

第七讲 异方差性

一、异方差的概念

二、异方差的类型

三、实际经济问题中的异方差性

四、异方差性的后果

五、异方差性的检验

六、异方差的修正

七、案例

一、异方差的概念

对于模型

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_k X_{ki} + \mu_i$$

如果出现

$$Var(\mu_i) = \sigma_i^2$$

即对于不同的样本点，随机误差项的方差不再是常数，而互不相同，则认为出现了**异方差性 (Heteroskedasticity)**。

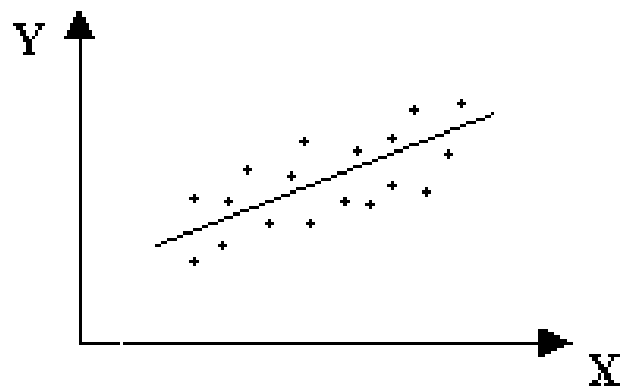
二、异方差的类型

同方差性假定： $\sigma_i^2 = \text{常数} \neq f(X_i)$

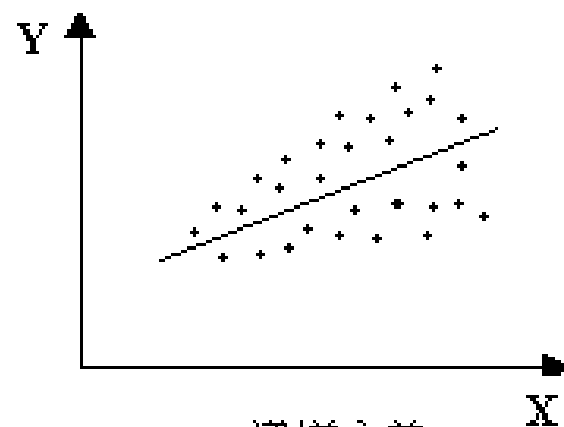
异方差时： $\sigma_i^2 = f(X_i)$

异方差一般可归结为三种类型：

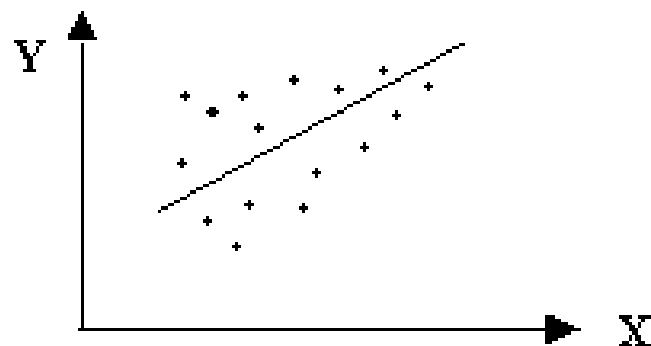
- (1) 单调递增型： σ_i^2 随X的增大而增大
- (2) 单调递减型： σ_i^2 随X的增大而减小
- (3) 复杂型： σ_i^2 与X的变化呈复杂形式



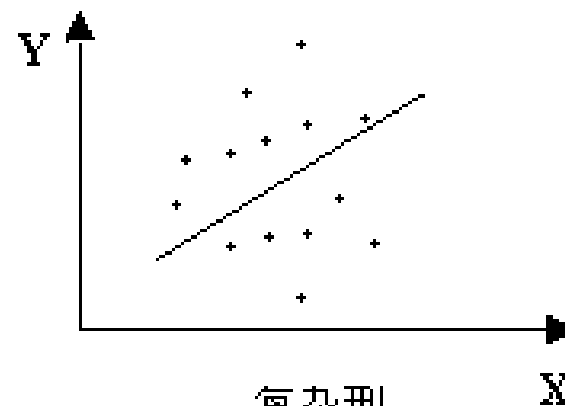
同方差



递增方差



递减方差



复杂型



三、实际经济问题中的异方差性

例4.1.1：截面资料下研究居民家庭的储蓄行为

$$Y_i = \beta_0 + \beta_1 X_i + \mu_i$$

Y_i :第*i*个家庭的储蓄额 X_i :第*i*个家庭的可支配收入

高收入家庭：储蓄的差异较大

低收入家庭：储蓄则更有规律性，差异较小

μ_i 的方差呈现单调递增型变化

例4.1,2，以绝对收入假设为理论假设、以截面数据为样本建立居民消费函数：

$$C_i = \beta_0 + \beta_1 Y_i + \mu_i$$

将居民按照收入等距离分成n组，取组平均数为样本观测值。

一般情况下，居民收入服从正态分布：中等收入组人数多，两端收入组人数少。而人数多的组平均数的误差小，人数少的组平均数的误差大。

所以样本观测值的**观测误差**随着解释变量观测值的不同而不同，往往引起异方差性。

例4.1.3，以某一行业的企业为样本建立企业生产函数模型

$$Y_i = A_i^{\beta_1} K_i^{\beta_2} L_i^{\beta_3} e^{\mu_i}$$

被解释变量：产出量 Y

解释变量：资本 K 、劳动 L 、技术 A ，

那么：每个企业所处的**外部环境**对产出量的影响被包含在随机误差项中。

每个企业所处的外部环境对产出量的影响程度不同，造成了随机误差项的异方差性。

这时，随机误差项的方差并不随某一个解释变量观测值的变化而呈规律性变化，呈现复杂型。

四、异方差性的后果

计量经济学模型一旦出现异方差性，如果仍采用OLS估计模型参数，会产生下列不良后果：

1、参数估计量非有效

OLS估计量仍然具有无偏性，但不具有有效性
因为在有效性证明中利用了

$$E(\mu\mu')=\sigma^2I$$

而且，在大样本情况下，尽管参数估计量具有一致性，但仍然不具有渐近有效性。

2、变量的显著性检验失去意义

变量的显著性检验中，构造了t统计量

$$t = \hat{\beta}_1 / S_{\hat{\beta}_1}$$

它是建立在 σ^2 不变而正确估计了参数方差 $S_{\hat{\beta}_1}$ 的基础之上的。

如果出现了异方差性，估计的 $S_{\hat{\beta}_1}$ 出现偏误（偏大或偏小），t检验失去意义。

其他检验也是如此。

3、模型的预测失效

一方面，由于上述后果，使得模型不具有良好的统计性质；

另一方面，在预测值的置信区间中也包含有参数方差的估计量 $S_{\hat{\beta}_1}$ 。

所以，当模型出现异方差性时，参数OLS估计值的变异程度增大，从而造成对Y的预测误差变大，降低预测精度，预测功能失效。

五、异方差性的检验

- 检验思路：

由于**异方差性**就是相对于不同的解释变量观测值，随机误差项具有不同的方差。那么：

检验异方差性，也就是检验随机误差项的方差与解释变量观测值之间的相关性及其相关的“形式”。

问题在于用什么来表示随机误差项的方差

一般的处理方法：

首先采用 OLS 法估计模型，以求得随机误差项的估计量（注意，该估计量是不严格的），我们称之为“近似估计量”，用 \tilde{e}_i 表示。于是有

$$Var(\mu_i) = E(\mu_i^2) \approx \tilde{e}_i^2$$

$$\tilde{e}_i = y_i - (\hat{y}_i)_{ols}$$

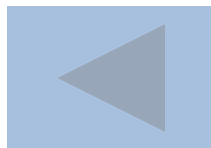
即用 \tilde{e}_i^2 来表示随机误差项的方差。

几种异方差的检验方法：

1、图示法

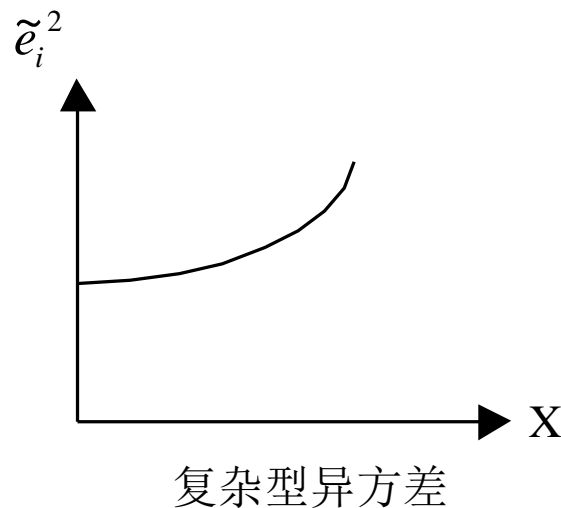
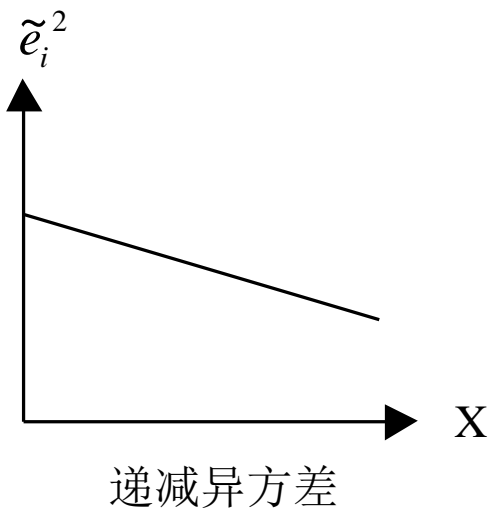
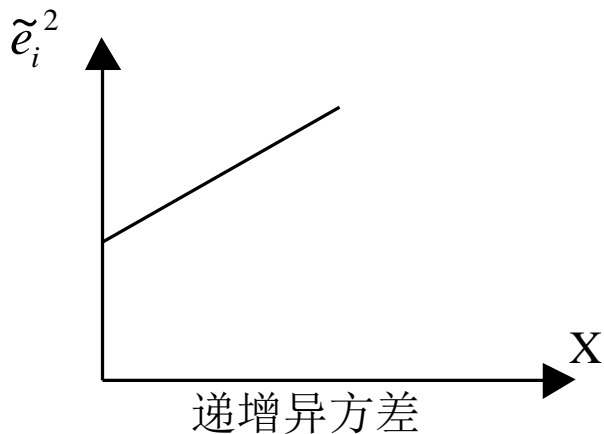
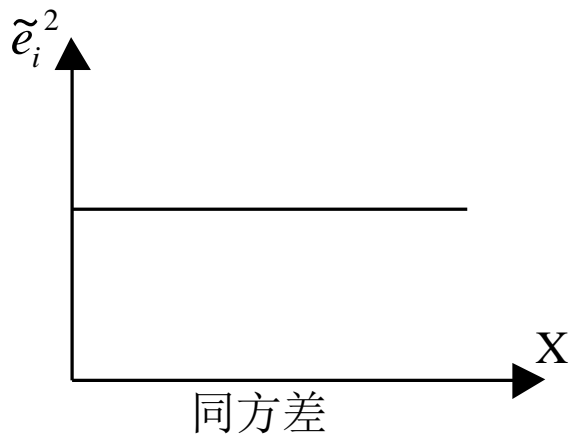
(1) 用X-Y的散点图进行判断

看是否存在明显的散点扩大、缩小或复杂型趋势（即不在一个固定的带型域中）



(2) $X-\tilde{e}_i^2$ 的散点图进行判断

看是否形成一斜率为零的直线



2、帕克(Park)检验与戈里瑟(Gleiser)检验

基本思想:

尝试建立方程:

$$\tilde{e}_i^2 = f(X_{ji}) + \varepsilon_i \quad \text{或} \quad |\tilde{e}_i| = f(X_{ji}) + \varepsilon_i$$

选择关于变量X的不同的函数形式，对方程进行估计并进行显著性检验，如果存在某一种函数形式，使得方程显著成立，则说明原模型存在异方差性。

如：帕克检验常用的函数形式:

$$f(X_{ji}) = \sigma^2 X_{ji}^\alpha e^{\varepsilon_i} \quad \text{或} \quad \ln(\tilde{e}_i^2) = \ln \sigma^2 + \alpha \ln X_{ji} + \varepsilon_i$$

若 α 在统计上是显著的，表明存在异方差性。

*3、戈德菲尔德-匡特(Goldfeld-Quandt)检验

G-Q检验以F检验为基础，适用于样本容量较大、异方差递增或递减的情况。

G-Q检验的思想：

先将样本一分为二，对子样①和子样②分别作回归，然后利用两个子样的残差平方和之比构造统计量进行异方差检验。

由于该统计量服从F分布，因此假如存在递增的异方差，则F远大于1；反之就会等于1（同方差）、或小于1（递减方差）。

G-Q检验的步骤:

- ①将n对样本观察值(X_i, Y_i)按观察值 X_i 的大小排队
- ②将序列中间的 $c=n/4$ 个观察值除去, 并将剩下的观察值划分为较小与较大的相同的两个子样本, 每个子样本容量均为 $(n-c)/2$
- ③对每个子样分别进行OLS回归, 并计算各自的残差平方和

分别用 $\sum \tilde{e}_{1i}^2$ 与 $\sum \tilde{e}_{2i}^2$ 表示较小与较大的残差平方和(自由度均为 $\frac{n-c}{2} - k - 1$);

④在同方差性假定下，构造如下满足F分布的统计量

$$F = \frac{\sum \tilde{e}_{2i}^2 / \left(\frac{n-c}{2} - k - 1 \right)}{\sum \tilde{e}_{1i}^2 / \left(\frac{n-c}{2} - k - 1 \right)} \sim F\left(\frac{n-c}{2} - k - 1, \frac{n-c}{2} - k - 1 \right)$$

⑤给定显著性水平 α ，确定临界值 $F_{\alpha}(v_1, v_2)$ ，
若 $F > F_{\alpha}(v_1, v_2)$ ，则拒绝同方差性假设，表明存在异方差。

当然，还可根据两个残差平方和对应的子样的顺序判断是递增型异方差还是递减异型方差。

4、怀特（White）检验

怀特检验不需要排序，且适合任何形式的异方差

怀特检验的基本思想与步骤（以二元为例）：

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \mu_i$$

先对该模型作OLS回归，得到 \tilde{e}_i^2

然后做如下辅助回归

$$\tilde{e}_i^2 = \alpha_0 + \alpha_1 X_{1i} + \alpha_2 X_{2i} + \alpha_3 X_{1i}^2 + \alpha_4 X_{2i}^2 + \alpha_5 X_{1i} X_{2i} + \varepsilon_i \quad (*)$$

可以证明，在同方差假设下：

$$nR^2 \rightsquigarrow \chi^2(h)$$

R^2 为(*)的可决系数， h 为(*)式解释变量的个数，

\rightsquigarrow 表示渐近服从某分布。

注意：

辅助回归仍是检验与解释变量可能的组合的显著性，因此，辅助回归方程中还可引入解释变量的更高次方。

如果存在异方差性，则表明确与解释变量的某种组合有显著的相关性，这时往往显示出有较高的可决系数以及某一参数的t检验值较大。

当然，在多元回归中，由于辅助回归方程中可能有太多解释变量，从而使自由度减少，有时可去掉交叉项。

六、异方差的修正

模型检验出存在异方差性，可用**加权最小二乘法**（**Weighted Least Squares, WLS**）进行估计。

加权最小二乘法的基本思想：

加权最小二乘法是对原模型加权，使之变成一个新的不存在异方差性的模型，然后采用OLS估计其参数。

$$\sum W_i e_i^2 = \sum W_i [Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_1 + \cdots + \hat{\beta}_k X_k)]^2$$

在采用OLS方法时：

对较小的残差平方 e_i^2 赋予较大的权数，

对较大的残差平方 e_i^2 赋予较小的权数。

例如，如果对一多元模型，经检验知：

$$Var(\mu_i) = E(\mu_i)^2 = \sigma_i^2 = f(X_{ji})\sigma^2$$

可以用 $\sqrt{f(X_j)}$ 去除该模型，得

$$\begin{aligned} \frac{1}{\sqrt{f(X_{ji})}} Y_i &= \beta_0 \frac{1}{\sqrt{f(X_{ji})}} + \beta_1 \frac{1}{\sqrt{f(X_{ji})}} X_{1i} + \beta_2 \frac{1}{\sqrt{f(X_{ji})}} X_{2i} + \cdots \\ &\quad + \beta_k \frac{1}{\sqrt{f(X_{ji})}} X_{ki} + \frac{1}{\sqrt{f(X_{ji})}} \mu_i \end{aligned}$$

新模型中，存在

$$Var\left(\frac{1}{\sqrt{f(X_{ji})}} \mu_i\right) = E\left(\frac{1}{\sqrt{f(X_{ji})}} \mu_i\right)^2 = \frac{1}{f(X_{ji})} E(\mu_i)^2 = \sigma^2$$

即满足同方差性，可用OLS法估计。

一般情况下:

对于模型

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$$

存在

$$E(\boldsymbol{\mu}) = 0$$

$$\text{Cov}(\boldsymbol{\mu}) = E(\boldsymbol{\mu} \boldsymbol{\mu}') = \sigma^2 \mathbf{W}$$

$$\mathbf{W} = \begin{bmatrix} w_1 & & & \\ & w_2 & & \\ & & \ddots & \\ & & & w_n \end{bmatrix}$$

即存在异方差性。

W是一对称正定矩阵，存在一可逆矩阵**D**使得

$$\mathbf{W} = \mathbf{D}\mathbf{D}',$$

用 \mathbf{D}^{-1} 左乘 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$

两边，得到一个新的模型：

$$\mathbf{D}^{-1}\mathbf{Y} = \mathbf{D}^{-1}\mathbf{X}\boldsymbol{\beta} + \mathbf{D}^{-1}\boldsymbol{\mu}$$

$$\mathbf{Y}_* = \mathbf{X}_*\boldsymbol{\beta} + \boldsymbol{\mu}_*$$

该模型具有同方差性。因为

$$\begin{aligned} E(\boldsymbol{\mu}_* \boldsymbol{\mu}_*') &= E(\mathbf{D}^{-1} \boldsymbol{\mu} \boldsymbol{\mu}' \mathbf{D}^{-1'}) = \mathbf{D}^{-1} E(\boldsymbol{\mu} \boldsymbol{\mu}') \mathbf{D}^{-1'} \\ &= \mathbf{D}^{-1} \sigma^2 \boldsymbol{\Omega} \mathbf{D}^{-1'} = \mathbf{D}^{-1} \sigma^2 \mathbf{D} \mathbf{D}' \mathbf{D}'^{-1} = \sigma^2 \mathbf{I} \end{aligned}$$

用OLS法估计新模型，记参数估计量为 $\hat{\beta}_*$ ，则

$$\begin{aligned}\hat{\beta}_* &= (\mathbf{X}'_* \mathbf{X}_*)^{-1} \mathbf{X}'_* \mathbf{Y}_* \\ &= (\mathbf{X}' \mathbf{D}^{-1'} \mathbf{D}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{D}^{-1'} \mathbf{D}^{-1} \mathbf{Y} \\ &= (\mathbf{X}' \mathbf{W}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{W}^{-1} \mathbf{Y}\end{aligned}$$

这就是原模型 $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$

的加权最小二乘估计量，是无偏、有效的估计量。

这里权矩阵为 \mathbf{D}^{-1} ，它来自于原模型残差项 $\boldsymbol{\mu}$ 的方差-协方差矩阵 $\sigma^2 \mathbf{W}$ 。

如何得到 $\sigma^2\mathbf{W}$?

从前面的推导过程看，它来自于原模型残差项 μ 的方差-协方差矩阵。因此

仍对原模型进行OLS估计，得到随机误差项的近似估计量 \tilde{e}_i ，以此构成权矩阵的估计量，即

$$\sigma^2 \hat{\mathbf{W}} = \begin{bmatrix} \tilde{e}_1^2 & & \\ & \ddots & \\ & & \tilde{e}_n^2 \end{bmatrix}$$

这时可直接以 $\mathbf{D}^{-1} = \text{diag}\{1/|\tilde{e}_1|, 1/|\tilde{e}_2|, \dots, 1/|\tilde{e}_n|\}$ 作为权矩阵。

注意：

在实际操作中人们通常采用如下的经验方法：

不对原模型进行异方差性检验，而是直接选择加权最小二乘法，尤其是采用截面数据作样本时。

如果确实存在异方差，则被有效地消除了；

如果不存在异方差性，则加权最小二乘法等价于普通最小二乘法

七、案例—中国农村居民人均消费函数

例4.1.4 中国农村居民人均消费支出主要由人均纯收入来决定。

农村人均纯收入包括(1)从事农业经营的收入,(2)包括从事其他产业的经营性收入(3)工资性收入、(4)财产收入(4)转移支付收入。

考察从事农业经营的收入(X_1)和其他收入(X_2)对中国农村居民消费支出(Y)增长的影响:

$$\ln Y = \beta_0 + \beta_1 \ln X_1 + \beta_2 \ln X_2 + \mu$$

表 4.1.1 中国 2001 年各地区农村居民家庭人均纯收入与消费支出相关数据（单位：元）

地区	从事农业经营 其他收入			地区	从事农业经营 其他收入		
	人均消费 支出 Y	的收入 X_1	X_2		人均消费 支出 Y	的收入 X_1	X_2
北 京	3552.1	579.1	4446.4	湖 北	2703.36	1242.9	2526.9
天 津	2050.9	1314.6	2633.1	湖 南	1550.62	1068.8	875.6
河 北	1429.8	928.8	1674.8	广 东	1357.43	1386.7	839.8
山 西	1221.6	609.8	1346.2	广 西	1475.16	883.2	1088.0
内 蒙 古	1554.6	1492.8	480.5	海 南	1497.52	919.3	1067.7
辽 宁	1786.3	1254.3	1303.6	重 庆	1098.39	764.0	647.8
吉 林	1661.7	1634.6	547.6	四 川	1336.25	889.4	644.3
黑 龙 江	1604.5	1684.1	596.2	贵 州	1123.71	589.6	814.4
上 海	4753.2	652.5	5218.4	云 南	1331.03	614.8	876.0
江 苏	2374.7	1177.6	2607.2	西 藏	1127.37	621.6	887.0
浙 江	3479.2	985.8	3596.6	陕 西	1330.45	803.8	753.5
安 徽	1412.4	1013.1	1006.9	甘 肃	1388.79	859.6	963.4
福 建	2503.1	1053.0	2327.7	青 海	1350.23	1300.1	410.3
江 西	1720.0	1027.8	1203.8	宁 夏	2703.36	1242.9	2526.9
山 东	1905.0	1293.0	1511.6	新 疆	1550.62	1068.8	875.6
河 南	1375.6	1083.8	1014.1				

普通最小二乘法的估计结果：

$$\ln \hat{Y} = 1.655 + 0.3166 \ln X_1 + 0.5084 \ln X_2$$

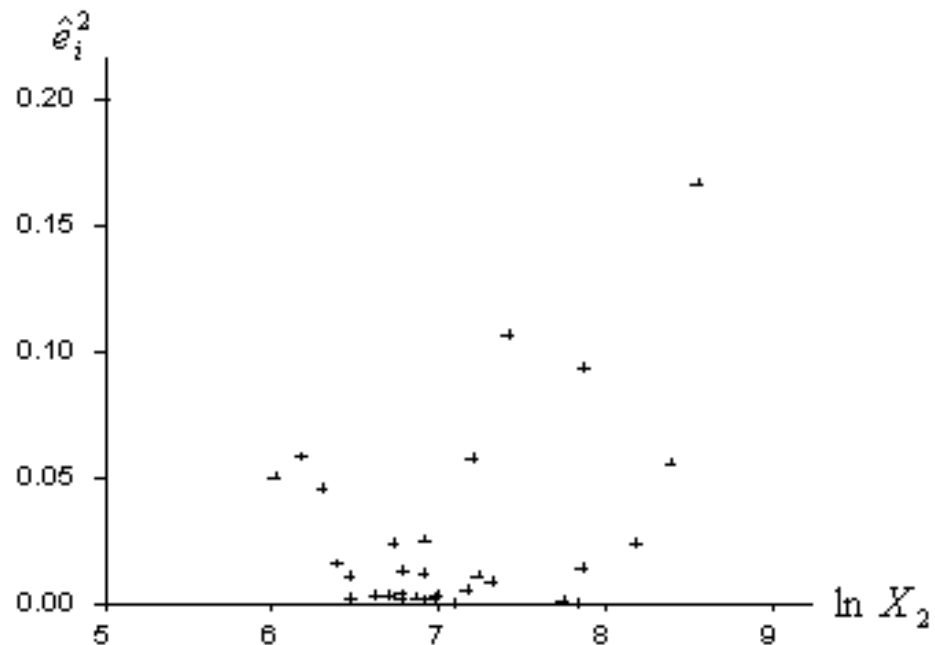
(1.87) (3.02) (10.04)

$R^2=0.7831$ $\bar{R}^2=0.7676$ DW=1.89 F=50.53 RSS=0.8232



异方差检验

OLS回归的
残差平方项
 \hat{e}_i^2 与 $\ln X_2$
的散点图



进一步的统计检验

*(1)G-Q检验

将原始数据按 X_2 排成升序，去掉中间的7个数据，得两个容量为12的子样本。

对两个子样本分别作OLS回归，求各自的残差平方和 RSS_1 和 RSS_2 ：

$$\begin{array}{l} \text{子样本1:} \quad \ln \hat{Y} = 4.061 + 0.343 \ln X_1 + 0.119 \ln X_2 \\ \quad \quad \quad (3.18) \quad (4.13) \quad \quad (0.94) \end{array}$$

$$R^2=0.7068, \quad RSS_1=0.0648$$

$$\begin{array}{l} \text{子样本2:} \quad \ln \hat{Y} = 0.791 + 0.138 \ln X_1 + 0.776 \ln X_2 \\ \quad \quad \quad (0.43) \quad (0.73) \quad \quad (6.53) \end{array}$$

$$R^2=0.8339, \quad RSS_2=0.2729$$

计算F统计量：

$$F = \text{RSS}_2 / \text{RSS}_1 = 0.2792 / 0.0648 = 4.31$$

查表

给定 $\alpha=5\%$ ，查得临界值 $F_{0.05}(9,9)=2.97$

判断

$$F > F_{0.05}(9,9)$$

否定两组子样方差相同的假设，从而该总体随机项存在递增异方差性。

(2) 怀特检验

作辅助回归：

$$\begin{aligned}\hat{e}^2 = & -0.17 + 0.102 \ln X_1 + 0.015 (\ln X_1)^2 - 0.055 \ln X_2 + 0.026 (\ln X_2)^2 \\ & (-0.04) \quad (0.10) \quad (0.21) \quad (-0.12) \quad (1.47) \\ & - 0.043 \ln X_1 \ln X_2 \\ & \quad (-1.11)\end{aligned}$$

$$R^2 = 0.4638$$

似乎没有哪个参数的t检验是显著的。但

$$n R^2 = 31 * 0.4638 = 14.38$$

$\alpha=5\%$ 下，临界值 $\chi^2_{0.05}(5)=11.07$ ，拒绝同方差性

去掉交叉项后的辅助回归结果

$$\hat{e}^2 = 3.842 - 0.570 \ln X_1 + 0.042 (\ln X_1)^2 - 0.539 \ln X_2 + 0.039 (\ln X_2)^2$$

(1.36) (-0.64) (0.64) (-2.76) (2.90)

$$R^2 = 0.4374$$

X_2 项与 X_2 的平方项的参数的t检验是显著的，且

$$n R^2 = 31 \times 0.4374 = 13.56$$

$\alpha=5\%$ 下，临界值 $\chi^2_{0.05}(4)=9.49$

拒绝同方差的原假设

原模型的加权最小二乘回归

对原模型进行OLS估计，得到随机误差项的近似估计量 \check{e}_i ，以此构成权矩阵 $\sigma^2 \mathbf{W}$ 的估计量；
再以 $1/|\check{e}_i|$ 为权重进行WLS估计，得

$$\ln \hat{Y} = 1.497 + 0.319 \ln X_1 + 0.527 \ln X_2$$

$$(5.12) \quad (5.94) \quad (28.94)$$

$$R^2=0.9999 \quad \bar{R}^2=0.9999 \quad DW=2.49 \quad F=924432 \quad RSS=0.0706$$

各项统计检验指标全面改善



本章STATA命令语句

1.White检验

Reg y x

estat imtest,white

2.加权最小二乘法

reg y x [aw=x^2]

3.稳健的标准误

reg y x,robust