2014.5.8

第二十三讲. 广义最小二乘(GLS)

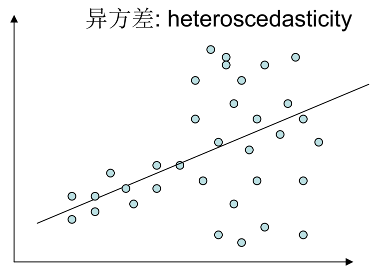
如果回归诊断发现线性回归模型假设不成立：

(1) 误差异方差(heteroscedasticity)：

解决方案：数据变换,GLS(广义最小二乘)，GLM(广义线性模型)

(2) 均值非线性 (非正态)

解决方案：数据变换，非线性回归, GLM



$Y = X\beta + \varepsilon$, $\varepsilon \sim$ 误差方差矩阵G.

(1) 如果G完全未知，仍可应用LS方法，LS估计无偏但方差过大。

(2) 如果G完全已知，则应用GLS方法;

(3) 如果G结构已知，仅含少数未 知参数,应用IRLS方法.

比如上图 $G = \mathrm{diag}(\sigma_1^2,\ldots,\sigma_1^2,\sigma_2^2,\ldots,\sigma_2^2)$

GLS (Generalized LS): 误差方差不等但已知情况

OLS：通常的(ordinary)最小二乘(等方差情形)

GLS：广义最小二乘法(Generalized Least Squares),

也称做加权LS, 应用于异方差heteroscedasticity)

或各次观察不独立情形。

例1. 若 $y_1,\ldots,y_n$独立, $y_i \sim N(\mu,\sigma_i^2)$, μ未知。假设σ_i^2已知，则极大似然估计

$$\max L \Leftrightarrow \min \sum \frac{(y_i-\mu)^2}{\sigma_i^2}, \text{解为} \hat{\mu}_{GLS} = \frac{\sum y_i/\sigma_i^2}{\sum 1/\sigma_i^2} \text{称为GLS}$$

但$\hat{\mu}_{OLS} = \bar{y} = \operatorname{argmin}\sum(y_i-\mu)^2$，称为OLS

两者都是无偏估计，但: $\mathrm{var}(\hat{\mu}_{GLS}) = \dfrac{1}{\sum 1/\sigma_i^2} \le \dfrac{\sum \sigma_i^2}{n^2} = \mathrm{var}(\hat{\mu}_{OLS})$

模型$:Y = X\beta+\varepsilon, E(\varepsilon \mid X)=0, \mathrm{var}(\varepsilon \mid X)=G\geq 0$ (G已知)，$\varepsilon \perp X$

GLS方法：

首先$Y = X\beta+\varepsilon$两边同乘$G^{-1/2}$：

$$Y^* = G^{-1/2}Y = G^{-1/2}X\beta + G^{-1/2}\varepsilon = X^*\beta+\varepsilon^* \qquad (*)$$

误差项$\varepsilon^* = G^{-1/2}\varepsilon$满足$:E(\varepsilon^* \mid X)=0, \mathrm{var}(\varepsilon^* \mid X)=I_n$

对模型(*)应用OLS方法,即

$$\min(Y^*-X^*\beta)'(Y^*-X^*\beta)=\min(Y-X\beta)'G^{-1}(Y-X\beta)$$

得$\hat\beta_{GLS}=(X^{*}{}'X^*)^{-1}X^{*}{}'Y^*=[(G^{-1/2}X)'(G^{-1/2}X)]^{-1}(G^{-1/2}X)'G^{-1/2}Y$

$$=(X'G^{-1}X)^{-1}X'G^{-1}Y$$

容易验证：

(1) $E(\hat\beta_{GLS} \mid X)=\beta$,

(2) $\mathrm{var}(\hat\beta_{GLS} \mid X)=(X'G^{-1}X)^{-1}$

另外一种做法是不管G是否具有形式$G=\sigma^2 I_n$,

直接应用通常的最小二乘法(OLS):

$$\mathrm{argmin}\,\|Y-X\beta\|^2 \Rightarrow \hat\beta_{OLS}=(X'X)^{-1}X'Y$$

(1) $E(\hat\beta_{OLS} \mid X)=\beta$, 无偏。

(2) $\mathrm{var}(\hat\beta_{OLS} \mid X)=(X'X)^{-1}X'\mathrm{var}(Y \mid X)X(X'X)^{-1}$

$=(X'X)^{-1}X'GX(X'X)^{-1}$

模型(*)满足Gauss-Markov假设，由GM定理知其OLS

估计$\hat\beta_{GLS}$是BLUE，即

$$\mathrm{var}(\hat\beta_{OLS} \mid X)\geq \mathrm{var}(\hat\beta_{GLS} \mid X) \quad \Leftrightarrow$$

$$(X'X)^{-1}X'GX(X'X)^{-1}\geq (X'G^{-1}X)^{-1}$$

直接证明该矩阵不等式：

只需证$(X'X)(X'G^{-1}X)^{-1}(X'X)\leq X'GX$

左端$=X'G^{1/2}\underline{G^{-1/2}X(X'G^{-1}X)^{-1}X'G^{-1/2}}G^{1/2}X\leq X'G^{1/2}G^{1/2}X=X'GX$

其中下划线部分(投影阵) $G^{-1/2}X(X'G^{-1}X)^{-1}X'G^{-1/2}\leq I_n$。

综上，线性回归模型的等方差假设不是本质的，即使不满足等方差假设，OLS估计也是无偏的，但OLS估计效率偏低。

G未知情形：迭代加权最小二乘方法 (Iteratively reweighted least squares, IRLS）

G通常是未知的。

如果G完全未知，因其含有$n(n+1)/2$个参数，不可估计。

所以需要根据具体问题假设G具有某种简单结构,即参数化。

一种最简单的方差结构 假设是$G=\sigma^2 I_n$（误差独立同分布），

此时$\hat\beta_{GLS}=(X'G^{-1}X)^{-1}X'G^{-1}Y=(X'(\sigma^2 I_n)^{-1}X)^{-1}X'(\sigma^2 I_n)^{-1}Y$

$=(X'X)^{-1}X'Y=\hat\beta_{OLS}$不依赖于未知参数$\sigma^2$。

此时GLS即OLS，即不需要估计σ^2也能估计β。

对于其它形式的G，$\hat{\beta}_{GLS}$涉及G，所以必须同时估计β和G。

同时估计β和G，计算上不再像OLS，GLS那么简单。
一种策略是分步估计：即先估计G，再计算GLS估计。

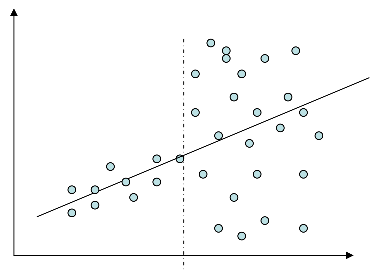
但估计G必定用到残差，而残差与β的估计有关..

IRLS方法是一个迭代算法：
给定初始**G**，
由**GLS**方法得到回归系数估计，并得到残差，
基于残差更新**G**的估计，
循环。

例1. 对于下图所示的独立样本情形，异方差（heteroscedasticity）模型

$$Y = X\beta + \varepsilon, \varepsilon \sim (0, G), G = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_n^2)$$

假设$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_m^2 = \sigma^2$(未知)$, \sigma_{m+1}^2 = \ldots = \sigma_n^2 = \tau^2$(未知)



*GLS*最小化：

$$\min_{\beta,\sigma^2,\tau^2} (Y - X\beta)' G^{-1} (Y - X\beta) = \min_{\beta,\sigma^2,\tau^2} \left(\sum_{i=1}^{m} \frac{(y_i - \tilde{x}_i \beta)^2}{\sigma^2} + \sum_{i=m+1}^{n} \frac{(y_i - \tilde{x}_i \beta)^2}{\tau^2} \right)$$

IRLS解法：

Step0. k = 0, 初始化$G^{(0)} = \sigma^2 I_n$.

Step1. k = k + 1,

计算GLS估计$\hat{\beta}^{(k)} = \hat{\beta}_{\mathrm{GLS}} = (X' G^{(k-1)^{-1}} X)^{-1} X' G^{(k-1)^{-1}} Y$,

计算残差$e = Y - X\hat{\beta}^{(k)}$

Step2. $\hat{\sigma}^2 = \sum_{i=1}^{m} e_i^2 / (m - p), \quad \hat{\tau}^2 = \sum_{i=m+1}^{n} e_i^2 / (n - m - p)$

$\Rightarrow G^{(k)} = diag(\hat{\sigma}^2, \ldots, \hat{\sigma}^2, \hat{\tau}^2 \ldots, \hat{\tau}^2)$

goto Step1, 重复至收敛。

例2. 假设模型$Y = X\beta + \varepsilon, \varepsilon \sim (0, G)$, 其中$G = \left(\sigma^2 \rho^{|i-j|}\right)_{1 \le i, j \le n}$

注意：与以往不同，这里我们各个观察之间是相依的

IRLS：

Step0. 由OLS方法得到 β 的无偏估计 $\hat{\beta}^{(0)}$, 计算 $e = Y - X\hat{\beta}^{(0)}$

Step1. $\hat{\sigma}^2 = \sum_{i=1}^{n} e_i^2 / (n - p), \quad \hat{\rho} = \sum_{i=1}^{n-1} e_i e_{i+1} / \sum_{i=1}^{n} e_i^2$

$\Rightarrow$ G的估计 $\hat{G} = \left(\hat{\sigma}^2 \hat{\rho}^{|i-j|}\right)_{1 \le i, j \le n}$

更新 β 的估计

$\hat{\beta}^{(1)} = (X' \hat{G}^{-1} X)^{-1} X' \hat{G}^{-1} Y$

Step2. 计算新的残差 $e = Y - X\hat{\beta}^{(1)}$

goto Step1，重复直至收敛。

矩估计，当然也可使用任何其它合理的估计

一般地, 如果G未知(但具有简单结构)，IRLS:

(0）首先由OLS得到β的一个无偏估计

$\hat{\beta}^{(0)} = \hat{\beta}_{OLS}$，计算残差$e = Y - X\hat{\beta}^{(0)}$

(1）考虑到G的特殊结构，由残差得到G的估计$\hat{G}^{(0)}$

(2) 由GLS方法更新β的估计:

$\hat{\beta}^{(1)} = (X'\hat{G}^{(0)^{-1}}X)^{-1}X'\hat{G}^{(0)^{-1}}Y$

(3）由此得到更新的残差$e = Y - X\hat{\beta}^{(1)}$

goto（1），重复至收敛

IRLS常用于如下优化问题：

$$\min \sum w_i(\beta,\theta)(y_i - x_i'\beta)^2$$

其中一般θ代表与方差有关的参数。比如稳健回归、广义线性模型(GLM)似然函数的最大化可以转化为上述优化问题。

IRLS:

(1) 初始化$\beta^{(0)}$：比如$\beta^{(0)} = $ OLS估计(视w为常数)

利用残差得到θ的估计$\theta^{(0)}$【根据具体问题设计, 如例1,2红色下划线部分】

(2) 计算$w_i(\beta^{(0)}, \theta^{(0)})$

记 $G^{(0)} = \mathrm{diag}\left(w_1(\beta^{(0)}, \theta^{(0)}),\ldots, w_n(\beta^{(0)}, \theta^{(0)})\right)$

(3)更新$\beta^{(1)} = \arg\min \sum w_i(\beta^{(0)}, \theta^{(0)})(y_i - x_i\beta)^2 = (X'G^{(0)}X)^{-1}X'G^{(0)}Y$,

(4)$\beta^{(0)} \leftarrow \beta^{(1)}$

goto (2)

例3 (线性模型的M-估计). 模型: $Y = X\beta + \varepsilon, \varepsilon \sim (0, \sigma^2 I_n)$,

$$\min \sum \rho(y_i - x_i\beta)$$

$\rho(e) \ge 0$关于0对称，$\rho(0) = 0$.

$\rho(e) = |e|^p$ 时:

$$\min \|Y - X\beta\|_p = \min \sum |y_i - x_i\beta|^p$$

称为最小L_p方法，$p = 1$时称为最小一乘法(LAD: least absolute deviation).

对于最小L_p方法, 改写目标函数:

$$\sum |y_i - x_i\beta|^p = \sum w_i(\beta)|y_i - x_i\beta|^2, \text{ 其中} w_i(\beta) = |y_i - x_i\beta|^{p-2}$$

Probit模型及广义线性模型

数据$(y_i, x_i), i = 1,.., n$. 响应变量取值 0,1，如何建立回归模型？

Probit 模型

假设均值函数$p_i = P(y_i = 1 | x_i) = E(y_i | x_i)$具有如下结构

$$p_i = \Phi(x_i'\beta)$$

Φ为$N(0,1)$的累积分布函数。

解释: 假设有一个隐变量(latent) z_i，连续, 但只能观察到

$$y_i = \begin{cases} 1, & z_i \ge 0 \\ 0, & z_i < 0 \end{cases}$$

假设$z_i = x_i'\beta + \varepsilon_i, \varepsilon_i$ iid $\sim N(0,1)$, 则

$$p_i = P(y_i = 1 | x_i) = P(x_i'\beta + \varepsilon_i \ge 0) = \Phi(x_i'\beta)$$

似然函数：

$$L(\alpha,\beta)=\prod p_i^{y_i}(1-p_i)^{1-y_i}=\prod \Phi(\mathrm{x}_i'\beta)^{y_i}(1-\Phi(\mathrm{x}_i'\beta))^{1-y_i}$$

$$\log L=\sum[y_i\log\Phi(\mathrm{x}_i'\beta)+(1-y_i)\log(1-\Phi(\mathrm{x}_i'\beta))]$$

极大化logL需要数值解法，比如 (1) Newton - Raphson 算法; (2) Fisher scoring 算法;(3) IRLS; (4) EM 算法

其中Newton算法最快，但可能会因 出现二阶导数矩阵不可 逆而不收敛

Fisher scoring 算法将二阶导数矩阵换 成Fisher信息阵（正定）

EM算法通常用于计算 latent variable或缺失数据 模型的极大似然估计，

因为Probit模型是一个 latent variable 模型，所以我们介绍 EM算法

EM算法基于下述两步：

（M）如果 z's能观察到，那么由 $z_i=\mathrm{x}_i'\beta+\varepsilon_i$，可得LS估计

$$\hat{\beta}=(\sum \mathrm{x}_i\mathrm{x}_i')^{-1}\sum \mathrm{x}_i z_i$$

（E）如果β 已知，那么我们可以预 测z_i：$\hat{z}_i=\mathrm{E}(z_i\mid \mathrm{x}_i,\mathrm{y}_i,\beta)$，细节如下：

由$z_i=\mathrm{x}_i'\beta+\varepsilon_i,\varepsilon_i$ iid $\sim N(0,1)$，得到z_i的条件分布：

$$\mathrm{P}(z_i\mid \mathrm{x}_i,\mathrm{y}_i,\beta)=\mathrm{P}(\mathrm{y}_i\mid z_i,\mathrm{x}_i,\beta)\mathrm{P}(z_i\mid \mathrm{x}_i,\beta)/\mathrm{P}(\mathrm{y}_i\mid \mathrm{x}_i,\beta)$$

$$=y_i\phi(z_i-\mathrm{x}_i'\beta)/\Phi(\mathrm{x}_i'\beta)^{y_i}(1-\Phi(\mathrm{x}_i'\beta))^{1-y_i}$$

$$=\begin{cases}\phi(z_i-\mathrm{x}_i'\beta)/\Phi(\mathrm{x}_i'\beta), & y_i=1\\ 0, & y_i=1\end{cases}$$

$$\Rightarrow \hat{z}_i=\mathrm{E}(z_i\mid \mathrm{x}_i,\mathrm{y}_i,\beta)=\begin{cases}\mathrm{x}_i'\beta+\phi(\mathrm{x}_i'\beta)/\Phi(\mathrm{x}_i'\beta), & y_i=1\\ \mathrm{x}_i'\beta-\phi(\mathrm{x}_i'\beta)/[1-\Phi(\mathrm{x}_i'\beta)], & y_i=0\end{cases}$$

*EM*算法：选取初始值 β_0，预测 z，使用 z求LS估计作为 β的更新，重复 (E) - (M) - (E) - (M)...., 至β收敛。

Logisitc回归模型

假设均值函数$p_i=P(y_i=1\mid \mathrm{x}_i)=E(y_i\mid \mathrm{x}_i)$具有如下结构

$$p_i=\frac{e^{\alpha+\mathrm{x}_i'\beta}}{1+e^{\alpha+\mathrm{x}_i'\beta}}$$

注：

(1) Probit模型容易解释，在社会 学、经济学、心理学中 较为常见，

但Logistic回归模型应用更为广泛 ;两个模型分析结果通常 差异不大。

(2) logistic回归的似然函数 :

$$L(\alpha,\beta)=\prod p_i^{y_i}(1-p_i)^{1-y_i},\text{其中}p_i=\frac{e^{\alpha+\mathrm{x}_i'\beta}}{1+e^{\alpha+\mathrm{x}_i'\beta}}$$

极大化似然函数求出参 数估计。