计量经济学基础

第五章　异方差

主要内容

一、异方差的概念

二、异方差的来源与后果

三、异方差的检验

四、异方差的修正

说 明

- 回归分析，是在对线性回归模型提出若干基本假设的条件下，应用普通最小二乘法得到了无偏的、有效的参数估计量。
- 但是，在实际的计量经济学问题中，完全满足这些基本假设的情况并不多见。
- 如果违背了某一项基本假设，那么应用普通最小二乘法估计模型就不能得到无偏的、有效的参数估计量，OLS法失效，这就需要发展新的方法估计模型。
- 如果随机误差项的同方差性假设被破坏，这种情况称之为异方差性。

第一节　异方差的概念

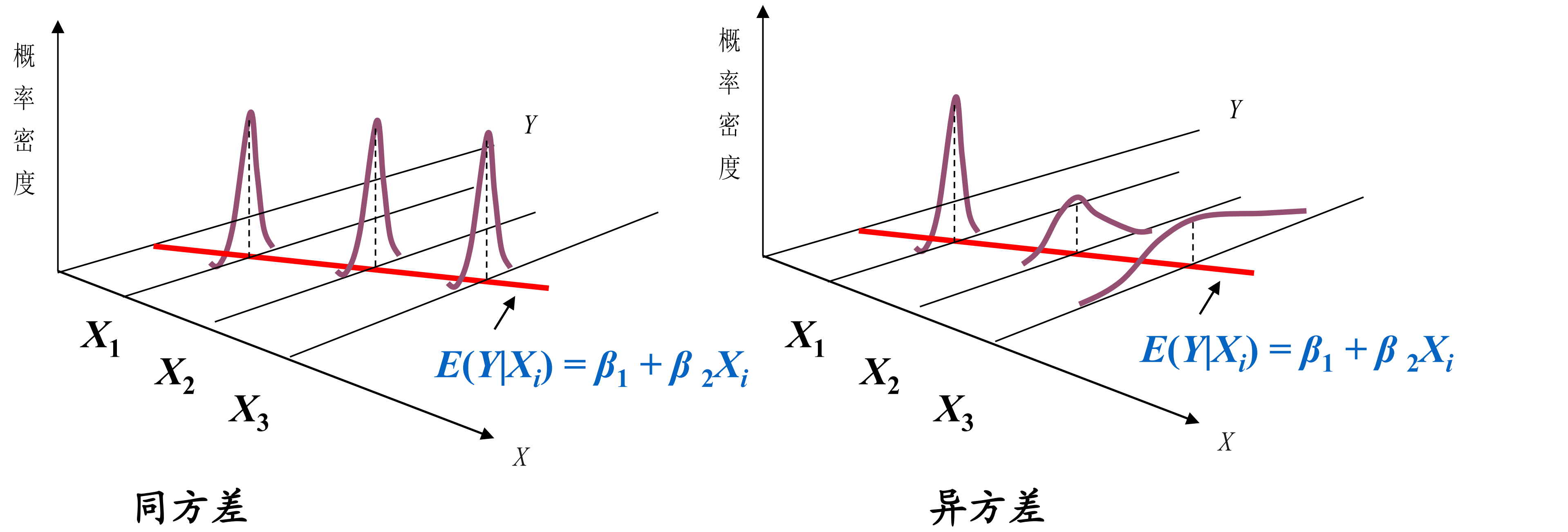
1、异方差性的概念

对于模型 $Y_i=\beta_1+\beta_2X_{2i}+\beta_3X_{3i}+\cdots+\beta_kX_{ki}+u_i$

如果出现 $Var(u_i|X_i)=\sigma_i^2$

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**)。

异方差的图解



上图显示了异方差本质：扰动项的方差不再是一个常数，而会随自变量取值的变化而变化

由此给出异方差的一个建设性定义： $Var(Y|X_i)=Var(u_i|X_i)=\sigma_i^2=f(X_i)$

该定义同时也给出了异方差检验的基本思路：即检验随机误差项的方差是否会随自变量取值的变化而变化？

异方差的矩阵表示

$$Var(U|X)=E(UU'|X)=\begin{array}{c c} & \begin{array}{cccc} u_1 & u_2 & \cdots & u_n \end{array} \\ \begin{array}{c} u_1 \\ u_2 \\ \vdots \\ u_n \end{array} & \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix} \end{array} \neq \sigma^2 I$$

2、异方差的类型

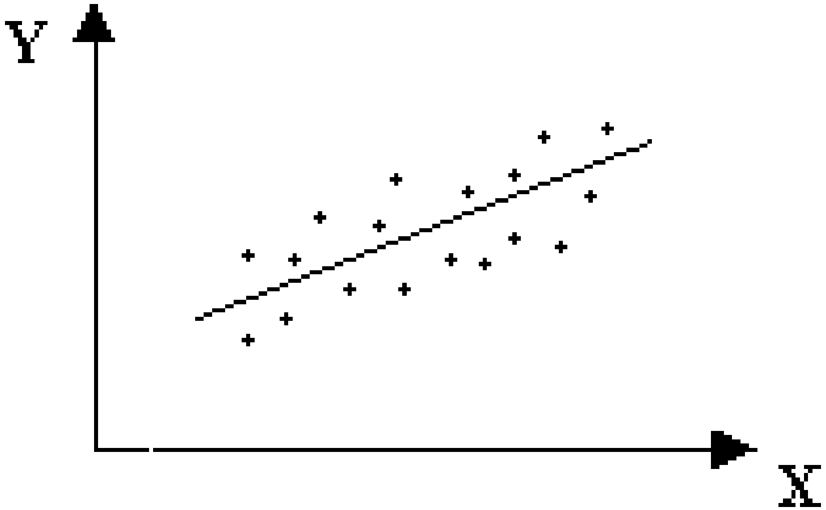
■异方差性主要产生于截面数据中，但金融时间序列数据也可能出现异方差。

■异方差一般可归结为三种类型：

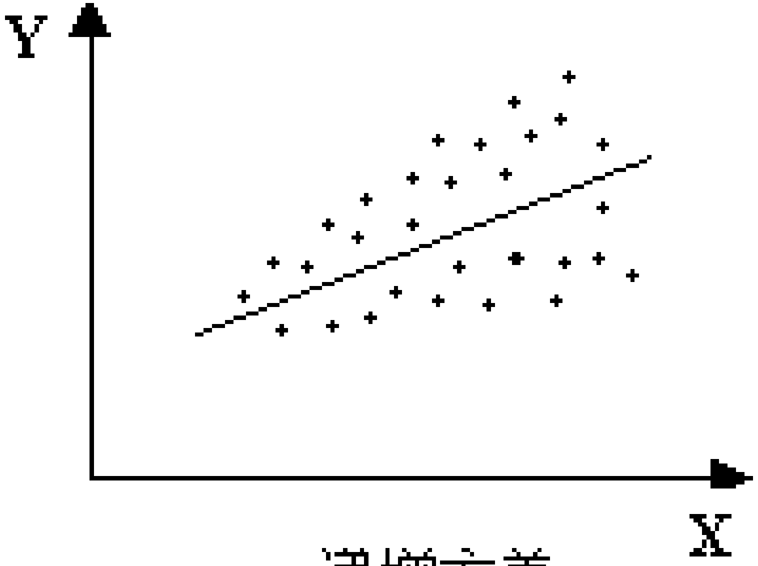
(1)单调递增型：　σ_i^2随X的增大而增大

(2)单调递减型：　σ_i^2随X的增大而减小

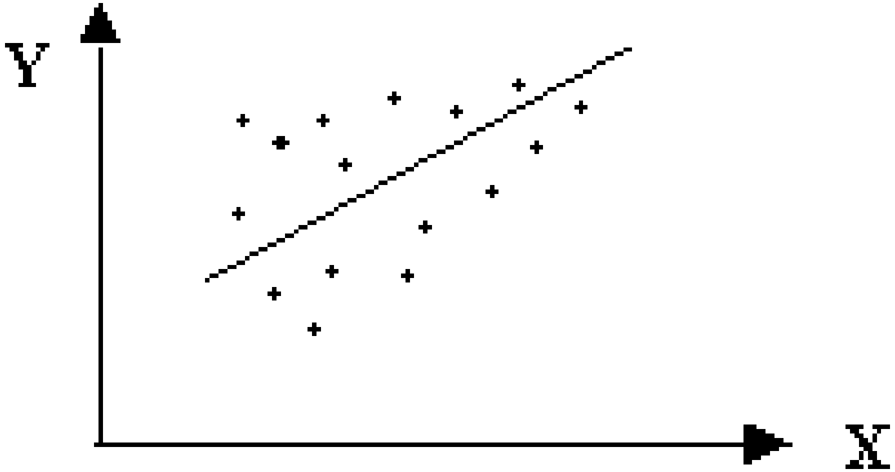
(3)复　杂　型：　σ_i^2与X的变化呈复杂形式



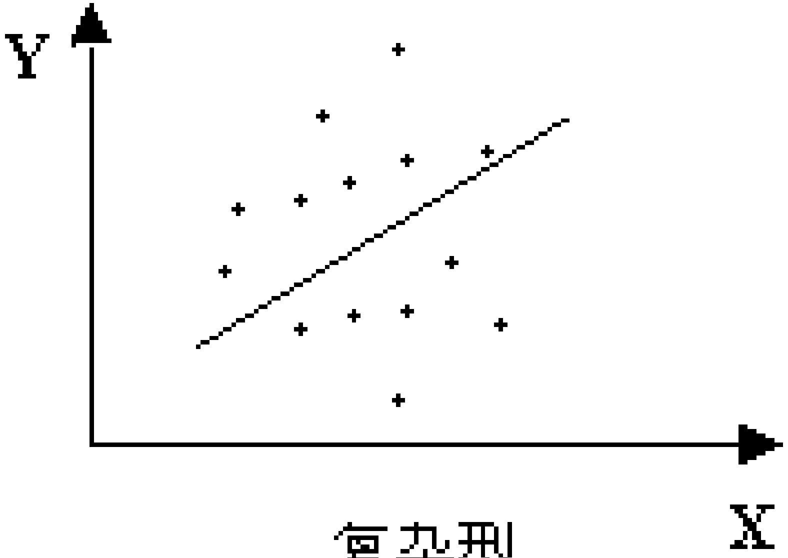
同方差



递增方差



递减方差



复杂型

第二节　异方差的来源

- 异方差来源于所研究的经济问题本身

例：在研究家庭收入（X_t)与储蓄（Y_t)关系的计量经济模型中：

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

➢高收入者比低收入者储蓄更多，且储蓄额的变化也更大，此时将产生递增型异方差；

➢经济问题中的异方差性大多是递增型的。

第二节 异方差的来源

■异方差来源于分组数据

例：在居民消费模型中

- 以绝对收入假设为理论假设、以截面数据作样本建立居民消费函数： $C_i = \beta_1 + \beta_2 I_i + u_i$
- 将居民按收入等距离分成*T*组，取组平均数为样本观测值。
- 不妨设居民收入服从正态分布，所以处于每组中的人数是不同的；中等收入组中的人数最多，高收入组和低收入组中的人数最少。
- 由于 $\sigma_{\bar{X}}^2 = \sigma^2 / n$ ，故随机误差项方差随着解释变量观测值的增大而呈先减后增型（*U*型异方差）。

第二节　异方差的来源

■异方差来源于解释变量的缺失

例：企业生产函数模型

以某一行业的企业为样本建立企业生产函数模型

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

产出量为被解释变量，选择资本、劳动、技术等投入要素为解释变量，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。

由于每个企业所处的外部环境(宏观经济政策等等)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，为复杂型的一种。

第二节 异方差的后果

计量模型一旦出现异方差性，若仍然采用OLS 估计模型参数，则会产生如下不良后果：

1、*OLS*估计量仍然满足线性性和无偏性，但不再具有有效性。

1）仍然满足线性性和无偏性：

对总体回归模型 $Y=X\beta+U$，在推导出β 的OLS估计量$\hat{\beta}$ 的过程中，并未用到同方差假定，因此是否存在异方差并不影响$\hat{\beta}$ 的表达式，即在存在异方差时，$\hat{\beta}$ 仍然为$\hat{\beta}=(X'X)^{-1}X'Y$，因此OLS估计量仍然满足线性性。

$$E\hat{\beta}=E\left[(X'X)^{-1}X'Y\right]=(X'X)^{-1}X'EY=(X'X)^{-1}X'E(X\beta+U)$$

$=(X'X)^{-1}X'X\beta+EU=\beta$。显然，为证明无偏性也无须用到同方差假定，故即使存在异方差，$E\hat{\beta}=\beta$仍然成立。

2）不再具有有效性：

由第三章知：$\hat{\beta}$ 的方差-协方差矩阵为 $Var\left(\hat{\beta}\middle|X\right)=\left(X'X\right)^{-1}X'E\left(UU'|X\right)X\left(X'X\right)^{-1}$

若随机扰动项同方差，则其方差－协方差矩阵 $E\left(UU'|X\right)=\begin{pmatrix}\sigma^2 & 0 & \cdots & 0\\ 0 & \sigma^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma^2\end{pmatrix}=\sigma^2 I$

若随机扰动项异方差，则 $E\left(UU'|X\right)=\begin{pmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \sigma_n^2\end{pmatrix}=\sigma^2\begin{pmatrix}\omega_1 & 0 & \cdots & 0\\ 0 & \omega_2 & \cdots & 0\\ \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & \cdots & \omega_n\end{pmatrix}=\sigma^2\Omega$

故 $Var\left(\hat{\beta}\middle|X\right)=\begin{cases}\sigma^2\left(X'X\right)^{-1} & \text{（同方差）}\\ \sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1} & \text{（异方差）}\end{cases}$

可以证明，在 β 的所有线性无偏估计量中，同方差下OLS 估计量的方差 $Var\left(\hat{\beta}\middle|X\right)=\sigma^2\left(X'X\right)^{-1}$ 是最小的。但异方差下OLS估计量的方差 $Var\left(\hat{\beta}\middle|X\right)=\sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}$ 不再是最小的了

若随机扰动项存在异方差，此时仍然采用OLS进行参数估计，则OLS默认所有经典假定满足，故将按照$Var\left(\hat{\beta}\middle|X\right)=\sigma^2\left(X'X\right)^{-1}$给出$S_{\hat{\beta}_j}$的值。

而在异方差情形下，正确的$S_{\hat{\beta}_j}$的值应该由$Var\left(\hat{\beta}\middle|X\right)=\sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}$给出。

故异方差下OLS估计量不再具有有效性的实质是：OLS会给出错误的$S_{\hat{\beta}_j}$。

由此将直接导致：

2、参数的显著性检验、参数的区间估计、方程的显著性检验和模型的预测全部失效

原因：错误的$S_{\hat{\beta}_j}$不仅导致错误的t统计量值，从而直接导致参数的显著性检验和区间估计以及模型预测不可信，且进一步导致$t=\dfrac{\hat{\beta}_j}{S_{\hat{\beta}_j}}$不再服从$t$分布，$F=\dfrac{ESS/(k-1)}{RSS/(n-k)}$不再服从$F$分布，从而使$F$检验失效。

第三节　异方差的检验

➢1、检验方法的共同思路

- 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？

- 一般的处理方法是:

首先采用OLS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”，记为e_i，于是有:

$$e_i = Y_i - \left(\hat{Y}_i\right)_{OLS} \qquad Var(u_i) = E\left(u_i^2\right) \approx e_i^2$$

即用 e_i^2来表示随机误差项的方差。

总结：异方差检验方法所遵循的共同思路:

(1)理论上是检验随机扰动项的方差是否随自变量取值的变化而变化。即检验$\sigma_i^2 = f\left(X_i\right)$是否显著成立。

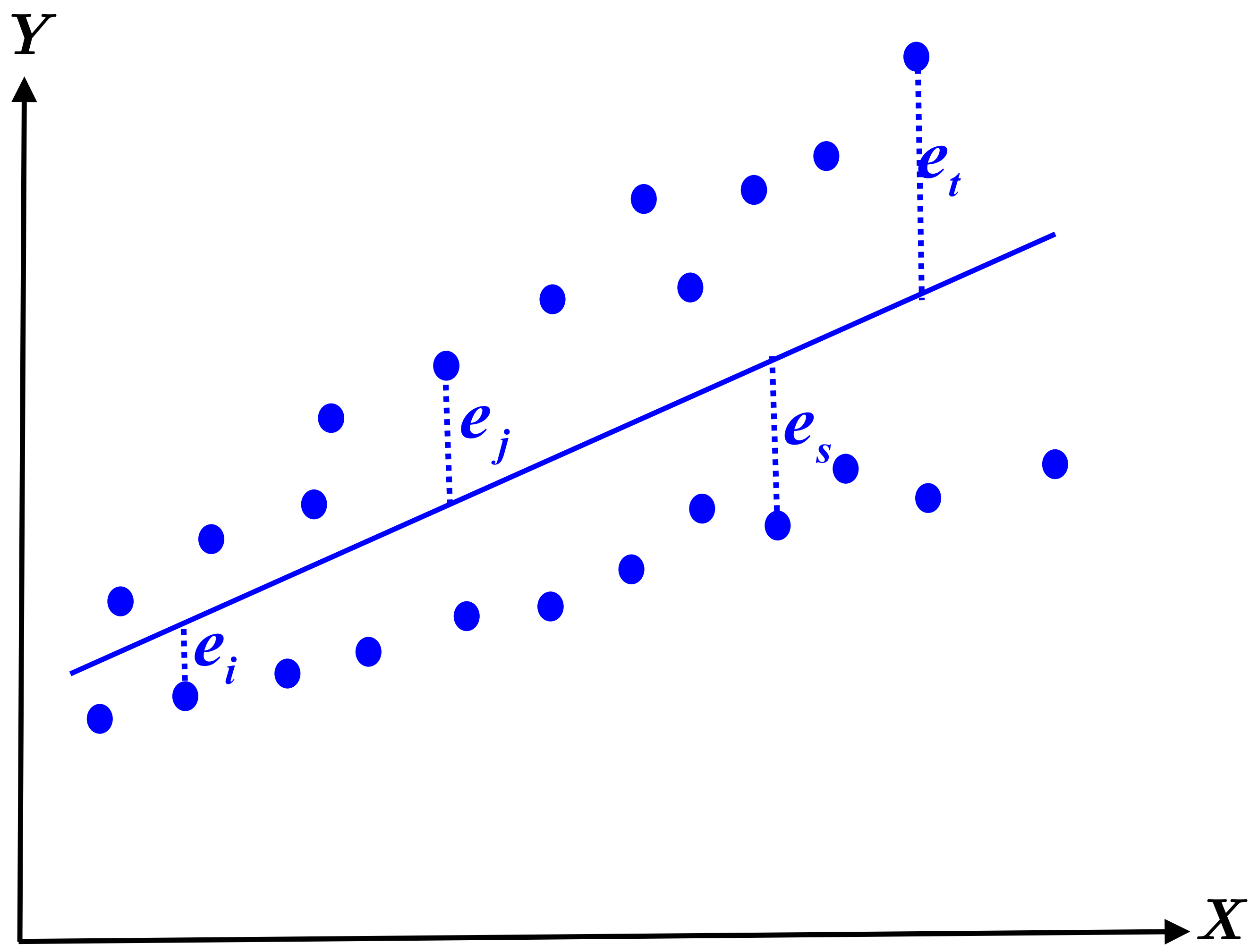
(2)实际检验时，则检验残差平方是否随自变量取值的变化而变化。即检验$e_i^2 = g\left(X_i\right)$是否显著成立。

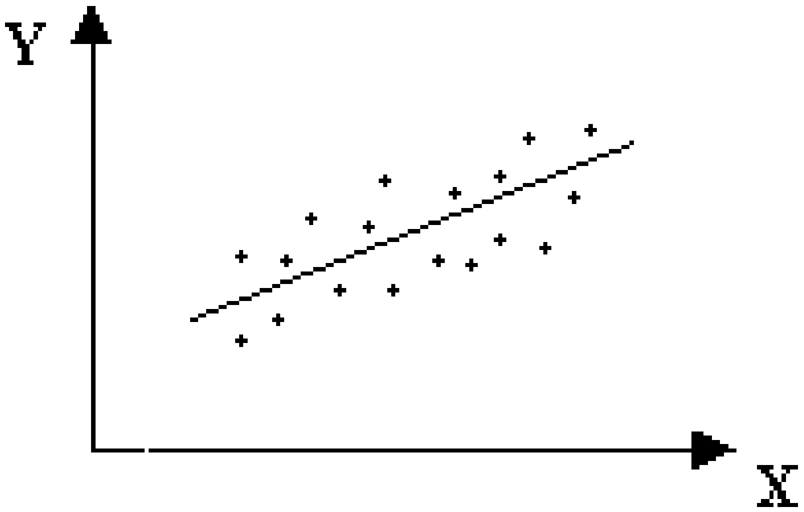
第三节　异方差的检验

➢2、图示检验法

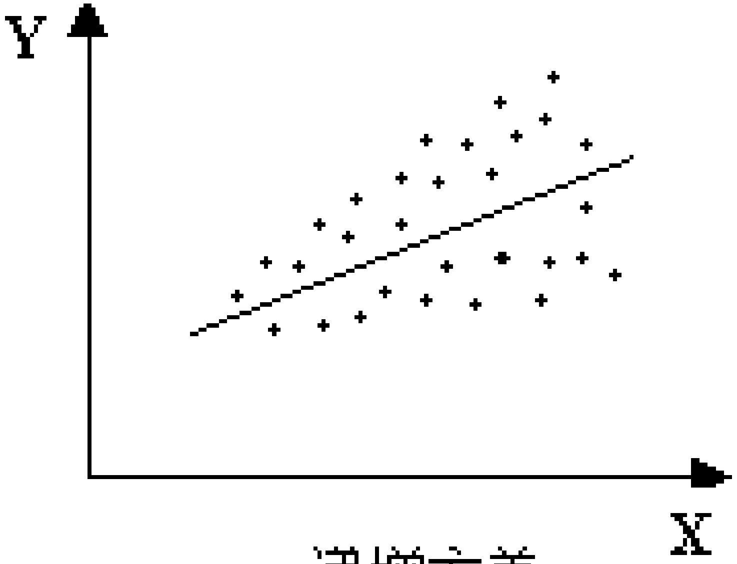
（1）用*X-Y*的散点图进行判断，看样本点对样本回归线的偏离程度是否存在明显**扩大、缩小或复杂的变动趋势**。

随着自变量取值的增加，样本点对样本回归线的偏离逐渐增加（或残差平方的值逐渐增加），因此怀疑随机扰动项存在递增型异方差。

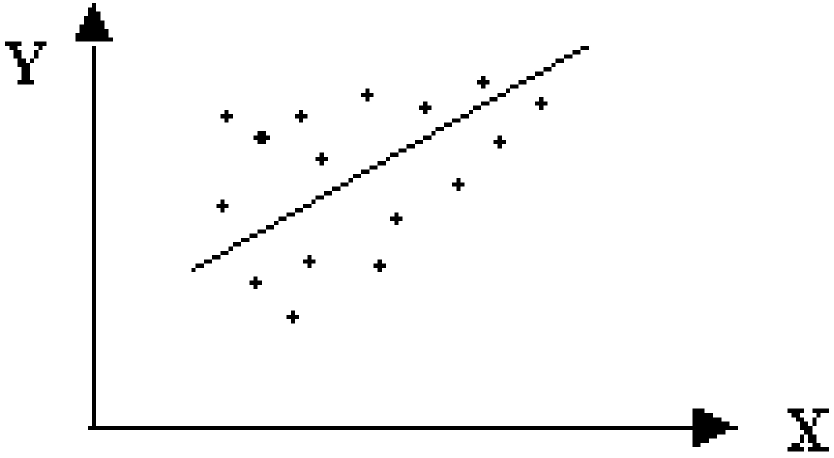




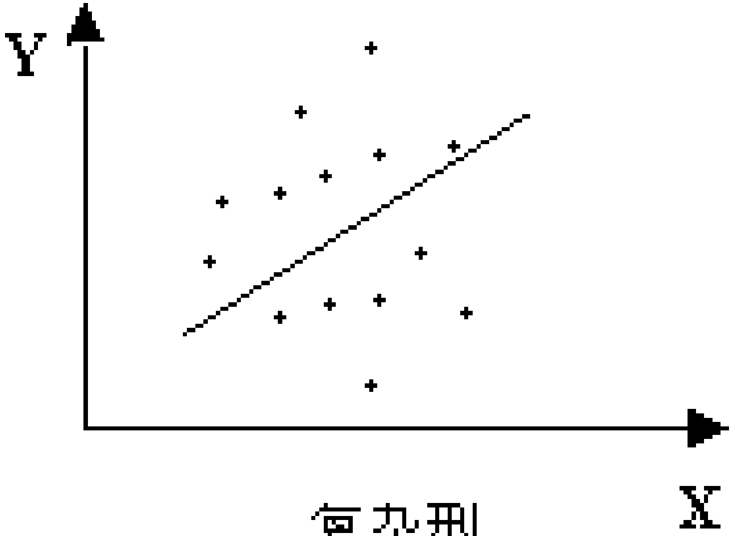
同方差



递增方差



递减方差



复杂型

（2）用 $X-e_i^2$ 的散点图进行判断

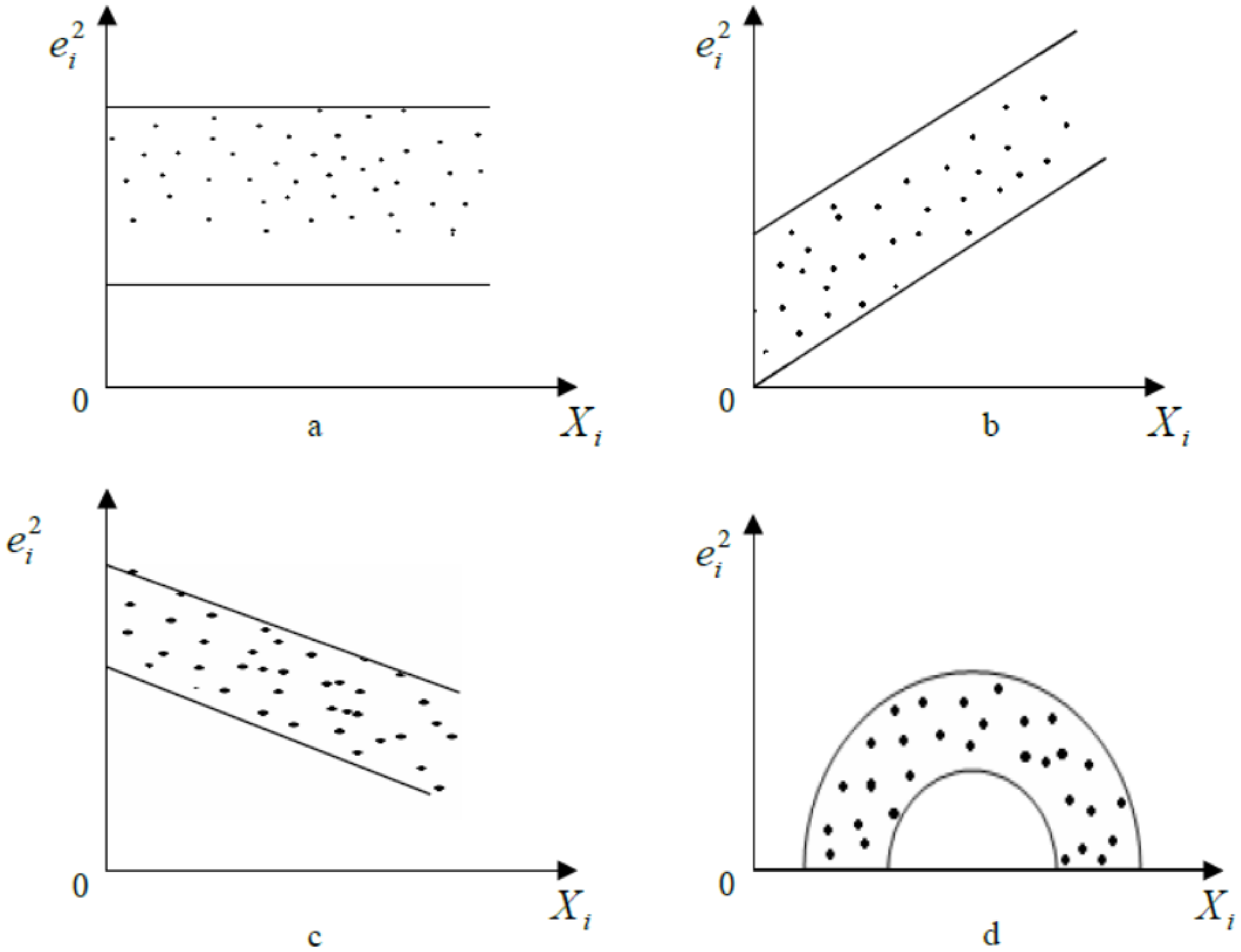
直接观察 $e_i^2=f(X_i)$ 是否显著成立：

(1)若 e_i^2 随 X 增加而增加，怀疑有递增型异方差。

(2)若 e_i^2 随 X 增加而减少，怀疑有递减型异方差。

(3)若 e_i^2 随 X 增加而忽增忽减，怀疑有复杂型异方差。

(4)若 e_i^2 不随 X 的变化而变化，怀疑随机扰动项同方差。



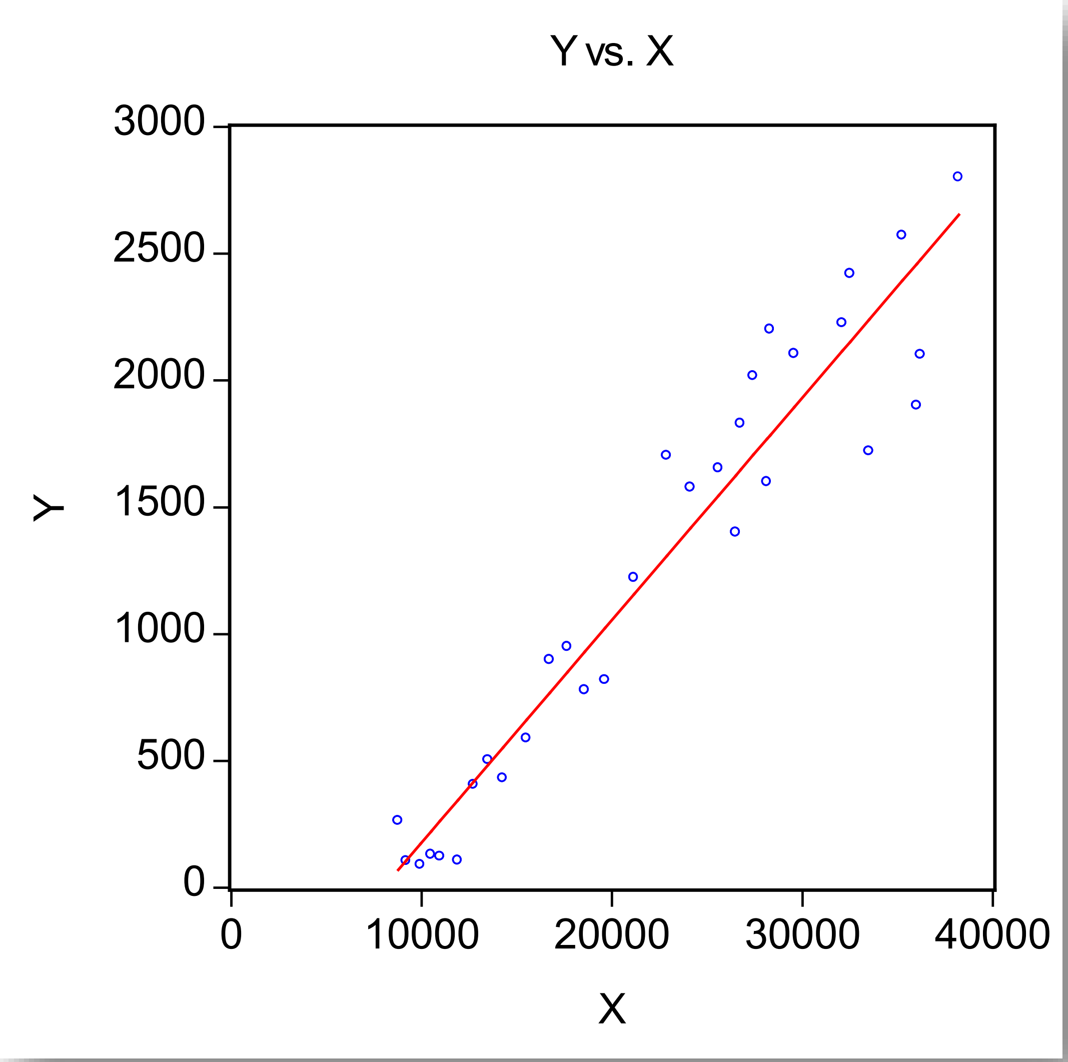
第三节　异方差的检验

例5.1：已知某地居民个人可支配收入（X）和个人储蓄（Y）的截面样本数据如下，建立回归模型$Y=\beta_1+\beta_2X+u$，试对其进行异方差检验。

序号	储蓄Y	收入X	序号	储蓄Y	收入X	序号	储蓄Y	收入X
1	264	8777	11	898	16730	21	2200	28300
2	105	9210	12	950	17663	22	2017	27430
3	90	9954	13	779	18575	23	2105	29560
4	131	10508	14	819	19635	24	1600	28150
5	122	10979	15	1222	21163	25	2250	32100
6	107	11912	16	1702	22880	26	2420	32500
7	406	12747	17	1578	24127	27	2570	35250
8	503	13499	18	1654	25604	28	1720	33500
9	431	14269	19	1400	26500	29	1900	36000
10	588	15522	20	1829	26760	30	2100	36200
						31	2800	38200

第三节　异方差的检验

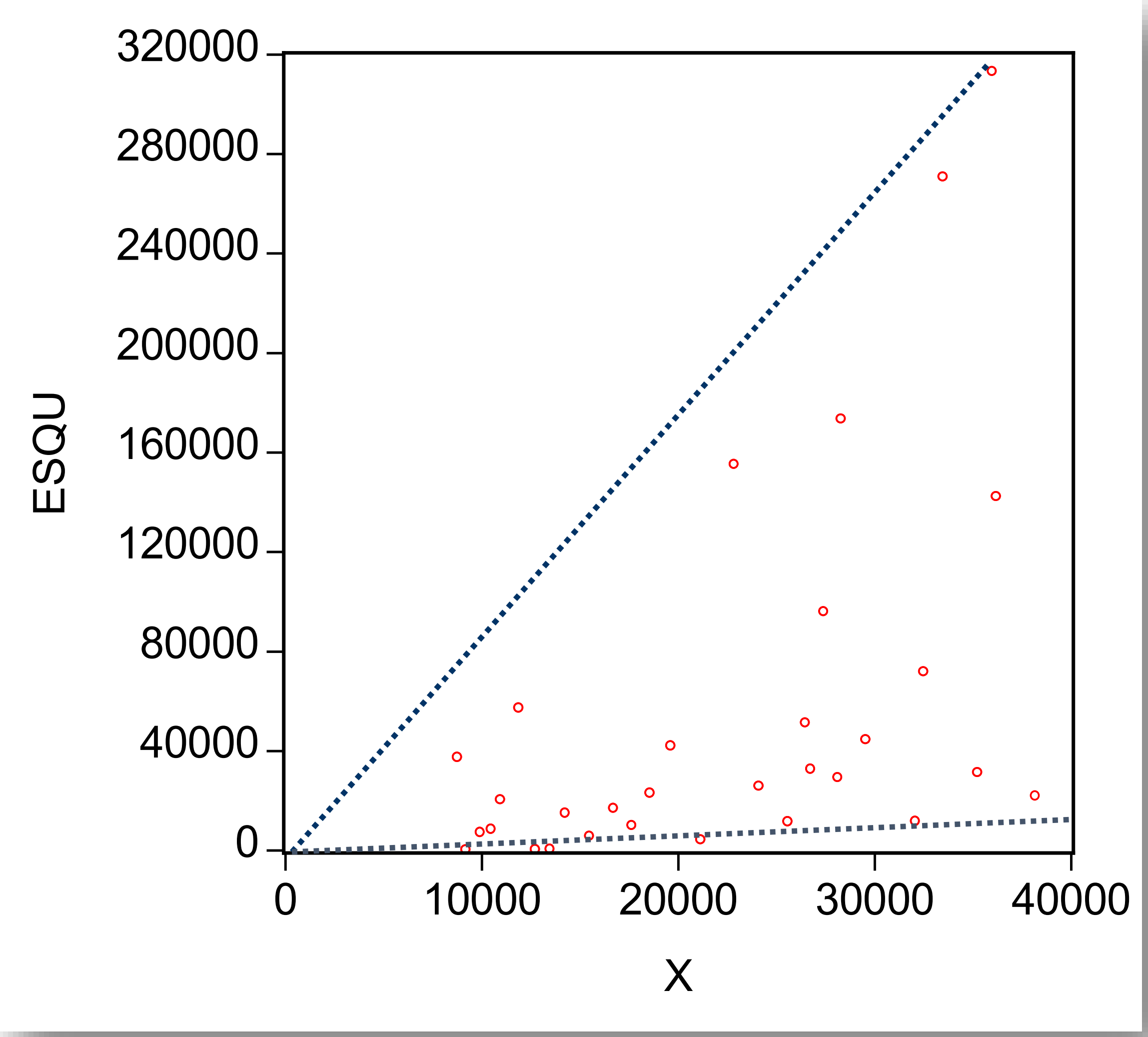
方法1：作X-Y散点图从图中可以看出，随着居民可支配收入X的提高，储蓄Y的离散程度增大，表明随机误差项存在递增型异方差。



第三节　异方差的检验

方法2：作X-e_i^2散点图

从图中可以看出，随着居民可支配收入X的提高，随机误差项平方e_i^2呈递增趋势。表明随机误差项存在递增型异方差。



第三节　异方差的检验

➢3、解析法

◆（1）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:

- 递增或递减型异方差（复杂型异方差无效）
- 样本容量较大
- 除了同方差假定不成立外，其它假定均满足。

1）戈德菲尔德-匡特检验

1、将样本按解释变量的升序重新进行排列：

$$X_1 < X_2 < \cdots < X_{s-1} < X_s < \cdots < X_{s+c} < X_{s+c+1} < \cdots < X_{n-1} < X_n$$

如果随机扰动项确实存在递增型异方差，则必有：

$$e_1^2 < e_2^2 < \cdots < e_{s\text{-}1}^2 < e_s^2 < \cdots < e_{s+c}^2 < e_{s+c+1}^2 < \cdots < e_{n-1}^2 < e_n^2$$

2、去掉中间的若干个样本点，将剩余样本分为两个子样：

$$X_1 < X_2 < \cdots < X_{s-1} < X_s < \cdots < X_{s+c} < X_{s+c+1} < \cdots < X_{n-1} < X_n$$

$$\underbrace{e_1^2 < e_2^2 < \cdots < e_{s\text{-}1}^2}_{\text{子样1}} < e_s^2 < \cdots < e_{s+c}^2 < \underbrace{e_{s+c+1}^2 < \cdots < e_{n-1}^2 < e_n^2}_{\text{子样2}}$$

3、如果随机扰动项确实存在递增型异方差，那么子样2 对应的RSS 应该显著大于子样1 对应的RSS。GQ 检验就是对这一猜测进行规范检验。

1）戈德菲尔德-匡特检验

✓ **G-Q检验的步骤：**

假设：H_0: $\sigma_1^2=\sigma_2^2=\ldots=\sigma_T^2$

H_1: $\sigma_1^2\le\sigma_2^2\le\ldots\le\sigma_T^2$

(1)将n对样本观察值(X_i, Y_i)按解释变量X_i的升序排列；

(2)将序列中间的c个观察值除去，并将剩下的观察值依次划分为两个子样本1和2，每个子样本的容量均为$(n-c)/2$;

1）戈德菲尔德-匡特检验

(3)对每个子样本分别进行OLS回归，并计算各自的残差平方和$RSS1$和$RSS2$

$$RSS_1=\sum e_{1i}^2\left(\text{自由度为}\frac{n-c}{2}-k\right),\quad RSS_2=\sum e_{2i}^2\left(\text{自由度为}\frac{n-c}{2}-k\right)$$

注：$\dfrac{RSS_1}{\sigma^2}=\dfrac{\sum e_{1i}^2}{\sigma^2}\sim\chi^2\left(\dfrac{n-c}{2}-k\right),\quad \dfrac{RSS_2}{\sigma^2}=\dfrac{\sum e_{2i}^2}{\sigma^2}\sim\chi^2\left(\dfrac{n-c}{2}-k\right)$

(4)在同方差性假定下，构造如下满足F分布的统计量：

$$F=\frac{\dfrac{\sum e_{2i}^2}{\sigma^2}\Big/\left(\dfrac{n-c}{2}-k\right)}{\dfrac{\sum e_{1i}^2}{\sigma^2}\Big/\left(\dfrac{n-c}{2}-k\right)}=\frac{\sum e_{2i}^2}{\sum e_{1i}^2}\sim F\left(\frac{n-c}{2}-k,\frac{n-c}{2}-k\right)$$

1）戈德菲尔德-匡特检验

(5)给定显著性水平α，查 F分布表得到临界值 $F_{\alpha}\left(\frac{n-c}{2}-k,\frac{n-c}{2}-k\right)$

(6)判断作出结论：若$F> F_{\alpha}$， 拒绝H_0，存在递增型异方差；反之接受H_0，不存在异方差。

附注：关于预测值省略数c的选择:

- G-Q检验的效果取决于c的大小，但理论上没有一个最优的c值存在;
- c越大，一方面会增大F统计量的值，从而增强检验功效；但另一方面又使每组中估计的自由度减少，从而降低检验功效;
- 哈维和菲利浦(1974)的研究表明，放弃的观测值数不应多于$n/3$;
- 经验值：当$n = 30$时，$c = 4$；当$n = 60$时，$c = 10$;

1）戈德菲尔德-匡特检验

- 两个子样的样本容量是否一致并不重要。因为它可以通过改变自由度和统计量计算公式进行调整;
- G-Q检验是否恰当取决于我们是否能正确地按递增型异方差排列样本观测值;
- 若随机误差项不服从正态分布，则G-Q统计量将不再服从F分布，G-Q检验失效。
- G-Q检验不能确定异方差的具体形式。

1）戈德菲尔德-匡特检验

例5.2　对例5.1中的数据进行异方差检验

解：将样本观测值按X的升序排列，剔除中间的5个样本，对剩余的26个样本平分为两个子样：

序号	储蓄Y	收入X
1	264	8777
2	105	9210
3	90	9954
4	131	10508
5	122	10979
6	107	11912
7	406	12747
8	503	13499
9	431	14269
10	588	15522
11	898	16730
12	950	17663
13	779	18575

序号	储蓄Y	收入X
19	1400	26500
20	1829	26760
21	2017	27430
22	1600	28150
23	2200	28300
24	2105	29560
25	2250	32100
26	2420	32500
27	1720	33500
28	2570	35250
29	1900	36000
30	2100	36200
31	2800	38200

1）戈德菲尔德-匡特检验

用子样1进行OLS回归：

$$\hat{Y} = -741.6171 + 0.08815X$$

$$t \qquad (-4.83) \qquad (7.73)$$

$$\sum e_{1i}^2 = 182659.1$$

用子样2进行OLS回归：

$$\hat{Y} = 184.0633 + 0.05974X$$

$$t \qquad (0.25) \qquad (2.60)$$

$$\sum e_{2i}^2 = 1140102$$

给定α=5%，查表知$F_{0.05}(11,11)=2.85$。而：

$$F = \frac{\sum e_{i2}^2}{\sum e_{i1}^2} = \frac{1140102}{182659.1} = 6.24 > 2.85$$

故拒绝不存在异方差的原假设。

2）怀特(White)检验

怀特检验的基本原理及步骤:

对多元线性回归模型 $Y = \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \beta_k X_k + u$ （注：X_1的值恒为1）

1、对原模型进行一次OLS回归，获取并固定原模型对应的e^2 序列

(注：EViews中可使用序列生成命令e2=Resid^2 固定e^2 序列)

2、以e^2为被解释变量，以X_1、$X_2 \cdots X_k$的平方项和两两交叉乘积项作为解释变量构造辅助如下回归模型并进行OLS回归，获取辅助模型的R^2 值:

$$e^2 = \alpha_1 + \underbrace{\underbrace{\sum_{j=2}^{k} \alpha_j X_j^2}_{\text{共有}k-1\text{项}}}_{\text{平方项：共}k-1\text{项}} + \underbrace{\underbrace{\sum_{l=2}^{k} \eta_l X_l}_{\text{共}k-1\text{项}} + \underbrace{\sum_{\substack{s,t=2 \\ s\neq t}}^{k} \gamma_{st} X_s X_t}_{\text{共}C_{k-1}^2=(k-1)(k-2)/2\text{项}}}_{\text{交叉乘积项：共}k(k-1)/2\text{项}} + \varepsilon$$

$$e^2=\alpha_1+\underbrace{\underbrace{\sum_{j=2}^{k}\alpha_j X_j^2}_{\text{共有}k-1\text{项}}}_{\text{平方项：共}k-1\text{项}}+\underbrace{\underbrace{\sum_{l=2}^{k}\eta_l X_l}_{\text{共}k-1\text{项}}+\underbrace{\sum_{\substack{s,t=2\\ s\neq t}}^{k}\gamma_{st}X_sX_t}_{\text{共}C_{k-1}^2=(k-1)(k-2)/2\text{项}}}_{\text{交叉乘积项：共}k(k-1)/2\text{项}}+u$$

3、提出假设：

$$H_0:\underbrace{\alpha_j}_{j=2\to k}=\underbrace{\eta_l}_{l=2\to k}=\underbrace{\gamma_{st}}_{s,t=2\to k\text{且}s\neq t}=0(\text{同方差})$$

H_1：至少有一个参数不等于零(异方差)

4、在H_0成立的前提下构造检验统计量：$WT(g)=nR^2\sim\chi^2(g)$

注：White检验统计量中的自由度g = 辅助回归模型中解释变量的个数

$$\text{即}g=\begin{cases}k-1 & (\text{不含交叉乘积项})\\ \dfrac{(k-1)(k+2)}{2} & (\text{包含交叉乘积项})\end{cases}$$

5、给定显著性水平α，查表求得临界值$\chi_{\alpha}^{2}\left(g\right)$，确定拒绝域$WT\left(g\right)>\chi_{\alpha}^{2}\left(g\right)$

6、计算检验统计量值$WT\left(g\right)=nR^{2}$，并与临界值比较：

1）若$nR^{2}>\chi_{\alpha}^{2}\left(g\right)$，落入拒绝域，拒绝原假设，故随机扰动项存在异方差

2）若$nR^{2}<\chi_{\alpha}^{2}\left(g\right)$，未落入拒绝域，则接受随机扰动项同方差的原假设

例5.3　对例5.1的回归模型进行White异方差检验（包含交叉乘积项）的结果如下：

Heteroskedasticity Test: White

F-statistic	5.819690	Prob. F(2,28)	0.0077
Obs*R-squared	9.102584	Prob. Chi-Square(2)	0.0106
Scaled explained SS	7.485672	Prob. Chi-Square(2)	0.0237

因$WT=nR^{2}$=9.1026>$\chi_{0.05}^{2}(2)=5.9915$，落入拒绝域，拒绝原假设，故随机扰动项存在异方差。

利用p值检验准则同样可知：White 异方差检验的伴随概率$p=1.06\%<\alpha=5\%$，故随机扰动项存在异方差

2）怀特(White)检验

- 不需要对异方差的性质作任何假设，这是怀特检验最大的优点;
- 怀特检验可能将模型的设定误差误判为异方差（如原模型错误地省略了变量的平方项）;
- 怀特检验是非建设性的，如果检验结果是拒绝同方差假设，则怀特检验对下一步应该做什么没有任何启示;
- 怀特检验的检验势比较低（即犯纳伪错误的可能性比较大，即接受同方差原假设时犯错误的可能性较大）

3）戈里瑟(Gleiser)检验

■基本思想及步骤:

➢1、用OLS来估计回归方程，求出残差;

➢2、尝试建立辅助回归方程 $|e_t| = f(X_t) + \upsilon_t$

（其中$f(X_t)$是X_t的幂函数）并进行OLS估计；然后，对上述辅助回归方程的估计结果进行（参数）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。

■这种方法不仅能够检验异方差的存在性，还可以检验出它的存在形式．但是，这种方法实质上是一种试错过程，这显然加大了正确检验的难度.

3）戈里瑟(Gleiser)检验

例5.4　使用戈里瑟检验对例5.1的回归模型进行异方差检验。

解：首先对模型进行OLS回归，并计算出残差序列的绝对值$|e_t|$;

建立辅助回归模型 $|e_t| = \alpha_0 + \alpha_1 X_t + \varepsilon_t$ ，回归结果如下：

$$|e_t| = -8.355 + 0.009X_t$$

$$p = (0.8784) \quad (0.0004)$$

$$adjR^2 = 0.3344 \quad F = 16.078 \quad (p = 0.00039)$$

表明$|e_t|$与自变量X显著正相关，即模型存在递增型异方差。

4）ARCH检验

（一）特征

大样本数据；时间序列数据；只能判断模型中是否存在异方差，而不能诊断出哪一个变量引起的异方差。

（二）ARCH 过程

设ARCH 过程为：$\sigma_t^2=\alpha_0+\alpha_1\sigma_{t\text{-}1}^2+\ldots+\alpha_p\sigma_{t\text{-}p}^2+v_t$

P为ARCH过程的阶数,并且v_t为随机误差，且$\alpha_0>0,\alpha_i>0\quad i=1,2,\ldots,p$

（三）检验的基本思想

在时间序列数据中，可认为存在的异方差性为ARCH过程，并通过检验这一过程是否成立去判断时间序列是否存在异方差。

（四）ARCH 检验的基本步骤

1.提出假设： $H_0: \alpha_1 = \alpha_2 = ... = \alpha_p = 0$； $H_1: \alpha_j \left(j = 1, 2, \ldots p\right)$ 不全为零

2.参数估计并计算:

对原模型作OLS估计，求出残差e_t，并计算残差平方序列$e_t^2, e_{t-1}^2, \ldots, e_{t-p}^2$

以分别作为对$\sigma_t^2, \sigma_{t-1}^2, \ldots, \sigma_{t-p}^2$的估计。

3.作辅助回归模型: $\hat{e}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 e_{t-1}^2 + \ldots + \hat{\alpha}_p e_{t-p}^2$

4.检验: 计算辅助回归的可决系数R^2与$n-p$的乘积$(n-p)R^2$。在H_0成立时，基于大样本，$(n-p)R^2$渐进服从自由度为p的χ^2分布。

给定显著性水平α，查χ^2分布表得临界值$\chi_\alpha^2(p)$，如果$(n-p)R^2 > \chi_\alpha^2(p)$，则拒绝原假设，表明模型中得随机误差存在异方差。

第四节 异方差的修正

- 由于异方差破坏了OLS估计量的有效性，因此通常的统计量检验和预测不再有效。故一旦检验到模型存在异方差，则必须进行修正。
- 异方差的修正措施可分为三类，即：
  - ➢使用加权最小二乘法（WLS）
  - ➢异方差稳健标准误
  - ➢重新设定模型

一、加权最小二乘法

■基本原理：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。

以简单线性回归模型为例：$Y_i = \beta_1 + \beta_2 X_i + u_i$。其中$Var\left(u_i \mid X_i\right) = \sigma_i^2$

以$\dfrac{1}{\sigma_i}$为权数对原模型进行加权变换，得到新模型：$\dfrac{Y_i}{\sigma_i} = \beta_1 \dfrac{1}{\sigma_i} + \beta_2 \dfrac{X_i}{\sigma_i} + \dfrac{u_i}{\sigma_i}$

新模型的随机扰动项为$\dfrac{u_i}{\sigma_i}$，其方差$Var\left(\left.\dfrac{u_i}{\sigma_i}\right| X_i\right) = \dfrac{Var\left(u_i \mid X_i\right)}{\sigma_i^2} = \dfrac{\sigma_i^2}{\sigma_i^2} = 1$，故加权变换之后的新模型满足同方差假定，可以使用OLS进行回归。

权数$\dfrac{1}{\sigma_i}$的确定:

1、由于$e_i^2 \approx \sigma_i^2 \Rightarrow \left|e_i\right| \approx \sigma_i \Rightarrow \dfrac{1}{\sigma_i} \approx \dfrac{1}{\left|e_i\right|}$。故可首选$\dfrac{1}{\left|e_i\right|}$作为权数进行WLS 回归。

2、若用$\dfrac{1}{\left|e_i\right|}$作为权数无法消除异方差，则利用Gleiser检验获得异方差的具体形式$\left|e_i\right| = f\left(X_i\right) \Rightarrow \dfrac{1}{\left|e_i\right|} = \dfrac{1}{f\left(X_i\right)}$，即使用$\dfrac{1}{f\left(X_i\right)}$作为权数进行WLS回归，直到异方差消除。

一、权重的选取

例. 对例5.1中的回归模型，使用$1/|e_i|$为权重的WLS估计结果如下：

$$\hat{Y}_i = -731.80 + 0.090724X_i$$

$$(p)\quad (0.0000)\quad (0.0000)$$

$$R^2 = 0.9907 \qquad F = 3095.555$$

怀特异方差检验结果如下：

Heteroskedasticity Test: White

F-statistic	9.069754	Prob. F(3,27)	0.0003
Obs*R-squared	15.55983	Prob. Chi-Square(3)	0.0014
Scaled explained SS	1.916352	Prob. Chi-Square(3)	0.5899

一、权重的选取

White检验表明，以$1/|e_i|$作为权重序列无法消除异方差，因此进一步使用戈里瑟检验的结果生成权重序列$\frac{1}{absef}=\frac{1}{|e_i|}=\frac{1}{-8.355+0.009X_i}$。

WLS估计结果如下：

$$\hat{Y}_i = \text{-}738.9402 + 0.0896X_i$$

$$(p)\quad (0.0000)\quad (0.0000)$$

$$R^2 = 0.0.9348 \qquad F = 415.462$$

怀特异方差检验结果如下：

Heteroskedasticity Test: White

F-statistic	4.123689	Prob. F(2,28)	0.0269
Obs*R-squared	7.053441	Prob. Chi-Square(2)	0.0294
Scaled explained SS	5.659882	Prob. Chi-Square(2)	0.0590

一、权重的选取

异方差依然存在，因此改变戈里瑟检验的形式，重新生成权重序列$\frac{1}{absef}=\frac{1}{|e_i|}=\frac{1}{78.9814+1.95\text{E-}07X_i^2}$。此时WLS估计结果如下：

$$\hat{Y}_i=-771.1721+0.09140X_i$$
$$(p)\quad(0.0000)\quad(0.0000)$$
$$R^2=0.09370\qquad F=431.5389$$

怀特异方差检验结果如下：

Heteroskedasticity Test: White

F-statistic	1.025474	Prob. F(3,27)	0.3969
Obs*R-squared	3.170891	Prob. Chi-Square(3)	0.3660
Scaled explained SS	1.926982	Prob. Chi-Square(3)	0.5877

二、异方差稳健标准误回归）

当模型存在异方差时，OLS将给出一个错误的，从而使参数估计量不再有效。但OLS并不影响参数估计量的无偏性，因此另一种修正异方差的思路就是：

仍然采用OLS进行参数估计，但OLS按公式$\sigma^2\left(X'X\right)^{-1}$所给出的错误$S_{\hat{\beta}_j}$修改为按公式$\sigma^2\left(X'X\right)^{-1}X'\Omega X\left(X'X\right)^{-1}$所给出的正确的$S_{\hat{\beta}_j}$，从而消除异方差所产生的不良后果。

当模型存在异方差时，OLS将给出一个错误的，从而使参数估计量不再有效。

例如：在简单线性回归模型中，若模型存在异方差，正确的$Var\left(\hat{\beta}_2\right)=\dfrac{\sum x_i^2\sigma_i^2}{\left(\sum x_i^2\right)^2}$；

而OLS给出的却是错误的$Var\left(\hat{\beta}_2\right)=\dfrac{\sigma^2}{\sum x_i^2}$。因此所谓的异方差稳健标准回归，

就是用e_i^2作为σ_i^2的近似，用$Var\left(\hat{\beta}_2\right)=\dfrac{\sum x_i^2e_i^2}{\left(\sum x_i^2\right)^2}$作为正确标准误估计值。

二、异方差稳健标准误法
（怀特异方差校正）

例5.8　对例5.1中的回归模型进行怀特异方差校正后的估计结果如下：

$$\hat{Y} = \text{-}700.4110 + 0.087831X$$
$$se = (95.76171)\ (0.005299)$$
$$adjR^2 = 0.9167 \quad F = 331.09$$

$$\hat{Y} = \text{-}700.4110 + 0.087831X$$
$$se = (116.6679)\ (0.004827)$$
$$adjR^2 = 0.9167 \quad F = 331.09$$

二、重新设定模型

■某些情况下，我们可以通过重新设定模型函数形式的方法达到消除异方差的目标。

■有时，异方差性产生于模型遗漏了重要的解释变量，此时，应该将这一变量纳入模型

■有时，通过压缩变量数据的测量尺度也可以减弱或消除异方差。常用的方式是将原来的线性模型变换为对数线性模型或其它形式的新模型。

二、重新设定模型

例5.9　在例5.1中，如果我们设定的模型为 $\sqrt{Y}=\beta_0+\beta_1\sqrt{X}+e$

则得到的OLS回归结果为：

$$\sqrt{Y}=-28.04003+0.416414\sqrt{X}$$

$$p=(0.0000)\qquad(0.0000)$$

$$adjR^2=0.9239\qquad F=365.41$$

怀特检验结果如下：

Heteroskedasticity Test: White

F-statistic	0.997129	Prob. F(2,28)	0.3817
Obs*R-squared	2.061127	Prob. Chi-Square(2)	0.3568
Scaled explained SS	1.145145	Prob. Chi-Square(2)	0.5641

注：模型变换必须有经济理论支撑，否则只是无意义的数据淘金游戏。

本例中变换后的模型虽然消除了异方差，但因为背离经济理论，却带来了更加严重的模型设定偏误问题。