

体育经济分析：理论与应用

专题一：回归与因果¹

周正卿

02 March 2022

引言

课程进度

上节课

- 经验研究
- 条件期望函数CEF

今天 + 1或2

- CEF \longrightarrow LPF \longrightarrow 回归 \longrightarrow 因果推断

重点阅读 *MHE* 3.1

回归分析

回归分析

为什么使用回归？

回归研究结果变量 Y 与一系列原因变量 X 之间关系的众多方法。(C&T, Ch4.1)

Q 为什么经验研究者偏好使用线性回归和最小二乘方法 (LS) ?

A 使用回归分析可以帮助研究者：(1) 因果识别 (causal identification) 和 (2) 推断 (inference)。另外，系数估计的科学性（无偏、一致、有效）和计算可操作性的兼顾也是重要考虑。

常见的线性回归和OLS背后有共同基础: 条件期望函数 (CEF)

回归分析

为什么使用回归?

暂不考虑因果关系，理解回归分析的步骤:

- 深入理解CEF（条件期望函数）
- 总体回归函数（PRF by MHE）或者 线性投影函数(LPF by Hansen)
- 参数估计值的样本分布

回归分析

CEF

上节课: 当涉及多个“条件”时，可以写作向量形式：

$$E[Y \mid X = x] = m(x)$$

CEF = $m(x) = E[Y \mid X = x]$ 是 $x \in \mathbb{R}^k$ 的函数，意味着“当 X 取值 x 时, Y 的平均值为 $m(x)$ ”，由于 X 可以取值任意的 x ，因此将CEF视为随机变量 X 的函数, **CEF也是随机变量**。

期望本身是总体概念，现实中面对的一个个样本，因而需要用样本信息推断总体信息（利用抽样分布与统计推断），**i.e** 用样本均值推断总体期望。拿到样本，关心的其实是总体。

CEF是从**总体视角**出发，那么样本如何呈现？

回归分析

CEF

通过实验或者抽样得到N个观测点， CEF重新写作: $E[Y_i | X_i]$

- X_i 是 $k \times 1$ 的列向量， X_i 每个元素是 x_{ki}
- 若 X_i 是1维的， 比如 $X_i = \text{学历}$ ， $E[Y_i | X_i = \text{研究生}]$ ， 总体中所有“研究生学历”的 Y 的均值；若 X_i 是2维的， 比如 $X_i = \{\text{学历}, \text{性别}\}$ ， $E[Y_i | X_i = \{\text{本科}, \text{男性}\}]$ ， 总体中所有“男性本科学历”的 Y 的均值

Key: 再次深刻理解条件期望函数是 x 的函数

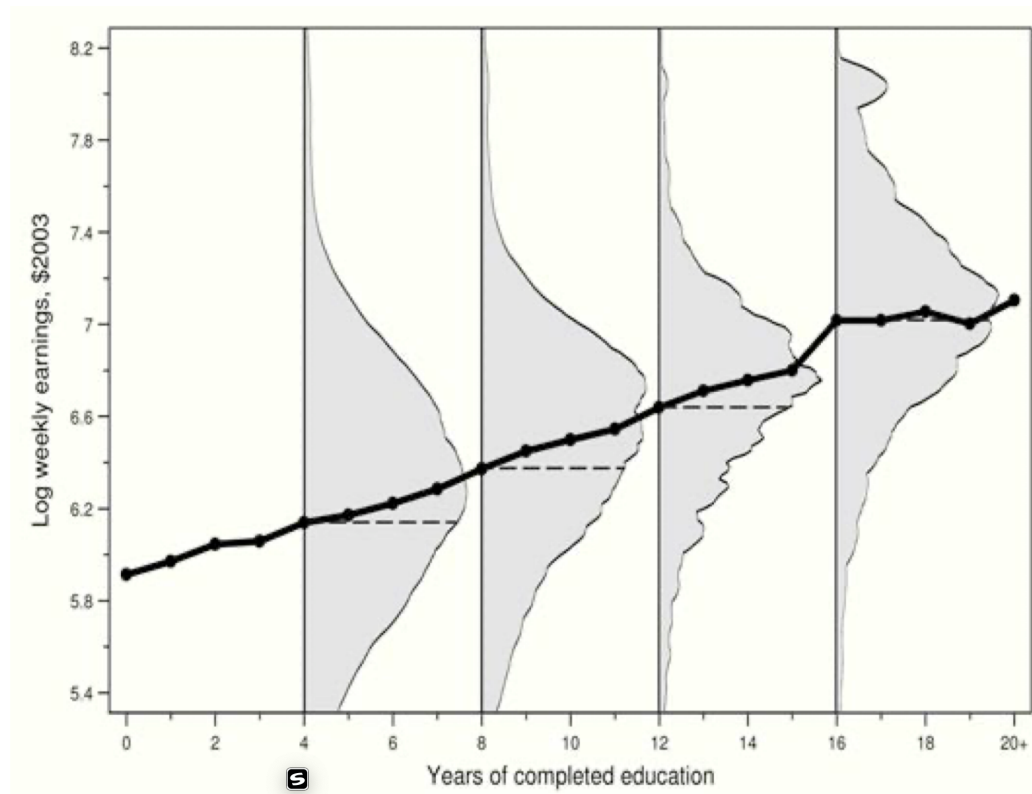
从图形上看CEF...

条件分布 Y_i , 对于8, ..., 22不同教育年限的 $X_i = x$.

条件期望函数, $E[Y_i | X_i]$ 其实是这些条件分布的均值

若只关注 CEF, $E[Y_i | X_i] \dots$

实际数据 (MHE)



© Princeton University Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

Figure 3.1.1: Raw data and the CEF of average log weekly wages given schooling. The sample includes white men aged 40-49. The data are from the 1980 IPUMS5 percent sample.

研究问题 → CEF

CEF基于数据出发，对于理解变量间至关重要，但经验研究的第一问题是如何将理论结果转换为实证模型， $E[\text{工资}_i | \text{运动技能}_i]$ ：

- step 1: 选取 Y 与 X (从研究问题出发)
- step 2: 在总体中重复抽样，获得样本
- step 3: 对 X "切片"，获得 $Y | X = x$ 的条件密度和条件分布
- step 4: 制作联合密度表格 $P(Y = y, X = x)$
- step 5: 计算边缘密度 $P(X = x)$
- step 6: 制作条件密度表格 $P(Y | X = x) = \frac{P(Y=y, X=x)}{P(X=x)}$
- step 7: 计算条件期望 $E(Y | X = x)$

回归分析

CEF 性质

分解清晰: CEF 将观测的结果变量分解成两部分

$$Y_i = E[Y_i | X_i] + e_i$$

1. 被 X_i 解释的部分(*i.e.*, CEF $E[Y_i | X_i]$)
2. 具有特殊性质的误差项[†]
 - i. e_i 均值独立于 X_i , *i.e.*, $E[e_i | X_i] = 0$
 - ii. e_i 与 X_i 的任何函数不相干

[†] 回忆第一讲的性质.

回归分析

CEF 性质

ANOVA 定理:

无条件方差与条件方差的关系：可将结果变量 Y_i 方差分解为两部分

$$Var(Y_i) = E[Var(Y_i | X_i)] + Var(E[Y_i | X_i])$$

1. 组内方差(的均值)(within group variance)。每个"等级"内Y的分布的方差的期望值(均值)。
2. 组间方差(across group variance)。条件期望值在"等级"间的分布的方差

或者MHE将上式分解为： CEF的方差(CEF可以解释) + 干扰项的方差(CEF无法可以解释)

回归分析

CEF 性质

良好预测: $m(X_i)$ 为 X_i 任意形式函数, CEF是最小均方误差 (**性质5**)

$$E[Y_i | X_i] = \underset{m(X_i)}{\operatorname{argmin}} E[(Y_i - m(X_i))^2]$$

CEF是给定 X_i 能够预测 Y_i 最好预测方式.

注意 m 可以是任意形式函数 (包含非线性) .

但问题是：CEF的函数形式非常复杂，有没有既形式简单，又良好预测，且分解清晰的办法呢？

但问题是：CEF的函数形式非常复杂，有没有既形式简单，又良好预测，且分解清晰的办法呢？

答案是：**LPF** 线性投影函数

但问题是：CEF的函数形式非常复杂，有没有既形式简单，又良好预测，且分解清晰的办法呢？

答案是：**LPF** 线性投影函数

深度阅读：为什么不是线性CEF，而是LPF呢？线性CEF(linear CEF model)也就是常见的线性回归模型(linear regression model)。其中一个原因是 $m(x_1, x_2)$ 其完整的线性形式写作 $m(x_1, x_2) = x_1\beta_1 + x_2\beta_2 + x_1^2\beta_3 + x_2^2\beta_4 + x_1x_2\beta_5 + \beta_6 \rightarrow \text{《E》}, 30-38$

线性回归是实证分析中应用最广泛的一种方法。

- LPF 与 CEF 什么关系?
- 什么条件下 LPF 具有因果解释?

回归分析

CEF 与 LPF

CEF是具有好的预测性质，那么“好”**准则**是什么？

定义**损失函数(loss function)**, 表达为常用的二次型形式：

$$L(Y, g(x)) = (Y - g(x))^2$$

其中 $L(\cdot)$ 是r.v., 取期望得**均值平方误差 (mean squared error, MSE)** , 简称**均方误**

$$R(Y, g(x)) = E[L(Y, g(x))] = E[(Y - g(x))^2]$$

回忆：CEF是MMSE(**性质5**)

回归分析

CEF 与 LPF

LPF

寻找使得MSE最小化的线性函数:

$$\beta = \underset{b}{\operatorname{argmin}} E \left[(Y_i - X_i' b)^2 \right]$$

其中 b 与 X_i 都是 $K \times 1$ 向量, Y_i 是标量, β 是 $K \times 1$ 向量, 是通过最小化MSE得到的总体最小二乘回归系数。

依据一阶条件: $E[X_i (Y_i - X_i' b)] = 0$ 得到 b 的最优解 $\beta = E[X_i X_i']^{-1} E[X_i Y_i]$

回归分析

CEF 与 LPF

LPF

$X_i'\beta$ 是 Y_i 在 X_i 上的最优线性投影 (best linear projection, BLP, MHE 3.1.5), 向量 β 是线性投影系数 (linear projection coefficient)

根据一阶条件重新构建 $E[X_i(Y_i - X_i'\beta)] = 0$, 也就是说 Y 的线性投影函数误差(linear projection function error, LPFE) $e_i = Y_i - X_i'\beta$ 与 X_i 不相关, 也就是说LPF具有 $E(X_i e_i) = 0$ (矩阵形式为 $E[Xe] = 0$) 的性质.

- **思考：**与CEFE的性质比较

回归分析

CEF 与 LPF

CEF与LPF

LPF是MMSE， CEF是MMSE。两者什么关系？继续使用最小化MSE**准则**：

$$\beta = \underset{b}{\operatorname{argmin}} E \left[\left(m(X_i) - X_i' b \right)^2 \right]$$

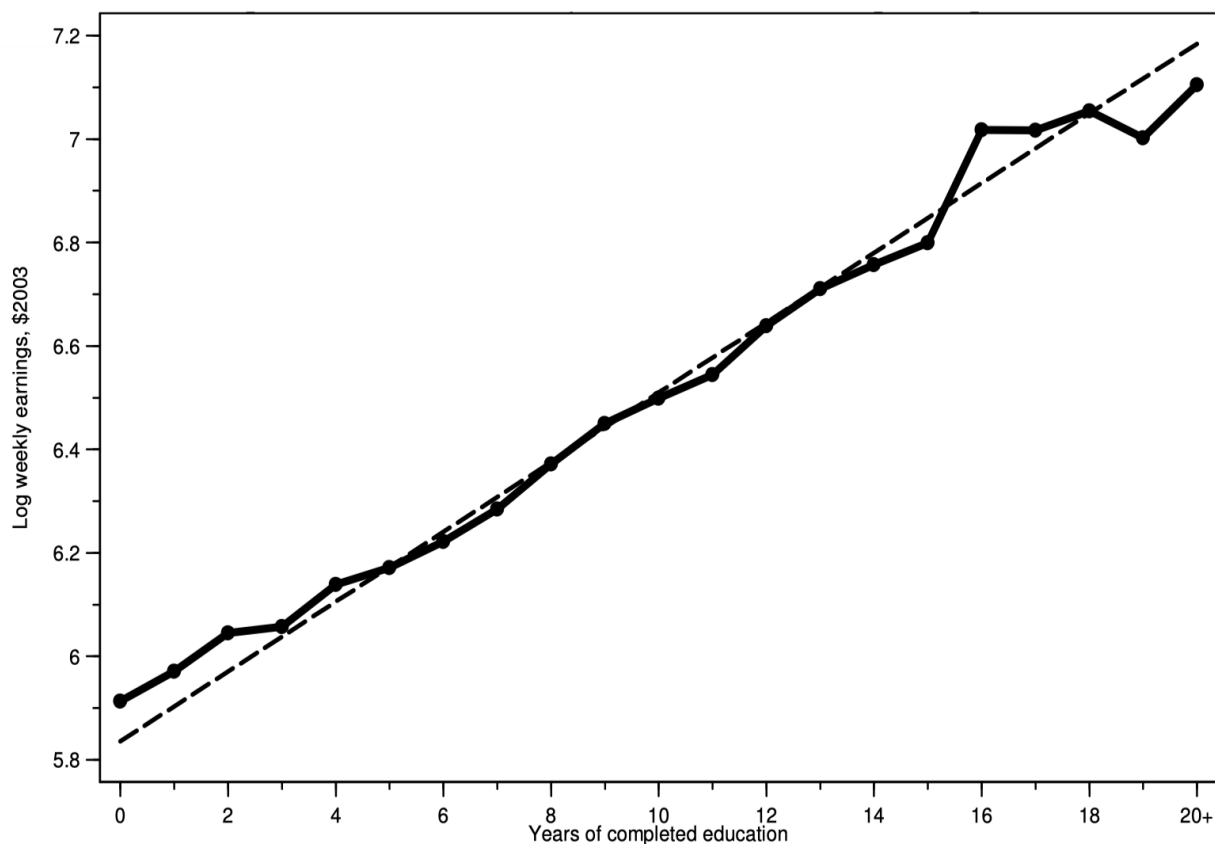
回归与条件期望函数定理（Regression-CEF Theorem, MHE 3.1.6），结论：

- **LPF就是CEF的MMSE**, 是CEF的BLP（MHE 3.1.6 的含义）
- 通常CEF不是LPF, 但若CEF是线性的, 那么 $\text{LPF} = \text{CEF}$ （MHE 3.1.4）

CEF

LPF

实际数据 (MHE)



© Princeton University Press. All rights reserved. This content is excluded from our Creative Commons license. For more information, see <http://ocw.mit.edu/help/faq-fair-use/>.

Sample is limited to white men, age 40-49. Data is from Census IPUMS 1980, 5% sample.

Figure 3.1.2: Regression threads the CEF of average weekly wages given schooling

LPF的参数识别

Original variables

$$x_{1,\text{original}} = [1, 3, 4, 7, 9, 9]$$

$$x_{2,\text{original}} = [9, 9, 6, 3, 1, 2]$$

$$y_{\text{original}} = [3, 5, 6, 8, 7, 10]$$

$$\mathbf{X} = \begin{bmatrix} -4.5 & 4 \\ -2.5 & 4 \\ -1.5 & 1 \\ 1.5 & -2 \\ 3.5 & -4 \\ 3.5 & -3 \end{bmatrix}$$

$$\mathbf{X}^T \mathbf{X} = \begin{bmatrix} 55.5 & -57.0 \\ -57.0 & 62 \end{bmatrix}$$

Centered variables

$$x_1 = [-4.5, -2.5, -1.5, 1.5, 3.5, 3.5]$$

$$x_2 = [4, 4, 1, -2, -4, -3]$$

$$y = [-3.5, -1.5, -0.5, 1.5, 0.5, 3.5]$$

$$\mathbf{y} = \begin{bmatrix} -3.5 \\ -1.5 \\ -0.5 \\ 1.5 \\ 0.5 \\ 3.5 \end{bmatrix}$$

$$\mathbf{X}^T \mathbf{y} = \begin{bmatrix} 36.5 \\ -36.0 \end{bmatrix}$$

- ▶ \mathbf{y} : $n \times 1$
- ▶ \mathbf{X} : $n \times k$
- ▶ \mathbf{b} : $k \times 1$
- ▶ \mathbf{e} : $n \times 1$

$\beta = E[X_i X_i']^{-1} E[X_i Y_i]$ 在实际观测点中怎样获得? 涉及到 β 的识别问题 (唯一且可估计)

Hansen : 保证 $k \times k$ 的 **设计矩阵 (design matrix)**

$Q_{XX} = E[XX']$ 是可逆的, 就可以识别LPF参数 β 。

回归分析

CEF与LPF系数含义的区别

- 线性CEF的参数

$$E[\log(wage) \mid Black, female] = -0.20Black - 0.24female + 0.10Black \times female + 3.06$$

线性CEF：黑人男性（相对于非黑人的男性）低20%，黑人女性（相对于非黑人女性）为高10%。

- LPF的参数

$$\mathcal{P}[\log(wage) \mid Black, female] = -0.15Black - 0.23female + 3.06$$

LPF：黑人打工者（相对于非黑人）低15%，忽略了性别的作用(计算净影响)，控制变量保证条件相同。

回归分析

CEF、LPF和因果关系

考虑到： LPF 是 CEF 的MMSE， 是 CEF 的 BLP

若CEF是相关关系的 \rightarrow LPF是相关的

回归分析

CEF、LPF和因果关系

考虑到： LPF 是 CEF 的MMSE， 是 CEF 的 BLP

若CEF是相关关系的 \rightarrow LPF是相关的

若CEF是因果关系 \rightarrow LPF是因果的

回归分析

CEF、LPF和因果关系

考虑到： LPF 是 CEF 的MMSE， 是 CEF 的 BLP

若CEF是相关关系的 \rightarrow LPF是相关的

若CEF是因果关系 \rightarrow LPF是因果的

问题是：怎样获得一个因果的 CEF？(客观)

回归分析

CEF、LPF和因果关系

考虑到： LPF 是 CEF 的MMSE， 是 CEF 的 BLP

若CEF是相关关系的 \rightarrow LPF是相关的

若CEF是因果关系 \rightarrow LPF是因果的

问题是：怎样获得一个因果的 CEF？(客观)

\rightarrow 必须依赖于理论认知(主观)

回归分析

CEF、LPF和因果关系

实际上，我们是从理论得到结论，进而在样本中验证理论关系是否成立。因此，直接作法是使用 LPF 进行建模。

由于只有 **线性的 CEF = LPF**，使用 LPF 建模(即便模型设定是正确的)

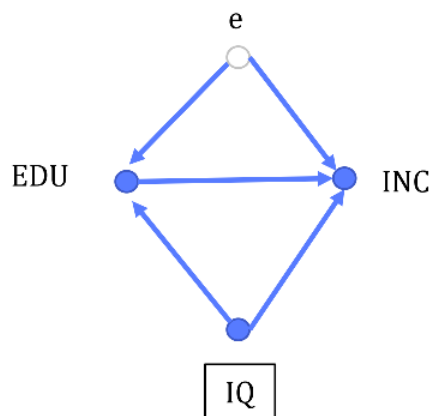
先天的 一部分信息就进入到了干扰项 e ，因此只有假定 **干扰项条件均值独立于解释变量**，即 $E(e | X) = E(e) = c$ ，才保证 LPF（依据理论建构的模型）= CEF（线性）

回忆下： CEF 天然就具备性质 $E(e | X) = E(e) = 0$

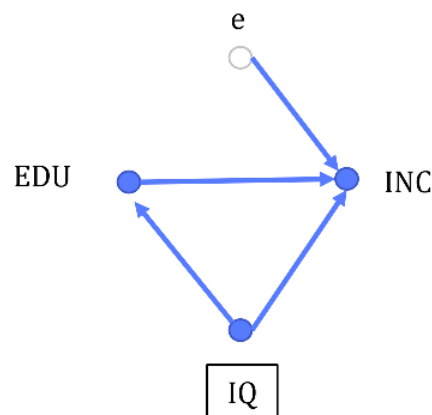
干扰项条件均值独立于解释变量的理解

假设 LPF:

$$INC = \alpha + \beta_1 EDU + \beta_2 IQ + e$$



干扰项与解释变量相关



干扰项与解释变量不相关

- 有向无环图(directed acyclic graph)
- 实心-可观测；空心-不可观测；单向箭头-因果关系；无法递归
- $EDU \rightarrow INC$ 直接因果路径
- $EDU \leftarrow IQ \rightarrow INC$ 混淆路径1
- $EDU \leftarrow e \rightarrow INC$ 混淆路径2
- 右图：干扰项条件均值独立于解释变量
- 左图：控制IQ无效

干扰项条件均值独立于解释变量的理解

$E(e | EDU, IQ) = E(e) = c$ 意味着：给定我们班{EDU = 研究生, IQ = 高} 干扰项的平均值一样(c)，假如性格（不可观测）进入 e ，干扰项条件均值独立于解释变量成立就表示符合上面两个条件的同学的性格是一样的。

这样，若EDU与IQ其中一个变化，另一不变，收入的条件平均值变化就可以归因于其中那个变化了的条件，获得“净”的因果效应。

因此，LPF中的参数 β_1 和 β_2 就是 EDU 和 IQ 对 INC 的因果效应。

将 $E(e | EDU, IQ) = c$ 的常数 c 并入 LPF常数项，就获得了与CEFE性质一样的 $E(e | EDU, IQ) = 0$ 重要假设。只不过该假设是从 LPF 出发建立的，也就是常在计量教材见到的线性回归方程的**干扰项条件均值为0**的假设。

干扰项条件均值独立于解释变量的理解

对LPF: $INC = \alpha + \beta_1 EDU + \beta_2 IQ + e$

两边取条件期望, 得到对应的线性CEF:

$$\begin{aligned} E(INC \mid EDU, IQ) \\ &= \alpha + \beta_1 EDU + \beta_2 IQ + E(e \mid EDU, IQ) \\ &= \alpha + \beta_1 EDU + \beta_2 IQ \end{aligned}$$

将线性CEF对EDU求偏导: 偏回归系数

$$\frac{dE(INC \mid EDU, IQ)}{dEDU} = \beta_1$$

β_1 表示在IQ固定不变, INC 的期望值 (均值) 随 EDU 如何变化。

相关关系CEF

若 $E(e \mid EDU, IQ) \neq 0$, 干扰项的存在如何影响因果关系的估计?

假设只观测到了INC和EDU, 将 LPF 写作:

$$INC = \alpha + \beta_1 EDU + \varepsilon, \quad \varepsilon = \beta_2 IQ + e$$

此时:

$$E(\varepsilon \mid EDU) = E(\beta_2 IQ + e \mid EDU) = \beta_2 E(IQ \mid EDU) \neq 0$$

此时的LPF为:

$$INC = \alpha + \beta_1 EDU + \varepsilon, \quad E(\varepsilon \mid EDU) \neq 0$$

相关关系CEF

假如此时我们错误地把 LPF 当成了 正确的、具有因果关系的线性CEF,

实际上相当于理解为 我们建构的模型 LPF^{*}: $INC = \gamma_0 + \gamma_1 EDU + u, \quad E(u \mid EDU = 0)$

的对应的"具有因果关系" (其实是相关关系) CEF^{*}: $E(INC \mid EDU) = \gamma_0 + \gamma_1 EDU$

求偏导:

$$\frac{dE(INC \mid EDU)}{dEDU} = \gamma_1$$

γ_1 反映了INC 的期望值随EDU如何变化, 但并没有控制IQ不变。

相关关系CEF

计算 γ_1 和 β_1 的关系:

$$\begin{aligned} E(INC \mid EDU) \\ &= E(\alpha + \beta_1 EDU + \beta_2 IQ + e \mid EDU) \\ &= \alpha + \beta_1 EDU + \beta_2 E(IQ \mid EDU) \end{aligned}$$

- 对EDU求导:

$$\frac{dE(INC \mid EDU)}{dEDU} = \beta_1 + \beta_2 \frac{dE(IQ \mid EDU)}{dEDU}$$

即

$$\gamma_1 = \beta_1 + \beta_2 \frac{dE(IQ \mid EDU)}{dEDU}$$

相关关系CEF

- 假设受教育程度于智商之间存在线性相关关系:

$$E(IQ \mid EDU) = \phi_0 + \phi_1 EDU$$

即

$$\frac{dE(IQ \mid EDU)}{dEDU} = \phi_1$$

则:

$$\gamma_1 = \beta_1 + \beta_2 \phi_1$$

相关关系CEF

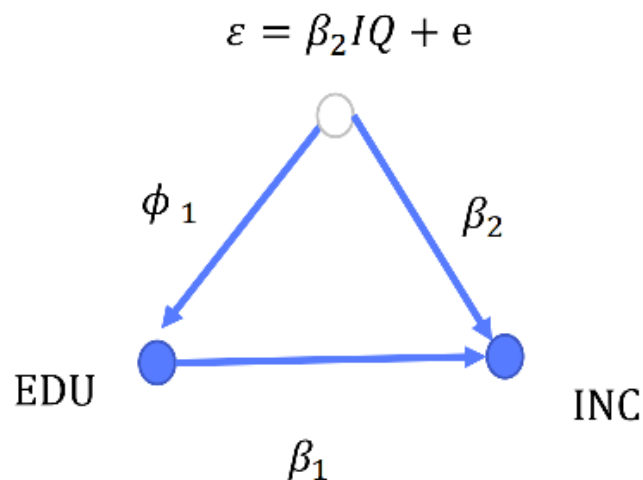
由此可见，错误的LPF*: $INC = \gamma_0 + \gamma_1 EDU + u, \quad E(u \mid EDU = 0)$

相关关系CEF

由此可见，错误的LPF^{*}: $INC = \gamma_0 + \gamma_1 EDU + u$, $E(u | EDU = 0)$

代表是错误的线性CEF^{*}，其估计系数 γ_1 反映是相关关系：

- $EDU \rightarrow INC$ 的因果影响 β_1
- EDU 与 IQ 的相关性 ϕ_1 乘以 IQ 对 INC 的因果影响 β_2



统计推断与大样本性质

统计推断与大样本性质

为什么需要推断?

之前 重点关注了CEF与LPF，都是总体的。刻画总体我们只能通过样本, 通过统计推断的方式进行。

本小节要回答以下问题：

1. 如何估计系数 $\hat{\beta}$?
2. $\hat{\beta}$ 的样本分布如何?
3. 如何解释估计系数 $\hat{\beta}$?

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_i \\ \dots \\ Y_N \end{bmatrix} = \begin{bmatrix} x_{11} & x_{21} & x_{31} & \dots & x_{k1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{k2} \\ \dots & \dots & \dots & \dots & \dots \\ x_{1i} & x_{2i} & x_{3i} & \dots & x_{ki} \\ \dots & \dots & \dots & \dots & \dots \\ x_{1N} & x_{2N} & x_{3N} & \dots & x_{kN} \end{bmatrix} \times \begin{bmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_i \\ \dots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_i \\ \dots \\ e_N \end{bmatrix}$$

进一步用 k 维向量 $X_i = (x_{1i} \ x_{2i} \ \dots \ x_{ki})'$, $\beta = (\beta_1 \ \beta_2 \ \dots \ \beta_k)'$ 简化为:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_i \\ \dots \\ Y_N \end{bmatrix} = \begin{bmatrix} X'_1 \\ X'_2 \\ \dots \\ X'_i \\ \dots \\ X'_N \end{bmatrix} \times \beta + \begin{bmatrix} e_1 \\ e_2 \\ \dots \\ e_i \\ \dots \\ e_N \end{bmatrix}$$

再用 $N \times k$ 矩阵 $\mathbf{X} = (X_1 \ X_2 \ \dots \ X_N)'$ 、 N 维向量 $Y = (Y_1 \ Y_2 \ \dots \ Y_N)'$
 $e = (e_1 \ e_2 \ \dots \ e_N)'$ 简化为:

总体最小二乘

将LPF 设定为: $Y = X'\beta + e$, $E(e | X) = 0$

展开: $Y = \alpha + \beta_1 X_1 + \cdots + \beta_k X_k + e$

该 LPF 对应的 线性CEF, 可以是因果关系的CEF, 也可以是相关关系的CEF:

$$E(Y | X) = X'\beta$$

利用最小二乘法求解系数 $\hat{\beta}_{ols}$, 就是最小化 Y 与 线性投影预测值 $\hat{Y} = X'b$ 的残差 $\varepsilon = Y - \hat{Y}$ 的MSE

$$\hat{\beta}_{ols} = \underset{b}{\operatorname{argmin}} E \left[(Y - X'b)^2 \right]$$

总体最小二乘

由一阶条件可得：

$$E \left[\mathbf{X} \left(Y - \mathbf{X}' \hat{\boldsymbol{\beta}} \right) \right] = 0$$

此条件同等与：

$$E \left[\mathbf{X} \left(Y - \mathbf{X}' \hat{\boldsymbol{\beta}} \right) \right] = E[\mathbf{X}\varepsilon] = 0$$

由此可见, LS的本质是求解系数 $\hat{\boldsymbol{\beta}}_{ols}$, 使得解释变量 X 与残差 ε 不相关

$$\hat{\boldsymbol{\beta}}_{ols} = E[\mathbf{X}\mathbf{X}']^{-1} E[\mathbf{X}Y]$$

总体最小二乘

将LPF代入上式:

$$\begin{aligned}\hat{\beta}_{ols} &= E[\mathbf{X}\mathbf{X}']^{-1}E[\mathbf{X}Y] = E[\mathbf{X}\mathbf{X}']^{-1}E[\mathbf{X}(\mathbf{X}'\beta + e)] \\ &= \beta + E[\mathbf{X}\mathbf{X}']^{-1}E[\mathbf{X}e]\end{aligned}$$

其中由于假设: $E(e | X) = 0$

- 故 $E[\mathbf{X}e] = E_{\mathbf{X}}[E(\mathbf{X}e | \mathbf{X})] = E_{\mathbf{X}}[\mathbf{X}E(e | \mathbf{X})] = \mathbf{0}$
- 故 $\hat{\beta}_{ols} = \beta$ 以上讨论说明, 最小二乘法的解对应的系数 $\hat{\beta}_{ols}$ 是LPF
 $Y = \mathbf{X}'\beta + e, E(e | X) = 0$, 即条件期望函数 $E(Y | X) = \mathbf{X}'\beta$ 的系数 β

意味着: 对因果关系的CEF使用最小二乘方法 \rightarrow 估计的系数 $\hat{\beta}_{ols}$ 具有因果效应

干扰项和残差

- 干扰项 e 包含了除 X 外的其他影响 Y 的因素, 与 X 是否相关无法检验, 靠经验和理论判断;
- 残差 ε 使用最小二乘法计算出来的, 总是与 X 不相关。

$$E \left[\mathbf{X} \left(Y - \mathbf{X}' \hat{\boldsymbol{\beta}} \right) \right] = E[\mathbf{X} \varepsilon] = 0$$

- 只有当干扰项 e 与 X 不相关时, 残差 ε 才是干扰项 e 的正确估计。

从总体到样本

最小二乘只是估计方法[†]，常见的估计方法还有矩方法、最大似然估计等

- 在总体中 $\hat{\beta}_{ols} = E[\mathbf{X}\mathbf{X}']^{-1} E[\mathbf{X}Y]$
- 在样本中 $\hat{\beta}_{ols} = (\sum_i X_i X_i')^{-1} (\sum_i X_i Y_i)$

[†] **MHE** 还提供矩方法(method-of-moments) 作为备选。根据大数定律和 中心极限定理使用样本矩 $\frac{1}{n} \sum_i X_i X_i'$ 估计总体矩 $E[X_i X_i']$. 还可以使用其他估计方法, *e.g.* Y_i 给定 X_i 去最小化 Y_i 的MSE.

假设真实的总体回归函数(永远未知)是 $Y = \mathbf{X}'\boldsymbol{\beta} + e$, 代入上式可得:

$$\hat{\boldsymbol{\beta}}_{ols} = \boldsymbol{\beta} + E[\mathbf{X}\mathbf{X}']^{-1}E[\mathbf{X}e]$$

其渐进协方差为: $E[\mathbf{X}\mathbf{X}']^{-1}\mathbf{X}'E[ee']\mathbf{X}E[\mathbf{X}\mathbf{X}']^{-1}$

$$E[ee'] = Var(e) = \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} & \cdots & \sigma_{1N} \\ \sigma_{21} & \sigma_2^2 & \sigma_{23} & \cdots & \sigma_{2N} \\ \sigma_{31} & \sigma_{32} & \sigma_3^2 & \cdots & \sigma_{3N} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{N1} & \sigma_{N2} & \sigma_{N3} & \cdots & \sigma_N^2 \end{bmatrix}$$

实际中用样本残差 $\varepsilon = Y - \hat{Y}$ + 同方差/异方差假设 $\longrightarrow Var(\hat{\boldsymbol{\beta}}_{ols})$

几个重要的大样本性质

1. 统计推断建立在随机变量与样本分布基础上，实际就是用 $\hat{\beta}$ 去推断真实值(未知) β
2. $\hat{\beta}$ 是一个r.v.,因此具有分布(均值和方差)
3. 在大样本下($n > 200$), OLS估计量 $\hat{\beta}$ 是总体回归系数 β 的一致估计 ($\text{plim } \hat{\beta} = \beta$)
4. 用样本残差代入中间项得到标准误差称为异方差一致性标准误差或稳健标准误差, STATA 中在回归后加上 `vce(robust)`
5. 在传统计量经济学中，对推断问题的处理从更强的假设开始。传统方式有时候被叫做经典正态回归模型，它假定自变量是**固定随机变量**、CEF是线性的、总体误差项符合正态分布以及同方差。这些更强假设可以得到：(1) OLS估计量无偏性（小样本性质），
(2) OLS估计值的样本方差公式, 在小样本和大样本中都成立, 即有效的。这种情况下，即便CEF非线性，随机化实验(RCTs)也可保证无偏性。

因为异方差是社会科学的常态, 所有Angrist and Pischke认为应该把很多统计工具的默认值切换到异方差(heteroskedasticity).

若是CEF是非线性的, 线性回归就会产生异方差.

$$\begin{aligned} E\left[(Y_i - X_i'\beta)^2 \mid X_i\right] \\ &= E\left[\left(\{Y_i - E[Y_i \mid X_i]\} + \{E[Y_i \mid X_i] - X_i'\beta\}\right)^2 \mid X_i\right] \\ &= Var(Y_i \mid X_i) + (E[Y_i \mid X_i] - X_i'\beta)^2 \end{aligned}$$

即使 $Var(Y_i \mid X_i)$ 是常数, 第二项也是异方差.

应用中

1. 异方差不是推断的**最大障碍**.

...如果异方差变化很大，比如使标准误上升了30%或者有任何显著的下降，那么你应该注意是不是存在可能的程序错误或者是其他问题。如果稳健标准误低于传统意义上的标准误[†](一般会大于)，可能标志着在稳健性计算中存在有限样本偏误。(MHE, p.47)

2. 现代微观实证建立在大样本假定下，已经避免了强的经典假设(正态分布、自变量非随机、线性 CEF、同方差)。

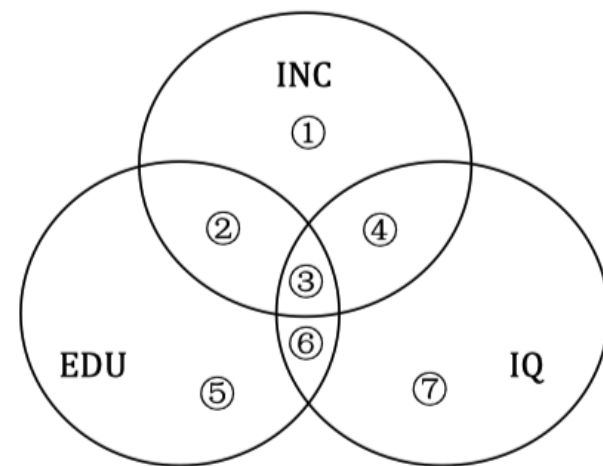
[†] 标准差(SD)是样本数据的方差的平方根,衡量样本数据的离散程度; 标准误(SE)是样本均值(是r.v.)的标准差，衡量样本均值的离散程度。实际中用样本均值来推断总体均值，那么样本均值的离散程度（标准误）越大，抽样误差就越大。所以用标准误（SE）来衡量抽样误差的大小，所以统计软件中报告是SE

韦恩图理解系数估计值

- 两两相交的部分指的是两个变量共同变化的部分，表明两个变量存在一定的线性相关关系

$$INC = \alpha + \beta_1 EDU + \beta_2 IQ + e$$

- $E(e | EDU, IQ) = 0$
 $Cov(EDU, IQ) \neq 0$
- β_1 反映的只有(2)的信息, β_2 反映的只有(4)的信息, 而信息(3) (反映了共同的影响) 被同时舍去
- 此时, 回归系数反映了EDU和 IQ 各自对INC的影响, 即因果关系



系数估计值的分解

基于OLS得到任何 X_i 的系数 $\hat{\beta}_i$ 都可以分解为两步求解

- 第一步: 将 X_i 作为因变量, 其他不包含 X_i 的解释变量作为自变量, OLS得到残差项 $\widetilde{X}_i = \hat{v}_i$, 即(2)+(5):

$$X_i = \hat{\gamma}_0 + \hat{\gamma}_1 X_1 + \cdots + \hat{\gamma}_{i-1} X_{i-1} + \hat{\gamma}_{i+1} X_{i+1} + \cdots + \hat{\gamma}_k X_k + \hat{v}_i$$

- 第二步: Y作因变量, \widetilde{X}_i 作自变量, OLS得到 $\hat{\beta}_i$, 即(2):

$$Y = \hat{\alpha} + \hat{\beta}_i \widetilde{X}_i + \hat{\varepsilon}$$

$$\text{其中 } \hat{\beta}_i = \frac{Cov(Y, \hat{v}_i)}{Var(\hat{v}_i)}$$

