II 核心
3 让回归变得有意义

胡智慧

基本无害讨论班，2019/10/10

3.1 回归的基本原理

3.1.1 经济关系与条件期望函数

- “系统性随机性”的例子：教育水平和收入的关系
  - 教育水平更高的人赚的更多，但这并不意味着教育水平的提高导致(cause)了收入提升
  - 教育水平能预测收入——使用CEF概括和总结这种预测能力

- **条件期望函数**(Conditional Expection Function, CEF)

$E[Y_i|X_i]$：给定X_i不变时，Y_i的期望或总体均值

① 连续Y_i：$E[Y_i|X_i=x]=\int t\underbrace{f_y(t|X_i=x)}_{\text{条件密度函数}}dt$

② 离散Y_i：$E[Y_i|X_i=x]=\sum_t t\underbrace{P(Y_i=t|X_i=x)}_{\text{条件概率质量函数}}$

3.1.1 经济关系与条件期望函数

- **迭代期望律**(the law of iterated expections, LIE)

$$E[Y_i] = E\{E[Y_i|X_i]\}$$

- 证明：$\mathrm{E}\{E[Y_i|X_i]\} = \int \underline{E[Y_i|X_i = u]} g_x(u)du$

$$= \int \left[\underline{\int t f_y(t|X_i = u)dt} \right] g_x(u)du$$

$$= \iint t\, f_y(t|X_i = u) g_x(u)dudt$$

$$= \int t \left[\int \underline{f_y(t|X_i = u) g_x(u)du} \right] dt$$

条件密度函数*边缘密度函数=联合密度函数

$$= \int t \left[\underline{\int \underline{f_{xy}(u,t)}du} \right] dt = \int t \underline{g_y(t)}dt = E[Y_i]$$

边缘密度函数

3.1.1 经济关系与条件期望函数

- **定理3.1.1 条件期望函数的分解性质**

$$Y_i = E[Y_i|X_i] + \varepsilon_i$$

(i) ε_i关于X_i均值独立，即$E[\varepsilon_i|X_i] = 0$

(ii) ε_i与关于X_i的任何函数都不相关，即$E[h(X_i)\varepsilon_i] = 0$

- 证明：

(i) $E[\varepsilon_i|X_i] = E[Y_i - E[Y_i|X_i]|X_i] = E[Y_i|X_i] - E[Y_i|X_i] = 0$

(ii) $E[h(X_i)\varepsilon_i] = E\{E[h(X_i)\varepsilon_i]|X_i\} = E\{h(X_i)E[\varepsilon_i|X_i]\} = 0$

3.1.1 经济关系与条件期望函数

- **定理3.1.2 条件期望函数的预测性质**

  令$m(X_i)$是关于X_i的任何函数，CEF是下面问题的解：

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

  条件期望函数就是给定X_i后，对Y_i的最小均方误预测

- 证明：

$$(Y_i - m(X_i))^2 = \left((Y_i - E[Y_i|X_i]) + (E[Y_i|X_i] - m(X_i))\right)^2$$

$$= \underbrace{(Y_i - E[Y_i|X_i])^2}_{\text{不包含}m(X_i)} + \underbrace{\left(E[Y_i|X_i] - m(X_i)\right)^2}_{m(X_i) = E[Y_i|X_i]\text{ 最小化}}$$

$$+ 2\underbrace{(Y_i - E[Y_i|X_i]) \times \left(E[Y_i|X_i] - m(X_i)\right)}_{\text{可以写成}h(X_i)\epsilon_i\text{，其期望为0}}$$

3.1.1 经济关系与条件期望函数

- **定理3.1.3 方差分析(ANOVA)定理**

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

其中$V(\cdot)$表示方差，$V(Y_i|X_i)$表示给定X_i下Y_i的条件方差

- 证明：

$$Y_i = E[Y_i|X_i] + \varepsilon_i$$

$$V(Y_i) = V(E[Y_i|X_i]) + \underline{V(\varepsilon_i)}$$

$$V(\varepsilon_i) = E\left(\varepsilon_i^2\right) = E\left[E\left[\varepsilon_i^2|X_i\right]\right] = E[V(Y_i|X_i)]$$

$$\varepsilon_i = Y_i - E[Y_i|X_i]$$

$$E\left[\varepsilon_i^2|X_i\right] = V(Y_i|X_i)$$

3.1.2 线性回归与条件期望函数

- **回归系数向量**：$(k \times 1)$

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

由一阶条件可得：

$$E[X_i(Y_i - X_i'b)] = 0$$

总体残差$e_i = Y_i - X_i'\beta$与回归元X_i不相关：

$$E[X_i(Y_i - X_i'\beta)] = E[X_i e_i] = 0$$

- 最小二乘估计的解：

$$\beta = E[X_i X_i']^{-1} E[X_i Y_i]$$

$$S(b) = E[(Y_i - X_i'b)^2] = E(Y_i^2) - 2b'E(X_i Y_i) + b'E(X_i X_i')b$$

$$\frac{\partial S(b)}{\partial b} = -2E(X_i Y_i) + 2E(X_i X_i')\beta = 0$$

3.1.2 线性回归与条件期望函数

- 单一回归元x_i， $\beta_1 = \frac{cov(Y_i,X_i)}{V(X_i)}$， $\alpha = E[Y_i] - \beta_1 E[X_i]$
- **解构回归**(regression anatomy)

  多变量时，第k个回归元的斜率系数是：$\beta_k = \frac{cov(Y_i,\tilde{x}_{ki})}{V(\tilde{x}_{ki})}$

其中，$\tilde{x}_{ki}$是将x_{ki}关于其他回归元回归后得到的残差项

- 证明：

$$Y_i = \alpha + \beta_1 x_{1i} + \cdots + \beta_k x_{ki} + \cdots + \beta_K x_{Ki} + e_i$$

$\tilde{x}_{ki}$是所有回归元的线性组合，与e_i不相关

$\tilde{x}_{ki}$是回归后的残差项，与其他回归元不相关

$\tilde{x}_{ki}$与x_{ki}的协方差就是$V(\tilde{x}_{ki})$

$$cov(Y_i, \tilde{x}_{ki}) = \beta_k V(\tilde{x}_{ki})$$

3.1.2 线性回归与条件期望函数

- **定理3.1.4 线性条件期望函数（回归的理由I）**

  如果CEF是线性的，那么总体回归方程就是它

- 证明：

  假设$E[Y_i|X_i]=X_i'\beta^*$，其中β^*是某$k\times 1$维向量

  由CEF分解性质：

$$E[X_i(Y_i-E[Y_i|X_i])]=0$$

  将$E[Y_i|X_i]=X_i'\beta^*$代入：

$$\beta^*=E[X_iX_i']^{-1}E[X_iY_i]=\beta$$

- 问题：在什么条件下，CEF是线性的？
  - 经典方案：加上联合正态分布假设（离散变量有局限性）
  - 另一线性化方案：饱和回归模型

3.1.2 线性回归与条件期望函数

- **定理3.1.5 最优线性估计量定理（回归的理由II）**

  在最小均方误的意义下，给定X_i，函数$X_i'\beta$是对Y_i的最优线性估计量

- 证明：

  $\beta = E[X_iX_i']^{-1}E[X_iY_i]$是总体的最小方差问题(3.1.2)的解

定理3.1.2：

给定X_i，在所有关于X_i的函数中，条件期望函数$E[Y_i|X_i]$能最好地(MMSE)预测Y_i

定理3.1.5：

在所有关于X_i的线性函数中，总体回归函数$X_i'\beta$能最好地(MMSE)预测Y_i

3.1.2 线性回归与条件期望函数

- **定理3.1.6 条件期望函数的回归定理（回归的理由III）**

  在最小均方误的意义下，函数$X_i'\beta$提供了对$E[Y_i|X_i]$的最优线性近似，即：

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

- 证明：

  观察到β是最小二乘估计的解，记：

$$\begin{aligned}(Y_i - X_i'b)^2 &= \big((Y_i - E[Y_i|X_i]) + (E[Y_i|X_i] - X_i'b)\big)^2 \\ &= (Y_i - E[Y_i|X_i])^2 + (E[Y_i|X_i] - X_i'b)^2 \\ &\quad +2(Y_i - E[Y_i|X_i]) \times (E[Y_i|X_i] - X_i'b)\end{aligned}$$

条件期望函数的线性近似问题和总体最小二乘问题相同

3.1.2 线性回归与条件期望函数

- 定理3.1.6——CEF回归定理的理解：

　　可以通过将$E[Y_i|X_i]$当做被解释变量进行回归来得到回归系数，不需要用Y_i

- 证明：

(i) 假定X_i是离散随机变量，其概率质量函数为$g_x(u)$：

$$E\{(\underline{E[Y_i|X_i]} - X_i'b)^2\} = \sum_u (E[Y_i|X_i = u] - u'b)^2 \, g_x(u)$$

　　可对$E[Y_i|X_i = u]$关于u进行加权最小二乘(WLS)来得到β，其中u取遍X_i的可能取值

(ii) 在关于β的公式中使用迭代期望律：

$$\beta = E[X_iX_i']^{-1}E[X_iY_i] = E[X_iX_i']^{-1}E[X_i\underline{E(Y_i|X_i)}]$$

3.1.3 渐进最小二乘推断

- 假设向量$W_i \equiv (Y_i \;\; X_i')'$在大小为$N$的样本中独立同分布

  总体一阶矩$E[W_i]$的样本估计：$\frac{1}{N}\sum_{i=1}^{N} W_i$ (大数定理)

  总体二阶矩$E[W_i W_i']$的样本估计：$\frac{1}{N}\sum_{i=1}^{N} W_i W_i'$

- **β的最小二乘(OLS)估计量：**

$$\hat{\beta} = \left[\sum_i X_i X_i'\right]^{-1} \sum_i X_i Y_i$$

- $\hat{\beta}$的渐进样本分布，完全依赖与对被估计量的定义和对数据来自随机样本的假设

3.1.3 渐进最小二乘推断

1. **大数定理(The law of Large Numbers)**

   样本矩依概率收敛于相应的总体矩

2. **中心极限定理(The Central Limit Theorem)**

   样本矩是渐进正态分布的（在减去总体矩，除以样本规模的平方根后），渐进协方差矩阵由相应随机变量的方差给出

3. **斯拉茨基定理(Slutsky's Theorem)**

   a_N是具有渐进分布的一个统计量，b_N是具有极限概率b的一个随机变量

   (1) $a_N + b_N$和$a_N + b$有相同的渐进分布

   (2) $a_N b_N$和$a_N b$有相同的渐进分布

3.1.3 渐进最小二乘推断

4. 连续映射定理(The Continuous Mapping Theorem)

函数$h(b_N)$的概率极限是$h(b)$，其中$plim\, b_N = b$，函数$h(\cdot)$在b处连续

5. 德尔塔法(The Delta Method)

如果随机变量b_N是渐进正态分布的，其协方差矩阵是Ω，$plim\, b_N = b$， $h(\cdot)$是连续可微的函数，具有梯度$\nabla h(b)$，那么$h(b_N)$的渐进分布是正态的，其协方差矩阵为$\nabla h(b)'\Omega\nabla h(b)$

“$h(b_N)$的渐进分布”实际上指$\sqrt{n}(h(b_N) - h(b))$的渐进分布

3.1.3 渐进最小二乘推断

- 使用斯拉茨基定理和中心极限定理：

$$Y_i = X_i'\beta + [Y_i - X_i'\beta] \equiv X_i'\beta + e_i$$

$$\beta = E[X_iX_i']^{-1}E[X_iY_i]$$

$$\hat{\beta} = \left[\sum_i X_iX_i{}'\right]^{-1}\sum_i X_iY_i$$

用$Y_i = X_i'\beta + e_i$替换$\hat{\beta}$中的Y_i：

$$\hat{\beta} = \beta + \left[\sum X_iX_i{}'\right]^{-1}\sum X_ie_i$$

$$\sqrt{N}(\hat{\beta} - \beta) = N\left[\sum X_iX_i{}'\right]^{-1}\frac{1}{\sqrt{N}}\sum X_ie_i \to E[X_iX_i']^{-1}\underline{\frac{1}{\sqrt{N}}\sum X_ie_i}$$

由CLT：渐进正态分布，均值0，协方差矩阵$E[X_iX_i{}'e_i^2]$

- $\hat{\beta}$是概率极限为β，协方差矩阵为

$$E[X_iX_i']^{-1}E\left[X_iX_i'e_i^2\right]E[X_iX_i']^{-1}$$

的渐进正态分布

3.1.3 渐进最小二乘推断

- **异方差修正标准误/怀特标准误/稳健标准误**

用估计残差$\hat{e_i} = Y_i - X_i'\hat{\beta}$，构造四阶矩矩阵$\sum[X_iX_i'\hat{e_i}^2]/N$

- 软件默认的标准误(同方差假设)：$E\left[e_i^2|X_i\right] = \sigma^2$

$$E\left[X_iX_i'e_i^2\right] = E\left(X_iX_i'E\left[e_i^2|X_i\right]\right) = \sigma^2 E[X_iX_i']$$

$\hat{\beta}$的渐进协方差矩阵简化为：

$$E[X_iX_i']^{-1}E\left[X_iX_i'e_i^2\right]E[X_iX_i']^{-1}$$
$$= E[X_iX_i']^{-1}\sigma^2 E[X_iX_i']E[X_iX_i']^{-1}$$
$$= \sigma^2 E[X_iX_i']^{-1}$$

- 将回归看作对CEF的近似，异方差：

$$E[(Y_i - X_i'\beta)^2|X_i]$$
$$= E\{Y_i - E[Y_i|X_i]) + (E[Y_i|X_i] - X_i'\beta)|X_i\}$$
$$= \mathrm{V}[Y_i|X_i] + (E[Y_i|X_i] - X_i'\beta)^2$$

3.1.4 饱和模型、主效应和回归的其他讨论

- **饱和(saturated)回归模型**

具有离散解释变量的回归模型，对解释变量的每一个可能取值，该模型都存在一个参数与之对应

- 考虑$S_i = 0,1,2,\cdots,\tau$，针对S_i的饱和模型是：

$$Y_i = \alpha + \beta_1 d_{1i} + \beta_2 d_{2i} + \cdots + \beta_\tau d_{\tau i} + \varepsilon_i$$

其中，$d_{ji} = 1[S_i = j]$是虚拟变量/示性函数，表示个体是否接受了j年的教育

β_j是受教育水平为j年级所带来的效应：

$$\beta_j = E[Y_i|S_i = j] - E[Y_i|S_i = 0]$$

其中，$\alpha = E[Y_i|S_i = 0]$

3.1.4 饱和模型、主效应和回归的其他讨论

- **主效应(main effect)：** 虚拟变量的系数
- **交互项：** 两个虚拟变量的乘积

- Eg. x_{1i}表示大学毕业， x_{2i}表示女性

$$E[Y_i|x_{1i}=0,x_{2i}=0]=\alpha$$
$$E[Y_i|x_{1i}=1,x_{2i}=0]=\alpha+\beta_1$$
$$E[Y_i|x_{1i}=0,x_{2i}=1]=\alpha+\gamma$$
$$E[Y_i|x_{1i}=1,x_{2i}=1]=\alpha+\beta_1+\gamma+\delta_1$$

有两个主效应和一个交互项的参数化过程：

$$E[Y_i|x_{1i},x_{2i}]=\alpha+\beta_1x_{1i}+\gamma x_{2i}+\delta_1(x_{1i}x_{2i})$$
$$Y_i=\alpha+\beta_1x_{1i}+\gamma x_{2i}+\delta_1(x_{1i}x_{2i})+\varepsilon_i$$

3.1.4 饱和模型、主效应和回归的其他讨论

- 将作为多值变量的教育变量与性别变量结合起来：

$$Y_i = \alpha + \sum_{j=1}^{\tau} \underline{\beta_j d_{ji}} + \underline{\gamma x_{2i}} + \sum_{j=1}^{\tau} \underline{\delta_j (d_{ji}\, x_{2i})} + \varepsilon_i$$

- 不包含交互项时，只有主效应相加的模型
- 只包含交互项，省略主效应的模型：

$$Y_i = \alpha + \gamma x_{2i} + \sum_{j=1}^{\tau} \delta_j (d_{ji}\, x_{2i}) + \varepsilon_i$$

- 饱和模型能够完美拟合CEF，这个性质与Y_i的分布无关(线性条件期望函数定理的重要特例)

3.2 回归与因果关系

3.2.1 条件独立假设

- 如果对于给定的总体，CEF刻画了平均潜在结果之间的不同，那么这个CEF具有因果性
- 条件独立假设(conditional independence assumption, CIA) 是能对回归赋予因果解释的核心假设
- 教育回报问题：虚拟变量C_i表示是否上大学

$$潜在结果 = \begin{cases} Y_{1i} & if\ C_i = 1 \\ Y_{0i} & if\ C_i = 0 \end{cases}$$

$$Y_i = Y_{0i} + \underline{(Y_{1i} - Y_{0i})}C_i$$

$$\underbrace{E[Y_i|C_i = 1] - E[Y_i|C_i = 0]}_{观察到的收入差距}$$

$$= \underbrace{E[Y_{1i} - Y_{0i}|C_i = 1]}_{处理的平均因果效应} + \underbrace{E[Y_{0i}|C_i = 1] - E[Y_{0i}|C_i = 0]}_{选择性偏误}$$

3.2.1 条件独立假设

- **条件独立假设(CIA)**指的是给定观察到的特点X_i，选择性偏误消失

$$\{Y_{0i}, Y_{1i}\} \perp |X_i$$

其中，符号“⊥”表示$\{Y_{0i}, Y_{1i}\}$与C_i之间相互独立

$$E[Y_i|X_i, C_i = 1] - E[Y_i|X_i, C_i = 0] = E[Y_{1i} - Y_{0i}|X_i]$$

- 将CIA拓展到因果变量可取多值：

受教育年数Si，个体收入函数$Y_{si} \equiv f_i(s)$

- 更一般的条件下，CIA变为：

$$Y_i \perp S_i|X_i, \ \forall s$$

在许多随机试验中，S_i在给定X_i下随机分配，CIA成立

3.2.1 条件独立假设

- 给定X_i，多接受一年教育带来的平均因果效应：

$$E[Y_i|X_i,S_i=s]-E[Y_i|X_i,S_i=s-1]$$
$$=E[f_i(s)-f_i(s-1)|X_i]$$

- 高中毕业的平均因果效应：

$$E[Y_i|X_i,S_i=12]-E[Y_i|X_i,S_i=11]$$
$$=E[f_i(12)|x_i,S_i=12]-E[f_i(11)|x_i,S_i=11]$$
$$=\underline{E[f_i(12)-f_i(11)|X_i,\ S_i=12]}$$
$$=E[f_i(12)-f_i(11)|X_i]$$

- 高中毕业的无条件因果效应：

$$E\{E[Y_i|X_i,S_i=12]-E[Y_i|X_i,S_i=11]\}$$
$$=E\{E[f_i(12)-f_i(11)|X_i]\}$$
$$=E[f_i(12)-f_i(11)]$$

3.2.1 条件独立假设

- 回归——自动将CIA转化成需要估计的因果效应

(i) 假设$f_i(s)$是关于s线性的，只有误差项η_i因人而异

(ii) $f_i(s)$未必关于s线性的，允许$f_i(s)$因人而异、非线性

- (i) 线性的、因果效应为常数的模型，假设：

$$f_i(s) = \alpha + \rho s + \eta_i$$

误差项η_i均值为0，用以捕捉决定潜在收入的其他不可观测因素，将观察到的s_i代入上式：

$$Y_i = \alpha + \rho S_i + \eta_i$$

s_i可能与潜在结果$f_i(s)$相关，即与η_i相关

3.2.1 条件独立假设

- (ii) 给定一系列可观察的协变量X_i，CIA成立，分解潜在收入水平的随机项：

$$\eta_i = X_i'\gamma + v_i$$

其中，γ是总体回归系数所成向量，假设$E[\eta_i|X_i] = X_i'\gamma$

v_i与X_i不相关，由CIA有：

$$E[f_i(s)|X_i, S_i] = E[f_i(s)|X_i]$$
$$= \alpha + \rho s + E[\eta_i|X]$$
$$= \alpha + \rho s + \mathrm{X_i'}\gamma$$

- 线性模型

$$Y_i = \alpha + \rho s_i + X_i'\gamma + v_i$$

的残差项与回归元s_i以及X_i都不相关，回归系数ρ就是我们感兴趣的因果效应

3.2.2 遗漏变量偏误公式

- 遗漏变量偏误(omitted variable bias)公式

描述的是当回归包含不同的控制变量时，回归结果之间存在的关系，也即长、短回归方程估计系数之间的关系

在不含控制变量的较短的回归方程中得到的系数被认为是有偏的(biased)

- 研究教育回报的回归方程：

$$Y_i = \alpha + \rho s_i + A_i'\gamma + e_i$$

其中α，ρ和γ是总体回归系数，e_i是回归残差，回归残差与所有回归元都无关，A_i是控制变量，简化为由家庭背景、智力和动机组成的“能力”

给定A_i，如果CIA成立，这里的ρ就是线性的、常数因果效应模型中的ρ

3.2.2 遗漏变量偏误公式

- **遗漏变量偏误公式(Omitted Variable Bias Formula)**

$$\frac{cov(Y_i, S_i)}{V(S_i)} = \rho + \gamma' \delta_{As}$$

其中， δ_{As}是对A_i关于S_i回归得到的参数

- 遗漏变量偏误与解构回归公式：

当遗漏变量和纳入回归方程的变量不相关时，长回归和短回归得到的系数是一样的

3.2.3 不合格的控制变量

- **不合格的控制变量：** 本身可作为被解释变量的变量
- **合格的控制变量：** 当选定回归元后，其取值已经固定给出的变量
- 不合格的控制变量带来的问题——选择偏误

考虑大学教育对收入的影响，同时还可以在白领和蓝领之间进行职业选择

职业选择和教育水平、收入都高度相关，它是不是遗漏变量？——在同一类职业中考察教育对收入的影响

存在的问题：考虑教育水平影响职业选择，即使教育水平随机分配，同一职业内不同教育水平下的工资差异不可相互比较

3.2.3 不合格的控制变量

- 大学教育/职业选择：

  W_i表示个体i是否为白领工人，收入水平为Y_i

  接受或不接受大学教育都带来收入水平和职业选择的两种不同潜在结果，分别记为$\{Y_{1i}, Y_{0i}\}$和$\{W_{1i}, W_{0i}\}$

$$Y_i = C_i Y_{1i} + (1 - C_i) Y_{0i}$$

$$W_i = C_i W_{1i} + (1 - C_i) W_{0i}$$

其中，$C_i = 1$表示大学毕业水平，$C_i = 0$为其他；假设C_i随机分配，它独立与所有的潜在结果

$$E[Y_i | C_i = 1] - E[Y_i | C_i = 0] = E[Y_{1i} - Y_{0i}]$$

$$E[W_i | C_i = 1] - E[W_i | C_i = 0] = E[W_{1i} - W_{0i}]$$

3.2.3 不合格的控制变量

- 给定白领职业($W_i = 1$)，考虑大学毕业生和非大学毕业生的收入差距：

$$E[Y_i|W_i = 1, C_i = 1] - E[Y_i|W_i = 1, C_i = 0]$$

$$= E[Y_{1i}|W_{1i} = 1, C_i = 1] - E[Y_{0i}|W_{0i} = 1, C_i = 0]$$

$$= E[Y_{1i}|W_{1i} = 1] - E[Y_{0i}|W_{0i} = 1]$$

$$= \underbrace{E[Y_{1i} - Y_{0i}|W_{1i} = 1]}_{\text{因果效应}} + \underbrace{\{E[Y_{0i}|W_{1i} = 1] - E[Y_{0i}|W_{0i} = 1]\}}_{\text{选择性偏误}}$$

- 不合格控制变量带来的问题：相比较的是不同事物
- 选择偏误项反映出大学学历会改变白领工人组成
- 这个例子的选择偏误符号不确定，但很可能为负

3.2.3 不合格的控制变量

- 使用代理变量做控制变量：

即纳入回归方程的变量可能部分的控制遗漏变量，但是它本身被我们感兴趣的变量影响

$$Y_i = \alpha + \rho S_i + \gamma a_i + e_i$$

控制变量为表示能力的数值变量a_i，可视为八年级时学生的智力检测分数，$E[S_i e_i] = E[a_i e_i] = 0$

a_i是在个体做出S_i决策之前做出的，是好的控制变量

另一度量能力的指标a_{li}是在个体接受完教育后得到的

$$a_{li} = \pi_0 + \pi_1 S_i + \pi_2 a_i$$

$$Y_i = \left(\alpha - \gamma \frac{\pi_0}{\pi_2}\right) + \left(\rho - \gamma \frac{\pi_1}{\pi_2}\right) S_i + \frac{\gamma}{\pi_2} a_{li} + e_i$$

γ、π_1和π_2都为正，只有π_1很小时，才接近因果效应

3.2.3 不合格的控制变量

- 对被解释变量添加不合格的控制变量：

  如果想在对教育回报的研究获得带有因果性的结果，就不要在回归方程中加入职业作为控制变量

- 遗漏变量偏误公式——当不存在控制变量时，关于S_i做回归产生的系数是$\rho + \gamma\delta_{as}$，其中δ_{as}是对a_i关于S_i进行回归得到的系数

- 使用代理控制变量：

  得到的系数比没有控制变量情况下更接近于ρ

- 挑选控制变量的准则：

  考虑控制变量被决定的时间，一般在我们感兴趣的变量产生前就被决定的变量都是好的控制变量