{\delta}_1$.

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

Example Return on Education

- Assume wage is determined by year of education and ability:

$$wage = \beta_0 + \beta_1 edu + \beta_2 ability + u.$$

- Because we do not observe *ability*, we do not include it in the regression.
- What's the impact on the estimation of β_1?

Example Return on Education

- Assume wage is determined by year of education and ability:

$$wage = \beta_0 + \beta_1 edu + \beta_2 ability + u.$$

- Because we do not observe *ability*, we do not include it in the regression.
- What's the impact on the estimation of β_1?
  - High-ability people often have a higher wage: $\beta_2 > 0$
  - High ability people often obtain higher education: $corr(edu, ability) > 0$
  - Thus, the bias is likely to be positive. In other words, we likely overestimate the impact of education on wages.

Case 2: Including Irrelevant Variables

- One (or more) of the independent variables is included in the model even though it has no partial effect on y in the population. (That is, its population coefficient is zero.)
- Mathematically, we say x_j is irrelevant if

$$E(y|x_1, ..., x_j, ..., x_n) = E(y|x_1, ..., x_{j-1}, x_{j+1}, ..., x_n).$$

  It means $\beta_j = 0$.
- Including x_j will not affect the unbiasedness of other estimators.

- Assume that the true model is $y = \beta_0 + \beta_1 x_1 + u$, instead, we regress y on x_1 and x_2
  - $\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$.
  - $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$, where $E(\hat{\beta}_2) = \beta_2 = 0$.
- Consider the variance of the two estimators:
  - $Var(\hat{\beta}_1) = \frac{\sigma^2}{SST_1(1-R_1^2)}$, where $SST_1 = \sum_{i=1}^{N}(x_{i1} - \bar{x_1})^2$, and R_1^2 is the R^2 obtained by regressing x_1 on x_2.
  - $Var(\tilde{\beta}_1) = \frac{\sigma^2}{SST_1}$.
- Unless x_1 is uncorrelated to x_2, $Var(\tilde{\beta}_1)$ is smaller than $Var(\hat{\beta}_1)$.
- We should not include irrelevant variables in the regression because it results in a larger variance.

Multicollinearity

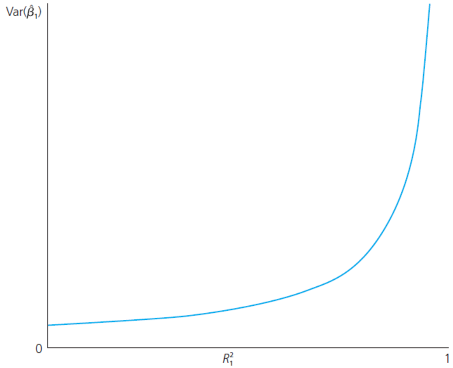
- Consider the variance of $\hat{\beta}_j$:

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{SST_j(1 - R_j^2)}.$$

- It depends on three factors:
  - σ^2: a larger σ^2 means larger variances for the OLS estimators. This is a feature of the population. It has nothing to do with the sample size.
  - SST_j: the larger the total variation in x_j is, the smaller the variance of $\hat{\beta}_j$. Increasing the sample size can increase SST_j.
  - The linear relationships among the independent variables, R_j^2: A large R_j^2 means other independent variables explains much of the variation in x_j.

Multicollinearity

- When R_j^2 increases, $Var(\hat{\beta}_j)$ also increases.
- If $R_j^2 \to 1$, then $Var(\hat{\beta}_j) \to \infty$.
- Multicollinearity: high (but not perfect) correlation between two or more independent variables



Recap

- Underspecifying the model (excluding a relevant variable): affect unbiasedness
- Overspecifying the model (including irrelevant variables in a regression model): affect the variance
- Multicollinearity: affect the variance

Summary

- Estimation
  - Sample analogue
  - Minimizing the sum squares of the residuals
  - FWL Theorem
- Interpretations: holding other factors fixed (partialling out)
- Gauss-Markov Theorem: BLUE
- Issues with the multiple linear regression model:
  - Underspecifying the model
  - Overspecifying the model
  - Multicollinearity

Proof of the Unbiasedness of $\hat{\beta}_1$

$$
\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_i \hat{r}_{i1} y_i}{\sum_i \hat{r}_{i1}^2} \\
&= \frac{\sum_i \hat{r}_{i1}(\beta_0 + \beta_1 x_{i1} + ... + \beta_k x_k + u_i)}{\sum_i \hat{r}_{i1}^2} \\
&= \beta_0 \frac{\sum_i \hat{r}_{i1}}{\sum_i \hat{r}_{i1}^2} + \beta_1 \frac{\sum_i \hat{r}_{i1}(\hat{x}_{i1} + \hat{r}_{i1})}{\sum_i \hat{r}_{i1}^2} + \beta_2 \frac{\sum_i \hat{r}_{i1} x_{i2}}{\sum_i \hat{r}_{i1}^2} + ... \\
&+ \beta_k \frac{\sum_i \hat{r}_{i1} x_{ik}}{\sum_i \hat{r}_{i1}^2} + \frac{\sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2} \\
&= \beta_1 + \frac{\sum_i \hat{r}_{i1} u_i}{\sum_i \hat{r}_{i1}^2}.
\end{aligned}
$$

$$
\begin{aligned}
E(\hat{\beta}_1|x_1, x_2, ..., x_k) &= \beta_1 + \frac{\sum_i \hat{r}_{i1} E(u_i|x_1, x_2, ..., x_k)}{\sum_i \hat{r}_{i1}^2} \\
&= \beta_1. \\
E(\hat{\beta}_1) &= E[E(\hat{\beta}_1|x_1, x_2, ..., x_k)] \\
&= E(\beta_1) = \beta_1.
\end{aligned}
$$

Here we can view $\hat{r}_{ij}$ as constants in the conditional mean because $\hat{r}_{ij}$ is a function of the independent variables.

▸ Back

Proof of the Variance of $\hat{\beta}_j$

We know that $\hat{\beta}_j = \frac{\sum_i \hat{r}_{ij} y_i}{\sum_i \hat{r}_{ij}^2}$

$$Var(\hat{\beta}_j|x) = Var(\frac{\sum_i \hat{r}_{ij} y_i}{\sum_i \hat{r}_{ij}^2}|x) = \frac{\sum_i \hat{r}_{ij}^2 Var(y|x)}{(\sum_i \hat{r}_{ij}^2)^2}$$
$$= \frac{\sum_i \hat{r}_{ij}^2 Var(u|x)}{(\sum_i \hat{r}_{ij}^2)^2} = \frac{\sigma^2 \sum_i \hat{r}_{ij}^2}{(\sum_i \hat{r}_{ij}^2)^2} = \frac{\sigma^2}{\sum_i \hat{r}_{ij}^2}.$$

Using the formula for R^2, we have

$$1 - R_j^2 = \frac{\sum_i \hat{r}_{ij}^2}{\sum_i (x_{ij} - \bar{x}_j)^2} = \frac{\sum_i \hat{r}_{ij}^2}{SST_j}.$$

So $\sum_i \hat{r}_{ij}^2 = (1 - R_j^2) SST_j$. Plug in we have

$$Var(\hat{\beta}_j|x) = \frac{\sigma^2}{(1 - R_j^2) SST_j}.$$

Back

Proof: Simple and Multiple Regression Coefficient

$$
\begin{aligned}
\tilde{\beta}_1 &= \frac{\frac{1}{N}\sum_{i=1}^{N}(x_{i1} - \bar{x_1})y_i}{\frac{1}{N}\sum_{i=1}^{N}(x_{i1} - \bar{x_1})^2} \\
&= \frac{\frac{1}{N}\sum_{i=1}^{N}(x_{i1} - \bar{x_1})(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \hat{u}_i)}{\frac{1}{N}\sum_{i=1}^{N}(x_{i1} - \bar{x_1})^2} \\
&= \hat{\beta}_1 + \hat{\beta}_2 \frac{\frac{1}{N}\sum_{i=1}^{N}(x_{i1} - \bar{x_1})x_{i2}}{\frac{1}{N}\sum_{i=1}^{N}(x_{i1} - \bar{x_1})^2} \\
&= \hat{\beta}_1 + \hat{\beta}_2 \tilde{\delta}_1.
\end{aligned}
$$

Back