

外生性下的线性回归

司继春

2025年10月



矩估计

在理解了外生性的概念之后，我们接下来讨论，如果外生性成立，我们该如何估计模型

$$y_i = x_i' \beta + u_i$$

需要注意的是，以下的估计方法都建立在独立同分布的假设条件下的：

假设（独立同分布）

假设 $\{[y_i, x_i', u_i]', i = 1, 2, \dots, N\}$ 为独立同分布的样本。

矩估计：识别条件

但是如果 $\mathbb{E}(x_i x_i')$ 不可逆，矩条件式将不会有唯一解，因而我们必须引入识别条件：

假设（识别条件，无完美共线性）

参数个数 K 是固定的，且 $\mathbb{E}(x_i x_i')$ 可逆。

矩估计

如果以上识别条件成立，根据矩估计的思想，如果我们分别使用均值：

$$\frac{1}{N} \sum_{i=1}^N x_i x_i' \text{ 和 } \frac{1}{N} \sum_{i=1}^N x_i y_i$$

代替总体均值，那么就得到了以上问题的矩估计：

$$\hat{\beta} = \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \frac{1}{N} \sum_{i=1}^N x_i y_i = \left[\sum_{i=1}^N x_i x_i' \right]^{-1} \sum_{i=1}^N x_i y_i$$

注意以上估计实际上就是最小二乘估计： $\hat{\beta} = (X'X)^{-1} X'Y$ 。

无偏性

无偏性意味着最小二乘估计量 $\hat{\beta}$ 估计真实参数 β 的平均（期望）误差为0。对于无偏性，由于：

$$\begin{aligned}\hat{\beta} &= (X'X)^{-1} X'Y \\ &= (X'X)^{-1} X'(X\beta + u) \\ &= \beta + (X'X)^{-1} X'u\end{aligned}$$

其中 $u = [u_1, \dots, u_N]'$ 为误差项向量。从而：

$$\begin{aligned}\mathbb{E}(\hat{\beta}|X) &= \beta + \mathbb{E}\left[(X'X)^{-1} X'u|X\right] \\ &= \beta + (X'X)^{-1} X'\mathbb{E}(u|X)\end{aligned}$$

在外生性假设下，有： $\mathbb{E}(\hat{\beta}|X) = \beta$ ，从而：

$$\mathbb{E}(\hat{\beta}) = \mathbb{E}\left[\mathbb{E}(\hat{\beta}|X)\right] = \beta$$

因而最小二乘估计量 $\hat{\beta}$ 是真实参数 β 的无偏估计

$\hat{\beta}$ 的方差估计

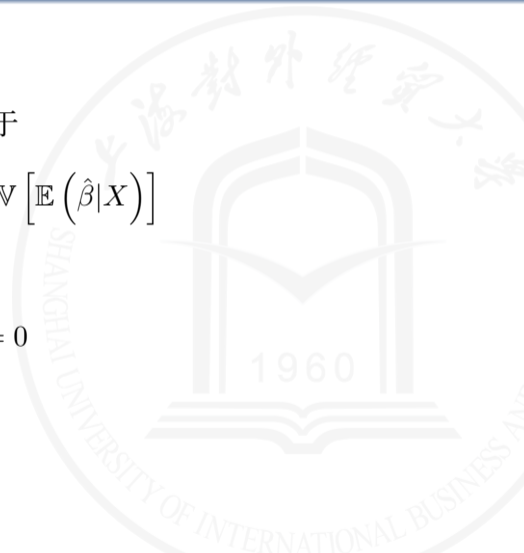
此外我们还可以计算最小二乘估计 $\hat{\beta}$ 的方差。由于

$$\mathbb{V}(\hat{\beta}) = \mathbb{E}[\mathbb{V}(\hat{\beta}|X)] + \mathbb{V}[\mathbb{E}(\hat{\beta}|X)]$$

而其中 $\mathbb{E}(\hat{\beta}|X) = \beta$ 从而

$$\mathbb{V}[\mathbb{E}(\hat{\beta}|X)] = 0$$

因而我们接下来将主要经历放在 $\mathbb{V}(\hat{\beta}|X)$ 上。



$\hat{\beta}$ 的方差估计

注意到由于 $\hat{\beta} - \beta = (X'X)^{-1} X'u$ ，因而 $\hat{\beta}$ 的条件协方差矩阵为：

$$\begin{aligned}\mathbb{V}(\hat{\beta}|X) &= \mathbb{E} \left([\hat{\beta} - \mathbb{E}(\hat{\beta}|X)] [\hat{\beta} - \mathbb{E}(\hat{\beta}|X)]' | X \right) \\ &= \mathbb{E} \left[(\hat{\beta} - \beta) (\hat{\beta} - \beta)' | X \right] \\ &= \mathbb{E} \left[(X'X)^{-1} X'uu'X (X'X)^{-1} | X \right] \\ &= (X'X)^{-1} X' \mathbb{E} [uu'|X] X (X'X)^{-1}\end{aligned}$$

了计算以上的方差，我们需要知道 $\mathbb{E} [uu'|X]$ 的结构。

同方差假设

- 根据独立同分布假设, 易知 $\mathbb{E}(u_i u_j | X) = 0$, 从而 $\mathbb{E}[u u' | X]$ 一定是一个对角矩阵。
- 此外, 我们还需要对对角线元素, 也就是 $\mathbb{V}(u_i | X) = \mathbb{V}(u_i | x_i)$ 的假设。

假设 (同方差, homoscedasticity)

假设 $\mathbb{V}(u_i | x_i) = \sigma^2$ 。

同方差假设

在同方差假定下:

$$\begin{aligned}\mathbb{E}(uu'|X) &= \begin{bmatrix} \mathbb{V}(u_1|X) & \mathbb{C}(u_1, u_2|X) & \cdots & \mathbb{C}(u_1, u_N|X) \\ \mathbb{C}(u_2, u_1|X) & \mathbb{V}(u_2|X) & \cdots & \mathbb{C}(u_2, u_N|X) \\ \vdots & \vdots & \ddots & \vdots \\ \mathbb{C}(u_N, u_1|X) & \mathbb{C}(u_N, u_2|X) & \cdots & \mathbb{V}(u_N|X) \end{bmatrix} \\ &= \begin{bmatrix} \sigma^2 & 0 & \cdots & 0 \\ 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma^2 \end{bmatrix}_{N \times N} = \sigma^2 I\end{aligned}$$

则上式可以化简为:

$$\begin{aligned}\mathbb{V}(\hat{\beta}|X) &= (X'X)^{-1} X' \mathbb{E}[uu'|X] X (X'X)^{-1} \\ &= (X'X)^{-1} X' \sigma^2 I X (X'X)^{-1} \\ &= \sigma^2 (X'X)^{-1}\end{aligned}$$

高斯-马尔可夫定理

- 可以证明，在 β 的所有线性无偏估计量中，最小二乘估计量是方差最小的估计量。
- 其中「线性性」意味着 β 的估计值是 Y 的一个线性函数，即可以写成 $\tilde{\beta} = CY$ 的形式， C 是一个不依赖于 Y （但是可以依赖于 X ）的矩阵。
- 现在不妨假设存在另外一个 β 的线性无偏估计量 $\tilde{\beta} = CY$ ，由于无偏性，有：

$$\begin{aligned}\beta &= \mathbb{E}(\tilde{\beta}) = \mathbb{E}(CY) \\ &= \mathbb{E}[C(X\beta + u)] \\ &= \mathbb{E}(CX)\beta + \mathbb{E}(Cu)\end{aligned}$$

上式意味着 $CX = I$ 以及 $\mathbb{E}(Cu) = 0$ 。

高斯-马尔可夫定理

- 注意 $\tilde{\beta}$ 的方差为： $\mathbb{V}(\tilde{\beta}|X) = \sigma^2 CC'$ 因而其两个估计量方差之逆的差为：

$$\begin{aligned}\mathbb{V}(\hat{\beta}|X)^{-1} - \mathbb{V}(\tilde{\beta}|X)^{-1} &= \frac{1}{\sigma^2} \left[X'X - (CC')^{-1} \right] \\ &= \frac{1}{\sigma^2} \left[X'X - X'C'(CC')^{-1}CX \right] \\ &= \frac{1}{\sigma^2} X' \left[I - C'(CC')^{-1}C \right] X\end{aligned}$$

注意到由于矩阵 $I - C'(CC')^{-1}C$ 为幂等矩阵，因而一定是半正定矩阵，从而：

$$\mathbb{V}(\hat{\beta}|X)^{-1} - \mathbb{V}(\tilde{\beta}|X)^{-1} \succeq 0$$

从而 $\mathbb{V}(\hat{\beta}|X) \preceq \mathbb{V}(\tilde{\beta}|X)$

高斯-马尔可夫定理

高斯-马尔可夫定理, Gauss-Morkov theorem

在外生性假设、独立同分布假设、识别假设、同方差假设的条件下，最小二乘估计量 $\hat{\beta}$ 是 β 的无偏估计量，此外最小二乘估计量是所有 β 的线性无偏估计量中方差最小的估计量，因而我们也称最小二乘估计量为最优线性无偏估计量（Best linear unbiased estimator, BLUE）。

标准误的估计

- 以上我们计算了在同方差假定下，最小二乘估计量的条件方差：

$$\mathbb{V}(\hat{\beta}|X) = \sigma^2 (X'X)^{-1} \quad (1)$$

其中 $(X'X)^{-1}$ 是可观测的，但是 σ^2 不可观测，所以为了计算条件方差的估计： $\mathbb{V}(\widehat{\hat{\beta}|X})$ ，我们还需要计算 σ^2 的估计。

- 由于 σ^2 是在同方差条件下误差项 u_i 的方差，一个很自然的想法是可以使用残差的方差：

$$\frac{1}{N-1} \sum_{i=1}^N \hat{u}_i^2 = \frac{\hat{u}'\hat{u}}{N-1} = \frac{\text{RSS}}{N-1}$$

对 σ^2 进行估计，然而该估计量不是无偏估计。

σ^2 的估计

- 为了得到 σ^2 的无偏估计, 注意到: $\hat{u} = MY = M(X\beta + u) = Mu$ 从而残差平方和的期望:

$$\begin{aligned}\mathbb{E}(\hat{u}'\hat{u}) &= \mathbb{E}(u'Mu) \\ &= \mathbb{E}[\text{tr}(u'Mu)] \\ &= \mathbb{E}[\text{tr}(Mu u')] \\ &= \text{tr}[\mathbb{E}(Mu u')]\end{aligned}$$

- 由于tr是一个线性运算, 从而

$$\begin{aligned}\mathbb{E}(\hat{u}'\hat{u}) &= \mathbb{E}[\text{tr}(Mu u')] \\ &= \text{tr}[\mathbb{E}(Mu u')] \\ &= \text{tr}[\mathbb{E}(\mathbb{E}(Mu u'|X))] \\ &= \text{tr}[\mathbb{E}(M\mathbb{E}(u u'|X))]\end{aligned}$$

σ^2 的估计

- 将 $\mathbb{E}(uu'|X) = \sigma^2 I$ 带入:

$$\begin{aligned}\mathbb{E}(\hat{u}'\hat{u}) &= \text{tr}[\mathbb{E}(M\mathbb{E}(uu'|X))] \\ &= \text{tr}[\mathbb{E}(\sigma^2 M)] = \sigma^2 \text{tr}(M)\end{aligned}$$

其中 $M = I - X(X'X)^{-1}X'$, 计算其迹:

$$\begin{aligned}\text{tr}(M) &= \text{tr}(I - X(X'X)^{-1}X') \\ &= \text{tr}(I) - \text{tr}(X(X'X)^{-1}X') \\ &= N - \text{tr}((X'X)^{-1}X'X) \\ &= N - K\end{aligned}$$

σ^2 的估计

- 从而： $\mathbb{E}(\hat{u}'\hat{u}) = \sigma^2(N - K)$ ，或者：

$$\mathbb{E}\left(\frac{\hat{u}'\hat{u}}{N - K}\right) = \sigma^2$$

从而对于 σ^2 的无偏估计为

$$\hat{\sigma}^2 = s^2 = \frac{\hat{u}'\hat{u}}{N - K} = \frac{\text{RSS}}{N - K}$$

我们称之为均方误差（mean squared error, MSE），从而残差平方和除以 $N - K$ 得到了 σ^2 的无偏估计，所以残差平方和的自由度为 $N - K$ 。

- 而均方根误差（root mean squared error, RMSE）即均方误差的开平方：

$$\text{RMSE} = \sqrt{\text{MSE}} = \sqrt{\frac{\text{RSS}}{N - K}}$$

标准误的估计

- 将以上的无偏估计带入到式(1)中，就可以得到条件方差的估计：

$$\widehat{\mathbb{V}}(\hat{\beta}|X) = s^2 (X'X)^{-1} = \frac{\text{RSS}}{N - K} (X'X)^{-1}$$

- 从而 $\hat{\beta}_k$ 的标准误的估计为：

$$\text{s.e.}(\hat{\beta}_k) = \sqrt{s^2 (X'X)^{-1}_{kk}}$$

其中 $(X'X)^{-1}_{kk}$ 代表矩阵 $(X'X)^{-1}$ 的第 k 行第 k 列的元素值。

σ^2 的估计

- 进一步，在正态性假设下，由于：

$$(N - K) s^2 = \hat{u}'\hat{u} = u'Mu$$

从而

$$\frac{(N - K) s^2}{\sigma^2} = \left(\frac{u}{\sigma}\right)' M \left(\frac{u}{\sigma}\right)$$

由于： $\frac{u}{\sigma} \sim \mathcal{N}(0, I)$ 为联合的标准正态分布，且 $\text{tr}(M) = N - K$ ，从而：

$$\frac{(N - K) s^2}{\sigma^2} = \left(\frac{u}{\sigma}\right)' M \left(\frac{u}{\sigma}\right) \sim \chi^2(N - K)$$

以上分布结论将是更进一步的假设检验结论的基础。

平方和的自由度

回忆

$$\text{TSS} = \text{ESS} + \text{RSS}$$

如果我们分别将其除以对应的自由度，就得到了相应的“均方”（mean squares），其中：

- TSS总平方和的自由度为 $N - 1$ ，均方 $\text{TSS}/(N-1) = s_y^2$ 为 y 的样本方差，在正态性假设下，如果假设 β 中除常数项外系数 $(\beta_2, \dots, \beta_K)$ 全都为0，那么 $y \sim \mathcal{N}(\beta_1, \sigma^2)$ ，此时有

$$\frac{(N - 1) s_y^2}{\sigma^2} = \frac{\text{TSS}}{\sigma^2} \sim \chi^2(N - K)$$

平方和的自由度

- RSS残差平方和的自由度为 $N - K$ ，均方 $\text{RSS}/(N - K) = s^2$ 为均方误差，在正态性假设下，有

$$\frac{(N - K) s^2}{\sigma^2} = \frac{\text{RSS}}{\sigma^2} \sim \chi^2(N - K)$$

- ESS的自由度通常记为 $(N - 1) - (N - K) = K - 1$ ，即总平方和与残差平方和自由度之差，理由是在正态性假设下，如果同样假设 $\beta_2 = \dots = \beta_K = 0$ ，有

$$\frac{\text{ESS}}{\sigma^2} \sim \chi^2(K - 1)$$

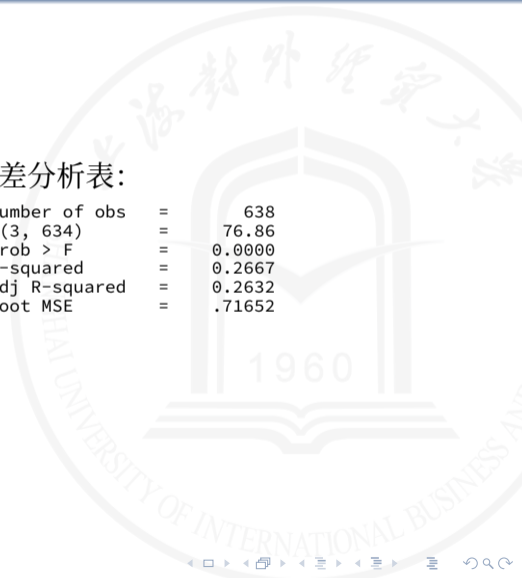
从而记ESS的自由度为 $K - 1$ ，即非常数项解释变量的个数。

方差分析表

如partitioned_regression.do中的回归结果的方差分析表:

Source	SS	df	MS	Number of obs	=	638
Model	118.386699	3	39.4622331	F(3, 634)	=	76.86
Residual	325.495238	634	.513399428	Prob > F	=	0.0000
				R-squared	=	0.2667
				Adj R-squared	=	0.2632
				Root MSE	=	.71652
Total	443.881937	637	.696831926			

以上数字之间的关联?



小样本分布

- 对以上分布进行标准化，得到：

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (X'X)^{-1}_{kk}}} \Bigg| X \sim \mathcal{N}(0, 1)$$

- 注意以上的标准正态分布是不依赖于任何未知参数的，即给定任意的 X ，上式都服从标准正态分布，从而我们可以把给定 X 去掉，得到：

$$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (X'X)^{-1}_{kk}}} \sim \mathcal{N}(0, 1)$$

一致性

- 一致性意味着当样本量趋向于无穷时，最小二乘估计量 $\hat{\beta}$ 与真实参数 β 的差距收敛到0。
- 注意到最小二乘估计量可以写为

$$\hat{\beta} = \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \frac{1}{N} \sum_{i=1}^N x_i y_i = \beta + \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \frac{1}{N} \sum_{i=1}^N x_i u_i$$

式中出现了两个平均，我们可以分别对以上两部分使用大数定律。

- 在独立同分布假设下，如果假设 $\mathbb{E}(x_{ik}^2) < \infty, k = 1, \dots, K$ ，根据大数定律，有

$$\frac{1}{N} \sum_{i=1}^N x_i x_i' \xrightarrow{P} \mathbb{E}(x_i x_i')$$

- 此外，如果假设 $\mathbb{E}|x_{ik} u_i| < \infty, k = 1, \dots, K$ ，根据大数定律，有

$$\frac{1}{N} \sum_{i=1}^N x_i u_i \xrightarrow{P} \mathbb{E}(x_i u_i) = 0$$

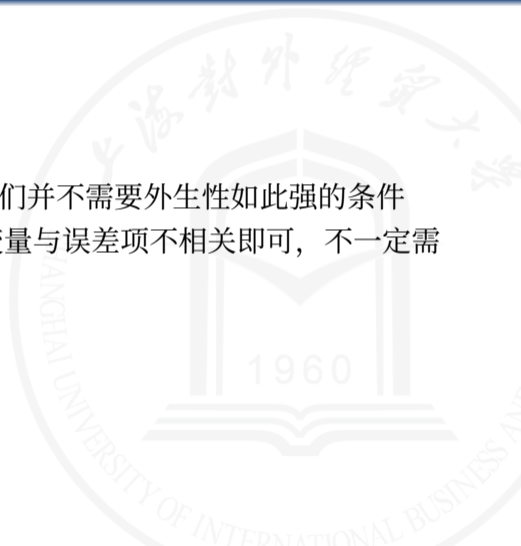
一致性

- 由于矩阵的逆也是连续映射，因而如果识别假设成立，那么

$$\begin{aligned}\hat{\beta} &= \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \frac{1}{N} \sum_{i=1}^N x_i y_i \\ &= \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \left\{ \frac{1}{N} \sum_{i=1}^N [x_i (x_i' \beta + u_i)] \right\} \\ &= \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \left[\frac{1}{N} \sum_{i=1}^N (x_i x_i') \beta + \frac{1}{N} \sum_{i=1}^N (x_i u_i) \right] \\ &\xrightarrow{p} [\mathbb{E}(x_i x_i')]^{-1} [\mathbb{E}(x_i x_i') \beta + \mathbb{E}(x_i u_i)] \\ &= \beta\end{aligned}$$

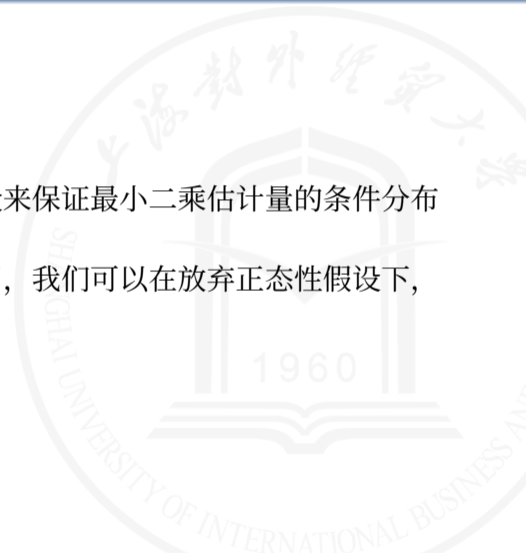
一致性

- 注意在以上一致性的证明过程中，实际上我们并不需要外生性如此强的条件
- 我们需要的仅仅是 $\mathbb{E}(x_i u_i) = 0$ 成立，即自变量与误差项不相关即可，不一定需要均值独立。
- 但是无偏性需要均值独立。



极限分布

- 在小样本统计性质中，我们通过正态性假设来保证最小二乘估计量的条件分布为联合正态分布。
- 在大样本条件下，由于中心极限定理的使用，我们可以在放弃正态性假设下，得到最小二乘估计量的渐进正态性。



极限分布

- 而第二部分可以使用中心极限定理，在独立同分布假设下，如果 $\mathbb{V}(x_i u_i)$ 存在，那么

$$\sqrt{N} \left[\frac{1}{N} \sum_{i=1}^N x_i u_i - \mathbb{E}(x_i u_i) \right] \stackrel{a}{\sim} \mathcal{N}(0, \mathbb{V}(x_i u_i))$$

- 由于 $\mathbb{E}(x_i u_i) = 0$ ，因而协方差矩阵：

$$\mathbb{V}(x_i u_i) = \mathbb{E}[(x_i u_i)(x_i u_i)'] = \mathbb{E}(u_i^2 x_i x_i')$$

因而：

$$\frac{1}{\sqrt{N}} \sum_{i=1}^N x_i u_i \stackrel{a}{\sim} \mathcal{N}(0, \mathbb{E}(u_i^2 x_i x_i'))$$

极限分布

- 将上式带入，我们可以得到：

$$\begin{aligned}\sqrt{N}(\hat{\beta} - \beta) &= \left[\frac{1}{N} \sum_{i=1}^N x_i x_i' \right]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N x_i u_i \\ &= [\mathbb{E}(x_i x_i')]^{-1} \frac{1}{\sqrt{N}} \sum_{i=1}^N x_i u_i + o_p(1) \\ &\stackrel{a}{\sim} \mathcal{N}\left(0, [\mathbb{E}(x_i x_i')]^{-1} \mathbb{E}(u_i^2 x_i x_i') [\mathbb{E}(x_i x_i')]^{-1}\right)\end{aligned}$$

极限分布

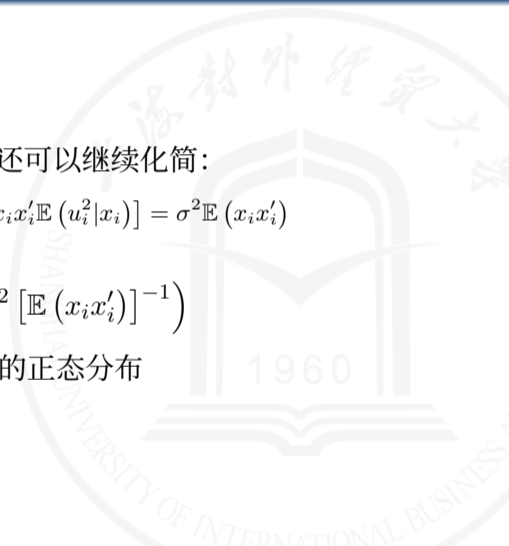
- 如果我们假设 u_i 是同方差的，以上渐进方差还可以继续化简：

$$\mathbb{E}(u_i^2 x_i x_i') = \mathbb{E}[\mathbb{E}(u_i^2 x_i x_i' | x_i)] = \mathbb{E}[x_i x_i' \mathbb{E}(u_i^2 | x_i)] = \sigma^2 \mathbb{E}(x_i x_i')$$

从而

$$\sqrt{N}(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(0, \sigma^2 [\mathbb{E}(x_i x_i')]^{-1})$$

注意以上渐进正态的结果并没有假设误差项的正态分布



标准误的估计

- 从而 σ^2 可以使用 s^2 替代, 即使用

$$\widehat{\mathbb{V}}(\hat{\beta}) = \frac{1}{N} s^2 \left(\frac{1}{N} \sum_{i=1}^N x_i x_i' \right)^{-1} = s^2 (X'X)^{-1}$$

对 $\hat{\beta}$ 的渐进方差进行估计。

- 该估计式与小样本条件下条件方差的估计式完全一样。进而标准误的估计为

$$\text{s.e.}(\hat{\beta}_k) = \sqrt{s^2 (X'X)^{-1}_{kk}}$$

异方差

- 以上方差的估计量是基于同方差假定下得到的，然而现实中同方差的假定仍然很强，异方差的存在，即 $V(u_i|x_i) = \sigma_i^2 \neq \sigma^2$ 使得以上对于渐进方差的估计变的不再可靠。
- 需要注意的是，异方差的存在并不影响参数的一致性。
- 异方差的两个后果：
 - 最小二乘估计量不再是最优线性无偏估计量，虽然无偏性仍然满足，但是有效性不再满足。
 - 影响渐进分布
 - $\sqrt{N}(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(0, \sigma^2 [E(x_i x_i')]^{-1})$ 不再成立
 - $\sqrt{N}(\hat{\beta} - \beta) \stackrel{a}{\sim} \mathcal{N}(0, [E(x_i x_i')]^{-1} E(u_i^2 x_i x_i') [E(x_i x_i')]^{-1})$ 仍然成立

异方差稳健标准误

- 解决方法：异方差稳健标准误——无需调整系数，调整标准误即可
- 估计 $[\mathbb{E}(x_i x_i')]^{-1} \mathbb{E}(u_i^2 x_i x_i') [\mathbb{E}(x_i x_i')]^{-1}$:
 - $[\mathbb{E}(x_i x_i')]^{-1}$ 可以使用： $\left[\frac{1}{N} \sum_{i=1}^N (x_i x_i')\right]^{-1}$ 替代
 - 而 $\mathbb{E}(u_i^2 x_i x_i')$ 中由于 u_i 不可观测，我们可以考虑使用残差 \hat{u}_i 替代。

