# 计量经济学STAT30021

## 第四讲：多元线性回归分析

肖志国

复旦大学管理学院

2025年10月

# 多元线性回归模型

总体模型：假设随机变量$Y$与$k$维随机向量$X = (X_1, \ldots, X_k)'$ 满足如下关系：

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_k X_k + u, \tag{1}$$

其中$u$满足

$$E[u|X_1, \ldots, X_k] = 0.$$

术语：

- $\beta_0$称为截距，$\beta_1, \ldots, \beta_k$称为斜率参数
- 其分析结果可以提供类似于“保持其他变量不变”情形下的解释
- $u$称为误差项或干扰项，代表除$X_1, \ldots, X_k$以外影响$Y$的因素

# 样本模型

设$(Y_i, X_i), i = 1, \ldots, n$为来自总体$(Y, X)$的一个简单随机样本。则有：

$$Y_i = \beta_0 + \beta_1 X_{i1} + \cdots + \beta_k X_{ik} + u_i, \tag{2}$$

其中$u_i$满足

$$E[u_i | X_i] = 0.$$

基本任务：构造参数$\beta_0, \beta_1, \ldots, \beta_k$的估计量。

# 方法一：矩估计法

我们有$k+1$个参数，需要构造$k+1$个矩条件方程。一个自然的矩条件是$E[u_i]=0$。另外$k$ 个矩条件为：$E[X_{i1}u_i]=0,\ldots,E[X_{ik}u_i]=0$。总的矩条件为：

$$E\begin{bmatrix} Y_i-\beta_0-\beta_1X_{i1}-\cdots-\beta_kX_{ik} \\ X_{i1}(Y_i-\beta_0-\beta_1X_{i1}-\cdots-\beta_kX_{ik}) \\ \cdots \\ X_{ik}(Y_i-\beta_0-\beta_1X_{i1}-\cdots-\beta_kX_{ik}) \end{bmatrix}=0$$

根据矩估计法的原理，$\beta_0,\beta_1,\ldots,\beta_k$的矩估计量满足：

$$\begin{aligned} \sum_{i=1}^{n}(Y_i-\beta_0-\beta_1X_{i1}-\cdots-\beta_kX_{ik}) &= 0 \\ \sum_{i=1}^{n}X_{i1}(Y_i-\beta_0-\beta_1X_{i1}-\cdots-\beta_kX_{ik}) &= 0 \\ \cdots & \\ \sum_{i=1}^{n}X_{ik}(Y_i-\beta_0-\beta_1X_{i1}-\cdots-\beta_kX_{ik}) &= 0 \end{aligned}$$

# 方法二：最小二乘法(OLS)

令$\beta = (\beta_0, \beta_1, \ldots, \beta_k)$。考虑最小化如下的目标函数：

$$S(\beta) = \sum_{i=1}^{n} \Big[ Y_i - (\beta_0 - \beta_1 X_{i1} - \cdots - \beta_k X_{ik}) \Big]^2$$

设$\hat{\beta} = \left( \hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_k \right)$为使得$S(\beta)$最小的值。我们称$\hat{\beta}$为$\beta$的最小二乘估计量。

易见，最小二乘估计量$\hat{\beta}$满足的方程组与矩估计量满足的方程组完全相同。故而两个估计方法得到的估计量相同。

# 最小二乘估计量的矩阵表达

令

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & X_{11} & \dots & X_{1k} \\ \vdots & \vdots & \vdots & \\ 1 & X_{n1} & \dots & X_{nk} \end{bmatrix}, \mathbf{u} = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}.$$

则样本点满足的方程(2)可以写成向量形式：

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}, \tag{3}$$

而$\beta$的OLS估计量可以表示为

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{Y} = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{u}. \tag{4}$$

# 线性回归中$\beta$系数的含义

1. $\beta_j$ $(j = 1, \ldots, k)$描述的是**当模型中使用的其他因素，也就是**$X_{-j} = (X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_k)$**固定不变**时，$X_j$的1个单位的改变对应的$Y$的平均值的改变。具体的：

$$\beta_j = \frac{\partial E[Y|X_1, \ldots, X_k]}{\partial X_j}.$$

2. 由此可以看到，多元回归的一个优点就是，不管我们使用的是实验数据还是非实验数据，回归系数天然的就具有因果效应解读的潜质，但注意，它并不必然代表的是因果效应。
3. 从理论上来说，数据反映的是一系列复杂事件的结果。但是因果效应涉及的却是这一系列复杂事件中间变量之间互相影响的过程。所以要想得到因果效应的判断，除了数据之外，我们还需要知道数据生成过程的特征。

# 线性回归中$\beta$系数的含义

1. 如果我们使用的实验数据，具体而言，如果对每一个固定的$X_{-j}$的值，$X_j$都是随机赋值的话，那么$\beta_j$代表的就是$X_j$对$Y$的因果效应。
2. 如果我们使用的是非实验数据，那么，$\beta_j$是因果效应的一个自然的充分条件是：
   - 第一，模型中使用的其他因素$X_{-j}$ 和变量$X_j$ 一起，构成了一个决定$Y$的变化的全部变量集。当然，实际情况中，我们使用的$X_{-j}$既可能遗漏重要变量，也可能多加了其它无关变量甚至干扰变量；
   - 第二，数据的生成过程满足某种典型的识别性特征（这一点我们后续会讨论）。
3. 以往的计量经济学研究主要是在讨论这个充分条件的第一部分，事实上，第二部分也很重要。

# 我们总是能够“维持其他因素不变”吗?

- 当我们解释$\beta_j$的含义时，我们隐含的一个假定是，我们能够固定其他因素不变
- 有时候这样的假设是不很合理的

$$wage_i = \beta_0 + \beta_1 edu_i + \beta_2 age_i + \beta_3 expr_i + u_i,$$

考虑*age*的影响的时候，固定*experience*未必是一个合理的假设。

- 有时候这样的假设是很不合理的

$$wage_i = \beta_0 + \beta_1 edu_i + \beta_2 age_i + \beta_3 age_i^2 + u_i,$$

考虑*age*的影响的时候，我们能让$age^2$固定吗？

# 分步回归：Frisch-Waugh-Lovell定理

公式(4)给出的是$\hat{\beta}$的整体解。对于单独的一个系数$\hat{\beta}_j$,下面的Frisch-Waugh-Lovell定理给出了其另外一种等价的计算办法：

## Theorem (Frisch-Waugh-Lovell定理)

对任意$j = 1, 2, \ldots, k$，$\hat{\beta}_j$可以通过下面的两步法得到：

1. 以$X_j$做被解释变量，$X_m, m \neq j$为解释变量做线性回归模型。记此回归模型的残差项为$\tilde{r}_{ij}$，$i = 1, \ldots, n$。也就是说：$\tilde{r}_j : X_j \sim \mathbf{X}_{-j}$
2. 以$Y$做被解释变量，$\tilde{r}_j$为解释变量做简单线性回归模型$Y \sim \tilde{r}_j$。由此得到的$\tilde{r}_j$的系数的最小二乘估计等于$\hat{\beta}_j$。也就是说：

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n}(\tilde{r}_{ij} - \bar{\tilde{r}}_j)(Y_i - \bar{Y})}{\sum_{i=1}^{n}(\tilde{r}_{ij} - \bar{\tilde{r}}_j)^2} = \frac{\sum_{i=1}^{n}\tilde{r}_{ij} Y_i}{\sum_{i=1}^{n}\tilde{r}_{ij}^2}$$

# Frisch-Waugh-Lovell定理另一种形式

## Theorem (Frisch-Waugh-Lovell定理)

对任意$j = 1, 2, \ldots, k$，$\hat{\beta}_j$可以通过下面的两步法得到：

1. 以$X_j$做被解释变量，$X_m, m \neq j$为解释变量做线性回归模型。记此回归模型的残差项为$\tilde{r}_{ij},\ i = 1, \ldots, n$。也就是说：$\tilde{r}_j :\ X_j \sim \mathbf{X}_{-j}$，并记$R^2_{X_j \mathbf{X}_{-j}}$
2. 以$Y$做被解释变量，$X_m, m \neq j$为解释变量做线性回归模型。记此回归模型的残差项为$\tilde{\tilde{r}}_{ij},\ i = 1, \ldots, n$。也就是说：$\tilde{\tilde{r}}_j :\ Y \sim \mathbf{X}_{-j}$，并记$R^2_{Y \mathbf{X}_{-j}}$
3. 以$\tilde{\tilde{r}}_j$为被解释变量，$\tilde{r}_j$为解释变量做简单线性回归模型$\tilde{\tilde{r}}_j \sim \tilde{r}_j$。由此得到的$\tilde{r}_j$的系数的最小二乘估计等于$\hat{\beta}_j$。也就是说：

$$\hat{\beta}_j = \frac{\sum_{i=1}^{n} \tilde{r}_{ij} \tilde{\tilde{r}}_{ij}}{\sum_{i=1}^{n} \tilde{r}_{ij}^2}$$

# 偏相关系数Partial Correlation Coefficient

给定一组变量$\mathbf{X}_{-j} = (X_m, m \neq j)$，$Y$和$X_j$的偏相关系数等于$\tilde{r}_j$和$\tilde{\tilde{r}}_j$的相关系数。也就是说：

$$\rho_{YX_j \cdot \mathbf{X}_{-j}} = \frac{\sum_{i=1}^{n} \tilde{r}_{ij} \tilde{\tilde{r}}_{ij}}{\sqrt{\sum_{i=1}^{n} \tilde{r}_{ij}^2} \sqrt{\sum_{i=1}^{n} \tilde{\tilde{r}}_{ij}^2}}$$

从而我们有

$$\hat{\beta}_j = \rho_{YX_j \cdot \mathbf{X}_{-j}} \sqrt{\frac{\sum_{i=1}^{n} \tilde{\tilde{r}}_{ij}^2}{\sum_{i=1}^{n} \tilde{r}_{ij}^2}} = \rho_{YX_j \cdot \mathbf{X}_{-j}} \frac{s_Y}{s_{X_j}} \sqrt{\frac{1 - R^2_{Y\mathbf{X}_{-j}}}{1 - R^2_{X_j\mathbf{X}_{-j}}}}$$

易见，上式是简单线性模型时($\mathbf{X}_{-j}$为常数)的公式$\hat{\beta}_1 = r_{XY} s_Y / s_X$的推广。

# FWL定理的解释

- $\hat{r}_{ij}$是$X_j$中和$X_m, m \neq j$不相关的部分：

$$\sum_{i=1}^{n} \hat{r}_{ij} X_{im} = 0, \forall m \neq j$$

- 或者说，$\hat{r}_{ij}$是$X_m, m \neq j$的效应被除尽之后的$X_j$

$$\hat{r}_{ij} = X_{ij} - \tilde{\gamma}_0 - \sum_{m \neq j} \tilde{\gamma}_m X_{im},$$

  其中$\tilde{\gamma}$为$X_j$对$X_m, m \neq j$回归的最小二乘估计
- 故而$\hat{\beta}_j$衡量的是当$X_m, m \neq j$的效应被除尽之后，$X_j$对$Y$的影响

# 拟合值，残差及其性质

定义$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_{i1} + \cdots + \hat{\beta}_k X_{ik}$ 为$Y_i$的拟合值，定义$\hat{u}_i = Y_i - \hat{Y}_i$为残差。

性质1 所有残差之和为零：

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

从而$\bar{\hat{Y}} = \bar{Y}$。

性质2 任一解释变量$X_j$与残差的样本协方差为零：

$$\sum_{i=1}^{n} X_{ij} \hat{u}_i = 0, \ \forall j = 1, \ldots, k.$$

性质3 拟合值$\hat{Y}$与残差的样本协方差为零：

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

# 波动恒等式

类似的，我们定义SST, SSE和SSR如下：

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

我们有：

$$SST = SSE + SSR$$

# $R^2$与相关系数

我们定义$R^2$为

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST} \tag{5}$$

令$r_{\hat{Y}Y}$为$\hat{Y}$与$Y$的相关系数，也就是：

$$r_{\hat{Y}Y} = \frac{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(\hat{Y}_i - \bar{\hat{Y}})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

则我们有如下结果：

$$R^2 = r_{\hat{Y}Y}^2$$

# 最小二乘估计量的小样本性质：假设

Gauss-Markov假设

1. 假设MLR.1 (线性模型)：总体变量$Y$,$X$满足线性模型关系
2. 假设MLR.2 (随机抽样)：样本数据为来自总体的简单随机抽样
3. 假设MLR.3 (不存在多重共线性)：$1, X_1 \ldots, X_k$不存在线性相关关系，或者说，矩阵$\mathbf{X}$为满秩矩阵
4. 假设MLR.4 (条件期望为零)：$E[u_i|X_i] = 0$
5. 假设MLR.5 (条件不相关性与同方差性)：$Var[\mathbf{u}|\mathbf{X}] = \sigma^2 I$，其中$I$为$n$维恒等矩阵

# 误差项的条件方差矩阵

误差项$\mathbf{u}$的条件方差矩阵$Var(\mathbf{u}|\mathbf{X})$的定义为

$$Var(\mathbf{u}|\mathbf{X}) = \begin{bmatrix} Var(u_1|\mathbf{X}) & Cov(u_1, u_2|\mathbf{X}) & \dots & Cov(u_1, u_n|\mathbf{X}) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(u_n, u_1|\mathbf{X}) & Cov(u_n, u_2|\mathbf{X}) & \dots & Var(u_n|\mathbf{X}) \end{bmatrix}$$

$$= \begin{bmatrix} Var(u_1|X_1) & Cov(u_1, u_2|X_1, X_2) & \dots & Cov(u_1, u_n|X_1, X_n) \\ \vdots & \vdots & \vdots & \vdots \\ Cov(u_n, u_1|X_1, X_n) & Cov(u_n, u_2|X_2, X_n) & \dots & Var(u_n|X_n) \end{bmatrix}$$

# 同方差性

我们称回归模型具有**同方差性**，如果假设MLR.5成立，也就是有如下情况：

$$Var(\mathbf{u}|\mathbf{X}) = \begin{bmatrix} \sigma^2 & & \\ & \ddots & \\ & & \sigma^2 \end{bmatrix} = \sigma^2 I. \tag{6}$$

也就是说，对于$i = 1, \ldots, n$，有

$$Var(u_i|\mathbf{X}) = \sigma^2, \tag{7}$$

且对于$i \neq j$，有

$$Cov(u_i, u_j|\mathbf{X}) = 0. \tag{8}$$

# 最小二乘估计量的无偏性

**Theorem (定理3.1)**

在假设条件*MLR.1-MLR.4*下，对于任意的$j = 0, 1, \ldots, k$，$\hat{\beta}_j$是$\beta_j$的无偏估计量*:*

$$E[\hat{\beta}_j | X_1, \ldots, X_n] = \beta_j.$$

从而我们有

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

# 添加无关变量的影响

假设我们的真实模型是

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + u, \tag{9}$$

其中

1. 假设条件MLR.1-MLR.4成立
2. $\beta_3 = 0$:控制了$X_1, X_2$之后，$X_3$对$Y$没有影响

在$\beta_3 = 0$假设条件下，MLR.4表示

$$E[Y|X_1, X_2, X_3] = E[Y|X_1, X_2] = \beta_0 + \beta_1 X_1 + \beta_2 X_2. \tag{10}$$

假设现在我们用模型(9)来拟合数据。这样得到的最小二乘估计量有什么性质？

# 添加无关变量的影响

或者说，真实模型是

$$E[Y|X_1, X_2, X_3] = \beta_0 + \beta_1 X_1 + \beta_2 X_2, \tag{11}$$

但是我们用方程(9)来拟合数据。

理论上：这样得到的$\beta_0, \beta_1, \beta_2$的最小二乘估计量仍然是无偏估计量。

为什么？

实际上：这样得到的$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$和不加入$X_3$得到的估计量的大小可能有很大的差别，当然也可能差异很小，这取决于数据的具体情况。

# 忽略有关变量的偏差

假设我们的真实模型是

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k X_k + u, \qquad (12)$$

且满足假设条件MLR.1-MLR.4。

也就是说，真实模型是

$$E[Y|X_1, \ldots, X_k] = \beta_0 + \beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k X_k, \qquad (13)$$

假设我们在回归方程中忽略一个变量，比如$X_k$。

这样得到的$\beta_0, \beta_1, \ldots, \beta_{k-1}$的最小二乘估计量有什么性质？

# 忽略有关变量的偏差（Omitted Variable Bias）

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

$$Y \quad \sim \quad X_1 + \cdots + X_k$$

回归得到的最小二乘估计量，令$\tilde{\beta}_j$ $(j = 0, 1, \ldots, k-1)$为

$$Y \quad \sim \quad X_1 + \cdots + X_{k-1}$$

回归得到的最小二乘估计量。令$\tilde{\delta}_j$ $(j = 1, \ldots, k-1)$为

$$X_k \quad \sim \quad X_1 + \cdots + X_{k-1}$$

回归中$X_j$的系数的最小二乘估计。则对于任意$j = 1, \ldots, k-1$，有

$$\tilde{\beta}_j = \hat{\beta}_j + \hat{\beta}_k \tilde{\delta}_j,$$

从而我们有

$$E\left[\tilde{\beta}_j | X_1, \ldots, X_k\right] = \beta_j + \beta_k \tilde{\delta}_j.$$

# OVB的解释

首先，OVB公式成立的前提条件是

$$E[Y|X_1,\ldots,X_k] = \beta_0 + \beta_1 X_1 + \cdots + \beta_{k-1} X_{k-1} + \beta_k X_k. \quad (14)$$

其次，$X_j$回归系数的忽略变量偏差

$$OVB_j = \beta_k \tilde{\delta}_j,$$

也就是说

$OVB_j =$ **被忽略变量**的效应 $\times$ 变量 j 对**被忽略变量**的回归效应.

例子：在如下回归中

$$\text{工资} \quad \sim \quad \text{教育水平},$$

“教育水平”的系数估计是高估了还是低估了?

# 回归模型中的控制变量

假设我们关心的是变量$X$对$Y$的影响，我们知道除了$X$之外，还有$Z_1, ..., Z_p$也会同时影响$Y$。这些$Z_1, ..., Z_p$通常叫做控制变量。

$$\begin{aligned} Y &= \alpha + \beta X + \gamma_1 Z_1 + \cdots + \gamma_p Z_p + u, \\ E[u|X, Z_1, \ldots, Z_p] &= 0. \end{aligned} \tag{15}$$

现在假设$X$与$Z_1, ..., Z_p$都不相关。

根据OBV公式，我们知道，在回归方程中忽略$Z_1, ..., Z_p$，不会影响到$\beta$的OLS估计量的无偏性。

那么为什么我们还要加入那些可能与$X$无关的控制变量呢？这是因为增加这些控制变量还有一个效果：降低残差项的方差$\hat{\sigma}^2$，从而会降低$\hat{\beta}$的标准误。也就是说，增加控制变量能够使得我们关注的核心参数$\beta$的估计变得更准确，即使这些控制变量和$X$无关。