外生性下的线性回归

司继春

2024年4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$\{(y_i, x_i, u_i), i = 1, 2, ..., N\}$为独立同分布的样本。

矩估计

- 对于回归方程：

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

其中x_i为K维向量，代表解释变量；β为解释变量的系数；u_i为误差项。

- 如果外生性假设即

$$\mathbb{E}\left(u_i|x_i\right) = 0$$

那么根据条件期望的性质，必然有

$$\mathbb{E}\left(x_i u_i\right) = \mathbb{E}\left[\mathbb{E}\left(x_i u_i|x_i\right)\right] = \mathbb{E}\left[x_i \mathbb{E}\left(u_i|x_i\right)\right] = \mathbb{E}\left[x_i \cdot 0\right] = 0$$

矩估计：识别

- 注意到不可观测误差项u_i可以写为$u_i = y_i - x_i'\beta$从而

$$\mathbb{E}\left[\left(y_i - x_i'\beta\right) x_i\right] = 0$$

- 注意由于$u_i = y_i - x_i'\beta$为标量，而x_i为$K \times 1$的向量，因而以上方程实际上包含着K个矩条件。对以上方程进行整理，得到：

$$\mathbb{E}\left(x_i x_i'\right) \beta = \mathbb{E}\left(x_i y_i\right)$$

如果假设$\mathbb{E}\left(x_i x_i'\right)$可逆，那么

$$\beta = \left[\mathbb{E}\left(x_i x_i'\right)\right]^{-1} \mathbb{E}\left(x_i y_i\right)$$

矩估计：识别条件

但是如果$\mathbb{E}\left(x_i x_i'\right)$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$\mathbb{E}\left(x_i x_i'\right)$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：$\hat{\beta}=\left(X'X\right)^{-1}X'Y$。

极大似然估计方法

- 如果在外生性假设上更进一步，不仅假设$\mathbb{E}(u_i|x_i)=0$，同时假设$u_i|x_i$的分布，比如正态分布，我们还可以得到β的极大似然估计。
- 如果假设$u_i|x_i \sim N(0,\sigma^2)$，那么$y_i|x_i \sim N(x_i'\beta,\sigma^2)$，因而其条件极大似然函数为：

$$L(Y|X)=\sum_{i=1}^{N}\left[-\ln(2\pi)-\ln\sigma-\frac{(y_i-x_i'\beta)^2}{\sigma^2}\right]$$

- 最大化以上似然函数，实际上等价于最小化目标函数：

$$\sum_{i=1}^{N}(y_i-x_i'\beta)^2$$

  即最小二乘的目标函数
- 因而在正态分布的假设下，使用条件极大似然方法得到的β的估计同样也是最小二乘估计量。

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：

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

其中$u = [u_1, ..., u_N]'$为误差项向量。从而：

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

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

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

因而最小二乘估计是真实参数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

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

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

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

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

$\hat{\beta}$的方差估计

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

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

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

同方差假设

- 根据独立同分布假设，易知$\mathbb{E}\left(u_i u_j|X\right)=0$，从而$\mathbb{E}\left[uu'|X\right]$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$\mathbb{V}\left(u_i|X\right)=\mathbb{V}\left(u_i|x_i\right)$的假设。

假设（同方差，homoscedasticity）

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

同方差假设

在同方差假定下：

$$\mathbb{E}\left(uu'|X\right)=\begin{bmatrix}\mathbb{V}\left(u_1|X\right) & \mathbb{C}\mathrm{ov}\left(u_1,u_2|X\right) & \cdots & \mathbb{C}\mathrm{ov}\left(u_1,u_N|X\right)\\ \mathbb{C}\mathrm{ov}\left(u_2,u_1|X\right) & \mathbb{V}\left(u_2|X\right) & \cdots & \mathbb{C}\mathrm{ov}\left(u_2,u_N|X\right)\\ \vdots & \vdots & \ddots & \vdots\\ \mathbb{C}\mathrm{ov}\left(u_N,u_1|X\right) & \mathbb{C}\mathrm{ov}\left(u_N,u_2|X\right) & \cdots & \mathbb{V}\left(u_N|X\right)\end{bmatrix}$$

$$=\begin{bmatrix}\sigma^2 & 0 & \cdots & 0\\ 0 & \sigma^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma^2\end{bmatrix}_{N\times N}=\sigma^2 I$$

则上式可以化简为：

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

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。

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：

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

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：$\widehat{\mathbb{V}\left(\hat{\beta}|X\right)}$，我们还需要计算$\sigma^2$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：

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

对σ^2进行估计，然而该估计量不是无偏估计。

σ^2的估计

- 可以证明：

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

从而对于σ^2的无偏估计为

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

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：

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

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：

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

- 从而$\hat{\beta}_k$的标准误的估计为：

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

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，有

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

以上分布结论将是更进一步的假设检验结论的基础。

小样本分布

- 对于统计推断来说，我们还需要给出估计量的抽样分布。
- 然而小样本情况下，如果对误差项u的分布不做任何假定，通常不能得到估计量的精确分布，因而必须对u的分布做进一步的假设。
- 一个常用的假设是正态性假设:

正态性假设

假设误差项u_i为独立同分布的正态分布，$u_i|X \sim N\left(0, \sigma^2\right)$，或者$u|X \sim N\left(0, \sigma^2 I\right)$。

小样本分布

- 在正态性假设基础之上，由于$\hat{\beta}-\beta=\left(X'X\right)^{-1}X'u$，因而：

$$\hat{\beta}|X \sim N\left(\beta, \sigma^2\left(X'X\right)^{-1}\right)$$

- 进而，单独某一个系数的分布也是正态分布，即：

$$\hat{\beta}_k|X \sim N\left(\beta_k, \sigma^2\left(X'X\right)^{-1}_{kk}\right)$$

小样本分布

- 对以上分布进行标准化，得到：

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

- 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：

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