- **简单回归分析有一个很大的缺陷：假定 SLR. 4在经验研究中很难成立：**

$$E(u_i|x_i) = 0$$

SLR.4要求u中包含的影响被解释变量y的其他因素与解释变量x均不相关

- 例：y：工资，x:教育，u中包含经验、智力等

- **此外，简单回归模型只能出现解释变量的一个函数（如x或log x等）**

- **接下来，我们讨论具有更好灵活性的多元回归模型**

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

$y = \beta_0 + \beta_1 x + \hat{u} \rightarrow \beta_2 x_2 + u$

$\sum \hat{u} = 0$

JEFFREY M. WOOLDRIDGE

Introductory Econometrics
A Modern Approach

SIXTH EDITION

Chapter 3 多元回归分析：估计

章节框架

- 在这一章中我们介绍多元回归模型的估计问题
- 首先，我们介绍多元回归模型
- 之后，我们提出多元回归模型的OLS估计方法
- 最后，我们讨论多元回归模型OLS估计值的性质：期望，方差和有效性

使用多元回归的动因

- **多元线性回归模型的定义**

"用变量 $x_1, x_2, \ldots, x_k$ 来解释变量 y "

截距项

斜率参数

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

因变量，被解释变量等等

自变量，解释变量等等

误差项，
扰动项，
不可见变量等等

使用多元回归的动因

- **多元回归的动机**
  - 将更多的解释因素纳入模型中
  - 固定其他因素的影响
  - 允许更灵活的函数形式
- **例子：工资等式**

在固定经验的条件下教育对工资的影响

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

其他因素…

每小时工资

教育时长（年）

劳动市场工作经验（年）

使用多元回归的动因

- **例子：平均考试成绩与每位学生的支出**

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

标准化的平均考试成绩

每位学生在校支出

在校学生的家庭平均收入

其他因素

  - 学生支出与家庭收入很大程度是相关的
  - 在回归中忽略家庭平均收入（avginc）会导致估计系数 β_1 有偏（支出对考试成绩的影响）
  - 在一个简单的回归模型中，学生人均支出的影响将部分包括家庭收入对考试成绩的影响

使用多元回归的动因

- **例子：家庭收入与家庭消费**

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

家庭消费

家庭收入

家庭收入的平方

其他因素

- 两个解释变量：收入与收入的平方
- 消费由收入的一个二次函数来解释

使用多元回归的动因

- **例子：CEO薪资、销售与CEO任期**

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

CEO薪资的对数

销售的对数

CEO任期的二次函数

$y = \beta_0 + \beta_1 x + \beta_2 x^2 + u$

  - 模型假设一个CEO薪资与其公司销售的不变弹性关系
  - 模型假设一个CEO薪资与他/她的任期的二次关系

- **“线性”回归的含义**
  - 模型是参数线性的，而不是变量

$\frac{\partial y}{\partial x} = \beta_1 + \beta_2 x$

$\beta_2 > 0$

使用多元回归的动因

$\frac{dy}{dx}$

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

- **多元线性回归模型系数的解释**

$\frac{\partial y}{\partial x_1}=\beta_1$

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

- $\frac{\Delta y}{\Delta x_1}=\beta_1$ **如果** $\Delta x_2=0$ ，… $\Delta x_k=0$ ， $\Delta u=0$

- **对于** β_j "Frisch – Waugh Theorem"

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

如果第j个自变量增加一个单位，保持所有其他自变量和误差项不变，因变量会发生多大变化

- 即使解释变量之间相关，多元回归模型也是固定其他解释变量来看其中一个的影响
- “其他条件不变（Ceteris paribus）”
- 仍然假设如果解释变量变化时误差项是不变的

使用多元回归的动因

- 例子：工资等式

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

- 保持exper和其他因素不变

$$\frac{\Delta wage}{\Delta educ} = \beta_1$$

- 例子：平均考试成绩与每位学生的支出

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

- 保持expend和其他因素不变

$$\frac{\Delta avgscore}{\Delta avginc} = \beta_2$$

使用多元回归的动因

- **但是，**

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

如果 x_2，⋯ x_k 中存在 x_1 的函数，可能导致

$$\frac{\Delta y}{\Delta x_1} \neq \beta_1$$

- **例子：家庭收入与家庭消费**

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

  - 解释系数时要小心：

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

如果收入增加一个单位，
消费会增长多少？

取决于已经有多少
收入了

普通最小二乘法的操作和解释

- 多元回归模型的OLS估计
- 随机样本

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

- 回归残差

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

- 最小化残差的平方和（as good as possible）

MSE

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

$$\frac{\partial MSE}{\partial \beta_0} = 0, \quad \frac{\partial MSE}{\partial \beta_1} = 0, \quad \ldots, \quad \frac{\partial MSE}{\partial \beta_k} = 0$$

普通最小二乘法的操作和解释

- **最小化残差的平方和：**

$$\min \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}\right)^2$$

- **一阶条件：**

$$\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}\right) = 0$$

$$\sum_{i=1}^{n} x_{i1}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}\right) = 0$$

$$\vdots$$

$$\sum_{i=1}^{n} x_{ik}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \cdots - \hat{\beta}_k x_{ik}\right) = 0$$

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \cdots + x_{iK}\beta_K + u_i, \qquad \boxed{\vec{x}_i = \begin{bmatrix} x_{i1} \\ \vdots \\ x_{iK} \end{bmatrix}}$$

$i = 1, \cdots, N$

N：观测数

K：自变量数、

$$y_i = \vec{x}_i{}'\beta + u_i. \qquad \beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}.$$

$$X = \begin{array}{c} \begin{array}{cccc} x_{i1} & x_{i2} & \cdots & x_{iK} \end{array} \\ \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1K} \\ \vdots & \vdots & \ddots & \\ x_{N1} & x_{N2} & \cdots & x_{NK} \end{bmatrix} \end{array} \qquad y = \begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix}, \qquad \underset{N\times 1}{y} = \underset{N\times K}{X}\,\underset{K\times 1}{\beta} + \underset{N\times 1}{u}$$

$N \times K$. $\qquad N\times 1 \qquad N\times 1$

$$y_i = x_{ik}'\beta_k + u_i \qquad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_K x_{iK} + u_i$$

$$\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_K x_{iK} + \hat{u}_i$$

$$y = X\beta + u$$

Sum of Squared Residuals (SSR)

$$\min_{\tilde\beta} SSR(\tilde\beta) = \min_{\tilde\beta} \sum_{i=1}^{N} u_i^2 = u'u \qquad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}$$

$$= (y - X\tilde\beta)'(y - X\tilde\beta)$$

$$= (y' - \tilde\beta' X')(y - X\tilde\beta)$$

$$= y'y - y'X\tilde\beta - \tilde\beta' X' y + \tilde\beta' X' X \tilde\beta$$

$$= y'y - \underbrace{2\underbrace{y'}_{1\times N}\underbrace{X}_{N\times K}}_{1\times K}\tilde\beta + \tilde\beta' X' X \tilde\beta$$

向量微分：

$$\underset{\sim}{\vec{a}} = \begin{bmatrix} a_1 \\ \vdots \\ a_k \end{bmatrix}, \quad \vec{a}'\tilde{\beta} = \sum_{i=1}^{k} a_i \tilde{\beta}_i = a_1\beta_1 + a_2\beta_2 + \cdots + a_k\beta_k$$

$\vec{a}': 1\times k$

$$\frac{\partial(\boxed{\vec{a}'}\tilde{\beta})}{\partial \tilde{\beta}_{k\times 1}} = \begin{bmatrix} \frac{\partial(\vec{a}'\tilde{\beta})}{\partial \beta_1} \\ \vdots \\ \frac{\partial(\vec{a}'\tilde{\beta})}{\partial \beta_k} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{bmatrix} = \underset{\sim}{\vec{a}}_{k\times 1}$$

① $\frac{\partial \vec{a}'\tilde{\beta}}{\partial \tilde{\beta}} = \vec{a}$

② A 为 K 阶对称矩阵，
$$\frac{\partial(\tilde{\beta}'A\tilde{\beta})}{\partial \tilde{\beta}} = \begin{bmatrix} \frac{\partial(\tilde{\beta}'A\tilde{\beta})}{\partial \tilde{\beta}_1} \\ \vdots \\ \frac{\partial(\tilde{\beta}'A\tilde{\beta})}{\partial \tilde{\beta}_k} \end{bmatrix} = 2A\tilde{\beta}$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \cdots & & & \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{bmatrix}$$

$$\tilde{\beta}' A \tilde{\beta} = \underline{\beta_1^2 a_{11} + \beta_1\beta_2 a_{12} + \cdots + \beta_1\beta_K a_{1K}}$$
$$+ \beta_2\beta_1 a_{21} + \beta_2^2 a_{22} + \cdots + \beta_2\beta_K a_{2K}$$
$$+ \cdots +$$
$$+ \beta_K\beta_1 a_{K1} + \beta_K\beta_2 a_{K2} + \cdots + \beta_K^2 a_{KK}$$

$$\frac{\partial}{\partial \beta_1} = 2a_{11}\beta_1 + \beta_2 \boxed{a_{12}} + \cdots + \beta_K a_{1K}$$
$$+ \beta_2 \boxed{a_{21}} + \beta_3 a_{31} + \cdots + \beta_K a_{K1}$$

$$= 2a_{11}\beta_1 + 2a_{12}\beta_2 + \cdots + 2a_{1K}\beta_K$$

$$y'y - 2y'X\tilde{\beta} + \tilde{\beta}'X'X\tilde{\beta}$$

$$\begin{bmatrix} 1 & x_{11} \\ \vdots & \vdots \\ 1 & x_{n1} \end{bmatrix} \quad X = \begin{bmatrix} 1 & \cdots & x_{1k} \\ \vdots & & \\ 1 & \cdots & x_{nk} \end{bmatrix}$$

$$\frac{\partial SSR}{\partial \tilde{\beta}} = -\cancel{2}X'y + \cancel{2}X'X\tilde{\beta} = 0$$

$$\sum (x_i - \bar{x})^2 > 0$$

$$\underset{k \times k}{(X'X)}\tilde{\beta} = X'y$$

① $$\boxed{\hat{\tilde{\beta}} = (X'X)^{-1}X'y}$$

条件： $X'X$ 可逆

|||

没有多重共线性

$\downarrow n$

N

$$\hat{\beta} = (X'X)^{-1}X'y, \quad X'X = \sum_{i=1}^{n} \vec{x}_i \vec{x}_i', \quad X'y = \sum_{i=1}^{n} \vec{x}_i y_i$$

$$\hat{\beta} = \left(\sum_{i=1}^{n} \vec{x}_i \vec{x}_i'\right)^{-1} \left(\sum_{i=1}^{n} \vec{x}_i y_i\right)$$

② $\hat{u} = y - \hat{y} = y - X\hat{\beta}$. $X'(X\hat{\beta} - y) = 0$.

$X'(y - X\hat{\beta}) = 0$. $X'\hat{u} = 0$ ✓

Normal Equation.

$$X = \begin{bmatrix} 1 & x_{11} & & x_{1k} \\ \vdots & \vdots & \cdots & \vdots \\ 1 & x_{n1} & & x_{nk} \end{bmatrix} \qquad \hat{u} = \begin{bmatrix} \hat{u}_1 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

$$\begin{cases} \sum_{i=1}^{n} \hat{u}_i = 0 \\ \sum_{i=1}^{n} x_{i1}\hat{u}_i = 0 \\ \cdots \\ \sum_{i=1}^{n} x_{ik}\hat{u}_i = 0 \end{cases} \quad \not\Leftarrow \quad E[u|X] = 0$$

$E[u] = 0$.

Ch 2. (2-5-2)

③ $\hat{\beta} = (X'X)^{-1}X'y = (X'X)^{-1}X'(X\beta + u) = \beta + (X'X)^{-1}X'u$ ☆

投影矩阵

$$\hat{y} = X\hat{\beta} = X((X'X)^{-1}X'y) = (X(X'X)^{-1}X')y$$

$$P \equiv X(X'X)^{-1}X'$$

P.S → X平面上

projection

消灭矩阵. Annihilator.

$$\hat{u} = y - \hat{y} = y - Py = (I_n - P)y = My$$

$$M \equiv I_n - P.$$

P和M的性质： 1) $PX = X$ 2) $P\hat{u} = 0$ 3) $MX = 0$

4) $P' = P$ 5) $M' = M$ 6) $P^2 = P$ 7) $M^2 = M$

Idempotent.

普通最小二乘法的操作和解释

- 定义$X_i=(1,x_{i1},x_{i2},\ldots,x_{ik})'$和 $\hat{\beta}=(\hat{\beta}_0,\hat{\beta}_1,\hat{\beta}_2,\ldots,\hat{\beta}_k)'$
- 一阶条件：

$$\sum_{i=1}^{n} X_i\left(y_i-X_i{}'\hat{\beta}\right)=0$$

- OLS表达式：

$$\hat{\beta}=\left(\sum\nolimits_{i=1}^{n} X_i X_i{}'\right)^{-1}\sum\nolimits_{i=1}^{n} X_i y_i$$

- 拟合值或预测值：

$$\hat{y}_i=X_i{}'\hat{\beta}$$

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

普通最小二乘法的操作和解释

- 对OLS估计值的解释：

$$\Delta\hat{y}_i = \hat{\beta}_1 \Delta x_{i1} + \hat{\beta}_2 \Delta x_{i2} + \cdots + \hat{\beta}_k \Delta x_{ik}$$

- $\hat{\beta}_1$具有偏效应（partial effect）或其他条件不变的解释：

如果

$$\Delta x_{i2} = \Delta x_{i3} = \cdots = \Delta x_{ik} = 0$$

我们有

$$\Delta\hat{y}_i = \hat{\beta}_1 \Delta x_{i1}$$

普通最小二乘法的操作和解释

- **例子：大学GPA的决定因素**

$$\underbrace{\widehat{colGPA}}_{\text{大学平均成绩}} = 1.29 + .453\,\underbrace{hsGPA}_{\text{高中平均成绩}} + .0094\,\underbrace{\boxed{ACT}}_{\text{能力测验分数}}$$

- **解释**
  - 固定ACT不变，hsGPA变化1分，colGPA变化0.453分
  - 或者：如果我们比较两个具有相同ACT的学生，如果A学生比B学生的hsGPA高1分，那么我们预测A学生比B学生的colGPA高0.453分
  - 固定hsGPA不变，ACT提高1分，colGPA变化不到0.01分

普通最小二乘法的操作和解释

- **例子：小时工资**

$$\widehat{\log(wage)} = .284 + .092\ educ + .0041\ exper + .022\ tenure$$
$$n = 526.$$

小时工资的对数

受教育年数

工作经历

任现职的任期

- 固定exper和tenure不变，educ变化1年，log wage变化0.092,即wage变化9.2%
- 或者：如果我们比较两个具有相同exper和tenure的工人，如果A工人比B工人的educ多1年，那么我们预测A工人比B工人的工资高9.2%
- 固定educ和tenure不变，exper变化1年，log wage变化0.0041，即wage变化0.41%
- **同时改变不止一个变量：如果**exper和tenure都增加一年，保持educ 不变：

$$\Delta\widehat{\log(wage)} = .0041\ \Delta exper + .022\ \Delta tenure = .0041 + .022 = .0261$$

普通最小二乘法的操作和解释

- **拟合值和残差**

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

拟合值或预测值

残差

- **OLS回归的代数性质**

$$\sum_{i=1}^{n} \hat{u}_i = 0 \qquad \sum_{i=1}^{n} x_{ij}\hat{u}_i = 0 \qquad \bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}_1 + \ldots + \hat{\beta}_k \bar{x}_k$$

与回归线的偏差总和为零（有截距项）

偏差和回归系数之间的协方差为零

y的样本均值与x的样本均值落在回归线上（有截距项）

普通最小二乘法的操作和解释

- 定义：$Y=(y_1,y_2,\dots,y_n)'$, $\boldsymbol{X}=(X_1,X_2,\dots,X_k)'$， $U=(u_1,u_2,\dots,u_n)'$

  和$\beta=(\beta_0,\beta_1,\beta_2,\dots,\beta_k)'$

- 多元回归模型的矩阵表达式：

$$Y=\boldsymbol{X}\beta+U$$

- OLS的矩阵表达式：

$$\hat{\beta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'Y$$

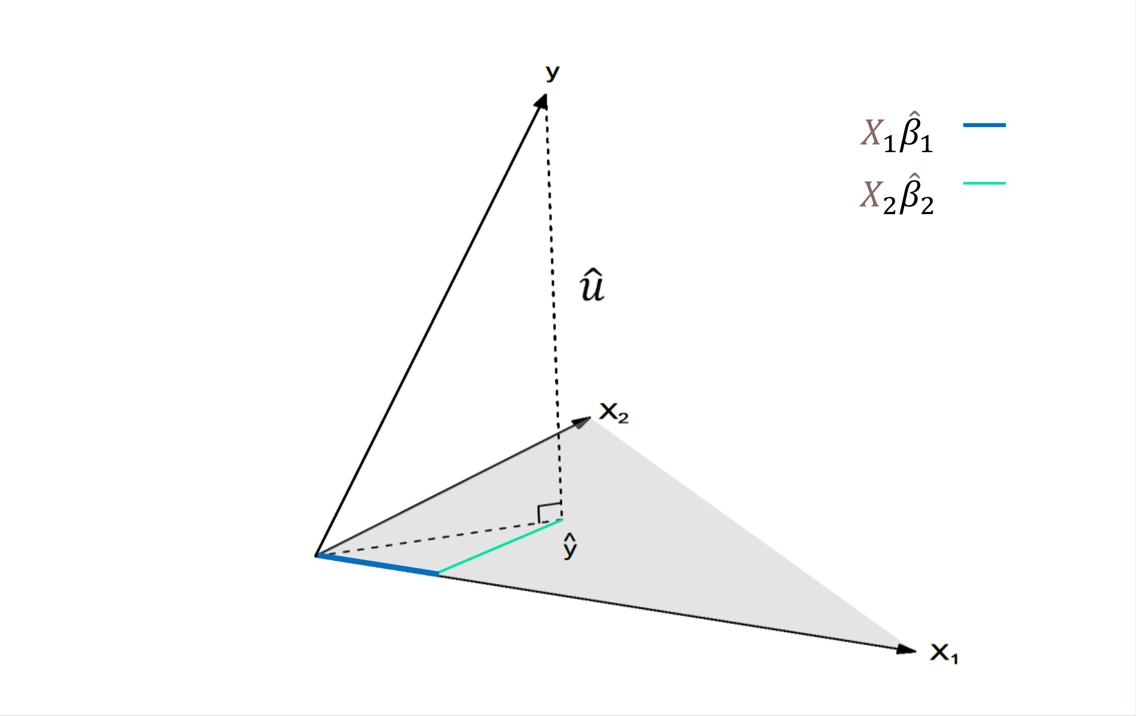
- 定义：$\hat{Y}=(\hat{y}_1,\hat{y}_2,\dots,\hat{y}_n)'$和$\widehat{U}=(\hat{u}_1,\hat{u}_2,\dots,\hat{u}_n)'$，我们有

$$\hat{Y}=\boldsymbol{X}\hat{\beta}=\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'Y$$

$$\widehat{U}=Y-\hat{Y}=\underbrace{(\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')}_{M}Y$$

普通最小二乘法的操作和解释

- **多元回归的“投影”解释：**



思考(1)什么情况下$\hat{\beta}_2 = 0$(2)什么情况下$\hat{\beta}_1 X_1$就是y直接对X_1的投影?

普通最小二乘法的操作和解释

- 简单回归与多元回归估计值的比较 P62 (3.23)式
- 在两个特殊情况下，y_i对x_{i1}做简单回归的$\tilde{\beta}_1$等于y_i对x_{i1}和x_{i2}做多元回归得到的$\hat{\beta}_1$：

（1） x_{i2}对y_i的偏效应为零（ $\hat{\beta}_2 = 0$）

（2） x_{i1}和x_{i2}不相关

注意：实际中$\hat{\beta}_1$与$\tilde{\beta}_1$几乎不可能完全相等

- 例子：养老金参与率（*prate*）与企业匹配率（*mrate* ）

$$\widehat{prate} = 80.12 + 5.52\, mrate + .243\, age$$

$$\widehat{prate} = 83.08 + 5.86\, mrate$$

其中，*mrate*与*age*的相关性仅为*0.12*

普通最小二乘法的操作和解释

- 多元回归“排除其他变量影响”的解释（**Frisch-Waugh theorem**）：
- 可以将多元回归的估计系数解释为如下两步：

$\hat{\beta}_1$

（1）用一个解释变量对其他解释变量做回归

$x_{i1}=\tilde{\delta}_0+\tilde{\delta}_1 x_{i2}+\boxed{\hat{r}_{i1}}$

（2）将被解释变量对（1）中残差做回归

$\beta=(X'X)^{-1}X'y$

例子，二元情况下：（1） x_{i1}对x_{i2}做简单回归得到残差 $\hat{r}_{i1}$

$y_i=\tilde{\beta}_0+\hat{\beta}_1\tilde{r}_{i1}+v_i$

（2） y_i对 $\hat{r}_{i1}$ 做简单回归得到$\hat{\beta}_1$

P62 (3.22)式

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

为什么？

- 第一步得到的残差$\hat{r}_{i1}$ 中包含x_{i1}中与x_{i2} 不相关的信息
- 第二步估计的系数$\hat{\beta}_1$代表x_{i1}对y_i的“单独”的影响

Hayashi (2000)

- First partition the data matrix in the following way:

$$X = \begin{bmatrix} x_{10} & x_{11} & \cdots & x_{1K-1} \\ \vdots & & & \\ x_{n0} & x_{n1} & \cdots & x_{nK-1} \end{bmatrix}$$

$$\boldsymbol{X} = \left[\boldsymbol{X}_1 \vdots \boldsymbol{X}_2 \right], \tag{81}$$

where $\boldsymbol{X}$ is the original $n \times K$ data matrix, and $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are $n \times K_1$ and $n \times K_2$ data matrices, respectively.

- Correspondingly, $\boldsymbol{\beta}$ is partitioned as

$$y = X\beta + \varepsilon$$

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \end{bmatrix} \tag{82}$$

$$y = [X_1 \; X_2] \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \varepsilon$$

- We are interested in the parameters of $\boldsymbol{\beta}_2$.
- The regression model can be written as

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{\epsilon} \tag{83}$$

x_{i5} $\qquad$ $M_1 x_5 \quad M_1 x_6 \cdots M_1 x_k$

- Define

$$X_2 = \begin{bmatrix} x_{15} & \cdots & x_{1k} \\ & \cdots & \\ x_{n5} & \cdots & x_{nk} \end{bmatrix}$$

$$\boldsymbol{P}_1 \equiv \boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1' \tag{84}$$

$$\boldsymbol{M}_1 \equiv \boldsymbol{I}_n - \boldsymbol{P}_1 \tag{85}$$

$$\tilde{\boldsymbol{X}}_2 \equiv \boldsymbol{M}_1\boldsymbol{X}_2 \tag{86}$$

(the kth column of $\tilde{\boldsymbol{X}}_2$ represents the vector of residuals when the kth column of $\boldsymbol{X}_2$ is regressed on $\boldsymbol{X}_1$)

$$\tilde{\boldsymbol{y}} \equiv \boldsymbol{M}_1\boldsymbol{y} \tag{87}$$

(the residuals from a regression of $\boldsymbol{y}$ on $\boldsymbol{X}_1$)

- The normal equations:

$\hat{\beta}$ $(X'X)^{-1}X'y$

$X = [X_1 \; X_2]$

$$X'Xb = X'y \tag{88}$$

$$\begin{bmatrix} X_1' \\ X_2' \end{bmatrix} \begin{bmatrix} X_1 \vdots X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1' \\ X_2' \end{bmatrix} y \tag{89}$$

$$\begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} \tag{90}$$

- Equation (90) can be written as

$$\begin{cases} X_1'X_1b_1 + X_1'X_2b_2 = X_1'y & (91) \\ X_2'X_1b_1 + X_2'X_2b_2 = X_2'y & (92) \end{cases}$$

- Our goal is to solve for b_2.

$X_1\cancel{(X_1'X_1)^{-1}X_1'X_1}b_1 + \underbrace{X_1(X_1'X_1)^{-1}X_1'}_{P_1}X_2b_2 = \underbrace{X_1(X_1'X_1)^{-1}X_1'}_{P_1}y$

- Premultiply equation (91) by $\boldsymbol{X}_1(\boldsymbol{X}_1'\boldsymbol{X}_1)^{-1}$ and solve for $\boldsymbol{X}_1\boldsymbol{b}_1$, we have

$$\boldsymbol{X}_1\boldsymbol{b}_1 = \boxed{-\boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{b}_2 + \boldsymbol{P}_1\boldsymbol{y}} \tag{93}$$

- Substitute this expression into (92):

$$\boldsymbol{X}_2'(-\boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{b}_2 + \boldsymbol{P}_1\boldsymbol{y}) + \boldsymbol{X}_2'\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'\boldsymbol{y} \tag{94}$$

$$\boldsymbol{X}_2' I_n \boldsymbol{X}_2\boldsymbol{b}_2 - \boldsymbol{X}_2'\boldsymbol{P}_1\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'\boldsymbol{y} - \boldsymbol{X}_2'\boldsymbol{P}_1\boldsymbol{y} \tag{95}$$

$M_1 = M_1'$

$M_1^2 = M_1$

$$\boldsymbol{X}_2'(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'(\boldsymbol{I}_n - \boldsymbol{P}_1)\boldsymbol{y} \tag{96}$$

$$\boldsymbol{X}_2'(\boldsymbol{M}_1^2)\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'(\boldsymbol{M}_1)\boldsymbol{y} \tag{97}$$

$M_1^2 = M_1$

$$\boldsymbol{X}_2'(\boldsymbol{M}_1'\boldsymbol{M}_1)\boldsymbol{X}_2\boldsymbol{b}_2 = \boldsymbol{X}_2'(\boldsymbol{M}_1'\cancel{\boldsymbol{M}_1})\boldsymbol{y} \tag{98}$$

$M_1' = M_1$

(since $\boldsymbol{M}_1$ is symmetric and idempotent)

$(M_1X_2)'(M_1X_2)b_2 = (M_1X_2)'M_1y$

$$\tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{X}}_2\boldsymbol{b}_2 = \tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{y}} \tag{99}$$

- Therefore, K_2

$$(X'X)^{-1}X'y$$

$$\underbrace{\boldsymbol{b}_2 = (\tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{X}}_2)^{-1}\tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{y}}}_{} \quad (100)$$

- $\tilde{\boldsymbol{X}}_2'\tilde{\boldsymbol{X}}_2 = \boldsymbol{X}_2'\boldsymbol{M}_1'\boldsymbol{M}_1\boldsymbol{X}_2$ is invertible by the full rank assumption.
  - Essentially, if $\boldsymbol{X}_2$ (and, by extension, $\boldsymbol{M}_1\boldsymbol{X}_2$, which is the component of $\boldsymbol{X}_2$ orthogonal to $\boldsymbol{X}_1$) does not have full column rank, $\boldsymbol{X}$ cannot have full column rank either.
- This solution for $\boldsymbol{b}_2$ is an important result known as the **Frisch-Waugh Theorem**.
- It states that you can obtain the coefficients of $\boldsymbol{b}_2$ in this way:
  1. Regress y on the variables in $\boldsymbol{X}_1$, and obtain the residuals.
  2. Regress each of the variables in $\boldsymbol{X}_2$ on all of the variables in $\boldsymbol{X}_1$, and obtain the residuals.
  3. Regress the residuals in (1) on the residuals in (2).

b_2是1维. $\tilde{\beta}_1 = (\sum_{i=1}^{n} \tilde{r}_{i1}^2)^{-1}(\sum_{i=1}^{n} \tilde{r}_{i1} y_i)$ → 一元

☆ $x_{i1} \sim 1, x_{i2}, \cdots, x_{ik-1}$ 做回归, 得残差 $\tilde{r}_{i1}$ 辅助回归

- Conceptually, this process of regressing $\boldsymbol{y}$ and the variables in $\boldsymbol{X}_2$ on $\boldsymbol{X}_1$, then taking the residuals is known as **partialling out** or **netting out** the effect of $\boldsymbol{X}_1$.
- For this reason, the coefficients in a multiple regression are sometimes called **partial regression coefficients**.
- As a corollary, when $\boldsymbol{X}_2$ and $\boldsymbol{X}_1$ are orthogonal (so that $\boldsymbol{M}_1\boldsymbol{X}_2 = \boldsymbol{X}_2$), no partialling out is necessary – you can just regress $\boldsymbol{y}$ on $\boldsymbol{X}_2$ directly and the variables in $\boldsymbol{X}_1$ will not make any difference at all.

1) 如果 x_1 和 $1, x_2, \cdots, x_{k-1}$ 都没关系. 此时辅助回归的 $\tilde{\delta}_0, \tilde{\delta}_2, \cdots, \tilde{\delta}_{k-1} = ?$

2) $\tilde{r}_{i1}$?

普通最小二乘法的操作和解释

- **拟合优度**
- **分解总体波动**

$$\overset{y}{SST} = \overset{\hat{y}}{SSE} + \overset{\hat{u}}{SSR}$$

总体波动　　被解释部分　　未被解释部分

$$SST \equiv \sum_{i=1}^{n} (y_i - \bar{y})^2 \qquad SSE \equiv \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2 \qquad SSR \equiv \sum_{i=1}^{n} \hat{u}_i^{\,2}$$

普通最小二乘法的操作和解释

- R2: 有截距项

$$R^2 \equiv SSE/SST = 1 - SSR/SST$$

$\hat{y}$ y $1-\frac{\hat{u}}{y}$

→ "SSR=SST(1-R²)"

- 注意：增加解释变量一定会提高R2

- 因为*SSR* 会变小：考虑OLS (k=1)：$\tilde{\beta}_0$, $\tilde{\beta}_1$和OLS (k=2)：$\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$

$$SSR(k=2) = min\sum_{i=1}^{n}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2})^2$$

$$\leq \sum_{i=1}^{n}(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1} - 0x_{i2})^2$$

$$= \sum_{i=1}^{n}(y_i - \tilde{\beta}_0 - \tilde{\beta}_1 x_{i1})^2$$

$$=SSR(k=1)$$

普通最小二乘法的操作和解释

- R2的另一种表达

$$R^2 = \frac{\left(\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})\right)^2}{\sqrt{\left(\sum_{i=1}^{n}(y_i - \bar{y})^2\right)}\sqrt{\left(\sum_{i=1}^{n}(\hat{y}_i - \bar{\hat{y}})^2\right)}}$$

- R2等于被解释变量的实际值y_i与其拟合值$\hat{y}_i$的样本相关系数的平方

$$(\mathrm{Corr}(y,\hat{y}))^2 = ①\ \frac{(y-\bar{y})'(\hat{y}-\bar{y})(y-\bar{y})'(\hat{y}-\bar{y})}{(y-\bar{y})'(y-\bar{y})(\hat{y}-\bar{y})'(\hat{y}-\bar{y})}$$

$$R^2 = \frac{(\hat{y}-\bar{y})'(\hat{y}-\bar{y})}{(y-\bar{y})'(y-\bar{y})} = ②\ \frac{(\hat{y}-\bar{y})'(\hat{y}-\bar{y})(\hat{y}-\bar{y})'(\hat{y}-\bar{y})}{(y-\bar{y})'(y-\bar{y})(\hat{y}-\bar{y})'(\hat{y}-\bar{y})}$$

分母分子都乘

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

①式分子： $(\hat{y}-\bar{y}+\hat{u})'(\hat{y}-\bar{y})(\hat{y}-\bar{y}+\hat{u})'(\hat{y}-\bar{y})$

$\hat{y}'\hat{u}=0$， $\bar{y}'\hat{u}=0$

普通最小二乘法的操作和解释

- **例子：解释拘捕记录**

1986年被拘捕的次数

86年前被捕导致定罪的比例（单位不是%）

86年在监狱的月数

86年被雇佣的季度数

$$\widehat{narr}86 = .712 - .150\ pcnv - .034\ ptime86 - .104\ qemp86$$

$$n = 2,725, \quad R^2 = .0413$$

- **Interpretation:**
  - 如果定罪比例提高0.5，则预计拘捕会减少7.5次
  - 监禁时间从0提高到12月，则预计拘捕次数少0.034×12=0.408次
  - 延长一个季度的合法就业，会使预计的拘捕次数减少0.104次

普通最小二乘法的操作和解释

- **例子：解释拘捕记录（续）**
  - 加入另一个解释变量：

x：是否进入处理组
直接回归 y~x

$$\widehat{narr86} = .707 - .151\, pcnv + .0074\, avgsen - .037\, ptime86 - .103\, qemp86$$

$$n = 2,725, \quad R^2 = .0422$$

此前定罪被宣判的平均时长

R2微弱增长

- Interpretation:
  - 判刑时长增加提高了被拘捕次数？
  - R2的微弱增加说明了该变量有限的额外解释能力

R^2.

- **对R2的注意**
  - 即使R2很小，这个回归仍然提供了一个对其他条件不变（ceteris paribus）效应的很好的估计（P65，3-2i前一段）

OLS估计量的期望值

无偏.

$E[u|X] = 0?$

- 多元回归模型的标准假设（高斯-马尔科夫假定）

$\begin{cases} E(\hat{\beta}) \overset{?}{=} \beta \\ \text{plim}\, \hat{\beta} \overset{?}{\longrightarrow} \beta \end{cases}$

- 假定 MLR.1 （线性于参数） Unbiased.

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

在总体中，y和x的关系是线性的

$Var(\hat{\beta})$ 大，小.

- 假定 MLR.2 （随机抽样） 默认.

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

数据从总体样本中随机抽样得到

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

每一个数据点都服从总体等式

OLS估计量的期望值

- **多元回归模型的标准假设（续）**

多重共线性

$\hat{\beta} = \boxed{(X'X)^{-1}} X'y$

- **假定 MLR.3 （不存在完全共线性）**

$\boxed{(X'X)^{-1} \text{ 存在.}}$

“在样本（因而在总体中），没有一个自变量是常数，自变量之间也不存在严格的线性关系”（注意：本书将截距项的1向量不看做自变量）

$N > K$

$X = [x_1 \cdots x_k]$ 满列秩.

- **对 MLR.3的评论**

若不可逆，

$\exists v \neq 0$,

s.t. $(X'X)v = 0$

$v'(X'X)v = 0$

$v \neq 0$

$x_j = \sum a_k x_k$,

a_k 不全为0.

  - 本假定只排除完全共线性；相关是允许的
  - 如果一个解释变量是其他解释变量的线性组合，则删掉它
  - 常数也不能作为解释变量（因为已经有截距项了）

$X'X$ 正定.

A

$\forall x \neq 0$, $x'Ax > 0$

$a'X'Xa > 0$, $a \neq 0$.

OLS估计量的期望值

- **完全共线性的例子：小样本**

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

在小样本中，avginc可能恰好是expend的精确的倍数；此时很难区分它们对被解释变量的影响

- **完全共线性的例子：解释变量间的关系**

$$voteA = \beta_0 + \beta_1 expendA + \beta_2 expendB + \beta_3 totexpend + u$$

expendA或expendB, totexpend应该删掉一个，因为: expendA + expend = totexpend

$$voteA = \beta_0 + \beta_1 shareA + \beta_2 shareB + u$$

shareA=expendA/totexpend 或shareB =expendB/totexpend应该删掉一个:
shareA + shareB = 1

OLS估计量的期望值

- **多元回归模型的标准假设（续）**
- **假定 MLR.4 （零条件均值）**

独立 α

$$E(u_i|x_{i1}, x_{i2}, \ldots, x_{ik}) = 0$$

解释变量必须不包含任何有关误差项均值的信息

  - 在多元回归模型中，零条件均值假设更可能成立，因为误差项中的因素更少了

- **例子：平均考试分数**

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

如果avginc没有作为解释变量，它将在误差项中；那么expend与u就相关，OLS估计值有偏

OLS估计量的期望值

- **零条件均值的讨论**
  - 解释变量与误差项的相关被叫做内生的（endogenous）；内生性违背了MLR.4
  - 解释变量与误差项不相关被叫做外生的（exogenous）；外生性服从MLR.4
  - 外生性是因果识别的关键假设，也是OLS估计值无偏的关键假设
- **定理 3.1（OLS的无偏性）**

$$\text{"}MLR.1\!-\!MLR.4\text{"} \quad \Rightarrow \quad E(\hat{\beta}_j) = \beta_j, \quad j = 0, 1, \ldots, k$$

OLS估计量的期望值

- 多元回归模型的矩阵表达式：

$$Y = \boldsymbol{X}\beta + U$$

$$y = X\beta + u$$

- OLS的矩阵表达式：

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'Y$$

$$E[\hat{\beta}] = \beta + E[(X'X)^{-1}X'u]$$

$$= \beta + E[E[(X'X)^{-1}X'u \mid X]]$$

$$= \beta + E[(X'X)^{-1}X' \underbrace{E(u \mid X)}_{=0}]$$

$$= \beta$$

- OLS的无偏性：

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\beta + U)$$

$$\hat{\beta} = \beta + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'U$$

$$E(\hat{\beta} \mid \boldsymbol{X}) = \beta$$

- 无偏性是重复样本的平均性质；在给定的样本中，估计值可能仍然远离真实值

OLS估计量的期望值

真实：$y=\beta_0+\beta_1x_1+\beta_2x_2+v$. $E[v|x_1,x_2]=0$

$E[y|x_1,x_2,x_3]=E[y|x_1,x_2]$

$=\beta_0+\beta_1x_1+\beta_2x_2$.

- **回归中包含无关变量**

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\beta_3x_3+u$$

= 0 在总体中

没问题，因为 $E(\hat{\beta}_3)=\beta_3=0$

$E[u|x_1,x_2,x_3]\overset{?}{=}0$

但是，

不用管→然而，包含无关变量可能会增加抽样的方差（3.4小节）

$=E[y-\beta_0-\beta_1x_1-\beta_2x_2|x_1,x_2]$

因此 $E(\hat{\beta}_3)=\beta_3=0$.

- **遗漏变量：简单情形**

$$y=\beta_0+\beta_1x_1+\beta_2x_2+u$$ 真实样本（包含x_1和x_2）

$$\tilde{y}=\tilde{\beta}_0+\tilde{\beta}_1x_1$$ 估计模型（x_2是遗漏的）

① $y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_{k-1} x_{ik-1} + \hat{\beta}_k x_{ik} + \hat{u}_i$

② $y_i = \tilde{\beta}_0 + \tilde{\beta}_1 x_{i1} + \cdots + \tilde{\beta}_{k-1} x_{ik-1} + \tilde{u}_i$

辅助回归：$x_{ik} = \tilde{\delta}_0 + \tilde{\delta}_1 x_{i1} + \tilde{\delta}_2 x_{i2} + \cdots + \tilde{\delta}_{k-1} x_{ik-1} + \tilde{r}_i$

结论：$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j$ $\quad j=0,\cdots,k-1$

对 $j=1$ 证明 $\tilde{\beta}_1 = \hat{\beta}_1 + \hat{\beta}_k \tilde{\delta}_1$

证明：辅助回归 $x_{i1} \sim 1, x_{i2}, \cdots, x_{ik-1}$，得残差 $\tilde{r}_{i1}$

$\check{\gamma}_0, \check{\gamma}_2, \cdots \check{\gamma}_{k-1}$

$\sum \tilde{r}_{i1} = 0$, (i) $\quad \sum x_{i2}\tilde{r}_{i1} = \cdots = \sum x_{ik-1}\tilde{r}_{i1} = 0$ (ii)

$$\tilde{\beta}_1 = \left(\sum_{i=1}^{n} \tilde{r}_{i1}^2\right)^{-1} \left(\sum_{i=1}^{n} \tilde{r}_{i1} y_i\right)$$

$\tilde{r}_{i1} = x_{i1} - \check{\gamma}_0 - \check{\gamma}_2 x_{i2} - \cdots - \check{\gamma}_{k-1} x_{ik-1}$

(iii)

$x_{i1} = \tilde{r}_{i1} + \check{\gamma}_0 + \check{\gamma}_2 x_{i2} + \cdots + \check{\gamma}_{k-1} x_{ik-1}$

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

$$\sum_{i=1}^{n} \tilde{r}_{i1} y_i = \sum_{i=1}^{n} \tilde{r}_{i1}\left(\hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + \hat{u}_i\right)$$

(i) $\sum \tilde{r}_{i1}\hat{\beta}_0 = 0$, (ii) $\sum \tilde{r}_{i1} x_{i2} = \sum \tilde{r}_{i1} x_{i3} = \cdots = \sum \tilde{r}_{i1} x_{ik-1} = 0$

$\sum \tilde{r}_{i1}\hat{u}_i = 0$，因为 $\hat{u}_i$ 和 $1, x_{i1}, \cdots, x_{ik}$ 有 Normal Equation

$$= \sum \tilde{r}_{i1}\hat{\beta}_1 x_{i1} + \sum \tilde{r}_{i1}\hat{\beta}_k x_{ik}$$

$$= \hat{\beta}_1\left(\sum \tilde{r}_{i1} x_{i1}\right) + \hat{\beta}_k\left(\sum \tilde{r}_{i1} x_{ik}\right)$$

(iii) $\sum \tilde{r}_{i1} x_{i1} = \sum \tilde{r}_{i1}^2$

故 $$\tilde{\beta}_1 = \left(\sum \tilde{r}_{i1}^2\right)^{-1}\cdot\left(\hat{\beta}_1 \sum \tilde{r}_{i1}^2 + \hat{\beta}_k\left(\sum \tilde{r}_{i1} x_{ik}\right)\right)$$

$$= \hat{\beta}_1 + \hat{\beta}_k \underbrace{\left(\sum \tilde{r}_{i1}^2\right)^{-1}\left(\sum \tilde{r}_{i1} x_{ik}\right)}_{} \equiv \tilde{\delta}_1 = \hat{\beta}_1 + \hat{\beta}_k \tilde{\delta}_1.$$

OLS估计量的期望值

- **遗漏变量偏误**

如果x_1和x_2是相关的，假设一个线性回归关系

$$x_2 = \delta_0 + \delta_1 x_1 + v$$

$$\Rightarrow \quad y = \beta_0 + \beta_1 x_1 + \beta_2(\delta_0 + \delta_1 x_1 + v) + u$$

$$= \boxed{(\beta_0 + \beta_2\delta_0)} + \boxed{(\beta_1 + \beta_2\delta_1)} x_1 + \boxed{(\beta_2 v + u)}$$

如果y只对x_1做回归，这就是截距项

如果y只对x_1做回归，这就是斜率项

误差项

- **结论：所有估计系数都是偏误的（All estimated coefficients will be biased）**

- 当遗漏变量x_2时β_1的估计偏误情况汇总：

	Corr$(x_1, x_2) > 0$ $(\delta_1 > 0)$	Corr$(x_1, x_2) < 0$ $(\delta_1 < 0)$
$\beta_2 > 0$	Positive bias	Negative bias
$\beta_2 < 0$	Negative bias	Positive bias

- 思考：

（1）如果x_1和x_2是相关的， 哪条基本假设不成立？

（2）如果x_1和x_2是不相关的呢？

OLS估计量的期望值

- **例子：工资等式中遗漏能力**

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

$$abil = \delta_0 + \delta_1 educ + v$$

都是正的

$$wage = (\beta_0 + \beta_2\delta_0) + \boxed{(\beta_1 + \beta_2\delta_1)} educ + (\beta_2 v + u)$$

β_1 将被过度估计（overestimated）因为 $\beta_2\delta_1 > 0$ 。这看起来好像受到多年教育的人会挣很高的工资，但这部分是由于平均来说受教育长的人能力也很强。

- **什么时候没有遗漏变量偏误？**
  - 如果遗漏变量与解释变量是不相关的

OLS估计量的期望值

- **遗漏变量偏误：更一般的情形**

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + u$$ 真实模型（包含x_1， x_2，和x_3）

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + w$$ 估计模型（x_3是遗漏的）

  - 很难得到偏误的方向，这是因为x间会两两相关
  - 如果x间不相关，那分析就和之前一样简单

- **例子：工资等式中遗漏能力**

$$wage = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 abil + u$$

如果exper与（educ，abil）都近似不相关，则遗漏变量的偏误方向就像之前两变量的简单情形一致：β_1将被过度估计

但实际上， exper与educ通常多少会负相关

OLS估计量的方差

- **多元回归模型的标准假设（续）**
- **假定 MLR.5 （同方差性）**

$$Var(u_i|x_{i1},x_{i2},...,x_{ik}) = \sigma^2$$

解释变量必须不能包含有关误差项波动性的信息

- **例子：工资等式**

$$Var(u_i|educ_i, exper_i, tenure_i) = \sigma^2$$

这个假设很难识别。一般肯定不会成立

- **向量表示**

解释变量集中于向量中

$Var(u_i|X_i) = \sigma^2$ with $X_i = (1, x_{i1}, x_{i2}, \ldots, x_{ik})'$

OLS估计量的方差

- **定理 3.2 （OLS斜率估计量的抽样方差）**

在MLR.1 – MLR.5之下：

误差项的方差

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

解释变量x_j的总样本波动

$$\sum_{i=1}^{n}(x_{ij}-\bar{x}_j)^2$$

R2是x_j对其他解释变量（包含常数项）进行回归得到的。

OLS估计量的方差

- **OLS方差的成分：**
- **1） 误差方差**
  - 误差方差增加会提高抽样方差，因为等式中的“噪音”变大了
  - 大的误差方差会使得估计不准确
  - 误差方差不会随着样本变大而降低
- **2） 解释变量的总样本波动**
  - 样本波动越大，估计越精确
  - 总体样本波动随着样本量的增加而渐进提高
  - 因此，增加样本量是提高估计精确性的方法

OLS估计量的方差

- **3） 自变量间的线性关系**

x_j 对其他自变量x进行回归（包含常数项）

该回归的R2越高，其他变量对x_j的线性解释能力越强

- $\widehat{\beta}_j$ 的抽样方差越高，解释变量 x_j 越能被其他解释变量所线性解释
- 几乎线性相关的解释变量问题称为多重共线性（即 $R_j \rightarrow 1$ ）

OLS估计量的方差

- **一个多重共线性的例子**

标准化的平均考试分数　　教师支出　　教材支出　　其他支出

$$avgscore = \beta_0 + \beta_1 teachexp + \beta_2 matexp + \beta_3 othexp + \ldots$$

不同的支出类别将有很强的相关性，因为如果一所学校拥有大量资源，它将在所有事情上花费大量资金。

很难估计不同支出类别的差异效应，因为所有支出要么高、要么低。要准确估计差异效应，需要了解支出类别变化差异的情况。

因此，估计的抽样方差将很大。

OLS估计量的方差

- **讨论多重共线性问题**
  - 在上面的例子中，最好把所有的支出分类放在一起，因为影响是无法分开的
  - 在其他情况下，删除一些自变量可能会降低多重共线性（但这可能会导致遗漏变量偏误）
  - 只有涉及多重共线性的变量的抽样方差会被夸大；对其他影响的估计可能非常精确
  - 请注意严格意义上多重共线性并不违反MLR.3
  - 多重共线性可以通过“方差膨胀因子”来检测

$$VIF_j = 1/(1 - R_j^2)$$

经验之谈，方差膨胀系数不应大于10

OLS估计量的方差

- **误设模型中的方差**
  - 通过分析偏差（bias）和方差（variance）之间的权衡，可以选择是否在回归中包含特定变量

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$ ← 真实总体模型

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2$$ ← 估计模型1

$$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1$$ ← 估计模型2

  - 在这种情况下，误设模型2中可能忽略的变量偏差可能会被较小的方差过度补偿

OLS估计量的方差

- 误设模型中的方差（续）

$$Var(\hat{\beta}_1) = \sigma^2 / \left[SST_1 (1 - R_1^2) \right]$$

已知x_1和x_2，模型2的方差总是小于模型1的

$$Var(\tilde{\beta}_1) = \sigma^2 / SST_1$$

- Case 1:

结论：不要包含无关解释变量

$$\beta_2 = 0 \;\Rightarrow\; E(\hat{\beta}_1) = \beta_1, E(\tilde{\beta}_1) = \beta_1, Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$$

- Case 2:

权衡偏差和方差；注意：即使在大样本中，偏差也不会消失

$$\beta_2 \neq 0 \;\Rightarrow\; E(\hat{\beta}_1) = \beta_1, E(\tilde{\beta}_1) \neq \beta_1, Var(\tilde{\beta}_1) < Var(\hat{\beta}_1)$$

OLS估计量的方差

- **估计 σ^2**

$$\hat{\sigma}^2 = \left(\sum_{i=1}^{n} \hat{u}_i^2\right) / [n - k - 1]$$

误差方差的无偏估计可以通过从观测值的数量中减去估计回归系数的数量来获得。观测值的数量减去估计参数的数量也被称为自由度。求和中的n个估计平方残差不是完全独立的，而是通过定义最小化问题一阶条件的k+1方程关联的。

- **定理 3.3 （σ^2的无偏估计）**

$$MLR.1 - MLR.5 \quad \Rightarrow \quad E(\hat{\sigma}^2) = \sigma^2$$

OLS估计量的方差

- **估计OLS估计值的抽样方差**

估计参数 β_j 的真实抽样方差

$$sd(\widehat{\beta}_j) = \sqrt{Var(\widehat{\beta}_j)} = \sqrt{\sigma^2 / \left[SST_j(1 - R_j^2)\right]}$$

用样本 $\widehat{\sigma}^2$ 来替换未知的 σ^2

估计参数 β_j 的样本抽样方差

$$se(\widehat{\beta}_j) = \sqrt{\widehat{Var}(\widehat{\beta}_j)} = \sqrt{\widehat{\sigma}^2 / \left[SST_j(1 - R_j^2)\right]}$$

- **请注意，这些公式只有在假定MLR.1-MLR.5下才成立（尤其必须具有同方差性homoskedasticity）**

OLS的有效性

- **OLS的有效性：高斯-马尔科夫定理**
  - 在假定MLR.1-MLR.5下，OLS是无偏的
  - 然而，在这些假定下，还有其他的估计值也是无偏的
  - 哪一个是具有最小方差的无偏估计？
  - 为了回答以上问题，人名经常限定于线性回归

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

线性回归估计值；OLS估计值也是这种形式

OLS的有效性

- **定理 3.4 （高斯-马尔科夫定理）**
  - 在假定MLR.1-MLR.5下，OLS估计值是最优线性无偏估计量（best linear unbiased estimators，BLUEs），即

$$Var(\hat{\beta}_j) \leq Var(\tilde{\beta}_j) \quad j = 0, 1, \ldots, k$$

对于所有无偏估计值 $\tilde{\beta}_j = \sum_{i=1}^{n} w_{ij} y_i$ （$E(\tilde{\beta}_j) = \beta_j, j = 0, \ldots, k$）均成立。

- **只有MLR.1-MLR.5都成立时，OLS才是最优估计量；例如，如果存在异方差，则有更好的估计值。**

OLS的有效性

- **例子：回归中包含无关变量**

$y = \beta_0 + \beta_1 x_1 + u$ ← 真实总体模型（MLR.1）

$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1$ ← 估计模型1

$\tilde{y} = \tilde{\beta}_0 + \tilde{\beta}_1 x_1 + \tilde{\beta}_2 x_2$ ← 估计模型2

- **$\hat{\beta}_1$满足高斯-马尔科夫条件，$\tilde{\beta}_1$不满足高斯-马尔科夫条件（MLR. 1）**
- **比较$\hat{\beta}_1$与$\tilde{\beta}_1$的（条件方差）：**

$$Var\left(\hat{\beta}_1\right) = \sigma^2 / SST_1$$

$$Var\left(\tilde{\beta}_1\right) = \sigma^2 / [SST_1 \left(1 - R_1^2\right)]$$

- **因此，$\hat{\beta}_1$方差更小（符合高斯-马尔科夫定理）**

小节

- 多元回归模型能使我们有效地在保持其他因素不变的情况下，考察一个特定的自变量对因变量的影响
- 每个斜率参数都度量了在保持所有其他自变量不变的条件下，相应的自变量对因变量的偏效应
- 普通最小二乘法被用于估计模型中的斜率和截距参数
- R2是因变量样本波动中能被自交量解释的部分，并用于度量回归的拟合优度
- 在前四个高斯-马尔科夫假定(假定 MLR.1到 MLR.4)下，OLS 估计量是无偏的
- 一方面，在模型中包含一个无关变量，对截距和其他斜率估计量的无偏性没有任何影响。另一方面，遗漏一个相关变量则导致 OLS 产生偏误
- 在五个高斯-马尔科夫假定下，OLS 斜率估计量的方差为 $Var(\hat{\beta}_j) = \dfrac{\sigma^2}{SST_j(1-R_j^2)}, \quad j = 1, \ldots, k$
- 高斯-马尔科夫定理说明OLS估计量是最优线性无偏估计量（有效性）