第 7 章　异方差与 GLS

7.1　异方差的后果

“异方差”(heteroskedasticity)是违背球型扰动项假设的一种情形，即$\text{Var}(\varepsilon_i \mid \boldsymbol{X})$依赖于$i$，不是常数。

在异方差的情况下：

(1) OLS 估计量依然无偏、一致且渐近正态，因为在证明这些性质时并未用到“同方差”的假定。

(2) OLS 估计量方差$\mathrm{Var}(\boldsymbol{b}\,|\,\boldsymbol{X})$的表达式不再是$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$，因为$\mathrm{Var}(\boldsymbol{\varepsilon}\,|\,\boldsymbol{X}) \neq \sigma^2\boldsymbol{I}$。因此，通常的$t$检验、$F$检验也失效了。

(3) 高斯-马尔可夫定理不再成立，OLS 不再是 BLUE。在异方差的情况下，GLS 才是 BLUE。

为何 OLS 不再是 BLUE？

假设$\mathrm{Var}(\varepsilon_i\,|\,\boldsymbol{X})$是某解释变量$x_i$的增函数，参见图 7.1。

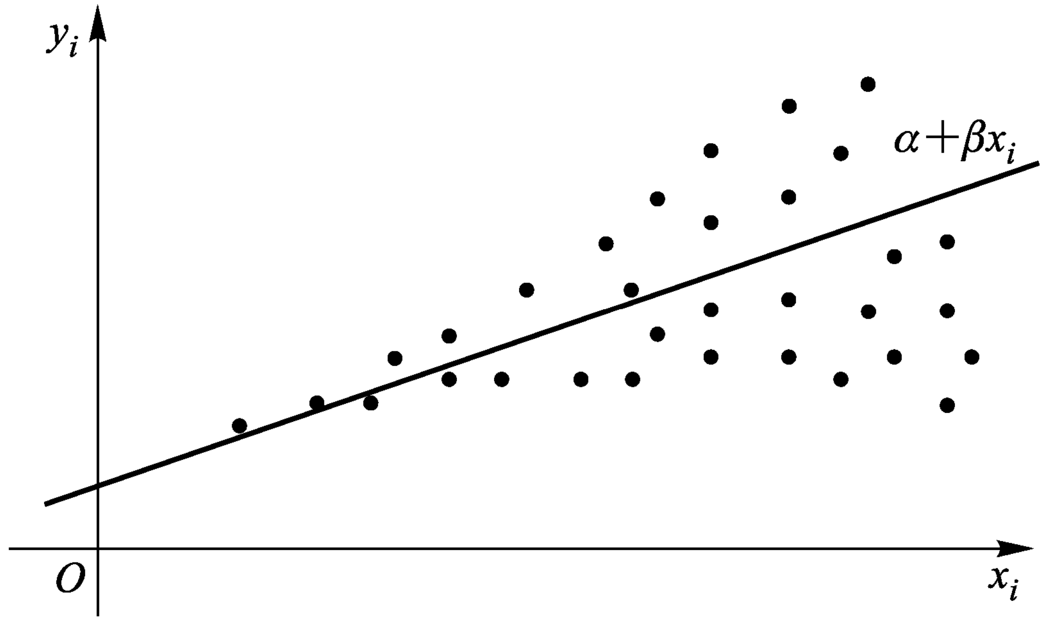


图 7.1　异方差的一种情形

GLS 及其特例“加权最小二乘法”(Weighted Least Square，简记 WLS)，通过对不同数据包含信息量的不同进行加权以提高效率。

7.2　异方差的例子

(1) 消费函数：

$$C_i = \alpha + \beta Y_i + \varepsilon_i$$

其中，C 为消费，Y 为收入。富人的消费计划较有弹性，而穷人的消费多为必需品。富人的消费支出难测量，包含较多测量误差。

(2) 企业的投资、销售收入与利润：大型企业的商业活动以亿元计，而小型企业以万元计，扰动项规模不同。

(3) 组间异方差：样本包含两组(类)数据，第一组为自我雇佣者(企业主、个体户)的收入，第二组为打工族的收入，前者的收入波动比后者大。

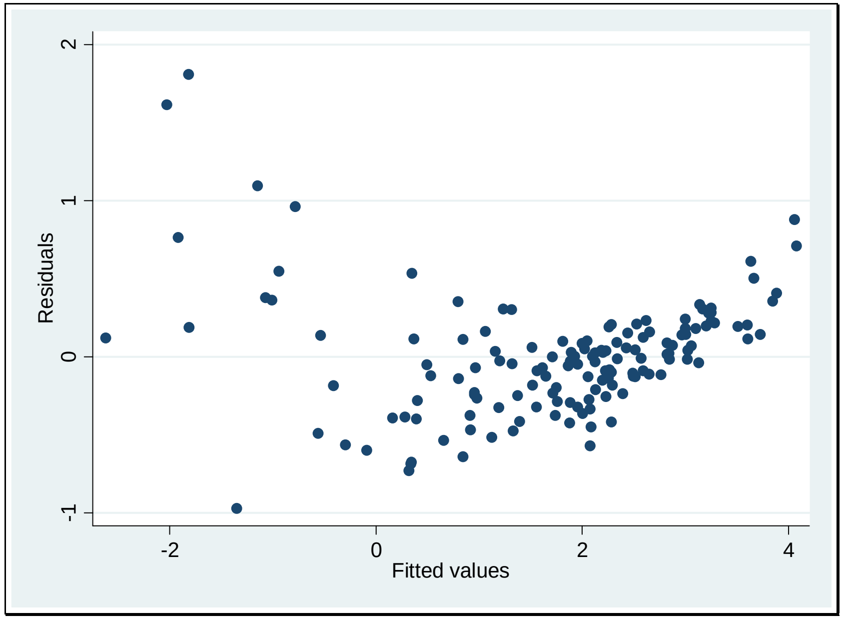
(4) 组平均数：如果数据为组平均数，则大组平均数的方差要比小组平均数的方差小。比如，全国各省的人均 GDP，人口多的省份其方差较小，方差与人口数成反比。

(5) 时间序列数据中也可能出现条件异方差，比如第 22 章的 ARCH 模型。

7.3 异方差的检验

1．看残差图(residual plot)

看“残差e_i与拟合值$\hat{y}_i$的散点图”(residual-versus- fitted plot)，或“残差e_i与某个解释变量x_{ik}的散点图”(residual-versus-predictor plot)。



2．怀特检验(White test)

在条件同方差下，稳健标准误还原为普通标准误，二者的差别

可用来度量条件异方差。怀特检验正是基于这一思想。

在同方差的原假设$H_0: \mathrm{E}(\varepsilon_i^2 \mid \boldsymbol{X}) = \sigma^2$下，稳健协方差矩阵与普通协方差矩阵之差收敛到一个零矩阵：

$$\hat{\boldsymbol{S}} - s^2 \boldsymbol{S}_{XX} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 \boldsymbol{x}_i \boldsymbol{x}_i' - s^2 \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' = \frac{1}{n}\sum_{i=1}^{n} (e_i^2 - s^2) \boldsymbol{x}_i \boldsymbol{x}_i' \xrightarrow{p} \boldsymbol{0}_{K\times K}$$

实际操作上，进行辅助回归：

$$e_i^2 \xrightarrow{\text{OLS}} \text{常数} + \boldsymbol{\psi}_i' \boldsymbol{\gamma}$$

并检验$\boldsymbol{\psi}_i$中所有变量的系数$\boldsymbol{\gamma}$均为 0。

如果R^2很低，意味着e_i^2无法由解释变量及其平方项与交叉项来解释，故倾向于接受同方差的原假设。可以证明：

$$nR^2 \xrightarrow{d} \chi^2(m)$$

如果nR^2很大(超过临界值)，则拒绝原假设H_0。

在大样本中，nR^2与检验整个方程显著性的F统计量渐近等价。

对于回归方程 $y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$，检验原假设$H_0: \beta_2 = \cdots = \beta_K = 0$，则$F$统计量：

$$F=\frac{R^2/(K-1)}{(1-R^2)/(n-K)}\sim F(K-1,n-K)$$

在大样本下，F分布与χ^2分布等价(第 5 章附录，见下)，即

$$(K-1)F=\frac{(n-K)R^2}{(1-R^2)}\xrightarrow{d}\chi^2(K-1)$$

在原假设成立的情况下，当$n\to\infty$时，$(n-K)\to n$，$(1-R^2)\to 1$，故$(K-1)F\to nR^2$，故F检验与nR^2检验在大样本下等价。

怀特检验的优点是，可以检验任何形式的异方差；缺点是，如果H_0被拒绝，并不提供有关异方差具体形式的信息。

A5.3　F 分布与 χ^2 分布在大样本下是等价的

命题　假设 $F \sim F(m, n-K)$ 分布，则当 $n \to \infty$ 时，$mF \xrightarrow{d} \chi^2(m)$。

证明： 因为 $F \sim F(m, n-K)$，故可设 $F = \dfrac{\chi^2(m)/m}{\chi^2(n-K)/(n-K)}$。

根据 χ^2 分布的性质，$\mathrm{E}\left[\chi^2(n-K)\right] = n-K$，而 $\mathrm{Var}\left[\chi^2(n-K)\right] = 2(n-K)$。

故 F 统计量分母的期望值为：

$$\mathrm{E}\left[\chi^2(n-K)/(n-K)\right] = 1$$

F 统计量分母的方差为：

$$\operatorname{Var}\left[\chi^2(n-K)\big/(n-K)\right]=\frac{2(n-K)}{(n-K)^2}=\frac{2}{n-K}\to 0 \quad (\text{当}\, n\to\infty\,\text{时})$$

故分母依均方收敛于 1。

因此，分母依概率收敛于 1，即 $\chi^2(n-K)\big/(n-K)\xrightarrow{p}1$。

所以，$F\xrightarrow{d}\chi^2(m)\big/m$。

故 $mF\xrightarrow{d}\chi^2(m)$。

3．BP 检验(Breusch and Pagan, 1979)

假设回归模型为$y_i = \beta_1 + \beta_2 x_{i2} + \cdots + \beta_K x_{iK} + \varepsilon_i$，检验以下原假设：

$$H_0: \ \mathrm{E}(\varepsilon_i^2 \mid x_2, \cdots, x_K) = \sigma^2$$

如果H_0不成立，则条件方差$\mathrm{E}(\varepsilon_i^2 \mid x_2, \cdots, x_K)$是$(x_2, \cdots, x_K)$的函数，称为“条件方差函数”(conditional variance function)。

BP 检验假设此条件方差函数为线性函数：

$$\varepsilon_i^2 = \delta_1 + \delta_2 x_{i2} + \cdots + \delta_K x_{iK} + u_i$$

原假设简化为

$$H_0: \delta_2 = \cdots = \delta_K = 0$$

由于ε_i不可观测，故使用e_i^2进行辅助回归：

$$e_i^2 = \delta_1 + \delta_2 x_{i2} + \cdots + \delta_K x_{iK} + error_i$$

使用nR^2统计量：

$$nR^2 \xrightarrow{d} \chi^2(K-1)$$

BP 检验与怀特检验的区别在于，后者还包括平方项与交叉项。

BP 检验的优点在于其建设性，可帮助确认异方差的具体形式。

7.4　异方差的处理

1. 使用“OLS + 稳健标准误”

一种处理方法是，仍进行 OLS 回归，但使用稳健标准误。

这是最简单，也是目前通用的方法。只要样本容量较大，即使在异方差的情况下，若使用稳健标准误，则所有参数估计、假设检验均可照常进行。

但还可能存在比 OLS 更有效的方法，比如 GLS。

2．广义最小二乘法(GLS)

假设$\text{Var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{V}(\boldsymbol{X}) \neq \sigma^2 \boldsymbol{I}_n$，其中$\boldsymbol{V}(\boldsymbol{X})$为对称正定矩阵且已知，可能依赖于$\boldsymbol{X}$。

GLS 的基本思想是，通过变量转换，使得转换后的模型满足球型扰动项的假定。

命题 对于对称正定矩阵$\boldsymbol{V}_{n\times n}$，存在非退化矩阵$\boldsymbol{C}_{n\times n}$，使得$\boldsymbol{V}^{-1} = \boldsymbol{C}'\boldsymbol{C}$。

在一维情况下，“$\boldsymbol{V}$ 正定”即要求$\boldsymbol{V}$为正数，故$\dfrac{1}{\boldsymbol{V}}$也是正数，可分解为$\dfrac{1}{\sqrt{\boldsymbol{V}}}\cdot\dfrac{1}{\sqrt{\boldsymbol{V}}}$；如果$\boldsymbol{V}$为负数，则无法进行此分解。

矩阵$\boldsymbol{C}$不唯一，但不影响 GLS 的最终结果。

将原回归模型$\boldsymbol{y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}$两边同时左乘矩阵$\boldsymbol{C}$：

$$\boldsymbol{Cy}=\boldsymbol{CX\beta}+\boldsymbol{C\varepsilon}$$

定义变量转换：

$$\tilde{\boldsymbol{y}}\equiv\boldsymbol{Cy},\ \tilde{\boldsymbol{X}}\equiv\boldsymbol{CX},\ \tilde{\boldsymbol{\varepsilon}}\equiv\boldsymbol{C\varepsilon}$$

可将模型写为

$$\tilde{\boldsymbol{y}} = \tilde{\boldsymbol{X}}\boldsymbol{\beta} + \tilde{\boldsymbol{\varepsilon}}$$

变换后的模型仍满足严格外生性：

$$\mathrm{E}(\tilde{\boldsymbol{\varepsilon}} \mid \tilde{\boldsymbol{X}}) = \mathrm{E}(\boldsymbol{C}\boldsymbol{\varepsilon} \mid \boldsymbol{C}\boldsymbol{X}) = \mathrm{E}(\boldsymbol{C}\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{C}\,\mathrm{E}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0}$$

球型扰动项的假定也得到满足：

$$\begin{aligned}\mathrm{Var}(\tilde{\boldsymbol{\varepsilon}} \mid \tilde{\boldsymbol{X}}) &= \mathrm{E}(\tilde{\boldsymbol{\varepsilon}}\,\tilde{\boldsymbol{\varepsilon}}' \mid \boldsymbol{X}) = \mathrm{E}(\boldsymbol{C}\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}'\boldsymbol{C}' \mid \boldsymbol{X}) = \boldsymbol{C}\,\mathrm{E}(\boldsymbol{\varepsilon}\,\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{C}' = \sigma^2\boldsymbol{C}\boldsymbol{V}\boldsymbol{C}' \\ &= \sigma^2\boldsymbol{C}(\boldsymbol{V}^{-1})^{-1}\boldsymbol{C}' = \sigma^2\boldsymbol{C}(\boldsymbol{C}'\boldsymbol{C})^{-1}\boldsymbol{C}' = \sigma^2\boldsymbol{C}\boldsymbol{C}^{-1}(\boldsymbol{C}')^{-1}\boldsymbol{C}' = \sigma^2\boldsymbol{I}_n\end{aligned}$$

故高斯-马尔可夫定理成立。

对变换后的模型使用 OLS 即得到 GLS 估计量：

$$\hat{\boldsymbol{\beta}}_{\text{GLS}} = (\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}})^{-1}\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{y}} = \left[(\boldsymbol{CX})'(\boldsymbol{CX})\right]^{-1}(\boldsymbol{CX})'\boldsymbol{Cy}$$

$$= (\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{CX})^{-1}\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{Cy} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

虽然$\boldsymbol{C}$不唯一，但$\hat{\boldsymbol{\beta}}_{\text{GLS}}$唯一，因为$\hat{\boldsymbol{\beta}}_{\text{GLS}}$不依赖于$\boldsymbol{C}$。

由于高斯-马尔可夫定理成立，故$\hat{\boldsymbol{\beta}}_{\text{GLS}}$是 BLUE，比 OLS 更有效。但前提是必须知道协方差矩阵$\boldsymbol{V}$。

3．加权最小二乘法(WLS)

假设仅存在异方差，无自相关，$\boldsymbol{V}(\boldsymbol{X})$为对角矩阵。

方差小的数据提供的信息量大。WLS 根据信息量大小进行加权。

假定$\mathrm{E}(\varepsilon_i^2 \mid \boldsymbol{x}_i) = \mathrm{Var}(\varepsilon_i \mid \boldsymbol{x}_i) = \sigma^2 v_i(\boldsymbol{X})$，即

$$\boldsymbol{V} = \begin{pmatrix} v_1 & & & 0 \\ & v_2 & & \\ & & \ddots & \\ 0 & & & v_n \end{pmatrix}, \quad \boldsymbol{V}^{-1} = \begin{pmatrix} 1/v_1 & & & 0 \\ & 1/v_2 & & \\ & & \ddots & \\ 0 & & & 1/v_n \end{pmatrix}$$

由于$\boldsymbol{V}^{-1}=\boldsymbol{C}'\boldsymbol{C}$，可知

$$\boldsymbol{C}=\boldsymbol{C}'=\begin{pmatrix} 1/\sqrt{v_1} & & & 0 \\ & 1/\sqrt{v_2} & & \\ & & \ddots & \\ 0 & & & 1/\sqrt{v_n} \end{pmatrix}$$

$$\tilde{\boldsymbol{y}}\equiv\boldsymbol{C}\boldsymbol{y}=\begin{pmatrix} 1/\sqrt{v_1} & & & 0 \\ & 1/\sqrt{v_2} & & \\ & & \ddots & \\ 0 & & & 1/\sqrt{v_n} \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}=\begin{pmatrix} y_1/\sqrt{v_1} \\ y_2/\sqrt{v_2} \\ \vdots \\ y_n/\sqrt{v_n} \end{pmatrix}$$

$$\tilde{\boldsymbol{X}} \equiv \boldsymbol{C}\boldsymbol{X} = \begin{pmatrix} 1/\sqrt{v_1} & & & 0 \\ & 1/\sqrt{v_2} & & \\ & & \ddots & \\ 0 & & & 1/\sqrt{v_n} \end{pmatrix} \begin{pmatrix} x_{11} & \cdots & x_{1K} \\ x_{21} & \cdots & x_{2K} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{nK} \end{pmatrix}$$

$$= \begin{pmatrix} x_{11}/\sqrt{v_1} & \cdots & x_{1K}/\sqrt{v_1} \\ x_{21}/\sqrt{v_2} & \cdots & x_{2K}/\sqrt{v_2} \\ \vdots & & \vdots \\ x_{n1}/\sqrt{v_n} & \cdots & x_{nK}/\sqrt{v_n} \end{pmatrix}$$

故权重为$1/\sqrt{v_i}$(标准差的倒数)。对于第 i 个观测值，回归方程为

$$\frac{y_i}{\sqrt{v_i}} = \beta_1 \frac{x_{i1}}{\sqrt{v_i}} + \beta_2 \frac{x_{i2}}{\sqrt{v_i}} + \cdots + \beta_K \frac{x_{iK}}{\sqrt{v_i}} + \frac{\varepsilon_i}{\sqrt{v_i}}$$

新扰动项为$\varepsilon_i/\sqrt{v_i}$，可将 WLS 视为最小化“加权的残差平方和”:

$$\min_{\tilde{\boldsymbol{\beta}}} \ \text{SSR} = \sum_{i=1}^{n} \left(e_i \Big/ \sqrt{v_i} \right)^2 = \sum_{i=1}^{n} \frac{e_i^2}{v_i}$$

从这个角度来看，权重为$1/v_i$，Stata 也是这样约定的。

4．可行广义最小二乘法(Feasible GLS，简记 FGLS)

必须先用样本数据估计$\boldsymbol{V}(\boldsymbol{X})$，然后才能使用 GLS，称为 FGLS 或“可行加权最小二乘法”(Feasible WLS，简记 FWLS)，即

$$\hat{\boldsymbol{\beta}}_{\text{FGLS}} = (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}$$

其中，$\hat{\boldsymbol{V}}$是$\boldsymbol{V}$的一致估计。

$\boldsymbol{V}(\boldsymbol{X})$包含很多参数。实践中，常考虑只有异方差，或只有一阶自相关的情形(参见第 8 章)。

以 FWLS 为例。在作 BP 检验时，通过辅助回归

$e_i^2 = \delta_1 + \delta_2 x_{i2} + \cdots + \delta_K x_{iK} + error_i$就可获得$\sigma_i^2$的估计值$\hat{\sigma}_i^2$。

为保证$\hat{\sigma}_i^2$为正，假设辅助回归为指数函数的形式：

$$e_i^2 = \sigma^2 \exp(\delta_1 + \delta_2 x_{i2} + \cdots + \delta_K x_{iK}) v_i$$

其中，v_i为乘积形式的扰动项。取对数后可得

$$\ln e_i^2 = (\ln \sigma^2 + \delta_1) + \delta_2 x_{i2} + \cdots + \delta_K x_{iK} + \ln v_i$$

得到对$\ln e_i^2$的预测值，记为$\ln \hat{\sigma}_i^2$，进而得到拟合值$\hat{\sigma}_i^2 = e^{\ln \hat{\sigma}_i^2}$，然后以$1/\hat{\sigma}_i^2$为权重进行 WLS 估计。

5．究竟使用“OLS + 稳健标准误”还是 FWLS

理论上，GLS 是 BLUE，但 FGLS 既非线性估计，也不是无偏估计，无资格参加 BLUE 的评选。

根据方程(7.22)，$\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}} = (\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\hat{\boldsymbol{V}}^{-1}\boldsymbol{y}$，而$\hat{\boldsymbol{V}}$是数据$(\boldsymbol{y}, \boldsymbol{X})$的非线性函数，故$\hat{\boldsymbol{\beta}}_{\mathrm{FGLS}}$是$\boldsymbol{y}$的非线性函数，一般来说是有偏的。

FWLS 的优点主要体现在大样本理论中。如果$\hat{\boldsymbol{V}}$是$\boldsymbol{V}$的一致估计，则 FWLS 一致，且在大样本下比 OLS 更有效。

FWLS 的缺点是必须估计条件方差函数$\mathrm{Var}(\varepsilon_i \mid \boldsymbol{x}_i)$，而通常不知道条件方差函数的具体形式。如果该函数的形式设定不正确，则

根据 FWLS 计算的标准误可能失效，导致不正确的统计推断。

使用“OLS + 稳健标准误”的好处是，对回归系数及标准误的估计都一致，不需要知道条件方差函数的形式。Stata 操作十分简单，只要在命令 `reg` 之后加选择项“`robust`”即可。

总之，“OLS + 稳健标准误”更为稳健，而 FWLS 更有效。必须在稳健性与有效性之间做选择。

由于“病情”通常难以诊断，故特效药也可能失效或起反作用。如果对 $\boldsymbol{V}$ 的估计不准，则 FGLS 的性能可能还不如 OLS。Stock and Watson (2011)推荐，大多数情况下应使用“OLS + 稳健标准误”。