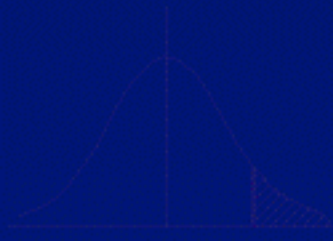


计量经济学

第五章

异方差性



引子：更为接近真实的结论是什么？

根据四川省2000年21个地市州医疗机构数与人口数资料，分析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。对模型估计的结果如下：

$$\hat{Y}_i = -563.0548 + 5.3735X_i$$

$$(291.5778) \quad (0.644284)$$

$$t \quad \quad \quad = (-1.931062)$$

$$(8.340265)785456 \quad \bar{R}^2 = 0.774146 \quad F = 69.56003$$

式中Y表示卫生医疗机构数（个），X表示人口数量（万人）。

模型显示的结果和问题：

- 人口数量对应参数的标准误差较小
- t 统计量远大于临界值
- 可决系数和修正的可决系数结果较好
- F 检验结果明显显著

表明该模型的估计效果不错，可以认为人口数量每增加1万人，平均说来医疗机构将增加5.3735个。

然而，这里得出的结论可能是不可靠的，平均说来每增加1万人口可能并不需要增加这样多的医疗机构，所得结论并不符合真实情况。

有什么充分的理由说明这一回归结果不可靠呢？
更为接近真实的结论又是什么呢？

Econometrics

第一节 异方差性的概念

一、异方差性的实质

(一) 同方差性的含义

同方差性：对所有的 i ($i=1, 2, \dots, n$) 有

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

(5.1)

因为方差是度量被解释变量 Y_k 的观测值围绕回归线

(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。

(二) 异方差性的含义

设模型为

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

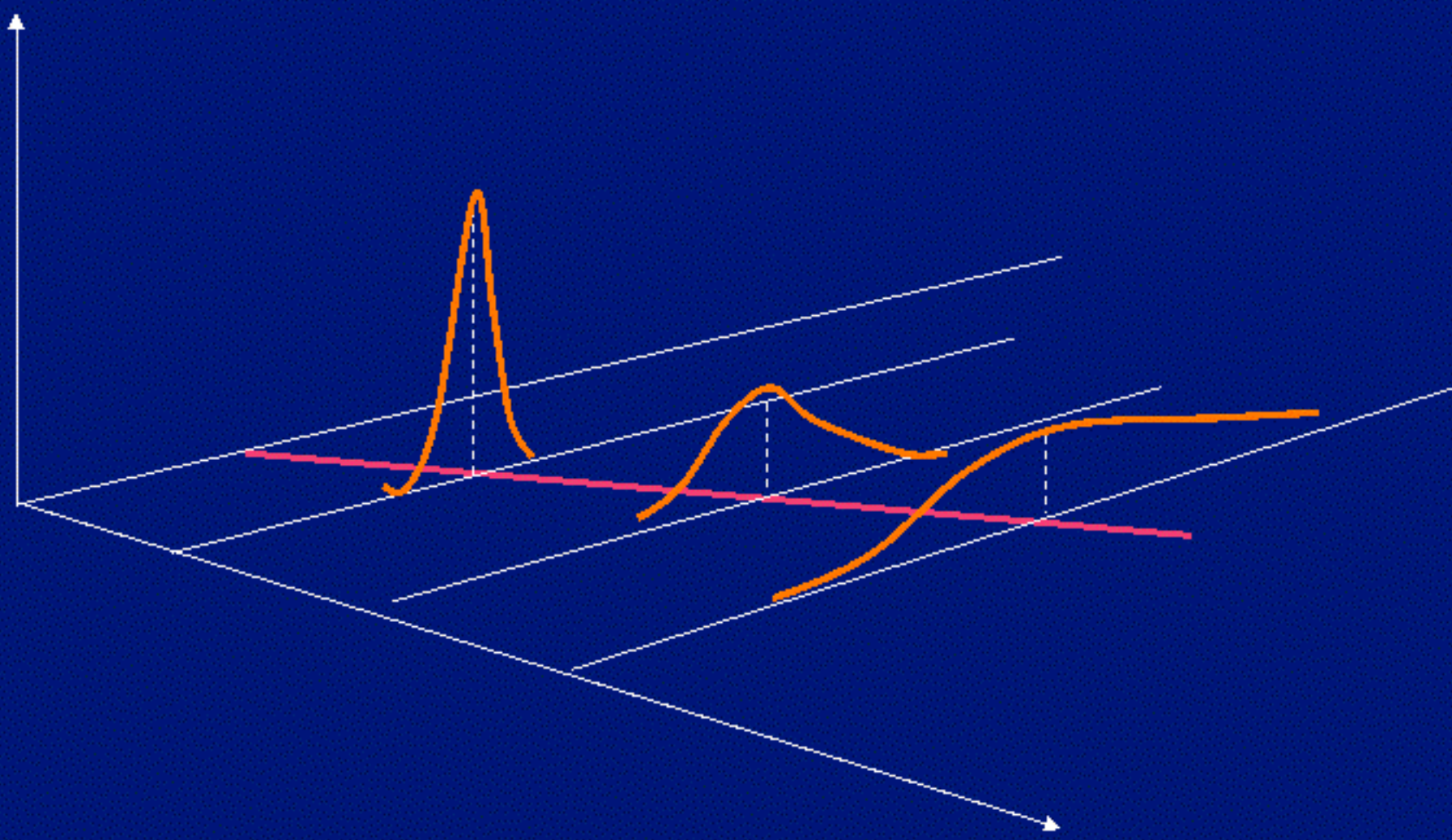
如果对于模型中随机误差项 u_i 有:

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

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:



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 X_{2i} 随 X_{3i} 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

模型的设定主要包括变量的选择和模型数学形式的确定。模型中略去了重要解释变量常常导致异方差，实际就是模型设定问题。除此而外，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) 测量误差的变化

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

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 $E(u_i) = 0$)。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。

二、对参数显著性检验的影响

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统计量值不能正确确定，所以，如果仍用t统计量值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

一、图形法

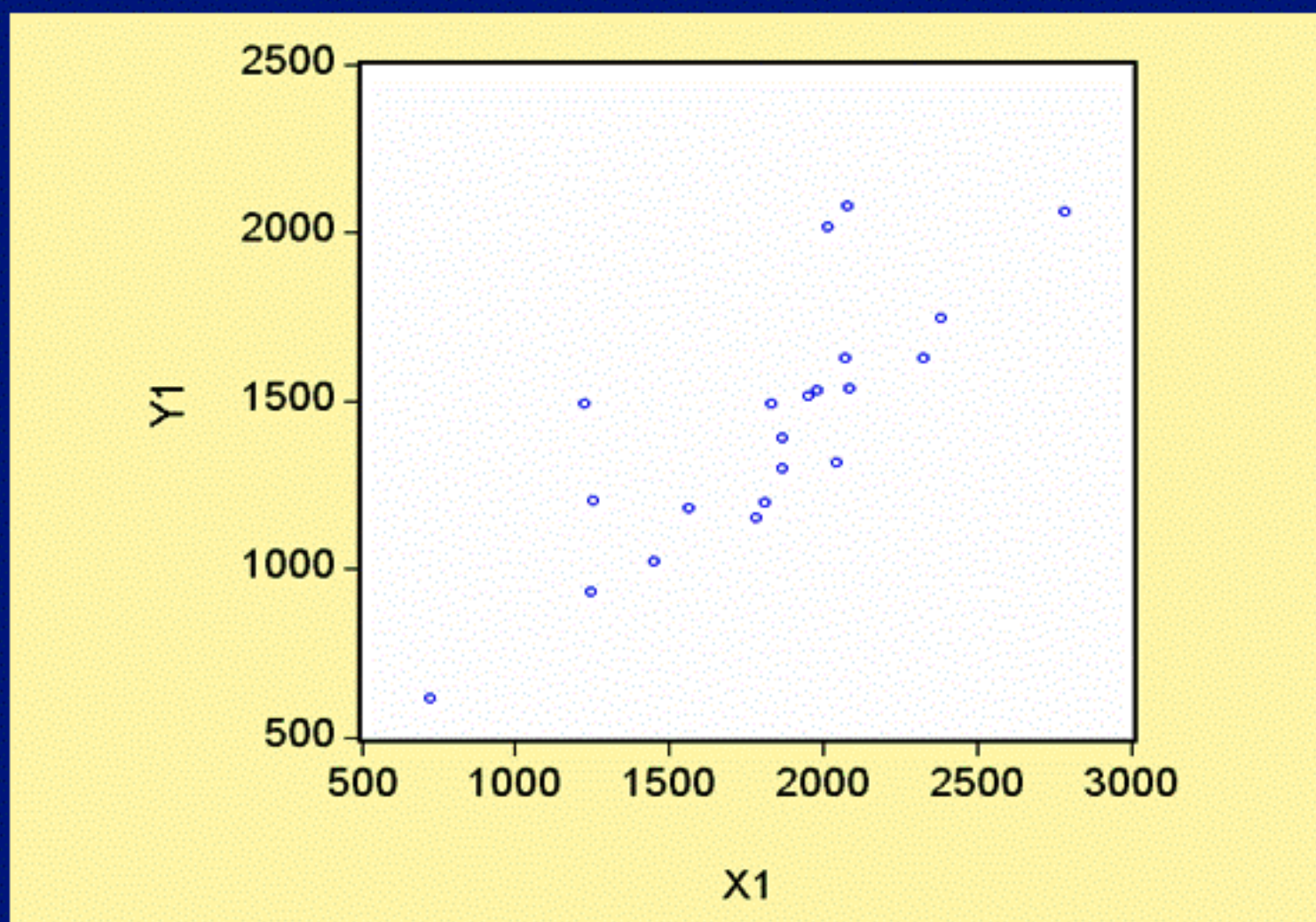
(一) 相关图形分析

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，可以初略地看到 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:

用1998年四川省各地市州农村居民家庭消费支出与家庭纯收入的数据，绘制出消费支出对纯收入的散点图，其中用 Y_1 表示农村家庭消费支出， X_1 表示家庭纯收入。



(二) 残差图形分析

设一元线性回归模型为： $Y_i = \beta_1 + \beta_2 X_i + u_i$

(5.10) 运用OLS法估计, 得样本回归模型为:

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

(5.11)

由式 (5.10) 和式 (5.11) 得残差

$$e_i^2 = Y_i - \hat{Y}_i$$

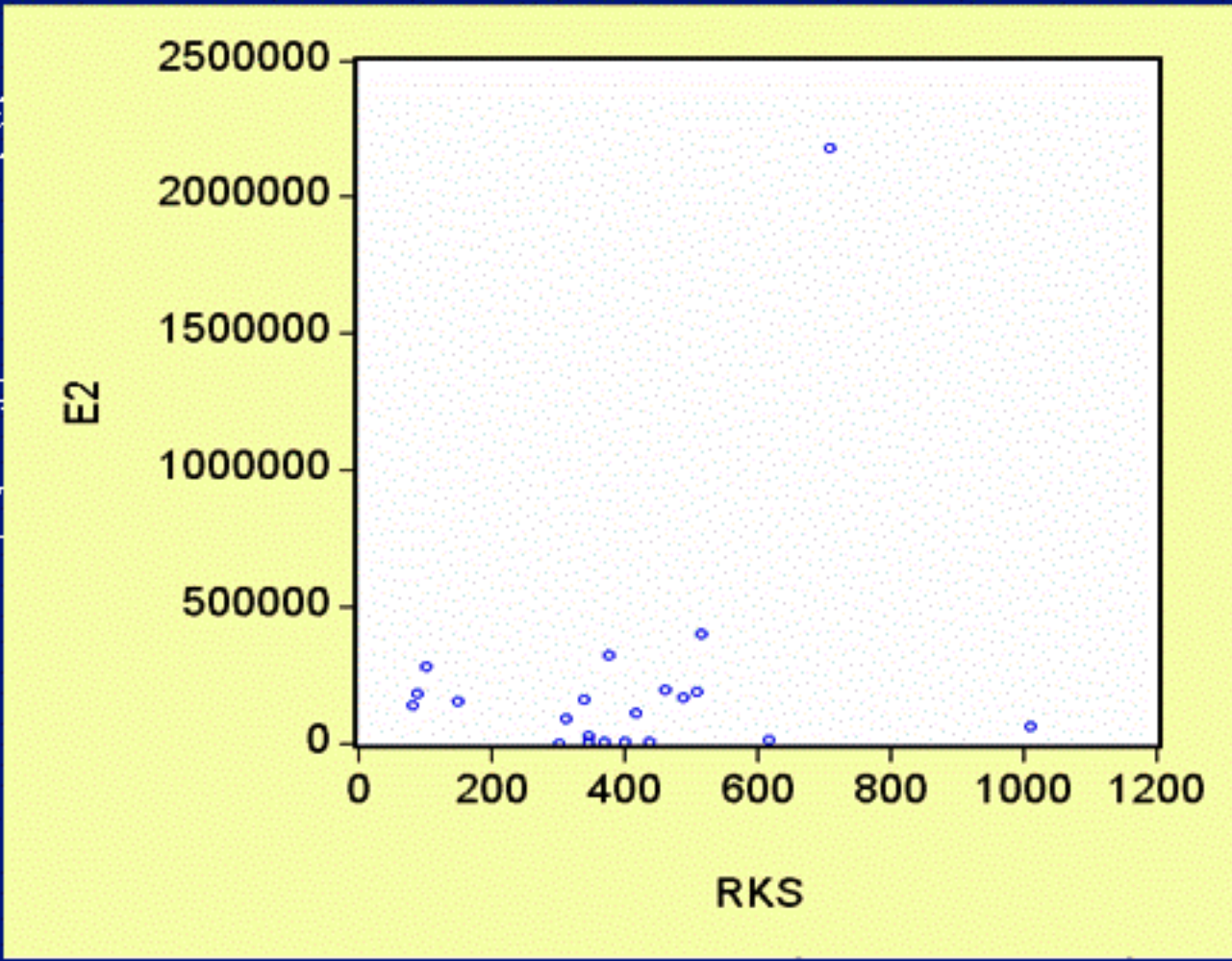
绘制出 e_i^2 对 X_i 的散点图

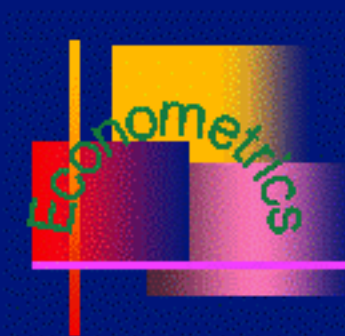
◆ 如果 e_i^2 不随 X_i 而变化

表明 e_i^2 不存在异方差,

◆ 如果 e_i^2 随 X_i 而变化,

表明 e_i^2 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。即：

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

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和为 $\sum e_{2i}^2$ 和 $\sum e_{1i}^2$ 。

$\sum e_{1i}^2$ 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 $\sum e_{2i}^2$ 和 $\sum e_{1i}^2$ 均服从 F 分布，可导出：
$$F = \frac{\sum e_{2i}^2 / [(n-c)/2 - k]}{\sum e_{1i}^2 / [(n-c)/2 - k]}$$

5、判断

给定显著性水平 α ，查F分布表得临界值 $F_{\left(\frac{c}{2}, \frac{n-c-k}{2}\right)}(\alpha)$
计算统计量 F^* ，如果 $F^* > F_{\left(\frac{c}{2}, \frac{n-c-k}{2}\right)}(\alpha)$

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，也可为递减型
- 检验结果与选择数据删除的个数 c 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

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

(5.14)

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 e_t^2 的平方。

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

关于 $X_{2t}, X_{3t}, X_{2t}^2, X_{3t}^2, X_{2t}X_{3t}$ 的辅助回归 $e_t^2 = \hat{\alpha}_1 + \hat{\alpha}_2 X_{2t} + \hat{\alpha}_3 X_{3t} + \hat{\alpha}_4 X_{2t}^2 + \hat{\alpha}_5 X_{3t}^2 + \hat{\alpha}_6 X_{2t}X_{3t}$ (5.15)

nR^2

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

$H_0: \alpha_2 = \dots = \alpha_6 = 0,$ $H_1: \alpha_j$
 ($j = 2, 3, \dots, 6$) 至少有一个不为零,

中

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 $\chi^2_{\alpha}(5)$, 如果 $nR^2 > \chi^2_{\alpha}(5)$, 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

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

(5.16)

$$\alpha_0 > 0, \alpha_i \geq 0, i = 1, 2, \cdots, p$$

v_t 为ARCH过程的阶数, 并且
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:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_p = 0; \quad H_1: \alpha_j$$

中至少有一个不为零

2、参数估计并计算

残差平方序列的估计。

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

以分别作为对

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

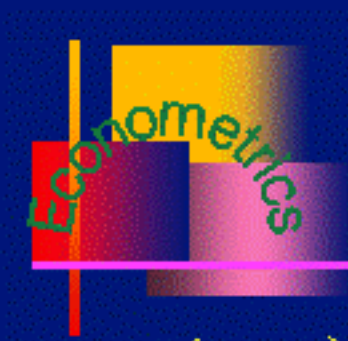
3、求辅助回归

4、检验

计算辅助回归的可决系数 $(n-p)R^2$ ，并且 H_0 成立时，基于大样本， $(n-p)R^2$ 渐进服从 $\chi^2(p)$ 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

- 变量的取值为大样本，并且是时间序列
- 只能判断模型中是否存在异方差，而不能诊断出哪



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

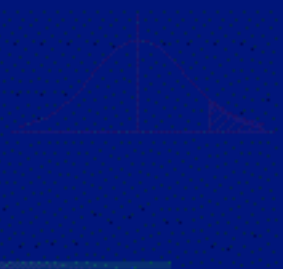
根据样本数据建立回归模型，并求残差序列 $e_i = Y_i - \hat{Y}_i$

2、寻找 $|e_i|$ 与 X_i 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

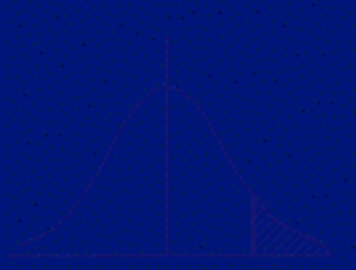
根据选择的函数形式作 $|e_i|$ 对 X_i 的回归，用 $|e_i|$ 作为 Y_i 的替代变量，对所选函数形式回归。用回归所得到的 R^2 、t、F 等信息判断，若表明参数 β 显著不为零，即认为存在异方差性。



◆ 模型变换法

◆ 加权最小二乘法

◆ 模型的对数变换



第四节 异方差性的补救措施

一、模型变换法

以一元线性回归模型为例：

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

(5.17)

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

经检验存在异方差，且 $\frac{X_i}{\sqrt{f(X_i)}}$ 是常数，其中

$\frac{X_i}{\sqrt{f(X_i)}}$ 是 $\frac{X_i}{\sqrt{f(X_i)}}$ 的某种函数。

变换模型时(用) $\frac{X_i}{\sqrt{f(X_i)}}$ 去除(5.17)式的两端，

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

(5.18)

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

记

(5.19) 式的随机误差项 u_i 的方差为

$$(5.20) \quad \text{var}(v_i) = \text{var}\left(\frac{u_i}{\sqrt{f(X_i)}}\right) = \frac{1}{f(X_i)} \text{var}(u_i) = \sigma^2$$

$$v_i = \frac{u_i}{\sqrt{f(X_i)}}$$

经变换的 (5.19) 式的随机误差项 v_i 是同方差。

已

情

常见的设定形式及对应的

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

二、加权最小二乘法

以一元线性回归模型为例：

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

(5.17)

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

经检验 X_i 存在异方差，
 且 σ^2 是常数， $f(X_i)$ 是 X_i 的某种函数。
 e_i^2 其中

(一) 基本思路 e_i^2

$$\sum e_i^2$$

σ_i^2 区别对待不同的 e_i^2 。对较小的 e_i^2 给予较大的权数，对较大的 e_i^2 给予较小的权数，从而使更好地反映 $\sum e_i^2$ 对残差平方和的影响。

(二) 具体做法

1、选取权数并求出加权的残差平方和

通常取权数 $(i=1,2,\dots,n)$ σ_i^2 w_i ，当 σ_i^2 越小时， w_i 越大。当 σ_i^2 越大时， w_i 越小。将权数与残差平方相乘以后再求和，得到加权的残差平方和 $\sum w_i e_i^2$

2、求使满足 $\min \sum w_i e_i^2$ 的 β_i^*

根据最小二乘原理，若使得加权残差平方和最小，则 $\hat{\beta}_1^* = \bar{Y}^* - \hat{\beta}_2^* \bar{X}^*$

$$\hat{\beta}_2^* = \frac{\sum w_i (X_i - \bar{X}^*)(Y_i - \bar{Y}^*)}{\sum w_i (X_i - \bar{X}^*)^2} \quad \text{其中: } \bar{X}^* = \frac{\sum w_i X_i}{\sum w_i}, \bar{Y}^* = \frac{\sum w_i Y_i}{\sum w_i}$$

三、模型的对数变换

在经济意义成立的情况下，如果对模型：

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

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

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。

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：

$$Y_i = \beta_1 + \beta_2 X_i + u_i \quad (5.18)$$

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)	医疗机构数 (个) Y	地区	人口数 (万人)	医疗机构数 (个) Y
成都	1073.3	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

进入EViews软件包，确定时间范围；编辑输入数据；选择估计方程菜单，估计样本回归函数如下：

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

$$\hat{Y}_i = -563.0548 + 5.3735X_i$$

$$(-1.9311) \quad (8.3403)$$

$$R^2 = 0.7855, s.e. = 508.2665, F = 69.56$$

三、检验模型的异方差

(一) 图形法

1、EViews软件操作

由路径：Quick/Qestimate Equation，进入 Equation Specification窗口，键入“y c x”，确认并“ok”，得样本回归估计结果，见表5.2。

(1) 生成残差平方序列。在得到表5.2估计结果后，立即用生成命令建立序列，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，即

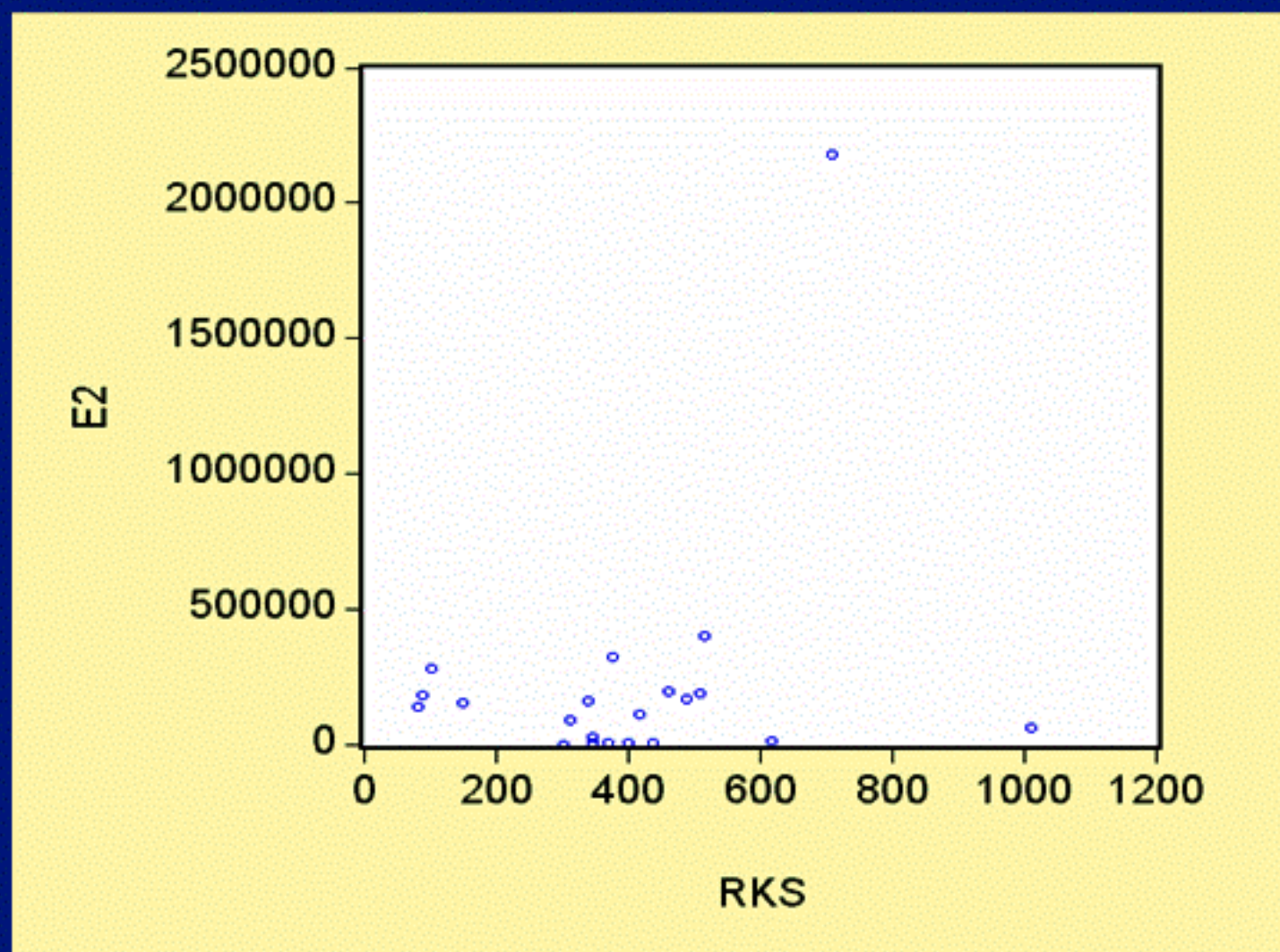
然后，在Generate Series by Equation对话框中（如图5.4），键入“ $e2 = (\text{resid})^2$ ”，则生成序列

The screenshot displays the EViews software interface. A dialog box titled "Generate Series by Equation" is open, showing the equation $e2 = (\text{resid})^2$ and the sample range 1 21. The background shows a list of variables and a regression results table.

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, 见下图。



2、判断。

由图5.5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 Series命令，出现排序对话框，如果以递增型排序，选Ascending，如果以递减型排序，则应选Descending，键入X，点ok。本例选递增型排序，这时变量Y与X将以X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， $n_1 = n_2 = 8$ 即

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5.3

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5.4

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5.3和表5.4中残差平方和的数据，即Sum squared resid的值。由表5.3计算得到的残差平方和为 $\sum e_{2i}^2 = 144958.9$ ，由表5.4计算得到的残差平方和为 $\sum e_{1i}^2 = 734355.8$ ，根

据Goldfeld-Quanadt检验，F统计量为

$$(5.19) \quad F = \frac{\sum e_{2i}^2}{\sum e_{1i}^2} = \frac{734355.8}{144958.9} = 5.066$$

(4) α 判断

在 $\alpha = 0.05$ 下，式(5.19)中分子、分母的自由度均为66，查F分布表得临界值为 $F_{0.05}(66, 66) = 4.28$ ，

因为 $5.066 > 4.28$ ，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

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_t + \alpha_2 x_t^2 + v_t \quad (5.20)$$

经估计出现White检验结果，见表5.5。

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， 18.0694 由White检验知，在 $\alpha = 0.05$ 不， $\chi^2_{0.02}(5) = 2.2012$ 查 nR^2 分布表得临界值 $\chi^2_{0.05}(5) = 5.9915$ 因为 $18.0694 > 5.9915$ 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

(一) 加权最小二乘法 (WLS)

分别选用权数 $w_{1t} = \frac{1}{X_t}$, $w_{2t} = \frac{1}{X_t^2}$, $w_{3t} = \frac{1}{\sqrt{X_t}}$, 权数生成过程是由图5.4, 在对话框中的Enter Equation处, 按如下格式分别键入:

$$w1 = 1/X$$

$$w2 = 1/X^2$$

$$w3 = 1/sqr(X)$$

经估计检验发现用权数_{2t} , 下面只给出用权数_{2t}的结果。

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)

$$R^2 = 0.9387, DW. = 1.7060, s.e. = 276.0493, F = 12.8838$$

结论: 运用加权小二乘法消

除了异方差性后, 参数的t检验

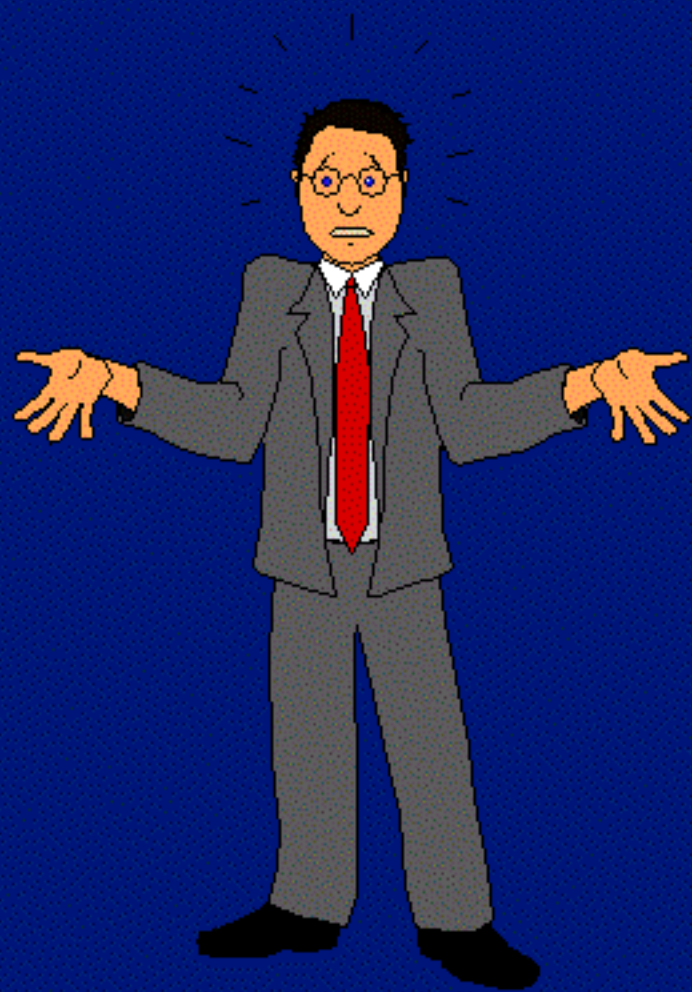
均显著, 可决系数大幅提高, F检验也显著, 并说明人口数量每增加1万人, 平均说来将增加2.953个卫生医疗机构, 而

不見引子中得出的增加5.3725

第五章小结

- ◆ 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。
- ◆ 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。
- ◆ 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。
- ◆ 检验异方差性的方法有多种，常用的有图形法、Goldfeld-Quandt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。
- ◆ 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