第六章 线性回归分析诸问题

之

异方差问题 篇

异方差

- 异方差(heteroscedasticy)就是对同方差假设(assumption of homoscedasticity)的违反。经典回归中同方差是指随着样本观察点X的变化，线性模型中随机误差项的方差并不改变，保持为常数，即

$$E(u_i^2) = \sigma^2$$

- 如果对不同的样本观察值各不相同，则称随机误差项具有异方差

$$E(u_i^2) = \sigma_i^2$$

为什么会产生异方差

- 因为随机误差项包括了测量误差和模型中被省略的一些因素对因变量的影响
- 来自不同抽样单元的因变量观察值之间可能差别很大
- 因此，异方差性多出现在横截面样本之中

考虑以下模型

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + u_i, \quad i = 1, 2, \ldots, n$$

$$E(u_i) = 0, \quad E(u_i^2) = \sigma_i^2, \quad E(u_i u_j) = 0 \quad (i \neq j)$$

上述模型称为具有异方差的模型。在这种情况下，高斯马尔科夫定理不成立。OLS估计量虽然具有无偏性，但不具有有效性，不是BLUE。

异方差检验（图示法）

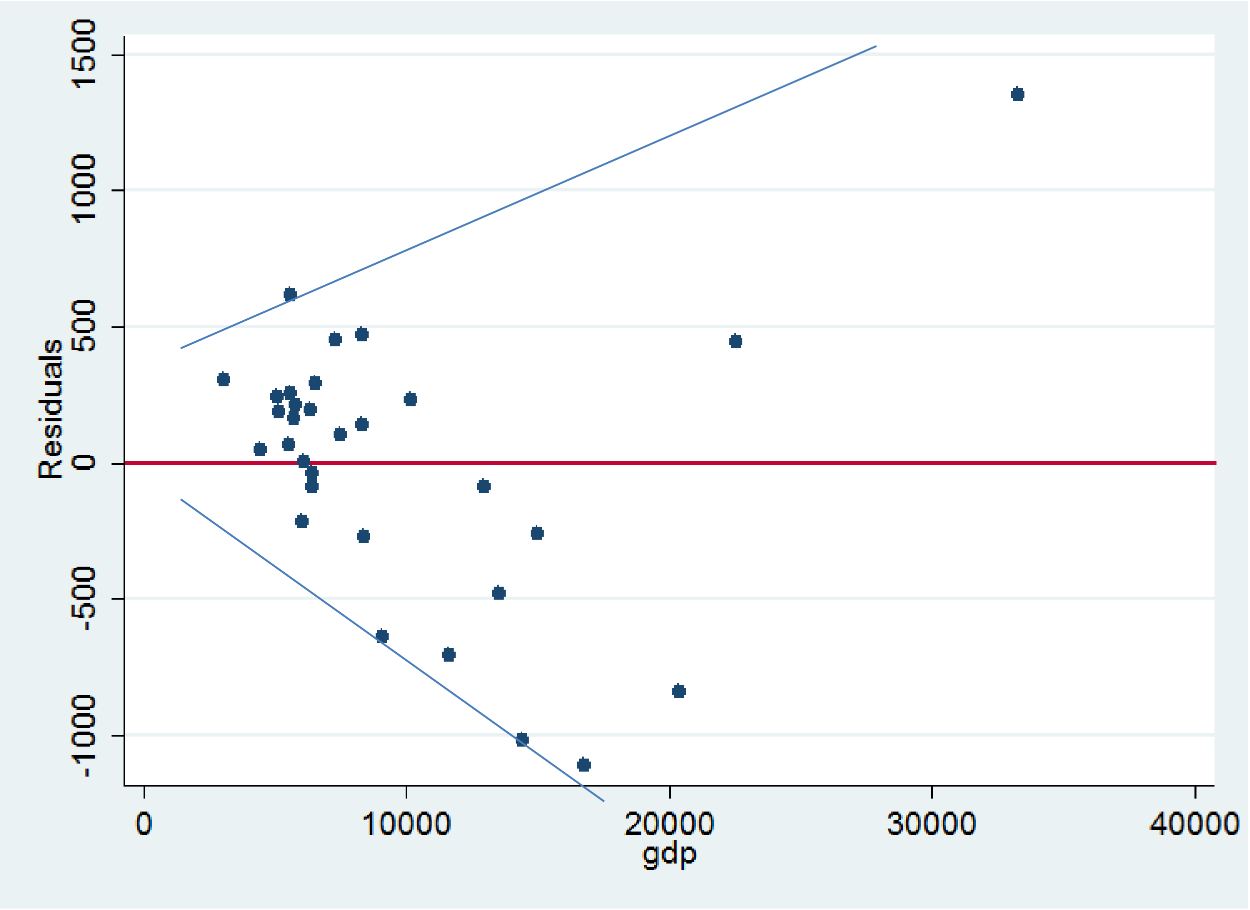

图3：残差和人均GDP的散点图

异方差检验（Gold-Quandt检验）

- Goldfeld-Quandt检验法是由S.M.Goldfeld和R.E.Quandt于1965年提出的。这种检验方法以F检验为基础，适用于大样本情形（n>30），并且要求满足条件：观测值的数目至少是参数的二倍；随机项没有自相关并且服从正态分布。
- 统计假设：零假设H_0：u_i是同方差（i=1,2,...,n）

  备择假设H_1：u_i 具有异方差

- Step 1.将某个解释变量的观测值按由小到大的顺序排列，将排列中间的约1/4的观测值去掉，除去的观测值的个数记作*c*，将其余的观测值分为两部分，每部分的观测值为(*n*–*c*)/2。
- Step 2.分别对两个样本进行回归分析，计算相应的残差平方和。RSS_1表示 较小值子样本的残差平方和,RSS_2表示 较大值子样本的残差平方和

$$F=\frac{\mathrm{RSS}_2/(\frac{n-c}{2}-k)}{\mathrm{RSS}_1/(\frac{n-c}{2}-k)}=\frac{\mathrm{RSS}_2}{\mathrm{RSS}_1}\sim F(\frac{n-c}{2}-k,\frac{n-c}{2}-k)$$

异方差检验（White检验）

Step1. 考虑下列回归式

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \ldots, n$$

Step2. 对于残差进行辅助回归

$$e_i^2 = \gamma_0 + \gamma_1 x_i + \gamma_2 x_i^2 + v_i,$$

Step3. 确立假说

$$H_0 : \gamma_1 = \gamma_2 = 0, \quad H_A : \gamma_1 \neq 0 \ \text{ or } \ \gamma_2 \neq 0$$

Step4. 计算统计量

$$nR^2 \sim \chi^2(2)$$

异方差的修正

- 异方差性虽然不损坏OLS估计量的无偏性和一致性，但却使它们不再是有效的，甚至不是渐近（即在大样本中）有效的。参数的显著性检验失效，降低了预测精度。故而直接运用普通最小二乘法进行估计不再是恰当的，需要采取相应的修正补救办法以克服异方差的不利影响。
- 其基本思路是变异方差为同方差，或者尽量缓解方差变异的程度。
- 在这里，我们将会遇到的情形分为两种：当误差项方差为已知和当为未知。

解决办法（方差已知的情况）

已知σ_i

$$\frac{y_i}{\sigma_i} = \beta_0 \frac{1}{\sigma_i} + \beta_1 \frac{x_{i1}}{\sigma_i} + \cdots + \beta_p \frac{x_{ip}}{\sigma_i} + \frac{ui}{\sigma_i}$$

此时

$$E[\frac{u_i}{\sigma_i}] = \frac{1}{\sigma_i} E(u_i) = 0$$
$$E[(\frac{u_i}{\sigma_i})^2] = \frac{1}{\sigma_i^2} E(u_i^2) = 1$$

使用加权最小二乘估计（Weight least squares method）

$$\sum_{i=1}^{n} \frac{1}{\sigma_i^2} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \cdots - \hat{\beta}_p x_{ip})^2$$

解决办法（方差未知的情况）

考虑下列回归式

$$y_i = \beta_0 + \beta_1 x_i + u_i, \quad i = 1, 2, \ldots, n$$

本来没有异方差的时候

$$V(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

现在

$$V(\hat{\beta}_1) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 \sigma_i^2}{(\sum_{i=1}^n (x_i - \bar{x})^2)^2}$$

由于σ_i^2未知，用 e_i 来代替，得

$$\hat{V}(\hat{\beta}_1) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 e_i^2}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}$$

此时标准误差称为，Heteroscedasticity Consistent Standard Error (HCSE)

$$HCSE(\hat{\beta_1}) = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 e_i^2}{(\sum_{i=1}^{n}(x_i - \bar{x})^2)^2}}$$

White修正的优点：即使异方差的形状未知，估计量具有一致性。对模型有异方差怀疑的时候，使用HCSE。

模型对数变换法

- 实际中用对经济变量取自然对数的方法也可以很好的克服异方差。

异方差检验的方法

（1）残差图观察法。由于异方差就是模型扰动项的方差是变化的，根据这个原理就可以观察模型残差拟合值的图形，根据图形的形状变化判断异方差是否存在。但是这种方法的严谨性稍差，并不是主流的检验方法。

（2）Goldfeld-Quandt Test

（3）White’s General Heteroscedasticity Test

（4） Breusch and Pagan Test

异方差的处理方法

- 稳健标准差法（HCSE）
- 只要样本容量足够大，在模型出现异方差的情况下，使用稳健标准差时参数估计、假设检验等均可正常进行，即可以很大程度上消除异方差带来的副作用。

避免异方差的方法

- 在实际应用中，避免异方差的两种方法。其一，使不同变量的测度单位接近。比如，不同国家的收入和消费数据。如果利用总收入和总消费进行分析，由于不同国家的总量相差非常巨大，因此模型中难免出现异方差。如果利用人均收入和人均消费进行分析，就可以使得减弱不同国家变量之间的测度差异，从而降低异方差的程度甚至消除异方差。
- 其二，对变量取自然对数。变量取对数降低了变量的变化程度，因此有助于消除异方差。