

计量经济学

第五章

异方差性

第五章 异方差性

本章讨论四个问题：

- 异方差的实质和产生的原因
- 异方差产生的后果
- 异方差的检测方法
- 异方差的补救

第一节 异方差性的概念

本节基本内容:

- 异方差性的实质
- 异方差产生的原因

一、异方差性的实质

同方差的含义

同方差性：对所有的 i ($i=1,2,\dots,n$)有：

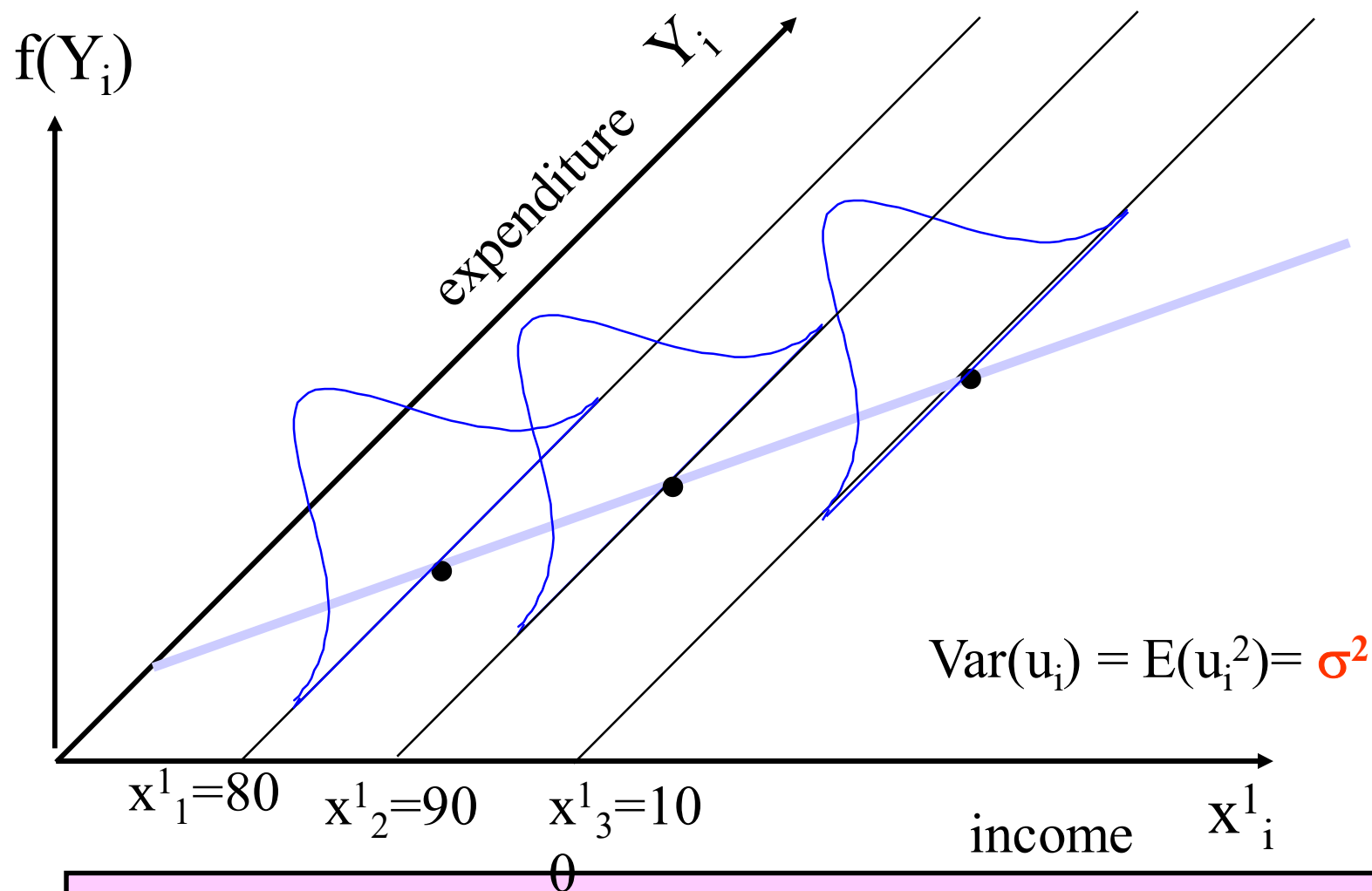
$$\text{Var}(u_i) = \sigma^2 \quad (5.1)$$

因为方差是度量被解释变量 Y 的观测值围绕回归线

$$E(Y_i) = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} \quad (5.2)$$

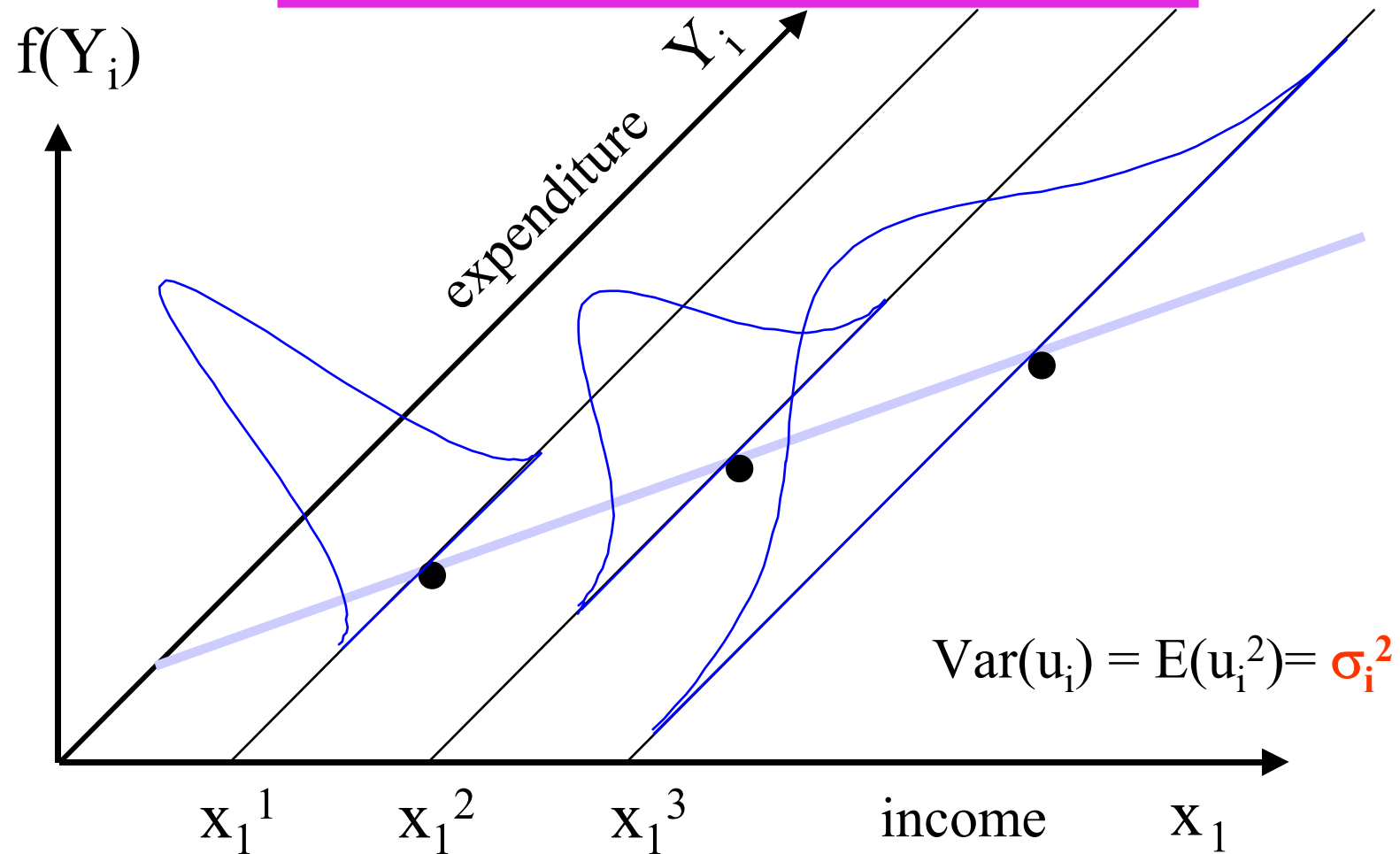
的分散程度，因此同方差性指的是所有观测值的分散程度相同。

Homoscedasticity Case



The probability density function for Y_i at two levels of family income, X_i , are identical.

Heteroscedasticity Case



The variance of Y_i increases as family income, X_i , increases.

异方差性的含义

设模型为

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \dots + \beta_k X_{ki} + u_i \quad i = 1, 2, \dots, n$$

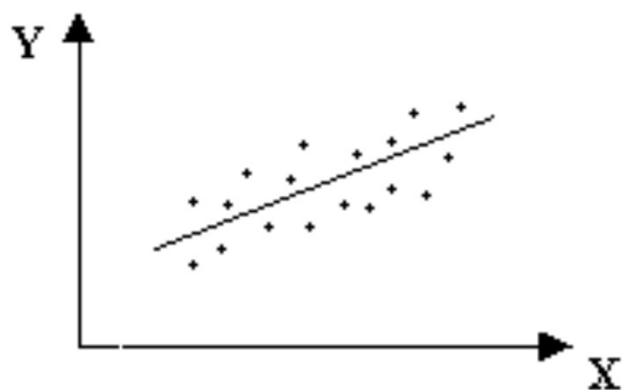
如果对于模型中随机误差项 u_i 有：

$$\text{Var}(u_i) = \sigma_i^2, \quad i = 1, 2, 3, \dots, n \quad (5.3)$$

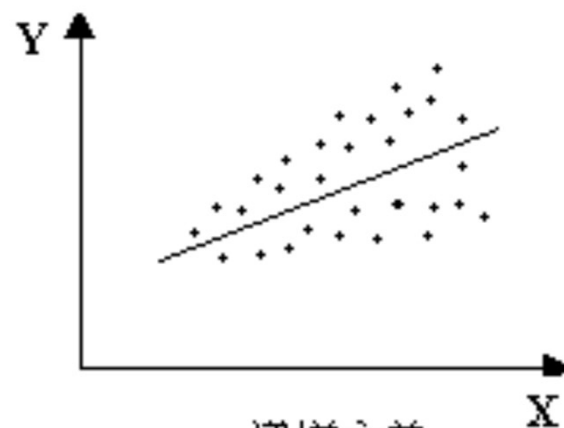
则称具有异方差性。进一步，把异方差看成是由于某个解释变量的变化而引起的，则

$$\text{Var}(u_i) = \sigma_i^2 = \sigma^2 f(X_i) \quad (5.4)$$

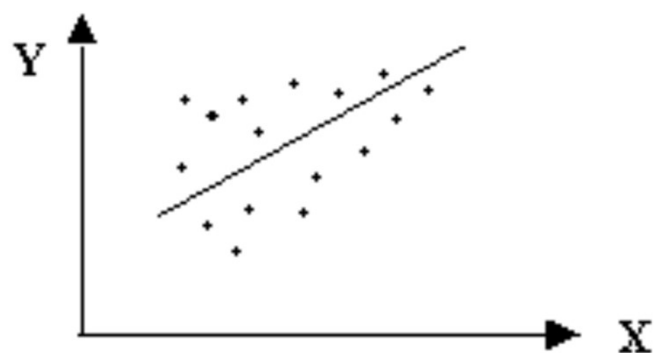
图形表示



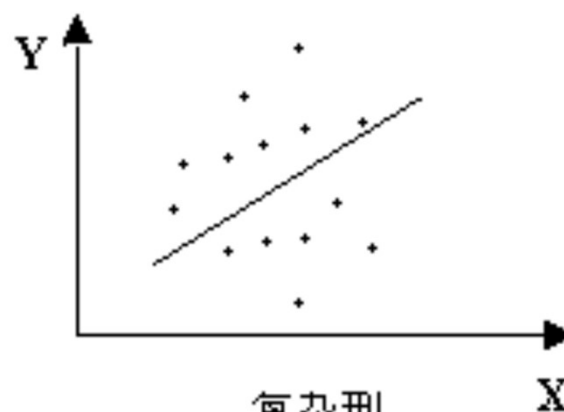
同方差



递增方差



递减方差



复杂型

二、产生异方差的原因

(一) 模型中省略了某些重要的解释变量
假设正确的计量模型是：

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

假如略去 X_{3i} ，而采用

$$Y_i = \beta_1 + \beta_2 X_{2i} + u_i^* \quad (5.5)$$

当被略去的 X_{3i} 与 X_{2i} 有呈同方向或反方向变化的趋势时，随 X_{2i} 的有规律变化会体现在 (5.5) 式的 u_i^* 中。

（二）模型的设定误差

模型的函数形式不正确，如把变量间本来为非线性的关系设定为线性，也可能导致异方差。

（三）数据的测量误差

样本数据的观测误差有可能随研究范围的扩大而增加，或随时间的推移逐步积累，也可能随着观测技术的提高而逐步减小。

（四）截面数据中总体各单位的差异

通常认为，截面数据较时间序列数据更容易产生异方差。这是因为同一时点不同对象的差异，一般说来会大于同一对象不同时间的差异。不过，在时间序列数据发生较大变化的情况下，也可能出现比截面数据更严重的异方差。

第二节 异方差性的后果

本节基本内容：

- 对参数估计统计特性的影响
- 对参数显著性检验的影响
- 对预测的影响

一、对参数估计统计特性的影响

（一）参数估计的无偏性仍然成立

参数估计的无偏性仅依赖于基本假定中的零均值假定（即 $E(u_i)=0$ ）。所以异方差的存在对无偏性的成立没有影响。

（二）参数估计的方差不再是最小的

同方差假定是**OLS**估计方差最小的前提条件，所以随机误差项是异方差时，将不能再保证最小二乘估计的方差最小。

为详细说明异方差使**OLS**参数估计量的无效性，我们考虑一元回归模型：

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

对于该模型，我们假定除同方差假设外，其他的高斯马尔科夫假设都成立。

如果模型随机误差项包含异方差，那么有

$$\text{Var}(\mu_i | X_i) = \sigma_i^2$$

这一异方差取决于 X_i 的值。

该模型参数的**OLS**估计量可以写为

$$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$$

在上述给定的异方差情况下,

容易证明 $\hat{\beta}_2$ 的方差为

$$Var(\hat{\beta}_2) = \frac{\sum_{i=1}^n x_i^2 \sigma_i^2}{\left(\sum_{i=1}^n x_i^2\right)^2}$$

(A)

而同方差假设下, $\hat{\beta}_2$ 的OLS

估计方差为

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$$

(B)

显然 (A) 式与 (B) 式不同, 只有在 $\sigma_i^2 = \sigma^2$ 时两者才是相同的。

二、对参数显著性检验的影响

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，从而正确估计了参数的标准误基础之上的。

由于异方差的影响，使得无法正确估计参数的标准误差，导致参数估计的 **t** 统计量的值不能正确确定，所以，如果仍用 **t** 统计量进行参数的显著性检验将失去意义。

三、对预测的影响

尽管参数的OLS估计量仍然无偏，并且基于此的预测也是无偏的，但是由于参数估计量不是有效的，从而对Y的预测也将不是有效的。

第三节 异方差性的检验

常用检验方法：

- 图示检验法
- Goldfeld-Quanadt检验
- White检验
- ARCH检验

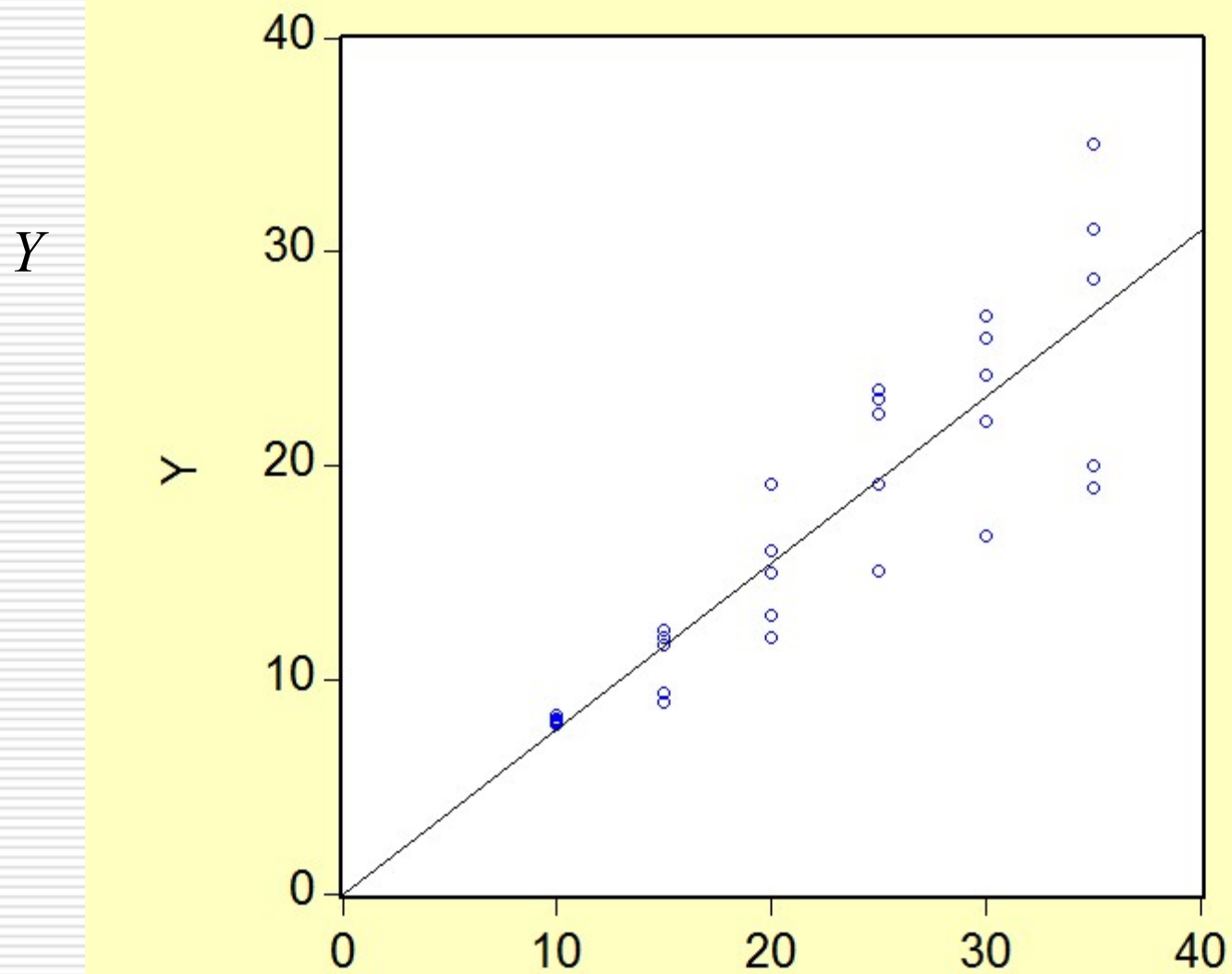
一、图示检验法

（一）相关图形分析

方差描述的是随机变量取值的（与其均值的）离散程度。因为被解释变量 Y 与随机误差项 u 有相同的方差，所以利用分析 Y 与 X 的相关图形，可以粗略地看到 Y 的离散程度与 X 之间是否有相关关系。

如果随着 X 的增加， Y 的离散程度为逐渐增大（或减小）的变化趋势，则认为存在递增型（或递减型）的异方差。

图形举例



(二) 残差图形分析

设一元线性回归模型为：

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

运用**OLS**法估计, 得样本回归模型为：

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

由上两式得残差： $e_i = Y_i - \hat{Y}_i$

绘制出 e_i^2 对 X_i 的散点图

- ◆如果 u_i 不随 X_i 而变化, 则表明**不存在异方差**;
- ◆如果 u_i 随 X_i 而变化, 则表明**存在异方差**。

二、Goldfeld-Quanadt检验

作用：检验递增性(或递减性)异方差。

基本思想：将样本分为两部分，然后分别对两个样本进行回归，并计算两个子样的残差平方和所构成的比，以此为统计量来判断是否存在异方差。

(一) 检验的前提条件

- 1、要求检验使用的为大样本容量。
- 2、除了同方差假定不成立外，其它假定均满足。

(二) 检验的具体做法

1. 排序

将解释变量的取值按从小到大排序。

2. 数据分组

将排列在中间的约1/4的观察值删除掉，记为 c ，再将剩余的分为两个部分，每部分观察值的个数为 $(n - c)/2$ 。

3. 提出假设

$$H_0 : \sigma_i^2 = \sigma^2, i=1,2,\dots,n; \quad H_1 : \sigma_1^2 \leq \sigma_2^2 \leq \dots \leq \sigma_n^2$$

4. 构造F统计量

分别对上述两个部分的观察值求回归模型，由此得到的两个部分的残差平方为 $\sum e_{1i}^2$ 和 $\sum e_{2i}^2$ 。

$\sum e_{1i}^2$ 为前一部分样本回归产生的残差平方和，

$\sum e_{2i}^2$ 为后一部分样本回归产生的残差平方和。它们的自由度均为 $[(n-c)/2]-k$ ， k 为参数的个数。

在原假设成立的条件下，因 $\sum e_{1i}^2$ 和 $\sum e_{2i}^2$ 自由度均为 $[(n-c)/2]-k$ ， χ^2 分布，可导出：

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

(5.13)

5. 判断

给定显著性水平 α ，查 **F**分布表得临界值

$F_{(\frac{n-c}{2}-k, \frac{n-c}{2}-k)}(\alpha)$ 计算统计量 F^* 。

如果

$$F^* > F_{(\frac{n-c}{2}-k, \frac{n-c}{2}-k)}(\alpha)$$

则拒绝原假设，接受备择假设，即模型中的随机误差存在异方差。

(三) 检验的特点

- 要求大样本
- 异方差的表现既可为递增型，也可可为递减型
- 检验结果与选择数据删除的个数 的大小有关
- 只能判断异方差是否存在，在多个解释变量的情况下，对哪一个变量引起异方差的判断存在局限。

三、White检验

(一) 基本思想:

不需要关于异方差的任何先验信息，只需要在大样本的情况下，将**OLS**估计后的残差平方对常数、解释变量、解释变量的平方及其交叉乘积等所构成一个辅助回归，利用辅助回归建立相应的检验统计量来判断异方差性。

(二) 检验的特点

要求变量的取值为大样本

不仅能够检验异方差的存在性，同时在多变量的情况下，还能判断出是哪一个变量引起的异方差。

(三) 检验的基本步骤:

以一个二元线性回归模型为例，设模型为：

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t$$

并且，设异方差与 X_{2t}, X_{3t} 的一般关系为

$$\sigma_t^2 = \alpha_1 + \alpha_2 X_{2t} + \alpha_3 X_{3t} + \alpha_4 X_{2t}^2 + \alpha_5 X_{3t}^2 + \alpha_6 X_{2t} X_{3t} + v_t$$

其中 v_t 为随机误差项。

1. 求回归估计式并计算 e_t^2

用**OLS**估计式 (5.14)，计算残差 $e_t = Y_t - \hat{Y}_t$ ，并求残差的平方 e_t^2 。

2. 求辅助函数

用残差平方 e_t^2 作为异方差 σ_t^2 的估计，并建立

$X_{2t}, X_{3t}, X_{2t}^2, X_{3t}^2, X_{2t}X_{3t}$ 的辅助回归，即

$$\hat{e}_t^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{2t}^2 + \hat{\alpha}_5 X_{3t}^2 + \hat{\alpha}_6 X_{2t}X_{3t} \quad (5.15)$$

3. 计算

利用求回归估计式 (5.15) 得到辅助回归函数的可决系数 nR^2 , n 为样本容量。

4. 提出假设

$$H_0 : \alpha_2 = \dots = \alpha_6 = 0, \quad H_1 : \alpha_j (j=2, 3, \dots, 6) \text{ 不全为零}$$

5. 检验

在零假设成立下，有 nR^2 渐进服从自由度为5的 χ^2 分布。给定显著性水平 α ，查 χ^2 分布表得临界值 $\chi_\alpha^2(5)$ ，如果 $nR^2 > \chi_\alpha^2(5)$ ，则拒绝原假设，表明模型中随机误差存在异方差。

四、ARCH检验

(一) ARCH 过程

设**ARCH** 过程为

$$\sigma_t^2 = \alpha_0 + \alpha_1 \sigma_{t-1}^2 + \dots + \alpha_p \sigma_{t-p}^2 + v_t$$

$$\alpha_0 > 0, \alpha_i > 0 \quad i = 1, 2, \dots, p$$

p 为**ARCH**过程的阶数,并且 v_t 为随机误差。

(二) 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。

(三) ARCH 检验的基本步骤

1. 提出原假设

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_p = 0; \quad H_1 : \alpha_j \text{不全为零}$$

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

3. 求辅助回归

$$\hat{e}_t^2 = \hat{\alpha}_0 + \hat{\alpha}_1 e_{t-1}^2 + \dots + \hat{\alpha}_p e_{t-p}^2 \quad (5.17)$$

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 χ^2 分布。

给定显著性水平 α ，查 $\chi^2(p)$ 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。

(四) 检验的特点

- 变量的样本值为大样本
- 数据是时间序列数据
- 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。

五、Glejser检验

（一）检验的基本思想

由**OLS**法得到残差，取得绝对值，然后将对某个解释变量回归，根据回归模型的显著性和拟合优度来判断是否存在异方差。

（二）检验的特点

不仅能对异方差的存在进行判断，而且还能对异方差随某个解释变量变化的函数形式 进行诊断。
该检验要求变量的观测值为大样本。

(三) 检验的步骤

1. 建立模型并求 e_i

根据样本数据建立回归模型，并求残差序列

$$e_i = Y_i - \hat{Y}_i$$

2. 寻找 $|e_i|$ 与 X 的最佳函数形式

用残差绝对值 $|e_i|$ 对 X_i 进行回归，用各种函数形式去试，寻找最佳的函数形式。

3. 判断

根据选择的函数形式作 X 对 $|e_i|$ 的回归, e_i^2 作为的替代变量, 对所选函数形式回归。用回归所得到的 β 、 t 、 F 等信息判断, 若参数 显著不为零, 即认为存在异方差性。

第四节 异方差性的补救措施

一、异方差稳健推断

二、加权最小二乘法

三、模型的对数变换

一、异方差稳健推断

定义：

如何调整标准差、 t 统计量、 F 统计量、 LM 统计量以使得他们在存在未知形式的异方差时仍然有效。这就意味着我们可以报告新的有效统计量，这种方法就是异方差稳健推断（**Heteroskedasticity-Robust Inference**）。

在大样本情况下，异方差稳健估计量是有效的。

我们先看看在异方差情况下，怎样推断参数估计量的方差 $Var(\beta_j)$ 。

对于一般的多元回归模型

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

假定随机误差项除了是**异方差**外，其他的高斯-马尔科夫假设满足，那么

$Var(\beta_j)$ 的**有效估计量**为

$$Var(\hat{\beta}_j) = \frac{\sum_{i=1}^n \hat{r}_{ij}^2 \hat{e}_i^2}{RSS_j^2}$$

这里 \hat{r}_{ij} 为来自用 X_j 对所有其他的解释变量进行回归得到的第 i 个残差，

RSS_j 是该回归的残差平方和。

上式的证明超过了本书的范围！（6-22）开方就是 $\hat{\beta}_j$ 的异方差稳健标准差。

- 一旦获得了异方差稳健标准差，就可以构造异方差稳健t统计量。
-

- 稳健标准差的优点在于：不需要知道总体模型是否存在异方差以及是何种形式的异方差。
- 异方差稳健标准差比普通的OLS标准差更有效。
- 在大样本下，截面数据分析中我们可以仅仅报告异方差稳健标准差，一般软件都提供。

-
- 运用EViews报告异方差稳健估计。
 - 打开OLS估计结果，Estimate, options, 在LS&TSLS中选择Heteroskedasticity consistent coefficient\white
 - 异方差稳健标准差通常大于OLS标准差。
 - STATA: `reg y x1 x2, vce(robust)`

二、加权最小二乘法(Weighted Least Square, WLS)

定义：

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**法估计其参数。

基本思想：

在采用OLS方法时，对较小的残差平方 e_i^2 赋予较大的权重，对较大的 e_i^2 赋予较小的权重，以对残差提供的信息的重要程度作一番修正，提高参数估计的精确程度。

不同形式的异方差要求用不同的加权方法来处理。

(一) 异方差为已知的解释变量的某一函数形式时

加权最小二乘估计

(二) 异方差形式未知时的估计——

可行的加权最小二乘法

(一) 异方差为已知的解释变量的某一函数形式时的加权最小二乘估计

以一元线性回归模型为例：

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

经检验 u_i 存在异方差，且

$$\text{var}(u_i) = \sigma_i^2 = \sigma^2 f(X_i)$$

其中 σ^2 是常数， $f(X_i)$ 是 X_i 的某种函数。

变换模型时，用 $\sqrt{f(X_i)}$ 除以模型的两端得：

$$\frac{Y_i}{\sqrt{f(X_i)}} = \frac{\beta_1}{\sqrt{f(X_i)}} + \beta_2 \frac{X_i}{\sqrt{f(X_i)}} + \frac{u_i}{\sqrt{f(X_i)}}$$

记 $Y_i^* = \frac{Y_i}{\sqrt{f(X_i)}}; X_i^* = \frac{X_i}{\sqrt{f(X_i)}}; \beta_1^* = \frac{\beta_1}{\sqrt{f(X_i)}}; v_i = \frac{u_i}{\sqrt{f(X_i)}}$

则有：

$$Y_i^* = \beta_1^* + \beta_2 X_i^* + v_i$$

随机误差项 v_i 的方差为

$$\text{var}(v_i) = \text{var}\left(\frac{u_i}{\sqrt{f(X_i)}}\right) = \frac{1}{f(X_i)} \text{var}(u_i) = \sigma^2$$

经变换的模型的随机误差项 $v_i = \frac{u_i}{\sqrt{f(X_i)}}$ 已是同方差,
 $f(X_i)$ 常见的设定形式及对应的 v_i 情况

函数形式	$\text{var}(u_i)$	v_i	$\text{var}(v_i)$
X_i	$\sigma^2 X_i$	$u_i / \sqrt{X_i}$	σ^2
X_i^2	$\sigma^2 X_i^2$	u_i / X_i	σ^2
$(a_0 + a_1 X_i)^2$	$\sigma^2 (a_0 + a_1 X_i)^2$	$u_i / (a_0 + a_1 X_i)$	σ^2

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。

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

W_i 为权数

• 一般情况下:

对于模型 $Y = X\beta + \mu$ 存在:

$$E(\mu) = 0$$

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

随机扰动项的方差-协方差矩阵

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

即存在异方差性。

\mathbf{W} 是一对称正定矩阵, 存在一可逆矩阵 \mathbf{D} 使得

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

$$\longrightarrow \boxed{W^{-1} = D^{-1'} D^{-1}}$$

用 \mathbf{D}^{-1} 左乘

$$Y = X\beta + \mu$$

两边，得到一个 **新的模型**：

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

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

该模型具有同方差性。因为

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

用OLS法估计新模型，记参数估计量为 $\hat{\beta}_*$ ，则

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

这就是原模型 $\mathbf{Y} = \mathbf{X}\beta + \mu$ 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 μ 的方差-协方差矩阵 $\sigma^2 \mathbf{W}$ 。

(二) 异方差形式未知时的估计—可行的加权最小二乘法

定义：

我们把估计的 $f(X_i)$ 记为 $\hat{f}(X_i)$ ，用 $\hat{f}(X_i)$ 而不是用 $f(X_i)$ 作为WLS估计中的权重函数得到的估计量称为**可行的加权最小二乘估计量（FWLS）**。

假定随机误差 μ_i 的方差为

$$\text{Var}(\mu_i | X_i) = \sigma^2 \exp(\delta_0 + \delta_1 X_{1i} + \delta_2 X_{2i} + \cdots + \delta_k X_{ki})$$

这里 X_1, X_2, \dots, X_k 是原回归模型中的解释变量， δ_j 是未知的参数，如果用前面的异方差表达式，那么这里

$$f(X_i) = \exp(\delta_0 + \delta_1 X_{1i} + \delta_2 X_{2i} + \cdots + \delta_k X_{ki})$$

$$f(X_i) = \exp(\delta_0 + \delta_1 X_{1i} + \delta_2 X_{2i} + \cdots + \delta_k X_{ki})$$

- 指数函数，我们需要估计 δ_j

$$\sigma_i^2 = \sigma^2 \exp(\delta_0 + \delta_1 X_{1i} + \cdots + \delta_k X_{ki}) v_i$$

取对数，并用 e_i^2 代替 σ_i^2 ：

$$\ln(e_i^2) = \alpha_0 + \delta_1 X_{1i} + \cdots + \delta_k X_{ki} + \eta_i$$

OLS 估计出 δ_j ，记为 $\hat{\delta}_j$ ，得到 *FWLS* 的权重函数：

$$\hat{f}(X_i) = \exp(\hat{\alpha}_0 + \hat{\delta}_1 X_{1i} + \cdots + \hat{\delta}_k X_{ki})$$

FWLS估计量的性质

$\hat{\beta}_{WLS}$ ：无偏、有效

$\hat{\beta}_{FWLS}$ ：有偏、一致、比 $\hat{\beta}_{OLS}$ 渐近有效

三、模型的对数变换

在经济意义成立的情况下，如果对模型：

$$Y_i = b_1 + b_2 X_i + u_i$$

作对数变换，其变量 Y_i 和 X_i 分别用 $\ln Y_i$ 和 $\ln X_i$ 代替，即：
$$\ln Y_i = b_1 + b_2 \ln X_i + u_i$$

对数变换后的模型通常可以降低异方差性的影响：

- ◆运用对数变换能使测定变量值的尺度缩小。
- ◆经过对数变换后的线性模型，其残差表示相对误差往往比绝对误差有较小的差异。

注意：对变量取对数虽然能够减少异方差对模型的影响，但应注意取对数后变量的经济意义。

第五节 案例分析

一、问题的提出和模型设定

为了给制定医疗机构的规划提供依据，分析比较医疗机构与人口数量的关系，建立卫生医疗机构数与人口数的回归模型。

假定医疗机构数与人口数之间满足线性约束，则理论模型设定为：

$$Y_i = b_1 + b_2 X_i + u_i$$

其中 Y_i 表示卫生医疗机构数， X_i 表示人口数。

四川省2000年各地区医疗机构数与人口数

地区	人口数（万人） X	医疗机构数 （个） Y	地区	人口数（万人） X	医疗机构数 （个） Y
成都	1013.3	6304	眉山	339.9	827
自贡	315	911	宜宾	508.5	1530
攀枝花	103	934	广安	438.6	1589
泸州	463.7	1297	达州	620.1	2403
德阳	379.3	1085	雅安	149.8	866
绵阳	518.4	1616	巴中	346.7	1223
广元	302.6	1021	资阳	488.4	1361
遂宁	371	1375	阿坝	82.9	536
内江	419.9	1212	甘孜	88.9	594
乐山	345.9	1132	凉山	402.4	1471
南充	709.2	4064			

二、参数估计

Dependent Variable: Y
 Method: Least Squares
 Date: 07/09/05 Time: 11:11
 Sample: 1 21
 Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-563.0548	291.5778	-1.931062	0.0685
X	5.373498	0.644284	8.340265	0.0000
R-squared	0.785456	Mean dependent var	1588.238	
Adjusted R-squared	0.774164	S.D. dependent var	1311.037	
S.E. of regression	623.0330	Akaike info criterion	15.79747	
Sum squared resid	7375233.	Schwarz criterion	15.89695	
Log likelihood	-163.8734	F-statistic	69.56003	
Durbin-Watson stat	0.429831	Prob(F-statistic)	0.000000	

估计结果为: $\hat{Y}_i = -563.0548 + 5.3735X_i$
 (-1.9311) (8.3403)

$$R^2 = 0.7855, se = 508.2665, F = 69.56$$

三、检验模型的异方差

(一) 图形法

1. EViews软件操作

由路径：**Quick/Qstimate Equation**，进入**Equation Specification**窗口，键入 $y \ c \ x$ ，点“**ok**”，得样本回归估计结果，见教材表**5.2**。

(1) 生成残差平方序列。

在得到表**5.2**估计结果后，用生成命令生成序列，记为 $e2$ 。生成过程如下，先按路径：

Procs/Generate Series，进入 **Generate Series by Equation** 对话框，键入下式并点“**OK**”即可：

$$e2 = \text{resid}^2$$

生成序列图示

File Proc Quick Options Window Help

SJG00 - (c:\windows\desktop\scws...) Save Label+/- Show Fetch Store Delete Genr Sa

Filter: * Default Eq: eq11

eq1
eq1:
eq1:
eq1:
eq1:
gdp
grol
grol
grol
grol
grol
grol
jgs
jgs2

Generate Series by Equation

Enter equation:
e2=(resid)^2

Sample:
1 21

OK Cancel

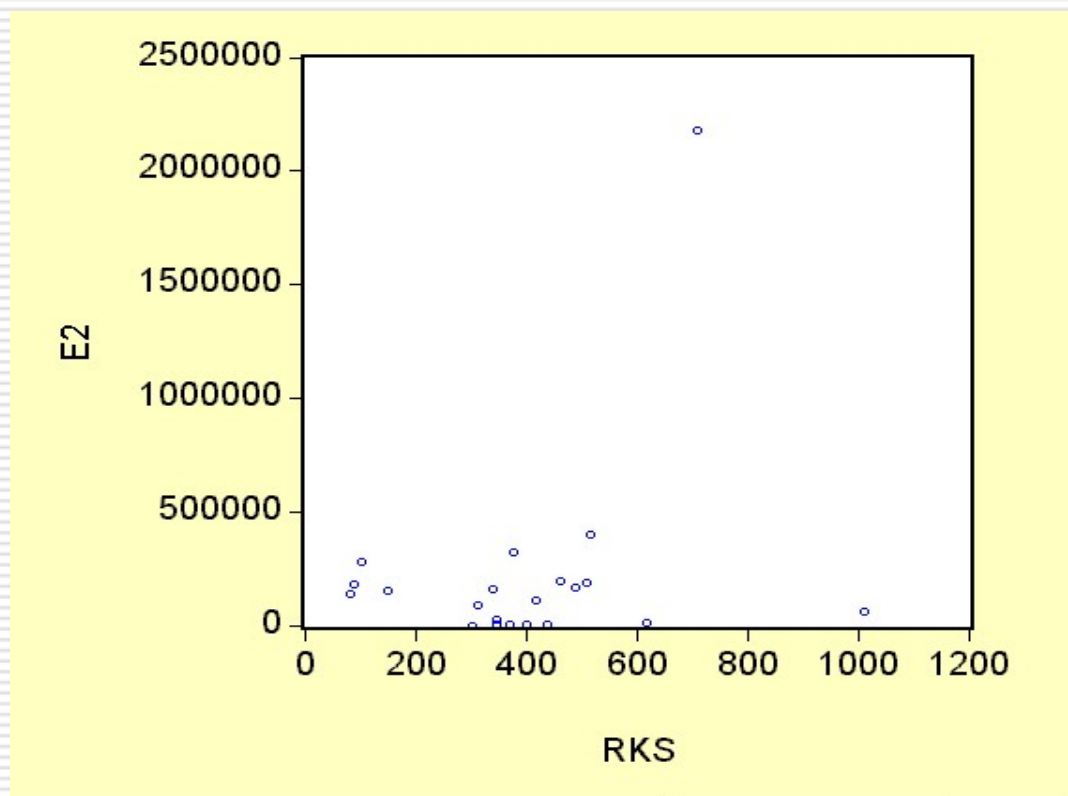
SJG00

Estimate Forecast St

Std. Error	t-Statistic
264.5140	-0.707956
0.808812	4.175037
0.661544	3.247941

R-squared 0.864732 Mean dependent var
Adjusted R-squared 0.849702 S.D. dependent var

(2) 绘制 e_t^2 对 X_t 的散点图。选择变量名 X 与 $e2$ 。（注意选择变量的顺序，先选的变量将在图形中表示横轴，后选的变量表示纵轴），进入数据列表，再按路径 **view/graph/scatter**，可得散点图，见右图：



2.判断

由图可以看出，残差平方 e_i^2 对解释变量 X 的散点图主要分布在图形中的下三角部分，大致看出残差平方 e_i^2 随 X_i 的变动呈增大的趋势，因此，模型很可能存在异方差。但是否确实存在异方差还应通过更进一步的检验。

(二) Goldfeld-Quanadt检验

1. EViews软件操作

(1) 对变量取值排序（按递增或递减）。在**Procs**菜单里选**Sort Current Page/Sort Workfile Series**命令，出现排序对话框，键入 X ，如果以递增型排序，选“**Ascending**”，如果以递减型排序，则应选“**Descending**”，点**ok**。本例选递增型排序，这时变量 X 将以 按递增型排序。

(2) 构造子样本区间，建立回归模型。在本例中，样本容量 $n=21$ ，删除中间1/4的观测值，即大约5个观测值，余下部分平分得两个样本区间：1—8和14—21，它们的样本个数均是8个，即 $n_1 = n_2 = 8$

在**Sample**菜单里，将区间定义为**1—8**，然后用**OLS**方法 求得如下结果(表**1**)

Dependent Variable: Y					
Method: Least Squares					
Date: 07/09/05 Time: 11:14					
Sample: 1 8					
Included observations: 8					
Variable	Coefficient	Std. Error	t-Statistic	Prob.	
C	598.2525	119.2922	5.015018	0.0024	
X	1.177650	0.490187	2.402452	0.0531	
R-squared	0.490306	Mean dependent var		852.6250	
Adjusted R-squared	0.405357	S.D. dependent var		201.5667	
S.E. of regression	155.4343	Akaike info criterion		13.14264	
Sum squared resid	144958.9	Schwarz criterion		13.16250	
Log likelihood	-50.57056	F-statistic		5.771775	
Durbin-Watson stat	1.656269	Prob(F-statistic)		0.053117	

在**Sample**菜单里,将区间定义为**14—21**，再用**OLS**方法求得如下结果(表2)

view	procs	objects	print	name	freeze	estimate	forecast	stats	resids
Dependent Variable: Y									
Method: Least Squares									
Date: 07/09/05 Time: 11:16									
Sample: 14 21									
Included observations: 8									
Variable		Coefficient	Std. Error	t-Statistic	Prob.				
C		-2941.087	430.3991	-6.833395	0.0005				
X		9.179365	0.692831	13.24907	0.0000				
R-squared		0.966949	Mean dependent var		2520.750				
Adjusted R-squared		0.961441	S.D. dependent var		1781.608				
S.E. of regression		349.8466	Akaike info criterion		14.76518				
Sum squared resid		734355.8	Schwarz criterion		14.78504				
Log likelihood		-57.06074	F-statistic		175.5379				
Durbin-Watson stat		1.812612	Prob(F-statistic)		0.000011				

(3) **求F统计量值**。基于表1和表2中残差平方和的数据，即**Sum squared resid**的值。由表1计算得到的残差平方和为 $\sum e_{1i}^2 = 144958.9$ ，由表2计算得到的残差平方和为 $\sum e_{2i}^2 = 734355.8$ 。根据**Goldfeld-Quanadt**检验，F统计量为

$$F = \frac{\sum e_{2i}^2}{\sum e_{1i}^2} = \frac{734355.8}{144958.9} = 5.066$$

(4) 判断

在 $\alpha = 0.05$ 下，式中分子、分母的自由度均为**6**，查**F**分布表得临界值为： $F_{0.05}(6,6) = 4.28$

因为 $F = 5.066 > F_{0.05}(6,6) = 4.28$ ，所以拒绝原假设，表明模型确实存在异方差。

(三) White检验

由表**5.2**估计结果，按路径**view/residual tests/white heteroskedasticity (no cross terms or cross terms)**，进入**White**检验。

根据**White**检验中辅助函数的构造，最后一项为变量的交叉乘积项，因为本例为一元函数，故无交叉乘积项，因此应选**no cross terms**，则辅助函数为：

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2 + v_t$$

经估计出现**White**检验结果，见表5.5。

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

White Heteroskedasticity Test:

F-statistic	55.49105	Probability	0.000000
Obs*R-squared	18.06936	Probability	0.000119

Test Equation:

Dependent Variable: RESID^2

Method: Least Squares

Date: 07/09/05 Time: 11:18

Sample: 1 21

Included observations: 21

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	823726.3	130406.0	6.316626	0.0000
X	-3607.112	554.1908	-6.508791	0.0000
X^2	4.743829	0.532983	8.900521	0.0000

R-squared	0.860446	Mean dependent var	351201.6
Adjusted R-squared	0.844940	S.D. dependent var	454283.3
S.E. of regression	178886.3	Akaike info criterion	27.15845
Sum squared resid	5.76E+11	Schwarz criterion	27.30767
Log likelihood	-282.1637	F-statistic	55.49105
Durbin-Watson stat	1.688003	Prob(F-statistic)	0.000000

从表5.5可以看出

$$nR^2 = 18.0694$$

由**White**检验知,在 $\alpha = 0.05$ 查 χ^2

分布表得临界值

$$\chi_{0.05}^2(2) = 5.9915$$

因为

$$nR^2 = 18.0694 > \chi_{0.05}^2(2) = 5.9915$$

所以拒绝原假设, 不拒绝备择假设, 表明模型存在异方差。

四、异方差的修正

加权最小二乘法 (WLS)

分别选用权数：

$$w_{1t} = \frac{1}{X_t}, w_{2t} = \frac{1}{X_t^2}, w_{3t} = \frac{1}{\sqrt{X_t}}$$

生成权数：

在**Genr/Enter equation**中分别键入：

$$w1 = 1/X \quad w2 = 1/X^2 \quad w3 = 1/\text{sqrt}(X)$$

经估计检验发现用权数 w_{2t} 较好，下面只给出用权数 w_{2t} 的结果。

方法:在 **Estimate equation** 中输入 “ $y \ c \ x$ ” ,
点

option,在对话框中点 **weighted LS**,在 **weighted**
中输入 “ ” 再点 **ok** ,即出现加权最小二乘结果。

表 5.7

View | Procs | Objects | Print | Name | Freeze | Estimate | Forecast | Stats | Resids

Dependent Variable: Y
 Method: Least Squares
 Date: 07/09/05 Time: 11:24
 Sample: 1 21
 Included observations: 21
 Weighting series: W2

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	368.6090	84.16870	4.379407	0.0003
X	2.952958	0.822688	3.589402	0.0020

Weighted Statistics

R-squared	0.938665	Mean dependent var	808.6991
Adjusted R-squared	0.935437	S.D. dependent var	1086.410
S.E. of regression	276.0493	Akaike info criterion	14.16943
Sum squared resid	1447861.	Schwarz criterion	14.26891
Log likelihood	-146.7790	F-statistic	12.88381
Durbin-Watson stat	1.705980	Prob(F-statistic)	0.001955

Unweighted Statistics

R-squared	0.625222	Mean dependent var	1588.238
Adjusted R-squared	0.605497	S.D. dependent var	1311.037
S.E. of regression	823.4555	Sum squared resid	12883501
Durbin-Watson stat	0.380523		

估计结果:

$$\hat{Y}_i = 368.6090 + 2.9530X_i$$

(4.3794) (3.5894)

$$R^2 = 0.9387, DW = 1.7060,$$

$$se = 276.0493, F = 12.8838$$

结论: 运用加权小二乘法消除了异方差性后, 参数的**t**检验均显著, 可决系数大幅提高, **F**检验也显著, 并说明人口数量每增加**1**万人, 平均说来将增加**2.953**个卫生医疗机构, 而不是引子中得出的增加**5.3735**个医疗机构。

第五章 小结

1. 异方差性是指模型中随机误差项的方差不是常量，而且它的变化与解释变量的变动有关。
2. 产生异方差性的主要原因有：模型中略去的变量随解释变量的变化而呈规律性的变化、变量的设定问题、截面数据的使用，利用平均数作为样本数据等。
3. 存在异方差性时对模型的**OLS**估计仍然具有无偏性，但最小方差性不成立，从而导致参数的显著性检验失效和预测的精度降低。

-
4. 检验异方差性的方法有多种，常用的有图形法、**Goldfeld-Qunandt**检验、**White**检验、**ARCH**检验以及**Glejser**检验，运用这些检验方法时要注意它们的假设条件。
 5. 异方差性的主要方法是加权最小二乘法，也可以用变量变换法和对数变换法。变量变换法与加权最小二乘法实际是等价的。

第五章 结束了!



THANKS